MASTER COPY - NO TOCAR

Serie Desarrollo de Recursos Humanos No. 50

BOOKS OF READINGS

EDUCATION OF OPERATIONAL RESEARCH AND SYSTEM ANALYSIS IN HEALTH SERVICES ADMINISTRATION PROGRAMS

PROGRAM OF EDUCATION IN HEALTH ADMINISTRATION
DIVISION OF HUMAN RESOURCES & RESEARCH
PAN AMERICAN HEALTH ORGANIZATION
WORLD HEALTH ORGANIZATION

C O N T E N T S

OPERATIONAL RESEARCH & SYSTEM ANALYSIS

	Page No.
INTRODUCTION..	*i*
FINAL REPORT	1

List of Articles

Fetter, R.B., Y.Shin y cols. - Case Mix definition by diagnosis-related groups. Medical Care (18)2 Supplement 1-52, 1980	34
Frerichs, R. y J. Prawda. - A Computer simulation model for the control of rabies in an urban area of Colombia Management Science 22(4).411-421, 1975.	90
Greenland, S., E. Watson y R. Neutra. - "The Case Control method in medical care evaluation. Med. Care 19(8),August 1981.	101
Lev, B., G. Revesz y cols. Patient flow analysis and the delivery of radiology service. Socio-Econ. Plan Sci 10. 159-166, 1976.	108
Meredith, J. Program evaluation techniques in the health services. Am J. Public Health 66(11): 1069-1073, 1976	116
Nutting, P., G. Shorr y B. Burkhalter. Assessing the performance of medical care systems: A method and its application. Med. Care 19(3), March, 1981	121
Reisman, A., J. Mello da Silva y J.B. Mantell. Systems and procedures of patient and information flow. Hosp Health Serv. Admin. Winter: 42-71, 1978	137
Shuman, L.J., H Wolfe y R. Dixon Speas, Jr. "The role of operations research in regional health planning. Operations Research 22:234-248, 1974	167
Vraciu, R. "Programming, budgeting, and control in health care organizations: the state of the art. Health Serv. Res 14(2), Summer, 1979	181
Duran, L. y A. Reisman. "Design of alternative provider team configurations: experience in both developed and developing countries. Technical memo 947, Department of Operations Research, Case Western Reserve University, Cleveland, Ohio, 1980	205

	Page No.
Hancock, W., D. Magerlein y cols. - "Parameters affecting hospital occupancy and implications for facility sizing Health Serv. Res 13(3), Fall. 1978.	229
Hindl, A., N. Dierckman y cols. - "Estimating the need for additional primary care physicians. Health Serv. Res 13(3), Fall, 1978	243
Reisman, A., B. V. - Dean y cols. "Physician supply and surgical demand forecasting: a regional manpower study. Management Science 19(12): 1345-1354, 1973.	255
Abernathy, W.J. y J. C. Hershey. "A Spatial allocation model for regional health services planning. Operations Research 20(3):629-642, 1972	268
Goldman, J. y H.A. Knappenberger. "How to determine the optimum number of operating rooms. Modern Hospital 111: 114-116, 1968	282
Revelle, C., D. Bigman y cols. "Facility location: A Review of context-free and EMS models. Health Serv. Res. 12(22):129-146, 1977	290
Brodheim, E. y G. P. Prastacos. "The Long Island blood distribution system as a prototype for regional blood management. Interfaces 9(5):3-20, 1979	308
Duraiswamy, N., R. Welton y A. Reisman. "Using computed simulation to predict ICU staffing needs. J. Nur Admin. February, 1981, pp. 39-44	324
Harrington, M. B. "Forecasting area-wide demand for health care services: A critical review of major techniques and their application. Inquiry 14-254-268, 1977	330
James B., Henry and Rodney, L. Roenfeldt "Cost analysis of Leasing Hospital Equipment, Inquiry, 15 No.1,pp 33-37, 1978	345
Klarman, H.E. "Application of cost-benefit analysis to the health services and the special case of technologic innovation. Int. J. Health Serv. 4(2):325-352, 1974	350
Warner, K.E. y R. C. Hutton. "Cost benefit and cost effectiveness analysis in health care. Med. Care 18(11) 1069-1084, 1980.	378

Page No.

Jackson, M.N., J. P. LoGerfo y cols. "Elective hysterectomy a cost benefit analysis. Inquiry 15(3):275-280, 1978.......... 394

FINAL REPORT OF THE MEETING (In Spanish)........................ 400

List of Participants... 419

FOREWORD *

These last years, PAHO with the support of Kellogg Foundation, has developed a series of workshops in Health Administration for Latin America and the Caribbean, such as Organizational Behaviour in Costa Rica, Evaluation and Health Planning in Barbados, Health Economics and Cost Control in Brazil, Epidemiological Approach to Decision Making in Chile, Management and Nurse Education in Panama, Strategic Management (or Planning) in Brazil, and as the latest Operational Research and System Analysis in Venezuela.

The materials prepared for these workshops, all in Spanish, have been periodically sent to all existing Administration Programs in the Region, including those in the Caribbean. Clearly, this Subregion has not benefited from those documents, and this fact has been stated by our colleagues in the English-speaking countries of the Caribbean.

Final reports from these workshops and other meetings were published also only in Spanish in PAHO's "Revista Educación Médica y Salud."

During these meetings, new approaches to curriculum development, exchange of research information, selection of best bibliographic references, and the choice of articles that should be disseminated to our Region were discussed. This was accomplished through the input of meeting participants.

I have no doubt that these articles discipline oriented, are the tip of an iceberg of hundreds of publications which are periodically published all over the world. These reprints are being published in Spanish as books of readings for Latin American countries to support continuing education of faculties and post-graduate students.

In discussions about these publications with professionals in charge of the Administration Programs in the Caribbean, they expressed their frustration due to the lack of information on PAHO's different programs in English. Bearing this in mind, the Health Care Administration Education Program is reproducing some original articles in English on Operational Research and System Analysis, selected during its Workshop in Venezuela,

Humberto de Moraes Novaes, M.D.
Regional Advisor in Health Care Administration Education

Workshop .. Education in Operational Research and System Analysis
Caracas, Venezuela
March 8-12,1982

I. Introduction and Background

A fundamental component of the World Health Organization's goal of health for all by the year 2000 is the development and support of health administration programs. To Further this objective in Latin America and The Caribbean, a series of workshops has been sponsored by the Pan American Health Organization (PAHO) and the W.K. Kellogg Foundation. These workshops have been concerned with organizational behavior, economics and financing, evaluation and planning, epidemiology, and administration. This particular workshop has dealt with the incorporation of Operations Research and System Analysis[1] techniques in training and continuing education programs for health systems and hospital administrators.

This workshop has been motivated by the fact that the health systems decision maker in Latin America is typically a physician. At present decisions are generally being made based on subjective grounds, without the use and, in many cases, the knowledge of the technical tools of operations research and system analysis. It is only recently that this situation has begun to change as quantitative methods are being included in both public health and continuing education curricula.

As late as 1970, health operations research was non-existent in Latin America. In 1971, PAHO sponsored a symposium on system

[1] Systems analysis and operations research should be broadly defined as encompassing those quantitative techniques of management science which can be used to solve real world problems. These techniques may range from classical industrial engineering to mathematical programming.

analysis which recommended that operations research at the hospital level become a priority.

Also at this time, a summer program at the Instituto Tecnoló gico de Estudios Superiores de Monterrey, México was initiated, in which Industrial Engineers students and professors worked together with health managers in the definition and solution of management hospital problems. Since that time diferents modalities of this type of program have been carried out in other universities of colombia, Costa Rica, Perú, México and Brazil.

At present thre are about 11 programs of industrial and systems engineering in which students and professors have participated in health services research studies and 54 academic programs that after some form of training in health administration.

The specific objectives of this particular workshop were to:

1. Exchange experiences among participants relative to the teaching of operations research and system analysis applied to health care problems.

2. Develop a basic plan for the teaching of operations research and system analysis.

3. Select a basic bibliography for use in updating instructors and health services administrators and for use in the educational process.

4. Design a basic program for continuing education in operations research and systems analysis for updating professors, researchers and administrators in the health sector.

The 19 participants at the conference represented six Latin American countries, the United States and PAHO. (see Appendix 1.).

The format of the workshop was organized to allow participants to expound on their perceptions of the problems involved in the training of Latin American health administrators in management science techniques. This enabled the group to form a concensus viewpoint from which specific problems could then be addressed, and recommendations formulated. These concerns can be summarized as follows:[2]

A. Health Concerns

1. Physicians, who are the health decision makers, do not have knowledge of the technical tools of systems analysis and operations research. As a result, decisions are made on subjective and political bases without benefit of quantitative methodologies and information.

2. There is a parallel need for the development and utilization of information systems in approaching health systems' problems. At present decisions are constantly being made without information at both the national and local levels.

3. Middle managers require a knowledge of descriptive statistics, systems analysis and evaluation techniques to improve their effectiveness.

4. Public health administrators have viewed those students being trained in management science techniques as a threat. This has led to the deterioration of training efforts in at least one country (Colombia.)

5. Effective teaching must be geared to demonstrating to health administrators how to apply methodologies to their

[2] For a specific example see: Carlos Perez, Improving the Managerial Capability of the Colombian Health System, May, 1981, included in the Appendix.

problems rather than straight lecturing. However, this requires a large amount of preparation time and such teaching aides as case studies and decision making laboratories. It is further constrained by the problem of a lack of proficient teachers who can communicate effectively with students. (See number B-1)

6. In educating administrators, particularly physicians, there is also the problem that, given a little knowledge obtained via a short course, they may assume they know a lot; that is, "a little knowledge is dangerous."

7. Health institutions have not created positions for teams who would apply management science techniques to their problems. Those that do, typically have only one position which is not satisfactory. Consequently, health institutions have not yet created a demand for operations research and system analysis techniques.

8. There has been a tendency by operations researchers to focus on the micro level problems, neglecting the more significant problems which must be addressed on a national scale. For example, the impact of improving the efficiency of a hospital is insignificant compared to that of improving the quality of water or nutritional level in the region which that hospital serves. This is particularly true in rural areas.

B. Educational Concerns

1. Most university professors do not have sufficient background in applying systems analysis and operations research in the health field to effectively motivate and educate health administrative students.

2. Programs for bringing together schools of engineering, business and medicine are not, a priori, guaranteed to succeed.

(Conversely, at Sao Paulo, Brazil, a joint program between a business school and private hospital has proven to be successful.)

3. An underlying problem in the educational process is that schools of administration are primarily oriented to the private sector. They have had little experience (and often little interest) in the public sector.

4. Most Latin American professors of public health are physicians. They have typically shown a resistance to the participation of other disciplines in the education process. This has been true in the overwhelming majority of programs in Latin America.

5. The Latin American universities typically don't have the financial resources to keep quality faculty. The result is that most leave for industry, consulting or jobs in the USA. Thus, obtaining and maintaining human resources may be the most pressing problem facing educational institutions.

6. There are both institutional and political barriers which prevent a public health school from developing a formal relationship with a management school. Without these formal relations, it is very difficult to develop any type of educational program. This type of intransiance may not be resolved until those who are responsible for budgetary decisions are cognizant of the situation and the necessity for its resolution.

As a means of addressing these concerns, the workshop participants formulated a series of recommendations. These recommendations are oriented towards the development of both continuing education as well as more formal education programs for health administrators in which the techniques of systems analysis and operations research are presented. In addition, since it is also necessary to train practitioners of those methodologies who will work in the health field, these types of programs have also been addressed. A fundamental belief of the participants is that these techniques, to be successfully applied in solving health problems, require an interdisciplinary team approach. They cannot be solved by individuals, neither administrators nor analysts, working independently. Consequently, the education of both administrators and practitioners must be addressed. Further, to be most effective, these education programs should overlap to as great as an extent as possible. Thus, as students, there is an opportunity for both future administrators and technicians to take courses together, work on group projects together, and develop the appreciation of each other's discipline and capabilities that will facilitate their working together in the future.

The next section of this report contains the recommendations of the workshop. It is followed by a section giving more specific details of the three levels of courses (short course, one year certification program, two year masters) that have been proposed.

II. Recommendations

A series of 22 recommendations has resulted from the workshop. These have been categorized into five subsets: 1) educational programs and additional workshops; 2) resources and teaching aides; 3) information exchange programs; 4) institutional curricula; and, 5) general recommendations. For each recommendation those individuals and institutions who should play the primary role in that recommendation's implementation have been noted using the following codes:

1. PAHO (the Pan American Health Organization.) Although PAHO has limited financial resources, it can continue to assume the important roles of facilitator and knowledge base. Specifically, it is a function of PAHO to bring together the appropriate individuals and institutions in order that the recommendations might be implemented. Part of this facilatator role would involve the continued development of financial resources from both foundations and the involved governments as required for program implementation.

2. EI (Educational Institutions.) This refers to those educational institutions in the different participating countries that would either provide the proposed programs directly or contract with other institutions for curricula and course development and offering.

3. Agencies refers to federal, regional and community level agencies that would provide students and funding for the proposed activities; serve as a source of data; and provide specific problem areas for study.

OR/SA (Operations Research and Systems Analysis.) This is directed at the participants of the workshop as well as selected key individuals in the various Latin American countries who would be involved with the implementation of the recommended courses and program options.

The specific recommendations are as follows. The recommended, responsible parties are indicated by ( ) after each recommendation.

A. Educational Programs

1. Three types of courses should be developed for teaching operations research and system analyses to Latin American Health Administration students and practitioners. These courses could be given as part of a specific program or in a continuing education mode. The three formats are summarized below; specific details are left for the next section.

a. Short courses geared primarily for administrators and other practitioners. The typical short course would be from 16 to 40 hours in duration and would be oriented towards showing the student how the particular methodology would be used in solving realistic problems of concern. A series of courses should be developed since it is unrealistic to expect that one course would be sufficient. The courses would have no methematical prerequisites. That would motivate the student towards an understanding of the system's perspective, an appreciation of information for making decisions and recognition that there are technically trained individuals who could apply management science techniques to the solution of health systems problems. Similar short courses would be geared to operations researchers and system analysts to enable them to more effectively function in the health field.

b. One-year certification program of approximately 240 hours oriented towards the practicing administrator. This program would be geared to giving the administrator a more indepth knowledge of management science techniques. However, it would require only mathematical proficiency in algebra as a prerequisite. The course should be designed as either an evening or weekend program to enable the administrator/student to continue to work fulltime.

c. A "two-year" masters program. This would be directed towards the administrator who wishes to acquire a more rigorous background in operations research and systems analyses. It would require a knowledge of calculus as a mathematical prerequisite. A second masters program should be developed to train a technically oriented individual (industrial engineer, computer scientist, mathematician or physicist) to function as a competent health systems problem solver. Based on the U.S. experience, such individuals would, over a relatively short period (e.g., five years) assume administrative positions in the health field and thus become a second source of quantitatively trained administrators.

As noted, suggested details of these programs are given in the next section. Specific examples of short courses are given in the Appendix. (PAHO,EI,OR/SA.)

2. Operations Research and Systems Analyses studies are dependent upon data. Informed decisions cannot be made without valid, reliable information. However, data and information are typically given little attention in most programs. Workshop participants strongly believe that PAHO should organize another conference on data and informatics which would complete the series of workshops. This latter conference would address: data sources, collection procedures, design and development of information systems, data entry procedures, editing, reduction, data processing, quality control, collation, simple data analysis, networking of data systems, information exchanges and other related topics. (PAHO.)

3. A series of directed workshops should be developed and held in each country. The workshops should bring together policy makers from government (and, where applicable, private industry) to interact with technical experts. These workshops should be aimed at addressing a specific problem of health services delivery. (PAHO, Agencies, EI.)

4. A follow-up workshop should be held to address the problem of implementation and update these recommendations in light of the progress made since the original workshop. Particular emphasis should be placed upon projects instituted and curricula developed following the first OR/SA workshop. The second workshop should emphasize the problems of program/project implementation with particular attention paid those political, cultural and financial constraints that effect implementation. Participants should include attendees from the first workshop

and selected individuals who have been involved with the implementation of operations research and systems analysis projects on a regional or national level. (PAHO.)

B. Resources and Teaching Aides

Prior to the start of the conference, two extensive literature searches (Spanish and English) were commissioned by PAHO to serve as resource material both during the workshop and for courses that would result from the conference's recommendations. The literature searches were primarily directed towards papers and documents which would provide the basis for case studies in the teaching of systems analysts and operations research techniques to health administrators. A secondary concern were papers that would acclimate technically trained individuals to health services delivery problems.

The Spanish (and Portugese) literature search was conducted by Dr. George Kastner who located 60 relevent articles, providing abstracts of 43. Given the limited number of publications found, no screening was done; rather, all articles were included for use during the workshop.

The English literature was reviewed by Dr. Richard Shachtman with assistance from Drs. Arnold Reisman and Larry Shuman. Unlike the Spanish literature, computer literature searches identified thousands of articles concerned with health system analysis and operations research.

The resultant set of 116 articles and 24 books were selected to encompass eight subcategories: project evaluation, manpower planning and scheduling, resource allocation, demand forecasting, cost benefit analysis, inventory control, technology assessment and catchment area analysis. Since a primary audience was to be administrators, the level of mathematical detail was a concern. Thus, articles of a purely theoretical nature were eliminated as were those that required a sophisticated mathematical background. While the more recent literature was emphasized, a number of earlier articles which could be considered "classics" were included. Also, articles that addressed issues of national priority were selected. These latter group included a number of articles concerned with the evaluation and cost benefit analysis of various programs.

5. Workshop participants reviewed the 116 English articles and have selected 42 that should be translated into Spanish. These articles would serve as resource documents and the bases for case studies to be used in conjunction with the three types of recommended courses. (The complete list of articles and the courses for which they would be applicable are given in the Appendix, along with a classification of those articles that were not initially recommended for translation.)

In addition, 14 of the 43 Spanish articles and 10 documents previously distributed by PAHO should be included in the final set of literature recommended for use by the health administration programs. Also included should be the list of 24 English language books and the supplemental bibliography (articles not selected for translation.) This will provide the instructor with a substantial set of resources. While the books and supplemental articles will be in English, they should be most applicable in the Master's programs where students will have a proficiency in English (at least at the reading level.) (PAHO,EI.)

6. In order to improve the quality of instruction and the relevancy of course content, PAHO should facilitate the development of case studies and computer software as teaching aides.

a. Several case studies, ranging in complexity from simple versions of material presently available in the literature to more complex cases which may require 8-16 hours of class and laboratory time to complete, should be developed. Problems chosen should be representative of those faced by Latin American health administrators. Source material may be provided by agencies within the countries themselves and/or analysts involved in health systems problems. The precedent for this type of case development has already been established in non-health related areas where faculty members have developed relevent cases under the sponsorship of private industry. Other case studies may come by expanding particular articles from the literature searches. Authors should be contacted about rewriting certain articles as case studies.

b. To enhance classroom lectures (as well as certain case studies) computer software should be developed for use in laboratory sessions. Sample data sets should be obtained which will enable students to utilize the software in order to better understand particular techniques and their use in problem solving.

As part of this effort, PAHO should encourage the interchange of information among Latin American universities. In particular, abstracts of those student projects and the theses directed at health systems problems should be circulated. Such projects could provide both case study and laboratory material.
(PAHO,EI,AGENCIES,OR/SA.)

7. An initial effort has been made during the conference to identify gaps in the published literature. As shown in the table below there are a number of areas in both the Spanish and English literature where participants felt deficiencies exist. These deficiencies should be further spelled out by representatives from Latin countries. PAHO should then commission papers or initiate studies which would reduce these deficiencies. (PAHO,OR/SA,Agencies.)

Status of Literature

	Type of Course					
	Short Course		Certificate Program		Master's Program	
Area	Span.	Eng.	Span.	Eng.	Span.	Eng.
Systems Anal.	D	D	S	D	D	D
Informatics	D	D	D	D	D	D
Decision Anal.	D	E	D	E	D	E
Oper. Research	D	E	D	E	E	E
General	D	D	D	D	D	D

D: Deficient
S: Satisfactory
E: Excellent

8. Participants and educational institutions should consider collaborating in the development of applicable software for selected microcomputers that could be used both in an educational environment as well as for real problem solving. This would require that participants work on compatible machines, using common languages. The recent advances in microprocessor technology will enable users to obtain computers capable of solving complex problems at relatively low cost. The cost of software

development will typically be more than the machine thus under-
-ing the importance of collaborative efforts (EI,OR/SA.)

9. There is considerable overlap among the various workshops that have been or will be sponsored by PAHO. PAHO should continue to provide future workshops with information from previous ones, particularly when such subject matter is relevent to more than one workshop. (PAHO.)

10. PAHO should investigate the feasibility of obtaining funds for the publication of a book in Spanish, concentrating on operations research applied to health services problems and directed at Latin American audiences. Problems of regional and national concerns should be emphasized. Such a book might contain some of the case studies noted above. (PAHO, OR/SA.)

11. PAHO should facilitate the preparation of an article describing the workshop process to appear in the appropriate professional public health journal . Such an article would help to encourage more activities of this type, and improve the communication among scientists of the different countries. (PAHO.)

12. PAHO should facilitate the dissemination of relevent working papers and theses prepared in the U.S. to Latin American educators. In particular, the Catalog of Hospital Management Engineering Technical Papers published by the clearinghouse for Hospital Management Engineering, American Hospital Association; selected citations from Hospital Management Abstracts, and brief summaries or abstracts of Master's theses related to health services problems (possibly obtained through the Association of University Programs in Hospital Administration) should

periodically be circulated to Latin American health administration programs. (PAHO,EI.)

C. Information Exchange Programs

13. PAHO should arrange for the routine surveying of the relevent English and Spanish health services literature. It should monitor major national and international meetings of the operations research societies and public health associations. This information should be distributed to Latin American health administration programs on a scheduled basis. Also included should be a listing of applicable short courses and professional meetings in both the U.S. and Latin American counties. (PAHO,OR/SA.)

14. The information exchange function of the workshops could be improved by having selected participants give formal presentations of specific projects on research activities. Such presentations could be in a seminar format. Such presentations might be scheduled prior to the actual start of the workshop. (PAHO.)

15. In order to most efficiently introduce operations research and systems analyses methodologies to health administrators while developing a cadre of sophisticated technical analysts, primary emphasis should be placed on developing the short courses and the two year master's program. The development of the certification program, while important, should receive less priority. As noted, a number of short courses should be developed with a variety of formats and content to satisfy identified target populations. (EI,OR/SA.)

16. All participants should promote the development of health operations research and systems analyses programs within industrial engineering, medicine, management and health administration departments/schools. An emphasis should be placed on instituting joint programs among the different disciplines, particularly when only one university is involved. (EI,OR/SA.)

17. An important component of both the certificate and masters programs should be the offering of at least one course in health operations research jointly with engineering, public health and/or the medical school. Such a course should involve at least one project in which students must function as an interdisciplinary team. (EI.)

18. While Latin American universities will become the predominant trainers of master's students, U.S. and Canadian universities will remain the primary training facilities for Ph.D.s. Consequently, it is in the best interest of Latin American universities to develop more formalized institutional arrangements with the appropriate U.S. schools and programs. Latin American universities should strive to assure that such U.S. programs remain relevent to and considerate of the educational needs of the Latin students. If possible, Ph.D. dissertations should be directed at problem areas that are also of importance in Latin America. (EI.)

19. As health administration programs continue to develop in Latin America, it is important that strong, formal ties are developed with both industrial engineering/operations research and management programs. If such ties are not developed,

management science techniques will not be effectively integrated into the health administration curricula. (EI,Agencies.)

E. General Recommendations

20. PAHO can further facilitate the development of research projects and collaborative programs between U.S. and Latin America by providing information to participants and other key individuals on priority problem areas, countries with favorable political climate and potential funding sources. (PAHO, EI,Agencies.)

21. It cannot be overemphasized that operations research and system analyses require interdisciplinary teams to successfully solve health services delivery problems. All programs and resultant curricula should carefully reflect this fact in their development and delivery. (PAHO,EI,Agencies,OR/SA.)

22. In order for operations research studies to lead to successful implementations, it is important that the decision maker be integrated into the planning process. This is why health administrators, at all levels, must be cognizant of the value and use of management science techniques. (PAHO,EI,Agencies, OR/SA.)

III. Specific Examples of Courses

Three types of courses were considered by the conference participants: short courses, certificate programs and full masters' degree programs. It was felt that the initial emphasis should be placed upon the development of short courses, followed by the masters' degree programs and then the certificate programs.

A. Short Courses

Short courses, by their nature, will typically be 16-40 hours in duration and either given intensively during a two to five day period, or spread over several weeks or one semester. Since it is impossible to cover much more than introductory material in one short course, the offering institution should design a series of related courses which may build upon one another and provide students with a more complete educational experience. In all instances, an attempt should be made to include problems for students to solve, both individually and in groups. These "laboratory" sessions may also enable students to have some computer interaction with either a game; e.g., simulation of a hospital or regional health system, or a simple resource allocation or forecasting model.

Examples of short courses include:

1. Stand alone course for hospital and health systems administrators. Typically, no mathematics background required, although some knowledge would be helpful. Course would present an introduction to the techniques of operations research and systems analysis with particular emphasis on understanding what these techniques can and cannot do. Students will gain a knowledge of the OR approach, cost/benefit analysis, information systems, data sources, interdisciplinary teams.* Students will learn how to recognize problems that could be solved by these methodologies, how to estimate the approximate time and effort

*These are suggested topics to be included in the optimal course. It is recognized that no one course may sufficiently cover all of them.

required for solution, how to recognize potential political barriers to implementation, and where to seek help in conducting such a study. Specific topics included in the course may be: "toy" or simplified decision problems, resource allocation problems, planning (PERT/CPM) and discounting. A team project in which an unstructured problem is formulated should also be included.

2. A stand alone course for industrial engineers, operations research, or other technical disciplines who will become involved in solving health delivery system problems. Students will become familiar with health systems and the types of problems typically encountered. Solution procedures for these problems will be presented and discussed.

Prototype courses:

a. 40 hour course - Introduction to Operations Research and System Analysis for health Professionals.

1. Introduction to the concepts of system analysis, planning, evaluation and decision making. This should include an overview, conceptual model of the systems approach as applied to health delivery problems. This framework should then be used and referred back to as the topics are presented in more depth.

2. Definition of a decision system, types of decisions that are made, types of models. Discussion of the levels of decision making.

3. Introduction to modeling; the art of modeling; example models of relevent health systems problems.

4. Presentation of a case to be developed during the course or a mini-project to be carried out by participants. Emphasis should be on problem formulation. Solution of part of the problem may be achieved by "hands on" use of canned computer programs during laboratory session.

5. Introduction to planning and evaluation approaches including flow charting, PERT (Program Evaluation and Review Technique), CPM (Critical Path Method), Gantt charts. Include actual examples with computer use optional. Also include discussion of process and outcome evaluations with appropriate examples.

6. Introduction to resource allocation models. Presentation of simple linear programming case study. Illustration of graphical solution technique. Discussion of assumptions in problem formulation. Computer usage desirable.

7. Introduction to data sources and data collection. Should review data management and processing concerns, data reliability issues, information systems and potential problem areas.

8. Preparation for cost-benefit analysis and capital budgeting. Discounting (net present value), project costing, simple cost-benefit analysis example, definition and development of internal rate of return, and benefit/cost ratio should all be covered.

9. Cost-Benefit Analysis. Definition of incremental benefits and costs; impact - incidence matrix; complete cost-benefit example, using hand calculations. Computer example desirable.

10. Marketing health services. Basic definitions, distinction between selling and marketing; strategic planning; specific health applications.

11. Introduction to uncertainty. Basic introduction to probability concepts, events, probability assessment and estimation. Concepts underlying Bayes rule, distributions, expectations.

12. Decision Analysis including decision trees, alternatives, outcomes, experiments, values and their construction. Analysis and interpretation of decision problems, simplified case studies. Laboratory session should include a full decision analysis case.

13. Forecasting and prediction techniques. Concepts of moving average, exponential smoothing, simple least squares. Interpretation of simple regression. Case study example; use of canned computer program desirable.

14. Inventory and material management. Relevent costs in inventory system. Basic inventory models -- lot size and order level.

15. Implementation of operations research/systems analysis studies. Interdisciplinary teams; working with decision makers; validation of study results; roadblocks to implementation.

b. 8-16 hour courses for health professionals*

*These courses utilize the following text: Arnold Reisman, Systems Analysis in Health Care Delivery; Lexington, Massachusetts; Lexington Books, 1979.

Topic	One Day Course	Two Day Course	Two Day Course With Math
Introduction to Systems Analysis	1 hr.	1 hr.	1 hr.
Systems Concepts	1	1	1
Graphical Techniques	2	5	5
Algebraic Techniques	-	-	8
Data Collection	1	3	3
Formal Aides to Creativity	.5	1	1
Art of Systems Analysis Strategies for Implementation	.5	1.5	1.5

c. Quantitative Methods Applied to the Administration of health care. (45 hours)

This course is designed for health administrators, typically physicians and nurses, who have limited mathematical and statistical background. The course, which is given at the University of Sao Paulo, Brazil, has been revised several times.

1. Elements of Mathematics - cartesian coordinates, linear equations, elementary graphical solution procedures.

2. Elements of Statistical Inference - set theory, probabilities, Bayesian probability, discrete distributions, continuous distributions, simple linear regression.

3. Elements of Operations Research - critical path method, linear programming, queueing.

d. Methods Improvements Techniques in Hospitals.

This is designed as a stand alone course for hospital department heads and administrative staffs. Students must define and solve a real problem effecting their department. (40 hours)

1. Introduction to Methods Improvement - scientific management; systematic approach to problem solving.

2. Philosophy of Work Simplification - extension to hospital setting; problem identification; use of personnel in work simplification; PERT.

3. Work Session in Problem Definition

4. Process Analysis - definitions and symbolic notation; flow diagrams; evaluation of flow diagrams and subsequent action.

5. Operations Analysis - operation charting; motion economy; micro-motion-video motion techniques.

6. Work Simplification - systematic examination of problems; what to look for; what to do; layout; automated systems.

7. Work Session on Problem Analysis

8. Work Sampling - sample size determination; procedure; interpretation of results.

9. Other Data Gathering Techniques - time study; link analysis; questionnaires.

10. Work Session on Measurement

11. Cost Analysis - cost measures; cost estimates; non-monetary criteria; comparing methods.

12. Elementary Statistics for Analyzing Differences - testing results; interpretation of data.

13. Work Session on Evaluating Results

14. Implementing the Results of Methods Improvement Studies

15. Work Session: Reports and Critiques of Projects

e. Health Systems Research Seminar

This is a 45 hour course designed for both health administrative students and industrial engineering/operations research students who want to learn about the actual application of quantitative methods to health services problems. In addition to classroom discussions, students form small, interdisciplinary teams and solve an actual hospital or health system problem. Discussion topics have included:

1. Introduction to Industrial Engineering/Operations Research in Health

2. Evaluation of a Unit Dose Drug Distribution System

3. Hospital Cost Models - step down cost allocation; microcosting.

4. Incentive Reimbursement - an industrial engineering approach.

5. Utilization Review Models

6. Nurse Staffing Methodologies

7. Predictive Reimbursement Models

8. Grouping Hospitals for Reimbursement and Cost Control

9. Operating Room Scheduling and Admissions System

10. Linear Programming Applications: menu planning; theraputic radiology

11. Regional Planning - location of a network of c'...ics.

12. Emergency Medical Services - information systems; simulation of EMS systems; evaluation of paramedic performance.

13. Health Economic Analyses and Models

14. Measuring the Effectiveness of Care

f. Information Systems and the Evaluation of Quality of Health

This 30 hour course is directed towards health services administrators and planners. It presents an overview of the use of information and data processing technology for control and utilization for administration and planning of health systems.

1. Origin, nature and utilization of information in health systems; medical documentation; administrative records.

2. Information for decision processes; importance and objectives of educating health personnel in informatics.

3. Objectives of the users of information; functional levels; levels of centralization or distribution of information systems; impact of information on organizations.

4. Needs of administrators; information for administration and planning.

5. Systems analysis; project, role of systems analysis; planning and control of projects; collection and distribution of data; process for implementation of information systems.

6. Computers and automation: basic concepts, electronic data processing; files and logical structures; data banks; computers and their components; data processing centers.

7. Options in data processing: centralized, decentralized, and distributed systems.

8. Applications of data processing to health systems; advantages and disadvantages for the use of computers in medicine.

9. Information systems for evaluation and control; methods for evaluating the utilization and quality of services; medical auditing.

g. Quantitative and Analytic Methods for Health Administration*

This 40 hour course seeks to: 1) create an appreciation for, and an understanding of, the basic processes of systems analysis modeling, decision making and control; 2) identify the intimate roles these processes play in the activities and responsibilities of the health administrator. Some of the more fundamental decision/control models and methodologies of systems analysis, management science, and economics are discussed. Strengths and limitations of these models for more effective health decision making are stressed. Prerequisites for this course include college algebra, basic linear algebra and elementary probability and statistics. Topics include:

1. Overview of the Systems Analysis and Operations Research/Management Science Health-Related Literature

2. Systems Approach and Systems Analysis - A Prologue for Health Care Decision Making

3. Quantitative Models and Their Role in Health Care Decision Making and Control

4. Simple Deterministic Decision Models: Inventory and Project Management

*Course developed by Barnett R. Parker, Ph.D., School of Public Health, University of North Carolina, Chapel Hill, N.C.

5. Complex Deterministic Decision Models: Mathematical Programming

6. Solving Linear Programming Problems by Computer; Sensitivity Analysis

7. Generation of Data Appropriate to the Health Care Decision Making Task: Techniques of Measurement

8. Goal Programming and Integer Programming Applications to the Health Sector

9. Cost-Benefit/Cost-Effectiveness Analysis: A fundamental Mode of Thinking in Health Care Decision Making

10. Feedback, Control and Program Evaluation for More Effective Health Care Management

11. Guest Speakers - Special Topics

B. Certificate Programs

This program would be directed towards the working health administrator who is seeking more indepth knowledge than is provided through a series of short courses. It is also for the individual who is seeking a career change. Middle level managers would take this program as one step in advancing to a higher management position. The program would typically be 240 hours. It would require a mathematical background through algebra.

The graduate of the program should be able to formulate and solve simple quantitative problems (linear programming and queuing); perform an elementary cost-benefit analysis; collect and reduce data (descriptive statistics); use "canned" computer programs for analysis; formulate (but not solve) more complex problems. The graduate should have a basic understanding of the costs involved in providing health services. The interdiscipline team concept should be stressed. A team project of 2-3 month's duration should be included.

Courses in the program should include an introduction to quantitative methods (e.g., short course c.); a more rigorous course in applying quantitative methods (e.g., short course e. or g.); an introductory course in health economics; an introductory course in biostatistics; and a course in information systems. Optional courses to be considered are: theory of health planning and evaluation; quantitative methods of planning and evaluation; and health organization behavior.

C. Masters' Program

Two forms of Masters' programs were discussed. The first would be directed towards hospital and health systems administrators who required more extensive educational training than offered under a certificate program. While students might be fulltime administrators, they would be expected to spend the necessary time in residence at an academic institution. The mathematical background should include college mathematics through calculus. This program would include all the courses listed under the certificate program. Also included would be an advanced course in statistics, a second level course in operations research methodologies (probabilistic models), and an introductory course in econometrics. Students would carry out two or three team projects during the program. These projects could be done as part of an interdisciplinary course (e.g., short course e.).

The second type of Masters' program would be for engineering undergraduates who wanted to solve health and hospital systems problems. The program would provide a strong emphasis in operations research methodologies. In addition, students should become familiar with cost accounting concepts, economic principles and information systems. Graduates of this program would be expected to perform a number of the following functions:

1. Solve linear programming and inventory problems;
2. forumulate and analyze stochastic models;
3. perform indepth statistical analyses (ANOVA and multiple regression);
4. design data collection studies;
5. program in a high level computer language (FORTRAN, Pascal or PL-1);
6. design program evaluation;
7. conduct computer simulations;
8. write technical reports;
9. interact with decision makers and work as part of an interdisciplinary team;
10. take unstructured problem, abstract it, formulate a mathematical model, obtain solutions and implement the best, acceptable solution;
11. conduct long range planning studies.

IV. Articles to Translate Into Spanish

As noted, an extensive literature search was prepared prior to the start of the conference. From this list, a total of 41 articles were recommended for translation. These are summarized below. Also given is that type of program (short course, certificate, masters') for which the article would be most beneficial.

Senior Author	Year Published	Applicable Program: Masters	Certificate	Short Course
Berkson	1979		X	X
Escudero	1980	X	X	X
Fetter	1980	X		
Frerichs	1975	X		
Greenland	1981	X	X[1]	
Hartunian	1980	X	X	X
Lev	1976	X		
Meredith	1976	X	X	X
Nutting	1981	X	X	X
O'Connor	1972	X	X	
Reisman	1978	X	X	
Shoenbaum	1976	X	X	X
Shuman	1974	X	X	X
Vracin	1979	X	X	
Duran	1980	X	X	X
Hancock	1978	X	X	X
Hindle	1978	X	X	X
Reisman	1973	X		
Abernathy	1972	X		
Goldman	1968			X
ReVelle	1977	X	X	
Brodheim	1979	X	X	X[1]
Duraiswany	1981	X	X	X
Harrington	1977	X	X	X[1]
Centerwall	1978	X	X	X
Couch	1981	X	X	X
Eisenberg	1978	X	X	X
Henry	1978	X	X	X
Klarman	1974	X	X	X[1]
McGregor	1978	X	X	X
Muller	1980	X	X	
Schwartz	1979	X	X	X[1]
Willems	1980	X	X	X
Warner	1980	X	X	
Kendall	1980	X	X	
Jackson	1978	X	X	X
Evans	1981	X	X	X
McNeil[2]	1975	X	X	X
Nemhauser	1975	X	X	X
Parker	1975	X	X	X

1. Optional
2. Series of three articles in NEMJ 293:211-226

In addition to these works in English, it was recommended that PAHO also distribute a number of the works already in Spanish. Included in this second list are several papers that have been published by PAHO. It was recommended that all of these PAHO publications be included in any distribution of papers. Recommended Spanish articles are:

Senior Author	Year Published	Applicable Program		
		Masters	Certificate	Short Course
Ackoff	1974	X	X	X
Barrenechea	1978	X	X	
Dunia	1976		X	X
Miembros Facultad de Ingenieria	1974	X	X	
Grundy, Reinke	1974	X	X	X
Marin	1978	X	X	X
Novaro	1973	X	X	
Rodrigues	1978		X	X
Schmidt (p.47)	1980		X	X
Schmidt	1981		X	
Schmidt (p.50)	1981	X	X	
Instituto Monterrey	1976	X	X	

Case Mix Definition by Diagnosis-Related Groups

ROBERT B. FETTER, YOUNGSOO SHIN, JEAN L. FREEMAN,
RICHARD F. AVERILL AND JOHN D. THOMPSON

Case Mix Definition by Diagnosis-Related Groups

ROBERT B. FETTER,* YOUNGSOO SHIN,† JEAN L. FREEMAN,‡
RICHARD F. AVERILL,§ AND JOHN D. THOMPSON‖

*From the Center for Health Studies,
Institution for Social and Policy Studies,
Yale University*

* *Professor, School of Organization and Management, and Institution for Social and Policy Studies.*
† *Assistant Professor, Department of Preventive Medicine, College of Medicine, Seoul National University, Seoul, Korea.*
‡ *Assistant in Research, Center for Health Studies, Institution for Social and Policy Studies.*
§ *Associate Director for Research, Health Management Program, School of Organization and Management.*
‖ *Professor of Public Health, School of Medicine, and Institution for Social and Policy Studies.*

The research reported herein was performed pursuant to Contracts No. HCFA 600-75-0180 and No. HCFA 600-77-0043, from the Health Care Financing Administration, Department of Health, Education and Welfare. Publication of this monograph was supported by Contract No. 600-75-0180, from the Health Care Financing Administration, Department of Health, Education and Welfare. The opinions and conclusions expressed herein are solely those of the authors and should not be construed as representing the opinions or policy of any agency of the United States government. Copies of a more detailed description of this research are available by writing to Al Esposito, Director, Division of Hospital Experimentation, Office of Demonstrations and Evaluations, Office of Research, Demonstrations, and Statistics, Health Care Financing Administration, Room 1E6, Oak Meadows Building, 6401 Security Blvd., Baltimore, MD 21235. Requests should be made for AUTOGRP Patient Classification Scheme and Diagnosis Related Groups (DRGs), *Health Care Financing Grants and Contracts Report Series, HEW Publication No. (HCFA) 03011 9-79. Inquiries regarding the software available for analyses with AUTOGRP and DRGs should be made in writing to Ms. Maria DeMusis, SOM, 12 Prospect Place, New Haven, CT 06520.*

Acknowledgment

The authors wish to acknowledge the support, insights, and cooperation during this research of Al Esposito, Project Officer, Health Care Financing Administration, Department of Health, Education and Welfare. The successful completion of this research relied heavily on the programming skills and patience of Enes Elia, Senior Systems Programmer, and the excellent administrative support of Ann Palmeri and Marla DeMusis. The assistance of the New Jersey Department of Health, especially Michael J. Kalison in preparing the description of the New Jersey prospective reimbursement system is greatly appreciated. The material presented on the case mix approach to hospital regional planning is based on the current doctoral research of Robert Chernow. The authors are indebted to the many physicians who, during the years of development of the DRGs, participated in their formulation and evaluation.

The research reported here represents the current stage of development of the Diagnosis-Related Groups. The initial work on the approach and associated technology for patient classification began nearly a decade ago. Significant contributions were made during the course of this research by Donald C. Riedel, Ronald E. Mills, Lesley Mills and Phyllis Pallett.

Supplement to *Medical Care* Volume 18, Number 2

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year, $36.00; special price U.S. students, $20.00; all other countries except Japan, one year, $42.00. Single copies $5.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

Preface

DURING THE PAST DECADE, health care financing researchers have sought to develop equitable methods to constrain the rate of increase in health care expenditures. The need for a successful national hospital cost containment program has been highlighted as a major step toward overall health care cost containment and a comprehensive national health plan.

In early efforts to compare and control hospital costs, researchers calculated product costs by unit of service, such as lab tests, radiology tests, or days of routine hotel service. Several incentive schemes were devised to encourage hospital efficiency and low unit cost. Experience has shown that focusing on unit cost alone encourages increased length of stay and ancillary utilization and argues that attention should be focused on medical-practice patterns.

Recently, much attention has been focused on the cost per patient stay or per case treated. Shifting to a single hospital product required that techniques be developed to adjust for variations in hospitals' patient or case mix.

To group hospitals with a similar case mix, researchers have generally concentrated on proxy measures, such as the mix of services or facilities available in each hospital. While these approaches are relatively easy to calculate, they normally do not address either the extent or the use of a hospital's available services or facilities.

What is required is a classification scheme that is both manageable in terms of the number of case types defined and reasonable in terms of the variation in resources needed to treat each case type. This should permit direct measurement of a hospital's case mix.

The development of the Diagnosis-Related Groups (DRGs) represents a significant step in case mix measurement and application for reimbursement purposes. DRGs classify 383 types of cases encountered in the hospital acute-care setting. Each DRG represents a class of patients requiring similar hospital services. Since DRGs are medically meaningful, they help provide a common basis for comparing cost effectiveness and quality of care delivered. DRGs also have the potential to assist the hospital administrator as he manages his institution and communicates with the medical staff.

Many PSROs have already explored the use of DRGs to review length-of-stay and treatment patterns. New York and Maryland have incorporated the DRG concept into their hospital cost containment programs and both New Jersey and Georgia will soon be incorporating DRG methodologies in new hospital cost containment programs. At the federal level, DRGs are being considered for incorporation in new reimbursement procedures for acute care hospitals. Given the current emphasis on hospital cost containment, the development of DRGs is both an important and timely advancement of the health care financing field.

JAMES M. KAPLE
Acting Director
Office of Research, Demonstrations and Statistics
Health Care Financing Administration
Department of Health, Education, and Welfare

MEDICAL CARE
February 1980, Vol. XVIII, No. 2, Supplement

1. Introduction

THE MAJOR FUNCTION of a hospital as an acute-care inpatient facility is to provide the diagnostic and therapeutic services required by physicians in the clinical management of their patients. In addition, the hospital also makes available certain hotel and social services such as meals, laundry and counseling, to patients while they are residents at the institution. Considered as an economic entity, the hospital's outputs are the specific services it provides in terms of hours of nursing care, medications and laboratory tests. Its inputs are the labor, material and equipment used in the provision of these services. Each patient receives a specific set of outputs or services which is referred to here as a product of the hospital.

For example, consider a patient under treatment for uncomplicated pneumonia. His hospital stay may last 6 days. Depending on his treatment process, he could receive several chest roentgenograms, a sputum Gram stain and culture, a blood culture, several complete blood counts, a urinalysis, routine blood chemistries, 2 days of oxygen therapy, 2 days of IV fluids, 2 days of IV penicillin followed by 4 days of IM penicillin and a supply of oral penicillin to be taken at home. During his stay, he will also require various amounts and levels of nursing care. This set of outputs constitutes a product ordered by the physician and provided by the hospital to the patient in addition to meals and other hotel services. Each department within the facility consumes a certain amount of inputs in terms of standard labor, material and overhead, depending on the output produced. For the radiology department, this would include the films, the technician time, and a portion of the departmental overhead expenses for the equipment and space utilized in the production of the various types of roentgenograms.

Since individual patients receive different amounts and types of services, the hospital may be viewed as a multiproduct firm with a product line that in theory is as extensive as the number of patients it serves. The particular product provided each patient is dependent upon his condition as well as the treatment process he undergoes during his stay. For example, in most instances an ulcer patient with a major surgical procedure such as an exploratory laparotomy or gastric resection requires more units of physician and nursing time, more medication, more ancillary services and remains hospitalized at least 10 days longer than a woman admitted for a normal delivery. The former case is generally referred to as more complex since it receives more of the institution's outputs in terms of nursing time, meals, laboratory tests, and special services than the latter case. During any specified time period, an institution admits and discharges a variety of cases representing different levels of complexity. The relative proportions of the different types of cases the hospital treats are collectively referred to as its case mix.

Historically, patient-days of care and the number of admissions or discharges have been used to describe hospital output while cost per patient-day, per cent occupancy and mortality rate have been used to evaluate hospital performance. Such aggregate measures presented out of context of the types of cases treated by an institution and their relative complexity is not particularly useful information to hospital management for internal assessment of efficiency and effectiveness nor to regulatory agencies for inter-institutional comparisons. In particular, the outcome and cost of the individual patient care processes that give rise to the hospital products cannot be adequately determined. As a consequence, administrators and clinicians cannot ascertain the quality and cost implications of the treatment plans practiced within the insitution nor can regulators assess the impact of alternative methods across institutions.

In order to evaluate, compare, and provide relevant feedback regarding hospital performance, it is necessary to identify the specific products that institutions provide. As defined above, a hospital product is a set of services provided to a patient as part of the treatment process controlled by his clinician. While each individual patient admitted to an institution is unique, he has certain demographic, diagnostic and therapeutic attributes in common with other patients that determine the type and level of services he receives. If these classes of patients with the same clinical attributes and similar processes of care can be identified, then the framework within which to aggregate patients into case types is established. Moreover, if these classes cover the entire range of inpatients in the acute-care setting, then collectively they constitute a classification scheme that provides a means for examining the products of the hospital, since patients within each class are expected to receive a similar product.

The means of defining hospital case mix for this purpose is the construction and application of a classification scheme comprised of subgroups of patients possessing similar clinical attributes and output utilization patterns. This involves relating the demographic, diagnostic and therapeutic characteristics of patients to the output they are provided so that cases are differentiated by only those variables related to the condition of the patient (e.g., age, primary diagnosis) and treatment process (e.g., operations) that affect his utilization of the hospital's facilities. These groups or patient classes may then be useful for certain applications in patient care monitoring, budgeting, cost control, reimbursement and planning.

Various definitions of case mix and measures of case mix complexity have been proposed and applied by a number of researchers.[3,15,17,18,22] The most frequently applied definition of case mix is based on a patient's primary diagnosis. Primary diagnoses are recorded using either of 2 coding schemes—the International Classification of Diseases, Adapted for Use in the United States, Eighth Revision (ICDA8) and the Hospital Adaptation of ICDA, Second Edition (HICDA2).[6,13]* These coding schemes provide a classification of conditions of morbidity and mortality for statistical reporting purposes as well as for information retrieval. Four-digit codes are assigned to diagnoses (3-digit major category descriptor and a fourth-digit modifier) and 3-digit codes to surgical procedures (2-digit major category descriptor and a third-digit modifier).

While partitioning into groups based on diagnosis or ranges of diagnoses alone has benefits in terms of uniform reporting of descriptive statistics about a population, it is not sufficient for defining cases with respect to output utilization. Other variables such as surgical procedures and age of patient in conjunction with diagnosis are necessary to describe adequately sets of patients with similar utilization patterns of inpatient facilities. For example, in one institution examined, it was determined that patients with a primary diagnosis in the general category Gastric and Peptic Ulcer (ICDA8 codes 531–5349) with no surgery or complicating secondary diagnoses were hospitalized on the average 6 days while those with a minor surgical procedure such as endoscopy stayed 12 days, and those

* The 9th Revision, International Classification of Diseases, Clinical Modification (ICD-9-CM) was implemented as of January 1, 1979.

with a major surgical procedure and multiple diagnoses stayed 21 days. Moreover, the first and second sets of patients incurred only 13 and 30 per cent, respectively, of the costs incurred by the latter. Thus, the type of surgery performed and the presence or type of secondary diagnosis make a considerable difference in the hospital product provided to the patient. For this reason, patient case types determined solely by primary diagnosis is not acceptable for use in defining hospital case mix, since it is not precise enough to account for either the condition of the patient or the complexity of the treatment protocol.

Another patient classification scheme in wide use is that developed by the Professional Activity Study (PAS) of the Commission on Professional and Hospital Activities (CPHA), which publishes tables of length-of-stay statistics based on data from participating hospitals.[7] These tables present summaries on the basis of primary diagnosis, the presence of additional diagnoses, the presence of any surgeries, and age. Specifically, the PAS classification divides all possible primary diagnoses into 349 mutually exclusive major diagnostic categories. Each of these major diagnostic categories is then divided on the presence or absence of a secondary diagnosis, presence or absence of any surgery and 5 age categories (0–19, 20–34, 35–49, 50–64, 65+).† This results in 20 subcategories for each of the 349 major diagnostic categories for a total of nearly 7,000 patient classes. The information is currently used extensively by Professional Standards Review Organizations (PSROs) in setting length-of-stay checkpoints as part of their concurrent review process.

Although the PAS scheme takes into account additional patient attributes such as secondary diagnosis, age and information related to the treatment process, such as operation status, these attributes still may be insufficient in certain instances to describe adequately a case in terms of output utilization. Refer to the example described above where the specific type of operation differentiated one set of surgical ulcer patients from another. Thus, because of its uniform structure throughout all 349 diagnosis groups, the PAS scheme tends to overspecify in some diagnostic categories where the extra variables may not be particularly relevant as regards utilization of facilities and underspecify in others where more precise information is required.

As an alternative patient classification scheme, the Diagnosis-Related Groups (DRGs) have been constructed, based on a procedure referred to as the significant attribute method. The fundamental purpose of the DRG approach is to identify in the hospital acute-care setting a set of case types, each representing a class of patients with similar processes of care and a predictable package of services (or product) from an institution. Using this approach, the entire range of diagnostic codes was initially divided into broad disease areas, such as Diseases of the Eye, Diseases of the Ear, Cerebrovascular Diseases, and Infectious Diseases. Each of these categories was further subdivided into groups based on values for those variables that demonstrated an effect in predicting output as measured by length of hospital stay. It was decided to use length of stay (LOS) as opposed to some other output measure since it is an important indicator of utilization, as well as being easily available, well standardized, and reliable. The use of LOS and its relation to other measures is discussed further in the next section. The 383 DRGs that resulted from this process are interpretable from a medical perspective as well as similar with respect to their patterns of length of stay.

This supplement presents a description of the approach implemented to construct the DRGs and a discussion of their application in a number of health care settings. Section 2 provides information regarding

† For major categories 342–349, pertaining exclusively to newborns, birthweight is used in place of age.

the data base, the statistical methodology, and the general process by which the methodology was applied to the data in the formation of the groups. At the end of Section 2 is an example illustrating the procedure, to which readers may initially refer for a general understanding of DRG construction. A discussion of the interpretation and application of the DRGs in utilization review, budgeting, cost control, prospective reimbursement and regional planning is found in Sections 4 through 6.

MEDICAL CARE
February 1980, Vol. XVIII, No. 2, Supplement

2. Construction of Diagnosis-Related Groups

THE PRIMARY OBJECTIVE in the construction of the DRGs was a definition of case types, each of which could be expected to receive similar outputs or services from a hospital. In order that the set of definitions be easily implemented in a wide range of settings as well as meaningful to medical and nonmedical users, it was considered important that the resultant classification scheme have the following attributes:

1. It must be interpretable medically, with subclasses of patients from homogeneous diagnostic categories. That is, when the patient classes are described to physicians, *they* should be able to relate to these patients and be able to identify a particular patient management process for them.

2. Individual classes should be defined on variables that are commonly available on hospital abstracts and are relevant to output utilization, pertaining to either the condition of the patient or the treatment process.

3. There must be a manageable number of classes, preferably in the hundreds instead of thousands, that are mutually exclusive and exhaustive. That is, they must cover the entire range of possible disease conditions in the acute-care setting, without overlap.

4. The classes should contain patients with similar expected measures of output utilization.

5. Class definitions must be comparable across the different coding schemes.

With respect to point 3, the number of classes considered manageable would clearly depend on the situation in which the patient classes were being applied. Any amount of aggregation involves a trade-off between loss of specificity and ease of review of statistical summaries for comparison purposes. It was decided to construct a set of approximately 500 groups to be used as the basic framework of case types. These could then be further refined or aggregated as needed by investigators depending on the nature of the application.

A number of procedures were tested as potential ways of constructing the groups. Initially, a normative approach was used, which involved having clinicians define case types using values of variables which they felt were important for determining the amount and type of services utilized. There was a tendency for their definitions to include an extensive set of specifications, requiring information which might not always be collected through a hospital's medical information system. If the entire range of patients were classified in this manner, it would ultimately lead to thousands of case types, most of which would describe patients seen infrequently at an institution. It was therefore recognized that the process of case definition would be facilitated if data from acute-care hospitals could be examined to determine the general characteristics and relative frequency of discharges. Also, statistical algorithms would be useful to suggest ways of forming patient classes that are homogeneous with respect to some aggregate output-utilization measure.

This type of approach, requiring the simultaneous inputs of physician judgment, efficient information processing and statistical algorithms, was adopted for the construction of the DRGs. An automated process was developed for screening a large amount of data with the aid of statistical algorithms. Physician review was required to insure that the classes formed by the process were medically meaningful. Length of stay (LOS) was used as the measure of output. While it may not be as accurate an indicator of the level of output as actual costs, it is still an important indicator of utilization as well as being easily available, well standardized and reliable.

The use of LOS as a measure of case complexity has been studied by other researchers. Luke in his work on case mix measurement[19] established the high degree of correlation between LOS and total charges rendered the patient. Lave and Leinhardt found significant correlation between LOS and measures of case mix complexity.[18]

The process of forming the DRGs was begun by partitioning the data base into mutually exclusive and exhaustive primary diagnostic areas, called Major Diagnostic Categories. Each major category was then examined separately and further subdivided into groups based on values of variables suggested by the statistical algorithm. Physician review of these recommended subdivisions often led to modification. Thus, at each stage of the process the subgroups were based both on statistical criteria as well as physician judgment. The precise variables that were included in class definitions varied across the major categories. For example, age was determined to be important in explaining utilization for hernia patients, but not an important factor for gastric ulcer patients. From each Major Diagnostic Category, a number of final patient classes was formed. These final patient classes are the DRGs. A more extensive discussion of the data base, the statistical algorithm, and the general strategy used in constructing the DRGs is presented in the following subsections.

2.1 Data Base

The data base used to construct the scheme contained approximately 500,000 hospital records from 118 institutions in New Jersey, 150,000 records from 1 Connecticut hospital and 52,000 records of federally funded patients from 50 institutions in a PSRO region. These records contained demographic information about each patient (e.g. sex, age) as well as clinical and diagnostic information related to his hospital stay (e.g. problems/diagnoses, surgical procedures, special services used).

Diagnostic information in the data base was coded with both classification systems, ICDA8 and HICDA2. Since there is not a direct match between the 2 schemes, data from all hospitals could not be combined in a unified data base. Thus, it was decided to construct the classification scheme using the more prevalent ICDA8 codes as the standard. The ICDA8 version was then translated to HICDA2. This translation was evaluated with hospital data and necessary modifications were made to insure the consistency of the classification across the 2 coding schemes. Both ICDA8 and HICDA2 record surgical procedures using 3-digit codes. These procedure codes cover not only operations performed but also some therapies and minor diagnostic procedures. For ICDA8, the ranges of codes that were considered to reflect actual operations were 010–999 and A10–A59. Likewise, for HICDA2, the actual code range for operations is considered to be 010–920 and 933–936. In constructing the patient classification scheme, only codes within these ranges were considered as surgical procedures.

2.2 Statistical Methodology

The particular statistical methodology employed is a variation of the Automated Interaction Detector (AID) method of Sonquist and Morgan, which has previously been applied in the analysis of complex sample survey data at the University of Michigan Survey Research Center.[25] The objective of this approach is to examine the interrelationships of the variables in the data base and to determine, in particular, which ones are related to some specified measure of interest, referred to as the dependent variable. This is accomplished by recursively subdividing the observations, through binary splits, into subgroups based on values of variables that maximize variance reduction or minimize the predictive error of the dependent variable. Subgroups are designated terminal groups when they cannot be partitioned further either because the sample sizes are too small or the remaining variation is either too low to be reduced further or unexplainable in terms of the variables in the data base. Each observation is contained in one and only one of these terminal groups, with a predicted value equal to the mean of the group. That is, if y_{kj} is the value of the dependent variable for the jth observation within the kth group, then

$$y_{kj} = \overline{Y}_k + e_{kj} \qquad (2.1)$$

where $\overline{Y}_k$ is the mean for all members in the kth group and e_{kj} is the error in using $\overline{Y}_k$ to predict or estimate y_{kj}. This procedure minimizes the sum of the $(e_{kj})^2$ over all observations. Thus, individual observations tend to have values close to the mean value of the terminal group to which they belong.

It was decided that the approach had be be implemented on an interactive basis to accommodate a high level of physician intervention. This was an important consideration since group formation using this algorithm is basically iterative in nature. Since no computer system existed that could handle large data bases efficiently in the interactive mode, a new technology was developed called AUTOGRP.[20] AUTOGRP supports a facility allowing one to invoke an algorithm that determines partitions based on the variance reduction criterion of the AID algorithm. This command, or capability, of the system is referred to as the CLASSIFY facility.

Mathematically, the algorithm can be described as follows[20]: Each observation in a data set has a value of the independent variable X and a value of the dependent variable Y. If there are N possible distinct values of the independent variable, then the subset of observations, or records, that has each value X_i $(1 \leq i \leq N)$ is called a category. If there are M_i observations in the ith category $(1 \leq i \leq N)$, the total sum of squares (TSSQ) of the data with respect to the dependent variable is defined as

$$TSSQ = \sum_{i=1}^{N} \sum_{j=1}^{M_i} (Y_{ij} - \overline{Y})^2 \qquad (2.2)$$

where Y_{ij} is the value of the dependent variable for the jth observation in the ith category of independent variable, and

$$\overline{Y} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij} \Big/ \sum_{i=1}^{N} M_i \qquad (2.3)$$

or the mean value of the dependent variable in the entire data set. The data set can be partitioned on the basis of the independent variable into G groups, where each group is the union of specified categories. That is, we can define the mapping of categories to groups with sets R_k $(1 \leq k \leq G)$, such that

$$R_k \cap R_{k'} = \emptyset, k \neq k'$$

$$\bigcup_{k=1}^{G} R_k = \{1,2,3,\ldots,N\}$$

The "within group sum of squares" (WGSSQ) is the total of the squared deviations (differences) of each group's observations from the group mean with respect to the dependent variable and can be expressed as

$$WGSSQ(k) = \sum_{i \in R_k} \sum_{j=1}^{M_i} (Y_{ij} - \overline{Y}_k)^2, \ 1 \leq k \leq G \qquad (2.4)$$

where

WGSSQ (k) = within group sum of squares for the kth group

R_k = set of all categories of the independent variable in the kth group and

$$\overline{Y}_k = \left(\sum_{i \in R_k} \sum_{j=1}^{M_i} Y_{ij} \right) \Big/ \left(\sum_{i \in R_k} M_i \right) \qquad (2.5)$$

is the mean value of the dependent variable in the kth group. The total within group sum of squares (TWGSSQ) for the G groups is the sum of the total squared deviations of each group's observations from the respective group mean and is given by

$$\text{TWGSSQ (G)} = \sum_{k=1}^{G} \sum_{i \in R_k} \sum_{j=1}^{M_i} (Y_{ij} - \overline{Y}_k)^2 \qquad (2.6)$$

For a given independent variable, the CLASSIFY algorithm partitions observations into the particular set of groups that results in the minimization of TWGSSQ for a specific dependent variable. Since TWGSSQ is proportional to the variance left unexplained by the independent variable, minimization of TWGSSQ results in the minimization of the unexplained variance of the data.

2.3 DRG Formation from Major Diagnostic Categories

To facilitate the analysis over the wide range of disease conditions in the acute-care setting all diagnoses were initially divided into 83 mutually exclusive and exhaustive Major Diagnostic Categories. Their formation was also motivated to insure diagnostic homogeneity. Thus, the final clusters do not contain patients that transcended these categories. For example, from the point of view of output utilization, it may be appropriate to form a patient class with hemorrhoids, hypertrophy of tonsils, and normal delivery. The output utilization of these patients is very similar, often requiring a relatively minor surgical procedure with a very short preoperative stay and a total hospitalization period of 2 or 3 days. However, the physicians who would treat these patients as well as the treatment processes of the problems they are presenting are quite different. Therefore, it was felt that including such patients in the same class would not define a medically meaningful category.

The specification of the Major Diagnostic Categories was performed by a committee of clinicians, following 3 general principles:

1. Major Diagnostic Categories must have consistency in terms of their anatomic, physiopathologic classification, or in the manner in which they are clinically managed.
2. Major Diagnostic Categories must have a sufficient number of patients.
3. Major Diagnostic Categories must cover the complete range of codes without overlap.

A list of these categories as defined by their ICDA8 and HICDA2 codes appears in Table 1. There is also an indication of the corresponding Professional Activity Study (PAS) diagnosis groups that correspond to each category. Note that the categories are very broad, such as Diseases of the Eye, Diseases of the Cardiovascular System and Infectious Diseases.

A consistent process was followed in partitioning each Major Diagnostic Category into DRGs. First of all, each category was refined by eliminating certain unwanted observations. Cases with dead patients or bad records, and those that were particularly deviant, were excluded from further analysis. Cases with dead patients were removed from consideration since their lengths of stay were probably atypical of the disease or problem under consideration. Records with obvious coding errors or missing data were also eliminated because their information could be misleading. Ob-

TABLE 1. Major Diagnostic Categories

Major Category	Initial Group Names	PAS Group No.	ICDA8 Codes	HICDA2 Codes
1	Infectious Diseases	1–8, 10–17	000–0689, 071–1360	001–0689 071–1360
2	Malignant Neoplasm of Digestive System	18–23	140–1590	140–1590
3	Malignant Neoplasm of Respiratory System	24–25	160–1639	160–1639
4	Malignant Neoplasm of Skin	27	172–1739	172–1739
5	Malignant Neoplasm of Breast	28	174–1740	174–1742
6	Malignant Neoplasm of Female Genital Organ	29–33	180–1849, 2340, 6211 6291	180–1849
7	Malignant Neoplasm of Male Genital Organ	34, 35	185–1879	185–1879
8	Malignant Neoplasm of Urinary System	36, 37	188–1899	188–1899
9	Malignant Neoplasm of Other and Unspecified Sites	26, 38–42	170–1719, 190–1991	179–1719, 190–1990
10	Neoplasm of Lymphatic and Hemopoietic Tissue	43–46	200–2090	200–2090
11	Benign Neoplasm of Female Genital Organ	53–56	218–2219	218–2219
12	Benign Neoplasm of Other Sites	47–51, 58–62, 64–65	210–2169, 222–2330, 2341–2399, 2552, 7434, 7571	210–2169, 2220–2221 2228–2264, 2266–2399
13	Diseases of Thyroid and Other Endocrine Glands	63, 66, 70, 71	240–2460, 251–2551., 2559–2589	240–2460, 251–2589 2265
14	Diabetes	67–69	250–2509	250–2507
15	Nutritional and Other Metabolic Diseases	72–75	260–2790	260–2790
16	Diseases of Blood and Blood Forming Organs	76–80	280–2890, 2894–2899, 6345	280–2899
17	Psychoses Not Attributed to Physical Conditions	86–92	295–2990	306–3099
18	Neuroses	93–95	300–3029	310–3129
19	Alcoholic Mental Disorder and Addiction	84, 96, 97	291–2919, 303–3039	302–3029, 313–3139
20	Other Mental Disorders	81–83, 85, 98–103	290–2901, 292–2949, 304–3159	290–2959, 296–3019, 303–3059 314–3189
21	Diseases of Central Nervous System	104–108	320–3499	320–3499
22	Diseases of Peripheral Nervous System	109–110	350–3580, 3589	350–3589
23	Diseases of Eye	111–121	360–3789	360–3789
24	Diseases of Ear and Mastoid Process	122–125	380–3879	380–3892
25	Hypertensive Heart Diseases	128, 129	400–4040	400–4050
26	Acute Myocardial Infarction	130	410–4109	410–4109
27	Ischemic Heart Diseases except AMI	131, 132	411–4149	411–4140
28	Arrhythmia and Slowed Conduction	133	3581, 4272–4279	415–4169
29	Heart Failure	135	4270–4271, 7824	427–4279
30	Carditis, Valvular, and Other Diseases	126, 127, 134, 136	390–3980, 420–4260, 428–4299	390–3980, 420–4269, 429–4299
31	Cerebrovascular Diseases	137–141	430–4389	430–4389
32	Diseases of Vascular System	142–147, 150, 152–154	2891–2893, 440–4431, 4438–4480,	440–4489, 452–4549, 456–4589

T 3LE 1. Continued.

Major Category	Initial Group Names	PAS Group No.	ICDA8 Codes	HICDA2 Codes
			452–4549, 456–4589	
33	Pulmonary Embolism	148	450–4500	450–4500
34	Phlebitis and Trombophlebitis	149	451–4519	451–4519
35	Hemorrhoids	151	455–4550	455–4559
36	Hypertrophy of Tonsil and Adenoid	163	500–5000	500–5000
37	Acute URI and Influenza	155, 156	460–4650, 470–4741	460–4650, 470–4700
38	Other Diseases of Upper Respiratory Tract	164–167	501–5089	501–5089
39	Pneumonia	157	480–4860	480–4860
40	Bronchitis	158, 159	466–4660, 490–4910	489–4919
41	Asthma	161	493–4930	493–4939
42	Other Lung and Pleural Diseases	160, 162, 168, 169	492–4920, 510–5199	492–4920, 494–4960 510–5199
43	Diseases of Oral Cavity, Salivary Glands and Jaws	170	520–5299	520–5299
44	Gastric and Peptic Ulcer	172–175	531–5349	531–5343
45	Upper GI Diseases except Gastric and Peptic Ulcer	171, 176, 177	530–5309, 535–5379	530–5309, 535–5379
46	Appendicitis	178–180	540–5430	540–5430
47	Hernia of Abdominal Cavity	181–184	550–5539	550–5539
48	Enteritis, Diverticula, and Functional Disorder of Intestine	185, 186	561–5649	561–5649
49	Diseases of Anus	187, 253	565–5660, 685–6850	565–5660, 685–6850
50	Miscellaneous Diseases of Intestine and Peritoneum	188	560–5609, 567–5699	560–5609, 567–5699
51	Diseases of Liver	9, 189, 190	070–0700, 0705–0709,9992 570–5739	070–0709, 570–5739
52	Diseases of Gallbladder and Bile Duct	191, 192	574–5769	574–5769
53	Diseases of Pancreas	193	577–5779	577–5779
54	Diseases of Kidney and Ureter	194–197, 200	580–5910, 593–5935, 792–7920	580–5910, 593–5939
55	Urinary Calculus	198, 199, 201	592–5920, 594–5940	592–5921, 594–5949
56	Cystitis and Other Urinary Diseases	202–204	595–5999	595–5999
57	Diseases of Prostate	206, 57, 205	600–6020	2222, 600–6029
58	Diseases of Male Genital Organs	207–210	603–6050, 607–6079	603–6079
59	Diseases of Female Genital Organs	213–220	612–6210, 6212–6270, 6290, 6292, 6294, 6296–6299	612–6299
60	Diseases of Breast	52, 212, 211	217–2170, 610–6119	217–2179, 610–6119
61	Abortion	228–231	640–6459	640–6469
62	Obstetrical Diseases of Antepartum and Puerperium	221–227,250	630–6344, 6346–6399 670–6730, 6739–6780	631–6399, 670–6789
63	Normal Delivery	232	650–6500	650–6500
64	Delivery with Complication	233–249	651–6620,6731	651–6649

TABLE 1. Continued.

Major Category	Initial Group Names	PAS Group No.	ICDA8 Codes	HICDA2 Codes
65	Diseases of Skin and Subcutaneous Tissue	251, 252, 254–255	680–6840, 686–7099	680–6840, 686–7099
66	Arthritis	256–258	710–7150	710–7150
67	Derangement and Displacement of Intervertebral Disc	265, 266	725–7259	725–7259
68	Diseases of Bone and Cartilege	261–264	7171–7180, 720–7249	7171–7180, 720–7249
69	Other Diseases of Musculo-Skeletal System	259, 260, 267 268–271	716–7170, 726–7389	716–7169, 719–7199, 726–7390
70	Congenital Anomalies	272–282	740–7433, 7438–7570, 7572–7599	740–7599
71	Normal Mature Born	342, 343	Y20–Y209, Y22–Y239, Y26–Y279	Y20–Y209, Y22–Y229 Y24–Y249
72	Certain Diseases and Conditions Peculiar to Newborn Infants	283, 344–349	Y21–Y219, Y24–Y259, Y28–Y299, Y005, 760–7799	Y21–Y219, Y23–Y239 Y25–Y299, Y40–Y489, 760–7689
73	Symptoms and Signs Referable to Nervous, Respiratory, and Circulatory Systems	284–287	4432, 780–7808, 7814–7815, 7817–7823, 7825–7834, 7836–7837	770–7709, 773–7769, 776–7799
74	Symptoms and Signs Referable to GI and Urinary System	288–290	784–7865	780–7839
75	Miscellaneous Signs, Symptoms, and Ill-defined Conditions	291–295	606–6060, 628–6280, 6293, 6295, 7810–7813, 7816, 7835, 7866–7889, 790–7910, 793–7969	771–7728, 777–7779, 784–7969
76	Fractures	296–310	800–8299	800–8299
77	Dislocation and Other Musculo-Skeletal Injury	311–316	830–8480	830–8489
78	Internal Injury of Cranium, Chest, and Other Organs	317–320, 326 332	850–8699, 950–9599, 9953, 9954	850–8699, 900–9049, 950–9599
79	Open Wound and Superficial Injury	321–325, 327–330	870–9390 996–9969	870–8979, 910–9391
80	Burn	331	940–9499	940–9498
81	Complication of Surgical and Medical Care	338, 339	997–9991, 9993–9999	996–9999
82	Adverse Effects of a Certain Substance	333–337	960–9952, 9955–9959	960–9959
83	Special Admissions and Examinations Without reported Diagnosis	340, 341	Y00–Y004, Y006–Y159, 379–3793, 388–3899, 789–7899	Y00–Y199, Y50–Y896

servations with disproportionately high values of length of stay were excluded since a few deviant records could have a marked effect on the stability of a group's distribution.

The screened set of records in each category was then used as input to the second stage, in which the CLASSIFY algorithm was applied to suggest groups of observations, on the basis of prespecified independent variables, that may be different with regard to length of stay. The set of independent variables selected as input to the algorithm was intentionally limited to those variables descriptive of the patient, his disease condition and his treatment process that would be readily accessible on most discharge abstracts, specifically diagnoses, surgical procedures, age, sex and clinical service. This constraint was applied for several reasons. First, mean lengths of stay are observed to vary across levels of these variables in descriptive statistical summaries of hospital discharge data.[23] Second, these variables are always recorded and entered in almost all hospital information systems. This increases the classification's potential for implementation in most research and applied healthcare settings. Including other items of information such as ancillary services used would limit its applicability to those systems that collect such data, which would exclude, for example, the PSRO PHDDS data base. Finally, restricting the variables to this set also simplifies the class definitions and controls to some extent the final number of categories. A classification with numerous groups and a complex definitional structure is unmanageable.

For each independent variable, output from the CLASSIFY procedure included the total number of observations in the data set, the total number of different values the variable assumed (i.e., number of categories or cells), the number of groups formed, the total sum of squares (TSSQ) and the per cent reduction in total sum of squares attained by such a grouping ((TSSQ-TWGSSQ) / TSSQ)* 100. Once these results were obtained, a clinician selected the most appropriate variable for division.

The interpretation of the partitioning suggested by the algorithm was a complex task, with many factors examined and weighed simultaneously. The decision to accept, to reject or possibly to revise the recommended partitioning was based on both the statistical evidence and the clinician's medical knowledge. The statistical results were examined in light of certain criteria. Variables yielding the highest percentage reduction in variance were prime candidates for dividing the data set. However, the number of cells or values for those variables and the number of groups formed were also considered. It is an artifact of the algorithm that many partitions can be created when the independent variable has a wide range of values. Moreover, too many groups formed at the first split become difficult to manage and are of questionable significance. For example, secondary diagnosis often had many different values and thus often produced significant variance reduction by forming many subgroups. However, such groups were difficult to interpret as a primary partition and tended to be of limited value. In all cases, groups were further examined in a more descriptive framework to determine if the statistical significance was supported by medical interpretability.

The fact that a variable that appeared to be powerful in explaining variance was not selected at a particular stage does not mean that the variable was ignored. Two possibilities still existed. If that variable were independent of other variables in explaining variance, it would appear in subsequent stages with the same power as at the earlier stage. If, however, it were correlated with a variable chosen at an earlier stage, then its explanatory power at subsequent stages would be lessened according to the strength of the correlation. For example, if secondary diagnosis was strongly related to age and age was selected as a partitioning variable, then secondary diagnosis would not appear powerful in subsequent clustering.

Groups were then generated based on the most appropriate variable, that is, the one that met as many of the criteria specified above as possible. In particular, it 1) exhibited a significant reduction in variance relative to most of the other variables, 2) created a manageable number of groups based on the relatively small number of values of the independent variable, and 3) created groups whose means were significantly different. Also, groups formed were homogeneous from a clinical perspective.

Once each Major Diagnostic Category was initially partitioned into subgroups based on the values of an independent variable, a decision was made whether or not to further subdivide each subgroup based on any of the other available variables or to end the partitioning process by treating them as terminal groups. The statistical basis for this decision was determined by a set of stopping rules. For any given group, the partitioning ceased when either one of the following conditions was met:

1. The group was not large enough to warrant another classification, that is, when the number of observations in the group was less than 100.

2. None of the variables reduced unexplained variation by at least 1%, or ((TSSQ-WSSQ)/TSSQ)*100 < 1 per cent.*

Otherwise, the group was further subdivided according to the criteria discussed previously for generating new subgroups. In some cases, however, the process was terminated for nonstatistical reasons regarding overall manageability (e.g. maintaining a low number of total groups) or medical interpretability.

* This 1 per cent bound was increased in certain Major Diagnostic Categories.

This grouping process resulted in the formation of 383 final groups or DRGs, each defined by some set of the following patient attributes: primary diagnosis, secondary diagnosis, primary surgical procedure, secondary surgical procedure, age, and in one case, clinical service area. While other variables such as sex, tertiary diagnosis or surgical procedure were examined, they were not found to be significant in explaining output utilization. A list of these groups with a brief narrative description of their contents appears in the Appendix. A more complete specification can be obtained from the Health Care Financing Administration.[1]

The DRGs vary considerably in their structure across the Major Diagnostic Categories. Some Major Diagnostic Categories are not further subdivided, such as Category 35, Hemorrhoids, in which no variable demonstrated a sufficient effect in further explaining output utilization. On the other hand, Appendicitis, Category 46, is further subdivided on the basis of specific primary diagnosis and the presence of a secondary diagnosis. This results in 4 DRGs: appendicitis (without peritonitis) and without a secondary diagnosis, appendicitis (without peritonitis) with a secondary diagnosis, appendicitis (with peritonitis) without a secondary diagnosis, and appendicitis (with peritonitis) with a secondary diagnosis. This symmetric breakdown suggests that the effects of primary diagnosis and the presence of a secondary diagnosis are additive in nature. Major Diagnostic Category 76, Fractures, has the most complex structure, resulting in 13 DRGs, indicating both the importance and interaction of 4 variables: primary diagnosis, secondary diagnosis, primary surgical procedure and age.

It should also be noted that when variables are highly correlated, very often only one appears in the classification for a specific major category. An extreme example of this is Major Diagnostic Category 36, Hypertrophy of Tonsils and Adenoid,

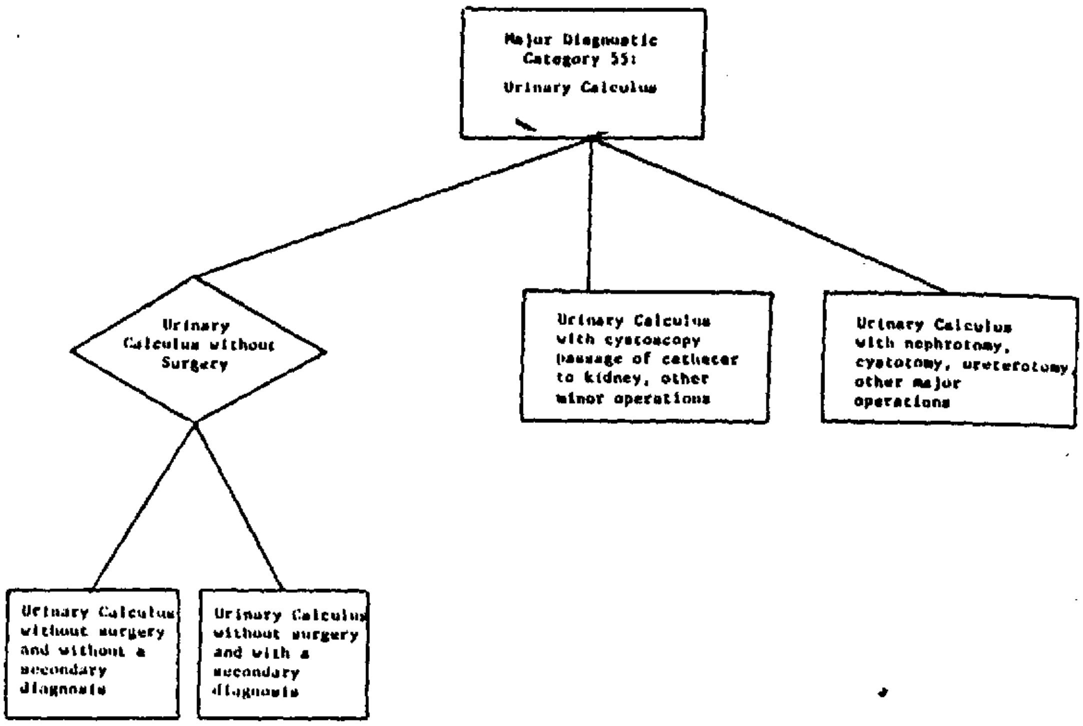


FIG. 1. Tree diagram illustrating partitioning of urinary calculus patients.

where almost everyone has a tonsillectomy and/or an adenoidectomy. The high correlation of primary surgical procedure with primary diagnosis results in no further variance reduction in the category that can be attributed to surgical procedures. Thus, although surgery is almost always used in the treatment of patients with Hypertrophy of Tonsils and Adenoids, since virtually every patient in this category had the same surgical procedure, the surgical information did not differentiate the utilization of these patients and was therefore not used in the formation of the DRGs for this category.

2.4 An Example

The iterative partitioning process used in forming the DRGs can best be illustrated in the context of an example—the classification of Major Diagnostic Category 55: Urinary Calculus. This category contains patients with a primary *diagnosis* (ICDA8 codes) of either 592, calculus of kidney and ureter, or 594, calculus of other parts of the urinary system. The formation of the DRGs from this Major Diagnostic Category is summarized in the tree diagram presented in Figure 1. First, this category is partitioned into 3 groups based on the variable primary surgical procedure. The first group contains nonsurgical patients, which are those with either no operation or with a procedure code (ICDA8) outside the range 010–999, A10–A59.† The second and third groups are formed on the basis of the specific procedure performed. In particular, the more complicated procedures performed on patients with a urinary calculus—nephrotomy, ureterotomy, cystotomy—are in the third group, while

† Operations coded outside these ranges are not considered actual surgical procedures since they represent minor procedures or therapies.

TABLE 2. Descriptive Statistics for the Partitioning of Major Diagnostic Category 55 (Urinary Calculus).

Size = 1425 Mean = 6.93 S.D. = 6.44

Independent Variables	Partial Variance Explained	DRG No.	Size	Mean Length of Stay	S.D.
Primary surgery	41.75				
Minor		241	428	6.36	4.30
Major		242	286	14.99	7.39
Secondary diagnosis, with no primary surgery	1.17				
None		239	449	3.28	2.88
1 or more		240	262	5.32	5.01

Total variation explained: 42.93.

relatively minor procedures associated with this diagnosis—cystoscopy, passage of catheter to kidney—are contained in the second. The nonsurgical group is partitioned further into 2 groups based on the presence or absence of a secondary diagnosis. In summary, the classification process resulted in the formation of 4 terminal groups, or DRGs 239–242, from the Major Diagnostic Category of Urinary Calculus:

239 Urinary calculus without surgery, and without a secondary diagnosis

240 Urinary calculus without surgery and with a secondary diagnosis

241 Urinary calculus with cystoscopy, passage of catheter to kidney, other operations

242 Urinary calculus with nephrotomy, cystotomy, ureterotomy, other major operations

A descriptive statistical summary of data coded in ICDA8 from the original data base used to construct the DRGs is presented in Table 2. The entire Major Diagnostic Category contains 1,425 observations, with a mean length of stay of 6.93 and a standard deviation of 6.44. The variables used in partitioning this group, primary surgery and secondary diagnosis, explain 42.93 per cent of the total variance, with 41.75 per cent attributed to the former and 1.17 per cent to the latter.

The actual process of forming these DRGs from the Major Diagnostic Category Urinary Calculus is summarized in the following steps:

Step 1

Fifteen records were eliminated: 3 with a discharge status of death, 10 with invalid surgical or diagnosis codes, and 2 with lengths of stay greater than 60 days. This reduced the size of the category from 1,440 to 1,425 observations.

Step 2

The algorithm was invoked on this refined data set to determine the basis for an initial split. The independent variables selected to define potential subgroups were primary surgical procedure (oper1),

TABLE 3.

Variable	No. of Groups	Per Cent Reduction
oper1	3	41.89
oper2	4	21.37
dx1	1	0.0
dx2	5	30.11
age	3	8.19
sex	2	1.63

TABLE 4. Suggested Partitioning (3 groups) of Urinary Calculus Patients on the Basis of Type of Primary Surgery

Group	Size	Mean	Independent Variable	Description
1	1	2.00	749	Other antepartum procedures to terminate pregnancy
	1	2.00	571	Meatotomy
	1	2.00	277	Venous anastomosis, intra-abdominal
	1	2.00	249	Other operations on peripheral vessels
	1	3.00	430	Incision of bile (hepatic) ducts
	1	3.00	862	Arthrocentesis
	1	3.00	601	Vasectomy
	1	3.00	921	Local excision of lesion of skin and subcutaneous tissue
	23	3.28	000	No code
	688	4.08		No code
	1	5.00	551	Ureterectomy
2	2	5.50	574	Repair and plastic operations on urethra
	7	5.71	559	Other operations on ureter
	218	6.25	A46	Cystoscopy and urethroscopy without effect upon tissue
	5	6.40	A45	Endoscopy of colon and rectum without effect upon tissue
	15	6.40	568	Removal of calculus and drainage of bladder without incision
	2	6.50	572	Excision or destruction of lesion of urethra
	146	6.59	557	Passage of catheter to kidney
	21	7.14	575	Dilation of urethra
	1	9.00	A16	Biopsy of thorax
3	2	10.00	A21	Biopsy (continued) of urinary tract
	2	11.50	566	Repair and other plastic operations on bladder
	1	12.00	A44	Esophagoscopy and gastroscopy without effect upon tissue
	1	13.00	549	Other operations on kidney
	1	13.00	556	Repair and plastic operations on ureter
	3	13.67	561	Local excision and destruction lesion of bladder, transurethrol
	[illegible]	14.00	562	Local excision and destruction of lesion of bladder, other
	3	14.00	582	Prostatectomy, transurethral
	1	14.00	583	Prostatectomy, other
	8	14.13	545	Nephrectomy, complete
	72	14.46	541	Pyelotomy
	40	14.47	560	Cystotomy
	101	14.63	550	Ureterotomy
	19	15.89	540	Nephrotomy
	1	16.00	513	Hemorrhoidectomy
	11	16.82	544	Nephrectomy, partial
	1	17.00	546	Repair and plastic operations on kidney
			570	Urethrotomy, external
	1	21.00	A27	Biopsy of bone
	1	21.00	563	Cystectomy, complete or partial
	1	22.00	685	Ligation and division of fallopian tubes, bilateral
	3	22.33	558	Ureterolysis
	1	29.00	543	Local excision and destruction of lesion of kidney

secondary surgical procedure (oper2), primary diagnosis (dx1), secondary diagnosis (dx2), age, and sex.

The number of groups formed by the algorithm and the corresponding per cent reduction in unexplained variation for each of the variables are shown in Table 3.

Since the greatest reduction in unexplained variation was achieved with oper1, and a limited number of groups (3) this variable was considered the prime candidate for initial subdivision of the category. The algorithm suggested 3 groups whose contents are described in Table 4. This table presents the different surgical procedures contained in each group, the corresponding number of observations (SIZE), and the mean length of stay (MEAN). Note that more than 98 per cent of the observations in the first group have no surgical procedure listed. The second group primarily contains observations with relatively minor procedures such as cystoscopy and urethroscopy (A46) and passage of catheter to kidney (557), while the third group includes somewhat more complex procedures as ureterotomy (550), cystotomy (560), and pyelotomy (541).

On the basis of these results, it was decided to divide the initial group of Urinary Calculus patients into 3 groups, similar to those suggested by the algorithm, namely, a group of nonsurgical patients, a group with relatively major procedures such as those listed under group 3 in Table 4, and finally, a group of all other procedures, which includes cases with minor procedures such as those listed under groups 1 and 2, and biopsy of urinary tract (A21) in group 3. While this latter group represents all other surgical procedures not explicitly listed under group 3, it is primarily represented by the 2 procedures cystoscopy and urethroscopy (A46) and passage of catheter to kidney (557).

Step 3

Each of the groups formed in Step 2 was then considered for further subdivision.

TABLE 5.

Variable	No. of Groups	Per Cent Reduction
dx1	1	0.0
dx2	4	22.66
age	4	14.18
sex	1	0.0

First of all, with respect to the nonsurgical patients, the number of groups formed by the algorithm and the corresponding per cent reduction in unexplained variation for each of the variables (except oper1 and oper2) were as shown in Table 5.

A closer examination was made of the characteristics of the 4 groups formed using the variable secondary diagnosis (dx2), since it exhibited the greatest per cent reduction in unexplained variation. The descriptive statistics for each of the groups are summarized in Table 6.

Several things were considered in evaluating the potential partitioning on secondary diagnosis. With respect to the distribution of observations, groups 3 and 4 were definitely too small (i.e., less than 100 observations) to be considered terminal groups and group 2 with 109 observations was marginal. Further, it was noted that more than 80 per cent of the observations in Group 1 had no secondary diagnoses listed and that the remaining cases in all 4 groups were distributed across 105 different secondary diagnosis codes, usually with less than 10 cases represented for each disease and with no apparent clinical pattern. Thus, it was decided that groups formed on the basis of specific secondary

TABLE 6.

Group	No. Observed	Mean	S.D.
1	534	3.22	2.71
2	109	4.87	2.70
3	50	7.68	4.93
4	18	12.83	12.37

Table 7.

Group	No. Observed	Mean	S.D.
No secondary	449	3.28	2.88
Secondary	262	5.32	5.01

diagnosis were not particularly meaningful, but that a more manageable and interpretable partition from a medical perspective would be 2 groups, based on the presence or absence of a secondary diagnosis. The descriptive statistics of these groups are shown in Table 7.

This alternative partition results in a markedly lower per cent reduction in unexplained variation—6.3 per cent lower. But, in terms of the overall objectives of the classification process, the increase in interpretability and manageability was considered more important than the sacrifice in predictive error.

Step 4

With respect to the other 2 groups formed in Step 2 on the basis of specific surgical procedure, the algorithm was applied using the variables secondary surgical procedure, primary diagnosis, secondary diagnosis, age, and sex. For the group with minor operations, the number of subgroups formed by the algorithm and the corresponding per cent reduction in unexplained variation for the variables are shown in Table 8.

Likewise the partitions with respect to these variables suggested for the group of

Table 8.

Variable	No. of Groups	Per Cent Reduction
oper2	2	13.36
dx1	1	0.0
dx2	4	34.62
age	2	4.73
sex	1	0.0

Table 9.

Variable	No. of Groups	Per Cent Reduction
oper2	3	18.36
dx1	2	1.26
dx2	4	43.03
age	2	3.85
sex	1	0.00

relatively major procedures have the characteristics shown in Table 9. In both cases it appeared that secondary diagnosis had the strongest effect and was selected as the potential variable to use in forming subgroups. However, after examining the contents of the suggested groups, it was found in both instances that at least half the observations had no secondary diagnosis listed and the others had secondary diagnoses distributed across at least 100 different codes, with no apparent clinical consistency. That is, the diagnoses were dissimilar and few were represented by more than 10 cases. Thus, like the nonsurgical cases discussed in Step 3, it did not appear that further division of these groups into subsets by specific secondary diagnosis was meaningful from a clinical perspective.

Partitioning each group on the basis of the presence or absence of secondary diagnosis was considered. This would achieve a 2.1 per cent reduction in unexplained variation for the minor surgical group and a 5.6 per cent reduction for the major surgical group. In both instances, it was decided that there was not sufficient medical justification for a further break-

Table 10.

Variable	No. of Groups	Per Cent Reduction
age	2	2.73
oper2	2	2.06
dx1	1	0.0
sex	1	0.0

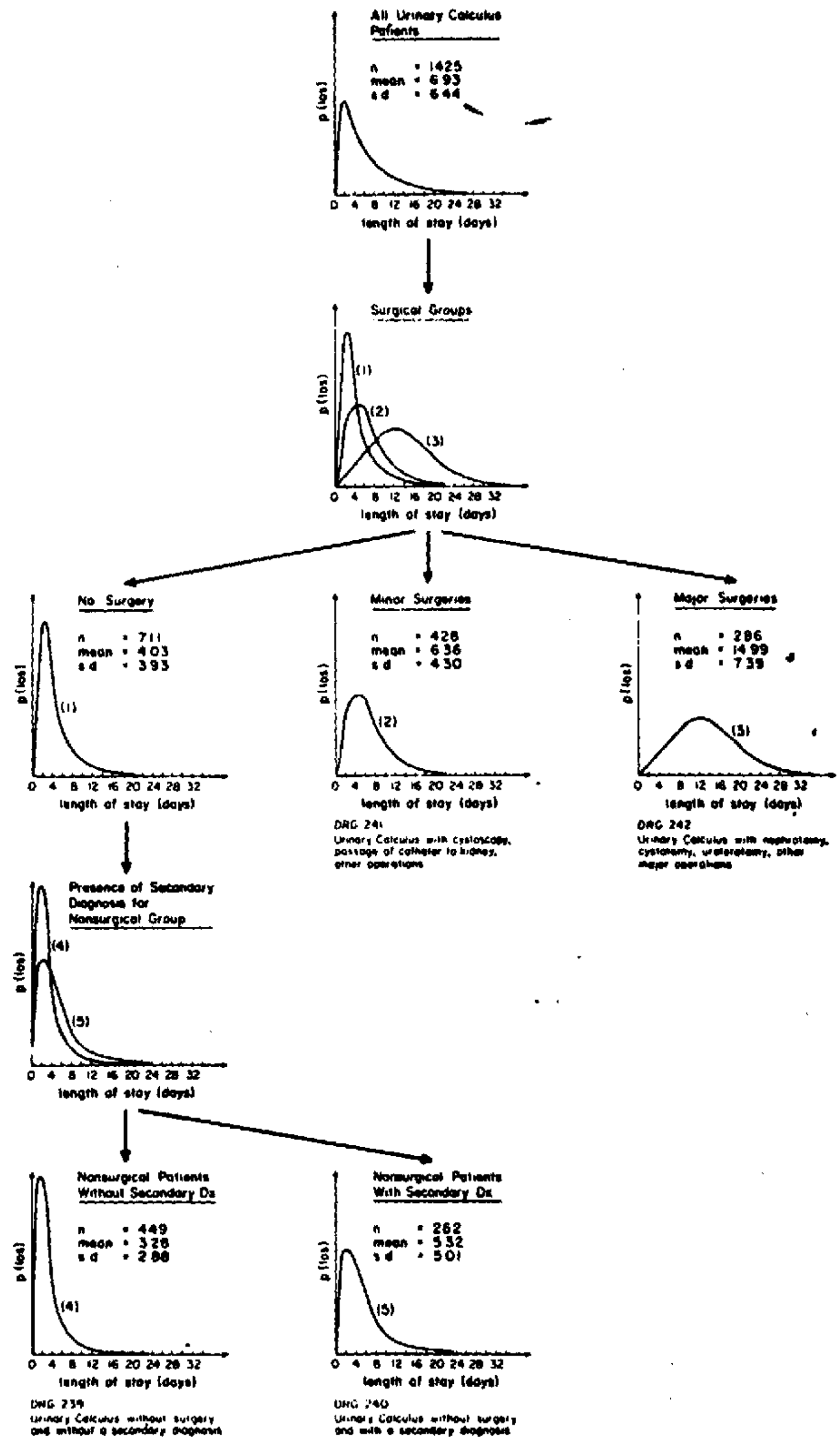


FIG. 2. Summary of length-of-stay distributions for groups formed in partitioning process.

TABLE 11.

Variable	No. of Groups	Per Cent Reduction
age	3	13.05
oper2	1	0.0
dx1	1	0.0
sex	1	0.0

down of the surgical groups on the basis of secondary diagnosis. Moreover, in light of one of the major objectives of keeping the total number of classes low, additional groups formed at this stage of the partitioning of Urinary Calculus patients would be of questionable value. Therefore, the 2 surgical groups were not divided further but were considered terminal groups.

Step 5

The 2 subgroups formed from the nonsurgical cases on the basis of presence or absence of other diagnoses were evaluated to determine if they should be partitioned further or left intact as terminal groups. The algorithm was applied and produced the results shown for the nonsurgical cases without multiple diagnoses. The algorithm produced the results listed in Table 11 for the nonsurgical cases with multiple diagnoses.

With respect to the nonsurgical cases without multiple diagnoses, both sets of groups form-d on the basis of age and secondary su...cal procedure, respectively, were determined unacceptable. In each instance, more than 95 per cent of the observations fell into the first group, leaving the second group with fewer than 25 cases.

For the nonsurgical cases with multiple diagnoses, the 3 groups formed using age levels were considered as potential subgroups. The age levels defining the boundaries of the groups were 66 and 70. This partition was rejected for reasons similar to those above, namely, the lopsided distribution of cases in the groups. Almost 90 per cent of the observations had an age under 66.

Thus, the nonsurgical groups with and without multiple diagnoses were considered terminal groups.

We conclude, then, that specific surgical procedures and the presence of multiple diagnoses were important variables in predicting length of stay for urinary calculus patients. The 4 DRGs formed were significantly different ($\alpha = 0.01$) with respect to their average lengths of stay and are clinically interpretable. To be sure, by overruling some of the partitions suggested by the algorithm, a certain amount of explanatory power was sacrificed. But, the trade-off was generating a reasonable number of subgroups or DRGs which could be interpreted from a medical perspective. Figure 2 presents a descriptive summary of the length-of-stay distributions for the groups formed as part of the partitioning process in this example.

MEDICAL CARE
February 1980, Vol. XVIII, No. 2, Supplement

3. Interpretation of the DRGs in Health-Care Studies

THE DRGs AS CONSTRUCTED met the original research objectives; that is, they provide a definition of case types by relative output utilization, where the definition of case type has the following properties:

1. The variables used in the definition of a case are limited to those referring to the patient's condition and treatment process that are commonly collected in a hospital information system, namely primary diagnosis, secondary diagnosis, age, primary surgical procedure, and secondary surgical procedure.

2. There are a manageable number of case types or DRGs (383) and they are mutually exclusive and exhaustive.

3. Each DRG has clinical interpretability with closely related diagnoses and operations.

4. Patients within a DRG have similar patterns of utilization as measured by length of stay.

Achieving these objectives meant in all cases a trade-off between maximizing explained variation and maintaining clinical interpretability. For example, there were some Major Diagnostic Categories which could have been partitioned into additional groups based on one of the variables used or the introduction of other variables. While this may have resulted in increased explained variation, the amount of increase was balanced against the potential loss in 1) medical interpretability, 2) the manageability of the resultant number of groups, and 3) the simplicity of case definition.

Since the statistical algorithms were primarily used as a screening device to suggest groups of patients with similar output utilization patterns, the process is to some extent subjective and dependent on the clinician making the decisions, the independent variables he was constrained to use, the dependent variable chosen to represent output utilization (length of stay), and the data base examined. Thus, individuals analyzing their own data within this framework may decide other definitions are more meaningful and significant for their purpose. For example, if cost information is readily available and well standardized, that would be a preferable alternative to length of stay as the output utilization measure. Likewise, additional independent variables may be incorporated, but the analyst should be aware of potential difficulties in this regard. Greatly expanding the set of variables that may be incorporated in a scheme tends to increase the number of final groups and restricts its application to within the immediate setting.

As they are now defined, the DRGs form a manageable, medically interpretable set of case types that allows one to control for differences in complexity attributable to patient characteristics as described by age, primary diagnosis, secondary diagnosis, primary surgical procedure and secondary surgical procedure. On the basis of values for these variables, practitioners can gain some understanding of the patient being identified and specify within reasonable limits expected services to be delivered, criteria to be applied in treatment, and expected outcomes. This is not to say that providers of such care will always agree on the treatment process or that all will agree with the identification. There are certainly disease entities that are not as well understood as others and treatment protocols for which there is lesser or greater consensus among practitioners. For example, senile cataracts are almost always treated by a surgical lens extraction but there is still observed great variation in the case regime and length of hospital stay for this procedure. Thus, while senile cataract patients with lens extraction form a patient class, differences among providers still remain with respect to length of stay and other output measures which cannot be accounted for by patient characteristics or surgical procedures.

The development of the patient classes is not intended to inhibit in any way the practice of medicine but to offer one the capability of examining reasons for variations in service utilization, treatment process, and outcome. In this context, the groups can provide a framework for the initiation of an ongoing process of comparative analysis of health care with the long-run goal of determining both the cost and value of any kind of care that might be delivered. With such information, meaningful dialogue among clinicians, administrators, planners and regulators can proceed in rationalizing of observed differences. Only in this way can strategy, policy and politics interact to the benefit of the communities served by each institution.

The classification of patient records into DRGs is a constantly evolving process. In fact, the group structure described here represents the third classification scheme developed using the methodology presented in the previous section. As coding schemes change and data are collected that are more current and representative of acute-care institutions in the United States, these groups will be re-examined and revised accordingly.

Currently the DRGs are being applied in a number of different practical and research settings in the health-care field. While these applications are discussed extensively elsewhere, a brief overview is presented in the following sections to give the reader a better understanding of the DRGs' potential utility in a variety of areas.

MEDICAL CARE
February 1980, Vol. XVIII, No. 2, Supplement

4. Applications of the DRGs in Utilization Review

4.1 Case Mix Analysis of Hospital Utilization

IN COMPARING HOSPITAL PERFORMANCE on the basis of patient-care–related measures such as length of stay, cost, and death rate, it is important to determine the extent to which observed differences can be attributed to case mix and to what extent they are related to differing treatment practices. An individual institution has limited control over the former, and one expects utilization and quality-of-care measures to vary across the different types of cases it treats. For example, while a 15 per cent mortality rate is not unusual for acute myocardial infarction patients in most inpatient acute-care facilities, it would be alarming for women with normal deliveries. Likewise, a 2-day stay for tonsillectomies is typical, but it is unusually short for appendectomies. Thus, a hospital with a higher proportion of relatively complex cases could be expected to have, on the average, longer lengths of stay, increased costs, and higher death rates. Any comparison, then, of this institution with another on the basis of such measures must take into account its more complex case mix.

As stated in the introduction, one of the original motivations for the development of the DRGs was to provide a framework of case types that reflects the relative complexity of patients seen at an institution. Given such a framework, case mix–adjusted measures may be computed to control for differences in case mix and used in comparing acute-care inpatient facilities or groups of facilities. In addition, observed differences in some variable between 2 hospitals or a hospital and a set of hospitals may be partitioned into 3 components: 1) the amount of difference attributed to hospital treatment practices; 2) the amount of the difference attributed to hospital case mix; and 3) the amount of the difference attributed to the interaction of hospital treatment practices *and* case mix. This information allows one to make comparisons of hospital utilization and quality of care taking into account case composition. A case mix analysis report has been developed and is currently being applied in several PSRO settings as part of their evaluation of institutional length-of-stay patterns.[4]

4.2 Case Mix Report

Table 12 contains a sample report presenting case mix–related statistics for a set of hospitals in a given region. Of particular interest here is a comparison of their inpatient utilization as measured by length of stay in days. First of all, as noted at the top of the report, all patients in the region, regardless of institution or case type, experienced an average stay of 6.772 days. The report's objective is to examine each hospital's utilization in light of both case mix and length-of-stay patterns and to compare it to the others using the region as the standard reference. The precise mathematical formulation of the statistics contained in the report is presented elsewhere,[4] but a brief descriptive summary is presented here in order to demonstrate the application of DRGs in case mix analyses. In addition, the reader is referred to Hill[11] and Kitagawa[14] for a further discussion of the 2 techniques applied in this context, standardization and separation into components.

A hospital identifier appears in the first column followed by the corresponding number of observations (OBS) or records from that institution used in the computation of the statistics in the remaining columns. Column 3 contains the overall average length of stay (AVERAGE LOS) for patients in that institution, which is the

TABLE 12. Sample Case Mix Report for 8 Institutions Representing a Given Region

(1) Hospit.	(2) OBS	(3) AVERAGE LOS	(4) CASE MIX ADJUSTED LOS	(5) LOS WEIGHTED CASE MIX	(6) AVERAGE LOS INDEX	(7) LOS INDEX	(8) CASE MIX INDEX	(9) AVERAGE LOS DIFFER	(10) DIFFER DUE TO LOS	(11) DIFFER DUE TO CASE MIX	(12) INTERACT DIFFER
1	12666	7.580	6.850	7.433	1.119	1.012	1.098	.808	.079	.661	.068
2	24049	6.533	6.591	6.643	.965	.973	.981	-.239	-.180	-.129	.070
3	14117	6.798	6.256	6.999	1.004	.924	1.034	.026	-.516	.227	.315
4	41643	7.479	7.320	6.960	1.105	1.081	1.028	.708	.548	.188	-.029
5	5499	6.395	6.068	7.072	.944	.896	1.044	-.377	-.704	.300	.027
6	8167	6.256	6.585	6.398	.924	.972	.945	-.516	-.187	-.374	.045
7	2828	6.493	6.434	7.230	.959	.950	1.068	-.279	-.338	.458	-.399
8	20494	4.943	5.457	5.785	.730	.806	.854	-1.828	-1.315	-.986	.473

See text for explanation.

number of bed days consumed per discharge. A case mix–adjusted length of stay (CASE MIX ADJUSTED LOS) appears in column 4 and is the mean length of stay the hospital would have experienced with the region's case mix. For the ith hospital, it is computed as

$$\sum_j a_{ij} P_j, \quad j = 1, \ldots, 383 \text{ (DRG)} \quad (4.1)$$

where a_{ij} is the average stay in the jth DRG in hospital i, and P_j is the proportion of the region's cases in DRG j. As such, it standardizes or holds constant case mix across all institutions and results in a measure that one can use to compare hospital utilization based on length of stay.

As a complement to the case mix–adjusted length of stay, column 5 contains the average length of stay the hospital would have experienced with its own case mix, but the region's average length of stay within each of the DRGs (LOS WEIGHTED CASE MIX). That is, for the ith hospital it is computed as

$$\sum_j A_j p_{ij}, \; j = 1, \ldots, 383 \text{ (DRG)} \quad (4.2)$$

where A_j is the region's average stay in the jth DRG, and p_{ij} is the proportion of the ith hospital's cases in the jth DRG. Since it uses a standard set of relative weights (A_j) for the DRGs, it can be interpreted as a relative measure of case mix complexity, with higher values indicating a more complex case mix, that is, a greater proportion of cases receiving a high amount of the hospital outputs. In interpreting this measure, it should be noted that complexity is a function of both the condition of patients as well as the treatment process selected by physicians. Columns 6 through 8 are indices computed by dividing the measures in columns 3 through 5, respectively, by the region's average length of stay.

Columns 9 through 12 present the observed difference between the institution's and region's average length of stay (AVERAGE LOS DIFFER) and indicate to what extent this difference can be attributed to DRG-specific average length of stay or the a_{ij} (DIFFER DUE TO LOS), the case mix of its patients or the p_{ij} (DIFFER DUE TO CASE MIX), and the combination of the 2 or their interaction (INTERACT DIFFER).

The separation of the difference into these components is based on the technique described by Kitagawa,[14] briefly described here for the reader's convenience. If a_i and A represent the average length of stay for the ith institution and the region, respectively, then

$$a_i = \sum p_{ij} a_{ij},\ j = 1, \ldots, 383 \text{ (DRG)} \quad (4.3)$$

$$A = \sum_j P_j A_j,\ j = 1, \ldots, 383 \text{ (DRG)} \quad (4.4)$$

The difference between a_i and A can be expressed as

$$\underbrace{a_i - A}_{\text{AVERAGE LOS DIFFER}} = \underbrace{\sum_j P_j (a_{ij} - A_j)}_{\text{DIFFER DUE TO LOS}} + \underbrace{\sum_j A_j (p_{ij} - P_j)}_{\text{DIFFER DUE TO CASE MIX}} + \underbrace{\sum_j (a_{ij} - A_j)(p_{ij} - P_j)}_{\text{INTERACT DIFFER}} \quad (4.5)$$

for $i = 1, \ldots, 383$ (DRG).

This may be rewritten as

$$a_i - A = (\sum_j P_j a_{ij} - \sum_j P_j A_j) + (\sum_j A_j p_{ij} - \sum_j A_j P_j) + \sum_j (a_{ij} - A_j)(p_{ij} - P_j) \quad (4.6)$$

for $j = 1, \ldots, 383$ (DRG),

from which we note that the first 3 differences are computed by subtracting the region's average length of stay from the measures in columns 3 through 5. The interaction component is the residual or the amount of the difference between the hospital and the region not accounted for by the LOS difference (column 10) or the case mix difference (column 11).

The utilization of the report lies in interpreting the differences observed between an individual institution and the set of institutions or region as a whole. For example, consider Hospital 4, the largest with 41,643 cases (column 2). Its average length of stay was one of the longest, 7.479 days (column 3) or 0.708 days longer than the region (column 9). This difference can be attributed in varying degrees to the following components: 1) differences in case composition or case mix; 2) differences in case-specific average lengths of stay; and 3) differences attributed to the interaction of case mix and case-specific average lengths of stay. With respect to the first component, if the hospital's case mix had been treated at the region's average length of stay per DRG, Hospital 4's average length of stay would have been 6.96 days (column 5), for a difference of 0.188 days over the region (column 11). On the other hand, if the case mix were standardized or all institutions were treating the same types of patients, Hospital 4 would have had an average length of stay of 7.320 days (column 4), about a half day (column 10) longer than the region. Moreover, relative to the others, it would have had the longest length of stay. Thus, institution 4 has an average stay that is .708 days longer than the region; .188 days is attributed to a more complex case mix and .548 days to longer DRG-specific average lengths of stay.

The sign and magnitude of the interaction difference (column 12) is an indication of the extent to which the observed difference between an institution's and the region's average length of stay is accounted for by both case mix and DRG-specific length of stay patterns jointly. It is that amount of the overall difference between the institutional and regional averages that cannot be allocated independently to case mix or DRG-specific utilization. A large positive interaction might arise, for example, if the hospital has an average length of stay higher than the region in those DRGs where it treats proportionately more patients than the region or a lower-than-average length of stay in those DRGs with proportionately less discharges. In this instance, the particular patterns of the hospital's deviation from the region in both case mix and utilization by DRG contibuted to it having a higher average length of stay.

The interaction difference is especially important to consider if its absolute value is large relative to the other differences (columns 10 and 11). Under these circumstances, one should not use the measures in the report for that hospital in comparison to the others. A large interaction indicates that hospital utilization patterns vary by case type and standardization would be misleading. Referring to the information in Table 12 Hospital 4 has a small-magnitude negative interaction component, −.029. On the other hand, consider Hospital 8, with an average length of stay of 4.943 days or 1.828 days shorter than the region. The bulk of this difference is attributed to lower DRG-specific utilization (−1.315 days), but also to a less complex case mix (−.986). However, note that these negative differences are partially offset by a half day (.473 days) attributed to the interaction of case mix and performance. In this instance, because of the relatively large interaction term, one should exercise caution in comparing this institution's standardized utilization measures with the other hospitals in the report.

Since discharge information on all patients is routinely collected by institutions for administrative purposes, census data of complete counts rather than samples are generally available for the report described above. However, because of its size or limitations in the number of services and staff available, an individual institution may not treat patients representing a large range of cases or DRGs. In this instance or in situations where there is known to be large interaction, as discussed in the preceding paragraph, an alternative strategy is to restrict one's comparison to a limited number of institutions with uniformly high numbers of cases or to perform comparisons on a limited set of DRGs, for example, those contained in a few Major Diagnostic Categories of particular interest.

4.3 Institutional and Practitioner Profiles in PSRO Evaluations

Professional Standards Review Organizations (PSROs) were mandated under an amendment to the Social Security Act of 1972 to assure the quality and appropriateness of health-care services delivered to federally funded patients (Medicare, Medicaid, Title V). To this end, one specific responsibility of each organization is "profile analysis"—that is, the construction and review of relevant summaries of aggregated data pertaining to the care and services received by patients and provided by practitioners and institutions in the PSRO area. It is basically a descriptive analysis of health care and service patterns using measures and formats which serve to

1) facilitate the identification of exceptional areas of performance and 2) provide a mechanism of the ongoing monitoring of the overall system, with particular attention to previously defined problem areas.

While considerable latitude is given the PSRO for the format and content of a profile, there are 3 basic components to its design: 1) *the patient group*, which defines the scope of the profile in terms of the population to be analyzed; 2) *the profile subjects*, which define the subset of the patient group whose data is to be profiled; that is, aggregated and summarized in terms of some measures for comparative purposes with other profile subjects or the entire patient group; and 3) *the profile elements*, which are the statistics or measures of interest to be displayed in the profile. These 3 components are chosen to meet the objectives of the analysis to be performed. Of particular importance is the selection of an appropriate patient group, whose definition should control, to the extent possible, those differences in the patient population that may affect the measures being displayed in the profile. For example, if a PSRO were interested in examining length-of-stay patterns for hospitals in its area, it would not be meaningful to form a group of patients from such diverse categories as normal deliveries, acute myocardial infaractions, gastric and peptic ulcers and hemorrhoids. In this case it would be impossible to determine from the aggregate summaries whether particularly deviant patterns for institutions were attributed to differences in the types of patients they treat or to differences in their length of stay.

The DRGs can therefore be an effective mechanism in profile analysis by providing a structure of consistent patient-class definitions within which institutional performance can be compared based on similar types of patients. As part of a project sponsored by the Health Standards and Quality Bureau, 7 PSRO sites have experimented with the use of institutional length-of-stay profiles based on DRGs for the retrospective monitoring of utilization patterns in their area.[10,12] Each site submitted tapes of their hospital abstract data, which were then assigned the appropriate DRG number based on information contained in the record. A series of reports were produced with summary statistics and institutional comparisons for each group.

An example of one report appears in Figure 3 and presents hospital length of stay performance patterns for DRG 121—Acute Myocardial Infarction. A hospital identifier appears on the left, adjacent to a grid containing the median (m), the interquartile range (25th [first x] to the 75th [second x] percentile), the 10th (first *) percentile, and 90th (second *) percentile of length of stay for each hospital. To the right of the grid are columns containing the number of discharges (OBS), the mean or average length of stay (MEAN), the standard deviation (S.D.) and the median (MEDIAN) length of stay, or 50th percentile. In this example, the median length of stay for the region as a whole was 2 weeks. However, there was considerable variation among hospitals. Hospital 8 had a median length of stay of 10 days, while Hospitals 4, 2 and 3 had medians of 15 days. Further investigation in this area through medical care evaluation studies may be warranted to determine if there are quality-of-care problems in the short-length-of-stays hospital or inappropriate utilization in the other 3 institutions.

As illustrated in the above example, these profiles are potentially useful in identifying areas of inappropriate utilization, which can lead to additional issues to address in quality of care. Since the overall number of DRGs is manageable and low compared to alternative classification by diagnosis or PAS group, sample sizes for individual hospitals in most cases are large enough to produce meaningful compari-

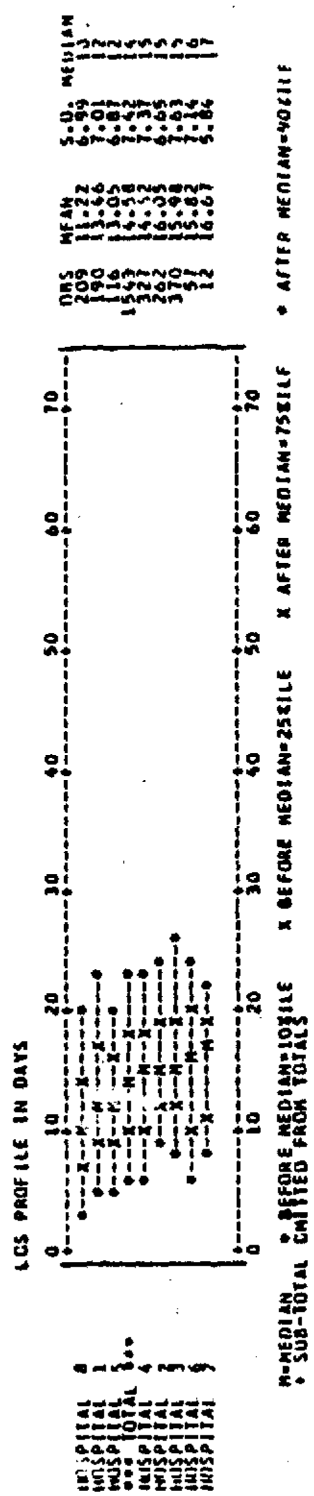


FIG. 3. Sample institutional length-of-stay profile for acute myocardial infarction patients.

sons. Moreover, each group can be reviewed by PSRO personnel and a set selected for follow-up analyses. To facilitate further the identification of aberrant utilization behavior, automated screens based on statistical algorithms have also been implemented that first flag those DRGs with relatively high inter-hospital variation, then determine for each hospital the groups in which its mean length of stay was significantly higher than the area and those in which it was significantly lower.

4.4 A PSRO Monitoring System

The utility of the DRGs in reviewing PSRO data has suggested their potential future application as a control mechanism to replace the present costly concurrent review process with a timely, cost-effective retrospective monitoring system. In this context, at the time of patient discharge, an abstract is produced and a DRG number assigned based on data in the record. When a small sample of records has accumulated in a DRG for an institution or set of institutions under examination, information from the sample can be used to estimate the parameters of the distributions of 1 or more dependent variables related to utilization or quality of care (e.g., length of stay, death rate, postoperative wound infection rate). Once acceptable patterns of practice have been established, the monitoring system can use standard statistical control procedures to detect changes in the process. When changes are detected, special studies may be justified to determine the cause for the change, which may be attributed to errors in the record, 1 or more particularly deviant cases, or an overall change in the treatment process that has affected all the cases. Work is currently under way to develop computer-based systems which could support such a control mechanism in a timely manner on either a hospital or PSRO level.

MEDICAL CARE
February 1980, Vol. XVIII, No. 2, Supplement

5. Case Mix Accounting in Hospital Budgeting, Cost Control and Prospective Reimbursement

5.1 Case Mix Accounting

AN IMPORTANT OBJECTIVE of hospital costing and budgetary systems is the understanding and control of hospital costs. In traditional organizational settings, cost control is most successful in those situations where well-defined products or services are provided with a predictable set of associated costs. The provision of different combinations of products or services results in different levels of total resource consumption and cost requirements. Cost control in such settings basically entails the monitoring of resources consumed and costs incurred during the production process to insure consistency with expected levels. Thus, for such a system to be operational within a hospital, there must be a precise definition of the services provided by the institution. In a general sense, hospitals provide "patient care," but more specifically, they provide patient care of various kinds and intensities over various durations based on the needs of the patients treated.

Since the DRGs form a classification of the patient population into classes with similar expected output utilization, they can provide a definition of the services provided by a hospital. As such, they allow the resources consumed and costs incurred to be related directly to the types of patients or case mix that the hospital treats.[2,8,9] This is important in a hospital setting, where it is not management (i.e., administrators) but rather individual physicians who are reasponsible for allocating resources through various services and departments in order to provide effective patient care. To a large extent, physicians act independently of each other and are not generally aware of the overall financial implications of their individual decisions. If hospital cost control is to be attained, effective communication between the financial systems of the hospital and its physicians must be achieved. By formulating the hospital budget in terms of patient classes with similar patterns of care, a direct linkage between the practices of individual physicians and the financial consequences for the hospital can be realized.

The goal, then, of a case mix accounting system is to provide a complete financial picture of the costs of treating specific types of patients, whose care is the basic service of a hospital. Under the traditional organizational structure of a hospital, there is no department whose responsibility is to insure that individual patients are financially well managed. Typically, the hospital's 2 accounting systems—financial and managerial— deal with patients in the aggregate and not on an individual basis. The financial system provides the basic financial description of the hospital in terms of the balance sheet, income statement and funds flow, while the managerial accounting system provides the financial information oriented at the department level (e.g. nursing, laboratory, medical records) for internal management purposes. Thus, hospital accounting systems have not provided the integrated picture of the financial consequences of the care delivered to individual patients that case mix accounting is designed to produce.

5.2 DRG Cost Model

The process of determining the cost of treating patients in each of the DRGs for an individual hospital or collection of hospitals is decribed elsewhere.[2] In summary, the types of accounts in a hospital chart of accounts can be categorized into 6 distinct service areas: 1) outpatient accounts; 2) overhead accounts not related to patient care; 3) overhead accounts related to patient care; 4) hotel and other general services accounts; 5) nursing accounts; and 6) ancillary services accounts.

The DRGs currently encompass only the inpatient population; hospital outpatient costs are not included in the DRGs' costs. Overhead accounts are costs incurred by the hospital in its general operation but are either not related or only indirectly related to the provision of patient care. Depreciation and interest charges are examples of overhead costs that are not related to patient care and therefore are not normally included in the DRG costs. Other overhead accounts such as housekeeping or laundry are indirectly related to the provision of patient care and are included in the DRG cost. The definition of the overhead accounts that are considered as patient-care–related versus non–patient-care–related can vary, depending on the goal of the case mix accounting system. For strictly internal management purposes it is reasonable to include as patient-care–related the various administrative services. However, if the case costs of a collection of hospitals are to be compared, then the administrative costs should not be included, since administrative costs can vary greatly across hospitals for reasons other than case mix. The remaining 3 types of accounts are all directly related to patient care and with the addition of the outpatient account, are referred to as the final cost centers. The services associated with these accounts can be directly related to individual patients, allowing the costs to be apportioned to each patient.

The direct costs of each final cost center and the portions of the cost of patient-care–related overhead accounts allocated to each final cost center (as determined by a special algorithm) represent the total cost of providing the services associated with each final cost center. An allocation statistic specific to each final cost center is used as the basis of apportioning the costs to the patients in each of the DRGs. For example, the cost of nursing is allocated to patients based on a DRG-specific per diem nursing weight which was derived through a study of the amount of nursing time spent with patients in each DRG. While all of the allocation statistics possess some defects, they are designed to reflect more equitably the quantity of an institution's resources consumed by the patients in each DRG. The end result of the DRG cost model is the determination of the unit cost of treating patients in each DRG.

5.3 Hospital Budgeting

The full case mix cost accounting approach has been applied to the budgetary process in 2 test hospitals. In the initial year, the unit costs (i.e., average cost per patient) in each DRG were determined. In order to establish the following year's budget, it was only necessary to project the hospital's case mix and apply the appropriate inflation factors. Deviations from the budget due to case mix were immediately detected and the diagnostic and service areas experiencing significant deviations from established unit costs were isolated.

The resulting unit costs in the test hospitals typically varied across DRGs by more than a hundredfold. The following DRGs illustrate this cost variation.

DRG	*Typical 1976 Unit Cost*
127—Ischemic heart disease except acute myocardial infarction with shunt or other major operation	$9,934
187—Gastric and peptic ulcer with gastric resection or other major operation with a secondary diagnosis present	7,362

112—Otitis media, chronic mastoiditis or otosclerosis without any operation	264
273—False labor without any operation	89

Even within a specific diagnostic area the DRGs provide a high degree of cost discrimination. For example, patients with a primary diagnosis of urinary calculus encompass 4 DRGs with the following typical 1976 unit costs:

Urinary Calculus DRGs	*Typical 1976 Unit Cost*
239—Without an operation or secondary diagnosis	$ 394
240—Without an operation with a secondary diagnosis	774
241—With minor operation such as cystoscopy or catheter to kidney	1,032
242—With major operation such as nephrotomy, cystotomy or ureterotomy	2,293

Thus, even within this narrow diagnostic area the unit costs across DRGs varied by a factor of nearly 6.

An example of a unit cost report for DRG 121—Acute Myocardial Infarction—appears in Figure 4. It compares the cost of treating AMI patients in the same hospital across 2 different years. The box at the top of the report summarizes the length of stay, charges, and costs experienced by AMI patients in the 2-year period. The bottom portion of the report breaks down the costs experienced in terms of the final cost centers of that hospital. For each item in the report both the absolute and per cent change across the 2 years are indicated. Such a report allows a hospital administrator to isolate both the diagnostic and service areas where there are significant differences across years or relative to other hospitals if comparable data from other hospitals is available. Once the potential problem areas have been identified, the administrator can begin a more meaningful dialogue with the physicians responsible for the identified patients and services.

5.4 Prospective Reimbursement

Traditionally, most health insurers have reimbursed hospitals retrospectively on the basis of reasonable and allowable costs. While this model of reimbursement guarantees coverage for most hospital expenditures, it provides little economic incentive to hospitals to control costs. Hospital prospective-reimbursement systems establish the rate of hospital reimbursement before the period over which the rate is to apply. The rewards and penalties inherent in a prospective system can potentially provide the motivation for hospitals to become more cost effective without sacrificing the quality of medical care. Under contract No. 600-77-022 from the Health Care Financing Administration, the State of New Jersey is in the process of moving from a per diem reasonable-cost–based reimbursement system to a cost-per-case incentive-based system.[21]

The Standard Hospital Accounting and Rate Evaluation (SHARE) system is the per diem cost-based reimbursement system currently in use in New Jersey. Under the SHARE system, costs are grouped into 31 cost centers according to uniform definitions of functional centers such as laboratory, radiology and the like. The inpatient costs are then regrouped within each cost center into 3 basic categories: 1) non-physician-controllable costs; 2) physician costs (e.g., physician and resident salaries and fees); and 3) other costs which are either not controllable by the hospital, or

DRG 121 ACUTE MYOCARDIAL INFARCTION

	1975	1976	CHANGE	% CHANGE
PATIENTS	267	315	48	17.98
TOTAL COST	736811.12	882896.69	146085.56	19.83
UNIT COST	2759.59	2802.85	43.25	1.57
DAILY COST	199.24	201.26	2.02	1.02
BED-DAYS	3698	4387	689	18.62
AVERAGE LOS	13.85	13.93	.08	.55
TOTAL CHARGES	820269.81	1042593.69	222323.87	27.10
AVERAGE CHARGE	3072.17	3309.82	237.65	7.74
COST/CHARGES	.89825	.84683		

FINAL COST CENTER	1975 UNIT COST	1976 UNIT COST	CHANGE UNIT COST	% CHANGE UNIT COST
DIETARY	127.46	131.44	3.98	3.12
ADMITTING	20.48	22.45	1.97	9.61
BILLING	90.40	97.27	6.87	7.60
HOTEL	159.28	180.74	21.45	13.47
NURSING	702.26	681.49	-20.77	-2.96
INTERNS & RESIDENTS	72.23	71.51	-.73	-1.00
MEDICAL RECORDS	22.01	26.39	4.38	19.87
SOCIAL SERVICE	16.36	17.50	1.14	6.95
NEW BORN CARE UNIT	.00	.00	.00	.00
INTENSIVE CARE UNIT	60.34	69.37	9.03	14.96
CORONARY CARE UNIT	636.73	509.58	-127.15	-19.97
OPERATING ROOM	.00	.00	.00	.00
RECOVERY ROOM	.00	.00	.00	.00
ANESTHESIOLOGY	.00	.00	.00	.00
DELIVERY ROOM	.00	.00	.00	.00
DIAGNOSTIC RADIOLOGY	70.77	86.87	16.10	22.75
RADIOISOTOPES	3.68	4.55	.87	23.55
RADIATION THERAPY	.00	.00	.00	.00
LABORATORY	276.29	330.85	54.56	19.75
EEG	.00	.00	.00	.00
EKG	58.22	55.07	-3.15	-5.41
MED-SURG SUPPLIES	101.24	136.18	34.94	34.51
PHYSICAL MEDICINE	68.74	84.43	15.69	22.83
RESPIRATORY THERAPY	77.25	82.49	5.24	6.79
IV THERAPY	39.82	49.30	9.47	23.79
PHARMACY	156.03	165.38	9.35	5.99
DIALYSIS	.00	.00	.00	.00
UROLOGY	.00	.00	.00	.00
KIDNEY TRANSPLANT	.00	.00	.00	.00

FIG. 4. Sample management report for acute myocardial infarction patients treated during a 2-year period.

which are not to be included in the determination of the SHARE reimbursement rate. After salary rates have been "equalized" in order to avoid distortions due to geographic wage differentials, controllable costs are grouped into 3 clusters of cost centers to allow for trade-offs in treatment modes. Hospital peer groups are formed and cost screens are set for each cluster of cost centers in each peer group. When a hospital's cluster of cost centers fails these screens, each cost center within each cluster is screened again within peer groups. Formulae then are applied to determine any unreasonable costs in the base period that should be disallowed. After a hospital's reasonable costs have been determined, a preliminary per diem reimbursement rate is determined for the hospital. The proposed SHARE per diem is re-examined in an informal review, often followed by appeals to the Department, and in some cases to the courts. Once the SHARE per diem is finalized, it provides the basis of reimbursement except for certain year-end adjustments made for volume variance, actual inflation, and cost increases because of legal and approved management changes, such as Certificates of Need and Blue Cross contractual adjustments.

The essential feature of the SHARE methodology is "peer grouping." This regulatory technique was developed in the utility field, wherein like institutions are classified according to comparable attributes in order to apply "fair" standards and develop reasonable rates. Since hospitals display more complexity than single-output industries such as electric, gas or telephone companies, the technique has been applied not at the aggregate level but at the cost center level. It should be noted that the form or unit of payment under the SHARE system is the per diem and that the basis of payment is cost.

Although the SHARE reimbursement methodology in New Jersey attained significant accomplishments in bringing a certain amount of order to cost reporting by hospitals in the state, there are basic elements to the structure of hospital reimbursement which remain problems and which SHARE is not equipped to address.

The fundamental problem in cost containment is defining the appropriate tools for measuring reasonable efficiency and effectiveness in the hospital setting. Accurate instruments are needed to measure the level of productivity and effectiveness in terms of both outputs and inputs and to respond with the appropriate financial incentives or disincentives. The attempt to measure and compare hospital efficiency at the cost-center level fails to recognize the role of case mix in determining hospital costs. Differences or lack of differences in hospital costs at the cost-center level can be the result of different case mix compositions and may not reflect differences in hospital productivity. Further, the quality or effectiveness of care must be properly identified and measured in order to evaluate accurately hospital productivity. Analyses by cost center may tend to obscure rather than clarify these problems.

The second limitation of the SHARE approach is in the existence of financial disincentives to the delivery of high-quality, appropriate and efficient care. It is no secret that the per diem form of payment tends to encourage longer lengths of stay. Furthermore, the length-of-stay incentive problem is compounded by the use of cost as a basis of reimbursement. The reasonable measure of a hospital financial manager's effectiveness is his success in maximizing reimbursement. The more cost is increased, the greater the revenues. Costs can be skillfully redistributed over different cost centers with the effect of escaping peer cost screens. Another problem for the regulator, and opportunity for the ambitious financial manager, results from a system in which some patients pay costs and others pay charges, compounded because the ratio of costs to charges can be manipulated. Some opportunities for business gamesmanship will doubtless be found in any reimbursement methodology. The reimbursement system must identify appropriate costs and design a financial incentive to monitor and minimize such practices.

Finally, the need for effective communication is perhaps the most serious problem facing regulation of the health-care industry. To be effective, regulators and managers must build a language to the physicians. It is a reflection of the state of the art that appeals in the hospital industry are managed by lawyers and accountants and therefore examine problems of allocation and finance. If the financial and medical information were merged, it would become possible to trace the relationship between the physicians' decisions and their effects on costs. From this base, the proper questions can be framed to deal equitably with issues of effectiveness, quality and efficiency in patient care.

Thus, the SHARE methodology was not equipped to address the problems of 1) defining and measuring hospital case mix, productivity and effectiveness; 2) providing incentives for better management; 3) avoiding business gamesmenship; and 4) fostering communications between the hospital financial systems and physicians. In order to deal with these problems, New Jersey is developing a prospective case-cost incentive-reimbursement system. The DRGs will provide the basis for the definition of case types and differential reimbursement rates will be established for each DRG. In order to develop the DRG reimbursement rates, it is necessary to obtain the patient abstract, billing and financial information for the case mix cost accounting model described in the previous section.

Since January 1976, New Jersey has been receiving medical discharge abstracts from all 104 acute-care hospitals in the state. In addition, 21 self-selected hospitals are also submitting their billing information. Hospital cost information is obtained through SHARE and is grouped into 2 categories: costs directly related to patient care, and institutional and other costs.

Costs that were judged to be relatively fixed in the short run, and not directly related to patient care, were categorized as institutional costs. Institutional cost centers include managerial services, facilities maintenance, and allied health, nursing and graduate medical education costs. A separate methodology was developed to determine reasonable institutional costs comprised of a fixed sum and variable amounts that relate to teaching commitment and to the total amount of patient-care dollars. The patient-care costs were allocated to the patients in each DRG using the DRG-cost accounting methodology.

The case mix cost accounting approach permits the development of the institutional productivity goals desired by the regulators and the proper recognition of case mix complexity differences across hospitals. Once the DRG costs have been determined, reimbursement standards by DRG provide an appropriate set of goals, which are balanced to reflect the interests of both the public and the industry. The preliminary rate design has set a rate per DRG for each hospital, composed of proportions of the hospital mean case cost per DRG and the state standard (the mean case cost per DRG across the sample of 21 hospitals). The hospital and state proportion would always sum to unity. Thus, for example, in the initial years of the system the rate per DRG might consist of 75 per cent hospital cost and 25 per cent state standard. Although the proportion of state standard would begin to approach 100 per cent over time, the early emphasis on the hospital actual cost will provide a reasonable opportunity for institutions themselves to make use of management information by DRG. Detailed management information by DRG will be provided to each hospital and will be organized to help the hospitals effectively focus on areas of concern in order to deal with problem areas.

In order to realize fully the potential of the system, the process of implementation currently is planned for 2 distinct stages. Stage I, the initial development, gives to each hospital both the opportunity and responsibility for reacting to the management information by implementing steps to remedy inefficiency, expanding efficiency, and opening an effective dialogue with their physicians. The motivation for discharging this burden is supplied by prospective, incentive-based reimbursement. Thus, in any given year, a hospital would retain the savings achieved by bringing in its cost per case under the prospective rate. Since the form of payment is per case rather than per diem, the unit of reimbursement no longer poses an incentive to increased lengths of stay. Recalculating the next year's prospective rate based on the previous year's actual achievement serves the public interest by embedding the results in an improved standard.

The second stage in the process of implementation will involve properly crystallizing questions of inter-hospital significance. Health-care issues will be placed in their DRG-specific and medical context, rather than the current collection of financially oriented appeals, thus permitting an examination of efficiency, quality and appropriateness. The stages permit the passage of time required to purge imperfections in the data base and to allow hospital reaction to successive generations of the rate. Beginning in early 1980, prospective DRG rates are scheduled to be set for 26 New Jersey hospitals.

MEDICAL CARE
February 1980, Vol. XVIII, No. 2, Supplement

6. A Case Mix Approach to Regional Planning for Acute Care Hospitals

WHILE THERE ARE ONGOING USES of the DRGs in the areas of utilization review and hospital budgeting, cost control, and reimbursement, the application of the DRGs in the area of hospital regional planning is only in its early stages of development.[5] Regional planning refers to the activity of organizing health-care resources in a defined geographic region to achieve a desired state of affairs in terms of the availability of health care of acceptable quality and cost. The primary thrust of the hospital planning activity has traditionally focused on hospital facilities, primarily beds. Through legislation such as the Hill-Burton Act, much of the planning activity prior to the 1970s emphasized the adequacy and distribution of hospital beds to meet the needs of the population. However, the rapid increase in sophisticated medical technology has resulted in a need to plan on a regional basis not only for hospital beds but also for specific hospital services and equipment. Thus, the planning for the quantity and distribution of major new equipment such as computed tomography scanners or specific hospital services such as open heart surgery has become an integral part of the planning activity. Since certain types of services and equipment are necessary to treat specific patient types, planning decisions will affect the case mix a hospital can treat.

The modification by the planning process of the case mix that an individual hospital can treat will inevitably affect the case mix of the other hospitals in the region. For example, if a new service is added in one hospital, then that hospital will begin to treat additional types of patients. This will likely result in a decrease in the number of those types of patients treated at the other hospitals in the region. Further, the new service may cause the other capacities of the hospital (e.g., beds or operating room time) to be exceeded, requiring that the hospital cease to treat patients to whom it previously provided care. The excess patients will have to be treated in the other hospitals in the region. Thus, the implications of a planning decision can be complex and difficult to predict. A case mix approach to regional planning would have as its central focus the patients being treated and the demands they place on hospital resources. The role of each hospital in the region would be defined in terms of the case mix it treats.

The basic hospital regional planning problem must consider a number of factors simultaneously:

1. All patients must have access to the necessary hospital services. Access would normally be defined in terms of travel time by ground transportation. An acceptable travel time could vary by case type, depending on the prevelance of the patient type and nature of the condition (i.e., chronic, emergency, etc.).

2. The bed and other capacities of the hospitals in the region must be sufficient to meet the demand for care.

3. Patients must only be treated in hospitals with adequate services, equipment and specialists. The type of hospital facilities and specialists necessary for proper care will vary with patient types. For example, it might be determined that patients younger than 14 years of age should only be treated at hospitals with pediatric units.

4. Minimal quality-of-care standards for each patient type must be met. The quality-of-care standards could be as basic as a minimal mortality rate for each patient type. For example, if the mortality rate for acute myocardial infarction patients in a hospital exceeded some percentage, such as 20 per cent, then the planning process must consider requiring that hospital to cease providing care to those patients. The quality-of-care standard could become

more sophisticated and include process and outcome measures.

5. The cost of providing care to the different types of patients must be reasonable. The information from projects such as the New Jersey prospective reimbursement experiment can establish the cost of providing care to each patient type in each hospital. If a hospital proves extremely cost ineffective in providing care to certain types of patients, then the planning process must consider whether a more cost-effective alternative exists.

Thus, the basic hospital regional planning problem can be described as a system of hospitals with the capacity to treat specified numbers and types of patients. In addition, there are a set of basic quality and access constraints that must be met. However, given that the capacity bounds and access and quality constraints can be met by a number of alternative configurations of case mix, the alternative requiring the least cost is likely to be the preferred alternative. Such a problem can be analyzed using linear programming techniques. The DRG can provide the basis for establishing hospital case mix. A case mix-based linear programming model of the regional planning problem should prove to have many immediate practical applications. For example, suppose a proposal to close a particular hospital had been made. The regional planning model could be used to suggest the most cost-effective means of distributing the closed hospital's case mix to the other hospitals in the region while still maintaining adequate access and quality and not exceeding the capacities of any hospital. Alternatively, if there were an existing recommendation for the distribution of the closed hospital's case mix, then the model could be used simply to simulate the impact of the case mix changes on the other hospitals and describe that impact in terms of various hospital parameters, such as occupancy rates and reimbursement levels. There are many other examples of the use of such a model. If there were significant changes in the composition of case types projected because of a new cure or an outbreak of an infectious disease, then the model could be used to evaluate the impact on the hospitals in the region. Thus, while a case mix regional planning model is only in its early stages of development, it does hold the potential to be a powerful tool for the hospital planning process.

MEDICAL CARE
February 1980, Vol. XVIII, No. 2, Supplement

7. Summary and Conclusions

PATIENTS treated in an acute-care facility can vary considerably in both the duration and intensity of services required to provide appropriate patient care. The relative amounts and types of hospital outputs utilized by individual patients are dependent on both the condition of the patient and the treatment process employed. By relating the demographic, diagnostic, and therapeutic characteristics of patients to the hospital outputs they utilize, a patient classification scheme can be developed which provides the framework for both the specification of hospital case mix and the measurement of the impact of case mix on hospital utilization and performance. The Diagnosis-Related Groups represent an attempt to provide such a patient classification scheme. As currently defined, the DRGs provide a manageable number of patient classes (383) that are exhaustive and mutually exclusive with respect to the types of patients seen in an acute-care setting. Further, the DRGs provide patient classes that are clinicially consistent and that have similar patterns of output utilization as measured by length of stay.

The comparison of patient data across institutions or providers will invariably reveal the existence of differential levels of utilization and performance. A comparative analysis by average length of stay, cost, or any other aggregate measure is not meaningful unless the impact of different case mix compositions can be determined. The DRGs can provide a framework for establishing the effects of case mix as well as for identifying diagnostic areas with potential problems. The goal of most comparative analyses is to isolate problem areas so that corrective measures can be initiated. If programs aimed at improving the performance of the hospital health-care system are to be successful, managers and regulators must establish an effective dialogue with those responsible for the delivery of services, the physician community. The DRGs provide the first step in such a dialogue since problems defined in the context of DRGs are understandable from a clinical perspective.

The various actual and potential applications of the DRGs in the areas of utilization review, hospital budgeting and cost control, prospective reimbursement and regional planning emphasize the central role of the patient. By focusing on the types of patients being treated, programs responsible for these activities will share a common conceptual basis even though they are concerned with different aspects of the health care system. While the applications to date have been implemented to meet the immediate needs of the individual programs, future work will be directed toward exploring the potential of the DRGs in achieving better integration and coordination of the different program goals and activities.

The current set of 383 DRGs were developed in light of the available data and its limitations at the time of their construction. As such, they represent just one implementation of an evolving series of patient classification schemes. As more comprehensive and reliable patient data become available and the practice of medicine changes, the DRGs must adapt to reflect these changes. To this end, it is felt that the technology and strategy used in forming the DRGs can be applied in the development of future generations of classification systems. Indeed, a major revision and evaluation of the DRGs will be undertaken as soon as ICD-9-CM data are available in sufficient quantities. Further, work has begun in extending the approach into other areas of health-care delivery, in particular, ambulatory care.[24]

References

1. Department of Health, Education, and Welfare. AUTOGRP patient classification scheme and diagnosis related groups (DRGs). Health care financing grants and contracts report series. Washington, D. C.: Department of Health, Education, and Welfare, 1978. (DHEW publication no. (HCFA)03011 9-79.)

2. Averill RF. Application of case mix cost accounting to hospital budgeting and prospective reimbursement presented at the 38th meeting of the Academy of Management, San Francisco, August 1978.

3. Bays CW. Case-mix differences between nonprofit and for profit hospitals. Inquiry 1977;14:17.

4. Center for Health Studies. Diagnosis related group reports for PSRO databases. New Haven: Yale University Center Health Studies, 1978.

5. Chernow R. A casemix approach to regional planning: an application of linear programming. Ph.D. dissertation. New Haven: Yale University, in progress.

6. Commission on Professional and Hospital Activities. Hospital adaptation of ICDA. 2nd ed. Ann Arbor: Commission on Professional and Hospital Activities, 1968.

7. Commission on Professional and Hospital Activities. Length of stay in PAS hospitals, by diagnosis. Ann Arbor: Commission on Professional and Hospital Activities, 1976.

8. Fetter RB, Mills R, Riedel DC, Thompson JD. The application of diagnostic specific cost profiles to cost and reimbursement control in hospitals. J Med Systems 1977;1:137.

9. Fetter RB, Thompson JD, Mills RE. A system for cost and reimbursement control in hospitals. Yale Biol J 1976;49:123.

10. Freeman JL. Applications of the AUTOGRP computer system as a patient care monitoring mechanism for PSRO's. Presented at the 38th meeting of the Academy of Management, San Francisco, August 1978.

11. Hill BA. Principles of medical statistics. New York: Oxford University Press, 1971.

12. Health Care Financing Administration. Implementation, evaluation and extension of a patient care monitoring mechanism. Final report. Contract 240-75-0051. Baltimore: Health Care Financing Administration, October 1977.

13. Public Health Service. International classification of diseases, eighth revision. Washington, D. C.: U.S. Government Printing Office, 1968. (Publication no. 1693.)

14. Kitagawa EM. Components of a difference between two rates. J Am Statistical Assoc 1955;50:296.

15. Lave JR, Lave LB. The extent of role differentiation among hospitals. Health Serv Res 1971;6:15.

16. Lave JR, Leinhardt S. The cost and length of a hospital stay. Inquiry 1976;13:327.

17. Lee ML, Wallace RL. Classification of diseases for hospital cost analysis. Inquiry 1972;9:69.

18. Lee ML, Wallace RL. Problems in estimating multiproduct cost functions: an application to hospitals. West Econ J 1973;11:350.

19. Luke RD. Dimensions in hospital case mix measurement. Inquiry 1979;16:38.

20. Mills RE, Fetter RB, Riedel DC, Averill RF. AUTOGRP: an interactive computer system for the analysis of health care data. Med Care 1976;14:603.

21. New Jersey State Department of Health. A prospective reimbursement system based on patient case-mix for New Jersey hospitals, 1976–1978. Annual Report. November 20, 1977.

22. Rafferty J. Patterns of hospital use: an analysis of short-run variations. J Political Econ 1971;79:154.

23. Ranofsky, AL. Utilization of short stay hospitals: annual summary for the United States, 1976. Rockville, MD: National Center for Health Statistics, 1978. (Vital and health statistics. Series 13, no. 26) (DHEW publication no. (PHS)78-1788.)

24. Social Security Administration. Progress report: development, testing and evaluation of a prospective case-payment reimbursement system. Washington, D. C., Social Security Administration, 1978.

25. Sonquist JA, Morgan JN. The detection of interaction effects. Ann Arbor: Institute for Social Research, University of Michigan, 1964.

.CAL CARE
bruary 1980, Vol. XVIII, No. 2, Supplement

Appendix
Diagnosis-Related Group Descriptions

Major Diagnosis Category			Diagnosis-Related Groups
01:	Infectious Diseases	001	Infectious Disease (Enteritis, Diarrhea) with Age less than 16
		002	Infectious Disease (Enteritis, Diarrhea) with Age greater than 15
		003	Infectious Disease (Viral Disease, VD, Meningitis) without Secondary Diagnosis
		004	Infectious Disease (Viral Disease, VD, Meningitis) with Secondary Diagnosis
		005	Infectious Disease (Blood Infection, TB, Salmonella) without Surgery
		006	Infectious Disease (Blood Infection, TB, Salmonella) with Surgery
02:	Malignant Neoplasm of the Digestive System	007	Cancer of the Mouth, Tongue, Large Intestine, Liver, Gallbladder without Surgery
		008	Cancer of the GI System (Esophagus, Stomach, Pancreas, Small Intestine, Rectum) without Surgery
		009	Cancer of the GI System with Surgical Procedure (Biopsy, Endoscopy, Local Excision, Centesis) without Secondary Diagnosis
		010	Cancer of the GI System with Surgical Procedure (Biopsy, Endoscopy, Local Excision, Draining) with Secondary Diagnosis
		011	Cancer of the GI System with Surgery (Gastric Resection, Colon Resection, Esophagus Resection)
03:	Malignant Neoplasm of the Respiratory System	012	Cancer of the Respiratory System (Trachea, Lung, Larynx, Thorax, Mediastinum) without Surgery without Secondary Diagnosis
		013	Cancer of the Respiratory System (Trachea, Lung, Larynx, Thorax, Mediastinum) without Surgery with Secondary Diagnosis
		014	Cancer of the Respiratory System with Surgical Procedure (Biopsy, Endoscopy, Excision of Lesion) without Secondary Diagnosis
		015	Cancer of the Respiratory System with Surgical Procedure (Biopsy, Endoscopy, Excision of Lesion) with Secondary Diagnosis
		016	Cancer of Respiratory System with Surgery (Lobectomy, Laryngectomy, Radical Resection)
04:	Malignant Neoplasm of the Skin	017	Cancer of the Skin except Malignant Melanoma without Secondary Diagnosis
		018	Cancer of the Skin except Malignant Melanoma with Secondary Diagnosis
		019	Cancer of the Skin—Malignant Melanoma with Surgical Procedure without Secondary Diagnosis
		020	Cancer of the Skin—Malignant Melanoma with Surgical Procedure with Secondary Diagnosis
05:	Malignant Neoplasm of the Breast	021	Cancer of the Breast without Surgery with Age less than 63
		022	Cancer of the Breast without Surgery with Age greater than 62
		023	Cancer of the Breast with Surgery without Secondary Diagnosis
		024	Cancer of the Breast with Surgery with Secondary Diagnosis
06:	Malignant Neoplasm of the Female Reproductive System	025	Cancer of the Female Reproductive System (Uterus, Cervix, Vagina, Ovary, Fallopian Tube) without Surgery without Secondary Diagnosis
		026	Cancer of the Female Reproductive system (Uterus, Cervix, Vagina, Ovary, Fallopian Tube) without Surgery with Secondary Diagnosis
		027	Cancer of the Female Reproductive System with Surgical Procedure (D&C, Biopsy, Excision of Lesion) without Secondary Diagnosis

Major Diagnosis Category		Diagnosis-Related Groups
	028	Cancer of the Female Reproductive System with Surgical Procedure (D&C, Biopsy, Excision of Lesion) with Secondary Diagnosis
	029	Cancer of the Uterus Body with Surgery (Removal of Uterus)
	030	Cancer of the Uterus, Cervix, Ovary with Surgery (Removal of Uterus or other Major Operation)
07: Malignant Neoplasm of the Male Reproductive System	031	Cancer of the Male Reproductive System (Penis, Prostate, Testicle) without Surgery
	032	Cancer of the Male Reproductive System with Surgical Procedure (Biopsy, Cystoscopy, Removal of Testicle) without Secondary Diagnosis
	033	Cancer of the Male Reproductive System with Surgical Procedure (Biopsy, Cystoscopy, Removal of Testicle) with Secondary Diagnosis
	034	Cancer of the Male Reproductive System with Surgery (Amputation of Penis, Removal of Prostate, Radical Excision of Lesion)
08: Malignant Neoplasm of the Urinary System	035	Cancer of the Urinary System (Bladder, Urethra, Kidney, Ureter) without Surgery
	036	Cancer of the Urinary System with Surgical Procedure (Cystoscopy, TUR, Excision of Lesion) without Secondary Diagnosis
	037	Cancer of the Urinary System with Surgical Procedure (Cystoscopy, TUR, Excision of Lesion) with Secondary Diagnosis
	038	Cancer of the Urinary System with Surgery (Removal/Excision of Bladder, Kidney, Ureter, Urethra)
09: Malignant Neoplasm of Other and Unspecified Sites	039	Cancer of the Bone, Thyroid, Connective Tissue, Nerves without Surgery
	040	Cancer of the Brain, Secondary Cancer, Multiple Cancer Sites without Surgery without Secondary Diagnosis
	041	Cancer of the Brain, Secondary Cancer, Multiple Cancer Sites without Surgery with Secondary Diagnosis
	042	Cancer of the Thyroid, Connective Tissue, Nerves with Surgical Procedure (Biopsy, Excision)
	043	Cancer of a Secondary Site, Multiple Sites with Surgical Procedure (Biopsy, Excision)
	044	Cancer of the Bone, Connective Tissue, Nerves, Secondary Site, Multiple Sites with Surgery
10: Neoplasm of the Lymphatic and Hemopoietic Tissue	045	Tumor of the Lymphatic System, Blood Making Tissue without Secondary Diagnosis with Age less than 16
	046	Tumor of the Lymphatic System, Blood Making Tissue with Secondary Diagnosis with Age less than 16
	047	Disease of the Lymphatic System, Hodgkins Disease, Sarcoma without Surgery without Secondary Diagnosis with Age greater than 15
	048	Disease of the Lymphatic System, Hodgkins Disease, Sarcoma without Surgery with Secondary Diagnosis with Age greater than 15
	049	Tumor of the Lymphatic System, Multiple Myeloma, Leukemia without Surgery with Age greater than 15
	050	Tumor of the Lymphatic System, Blood Making Tissue with Surgical Procedure (Excision of Node) without Secondary Diagnosis with Age greater than 15
	051	Tumor of the Lymphatic System, Blood Making Tissue with Surgical Procedure (Excision of Node) with Secondary Diagnosis with Age greater than 15
	052	Tumor of the Lymphatic System, Blood Making Tissue with Surgery (Splenectomy, Radical Resection) with Age greater than 15

	Major Diagnosis Category		Diagnosis-Related Groups
11:	Benign Neoplasm of the Female Reproductive System	053	Benign Tumor (Papilloma, Polyp) of the Uterus, Vagina, Vulva without Secondary Diagnosis
		054	Benign Tumor (Papilloma, Polyp) of the Uterus, Vagina, Vulva with Secondary Diagnosis
		055	Benign Tumor (Fibroma) of the Uterus, Ovary without Surgery
		056	Benign Tumor (Fibroma) of the Uterus, Ovary with Surgical Procedure (D&C, Excision of Lesion) without Second Surgery
		057	Benign Tumor (Fibroma) of the Uterus, Ovary with Second Surgery
		058	Benign Tumor (Fibroma) of the Uterus, Ovary with Surgery (Removal of Ovary)
		059	Benign Tumor (Fibroma) of the Uterus, Ovary with Surgery (Removal of Uterus)
12:	Benign Neoplasm of Other Sites	060	Benign Tumor of the Intestines, Urinary System, without Surgery
		061	Benign Tumor of the Brain, Pituitary Gland without Surgery
		062	Benign Tumor of the Skin, Bone, Urinary System (Kidney, Bladder), Connective Tissue with Surgery without Secondary Diagnosis
		063	Benign Tumor of the Skin, Bone, Urinary System (Kidney, Bladder), Connective Tissue with Surgery with Secondary Diagnosis with Age less than 43
		064	Benign Tumor of the Skin, Bone, Urinary System (Kidney, Bladder), Connective Tissue with Surgery with Secondary Diagnosis with Age greater than 42
		065	Benign Tumor of the Intestines, Nerves with Surgical Procedure (Excision, Other) without Secondary Diagnosis
		066	Benign Tumor of the Intestines, Nerves with Surgical Procedure (Excision, Other) with Secondary Diagnosis
		067	Benign Tumor of the Intestines, Nerves with Surgery (Colon Resection, Craniotomy Radical Resection, Other Major Operation)
		068	Benign Tumor of the Stomach, Brain, Respiratory System, Esophagus, Pituitary Gland with Surgery
13:	Diseases of Thyroid and Other Endocrine Glands	069	Disease of the Thyroid (Non-Toxic, Simple), Other Endocrine Glands (Adrenal, Pancreas) without Surgery
		070	Disease of the Thyroid (Toxic), Low Function Pituitary without Surgery
		071	Endocrine Disorder with Surgical Procedure (Thyroidectomy, Other)
		072	Endocrine Disorder with Surgery
14:	Diabetes	073	Diabetes without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis with Age less than 36
		074	Diabetes without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis with Age greater than 35
		075	Diabetes without Surgery with Major Secondary Diagnosis
		076	Diabetes with Surgical Procedure (Endoscopy, Biopsy)
		077	Diabetes with Surgery (Amputation of Extremity, Other Major)
15:	Nutritional and Other Metabolic Diseases	078	Metabolic Disorder (Gout, Blood Globulin) without Secondary Diagnosis
		079	Metabolic Disorder (Gout, Blood Globulin) with Secondary Diagnosis (Nutrition Deficiency)
		080	Metabolic Disease (Cystic Fibrosis, Sprue, Unspecified)
		081	Metabolic disease (Obesity, Malabsorption, Unspecified)
16:	Diseases of the Blood and Blood Forming Organs	082	Mediterranean Anemia, Hemophilia without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis with Age less than 11

Major Diagnosis Category		Diagnosis-Related Groups	
		083	Mediterranean Anemia, Hemophilia without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis with Age greater than 10
		084	Disease of Blood Hemoglobin without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
		085	Disease of the Blood (Anemias), Blood Forming Organs (Spleen) without Surgery with Major Secondary Diagnosis
		086	Disease of the Blood (Anemias), Blood Forming Organs with Surgery with Age 2–52
		087	Disease of the Blood (Anemias), Blood Forming Organs with Surgery with Age less than 1 or greater than 53
17:	Psychoses Not Attributed to Physical Conditions	088	Schizophrenia (Paranoid, Catatonic, Unspecified) Involutional Melancholia with Psychiatric Service
		089	Schizophrenia (Paranoid, Catatonic, Unspecified) Involutional Melancholia without Psychiatric Service
		090	Schizophrenia (Affective, Acute Episode), Manic-Depressive Psychosis
18:	Neuroses	091	Neurosis (Anxiety, Hysterical, Phobic, Hypochandriacal Unspecified)
		092	Neurosis (Obsessive-Compulsive, Depressive), Personality-Disorders
19:	Alcoholic Mental Disorder and Addiction	093	Alcoholism without Secondary Diagnosis or with Minor Secondary Diagnosis
		094	Alcoholism with Major Secondary Diagnosis (Liver Cirrhosis, Delirium Tremens, Other)
20:	Other Mental Disorders	095	Drug Dependence, Physical Disorder (Probably Psychiatric Origin), Cephalgia
		096	Psychosis, Non-Psychosis Related Brain Condition
21:	Diseases of the Central Nervous System	097	Epilepsy, Migraine, Brain Disorder (Unspecified) without Surgery without Secondary Diagnosis
		098	Epilepsy, Migraine, Brain Disorder (Unspecified) without Surgery with Secondary Diagnosis
		099	Multiple Sclerosis, Paralysis Agitans, Meningitis, Hemiplegia without Surgery
		100	Disease of the Central Nervous System with Surgical Procedure (Nerve Block, Other)
		101	Disease of the Central Nervous System with Surgery (Laminectomy, Spinal Fusion Ventricular Shunt)
22:	Diseases of the Peripheral Nervous System	102	Facial Paralysis, Neuralgia (Trigeminal, Other Unspecified) without Surgery
		103	Sciatica, Polyneuritis without Surgery
		104	Disease of the Median Nerve with Surgery
		105	Disease of the Peripheral Nerves except Median with Surgical Procedure (Nerve Block, Other Unspecified)
		106	Disease of the Peripheral Nerves except Median with Surgery (Spinal Cord, Nerve Roots)
23:	Diseases of the Eye	107	Cross Eyedness, Cataract, Cyst of the Eyelid without Surgery
		108	Glaucoma, Corneal Inflammation/Ulceration, Disease of the Iris, Retina without Surgery
		109	Disease of the Eye with Surgical Procedure (Muscle Repair of Eyelid, Other)
		110	Disease of the Eye with Surgical Procedure (Removal of Lens, Incision into Sclera
		111	Disease of the Eye with Surgical Procedure (Reattachment of Retina, Repair of Cornea)

	Major Diagnosis Category		Diagnosis-Related Groups
24:	Disease of the Ear and Mastoid Process	112	Disease of the Middle Ear (Inflammation, Chronic Mastoid Bone Inflammation) without Surgery
		113	Disease of the Inner Ear (Inflammation, Menieres Disease) without Surgery
		114	Disease of the Ear with Surgical Procedure (Incision of Membrane, Removal of Adenoids, Other)
		115	Disease of the Middle Ear with Surgery (Removal of Bone, Repair of Membrane)
		116	Disease of the Ear with Surgery (Removal of Mastoid Bone, Excision of Middle Ear, Other)
25:	Hypertensive Heart Diseases	117	Hypertensive Heart Disease without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
		118	Hypertensive Heart Disease without Surgery with Major Secondary Diagnosis
		119	Hypertensive Heart Disease (Fatal) with Kidney Involvement without Surgery with Major Secondary Diagnosis
		120	Hypertensive Heart Disease with Surgery
26:	Acute Myocardial Infarction	121	Disease of the Heart—Acute Myocardial Infarction
27:	Ischemic Heart Diseases Except AMI	122	Disease of the Heart, Ischemia (Blood Deficiency) except AMI without Surgery without Secondary Diagnosis
		123	Disease of the Heart, Ischemia (Blood Deficiency) except AMI without Surgery with Minor Secondary Diagnosis
		124	Disease of the Heart, Ischemia (Blood Deficiency) except AMI without Surgery with Major Secondary Diagnosis
		125	Disease of the Heart, Ischemia (Blood Deficiency) except AMI with Cardiac Catheterization
		126	Disease of the Heart, Ischemia (Blood Deficiency) except AMI with Surgical Procedure (Endoscopy, Insertion of Electronic Device)
		127	Disease of the Heart, Ischemia (Blood Deficiency) except AMI with Surgery (Shunt, Other Major)
28:	Arrhythmia and Slowed Conduction	128	Disease of the Heart, Irregular Heart Rhythm, Slowed Conduction without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
		129	Disease of the Heart, Irregular Heart Rhythm, Slowed Conduction without Surgery with Major Secondary Diagnosis
		130	Disease of the Heart, Irregular Heart Rhythm, Slowed Conduction with Replacement of Heart Device or Cardiac Catheterization
		131	Disease of the Heart, Irregular Heart Rhythm, Slowed Conduction with Insertion of Electronic Heart Device
29:	Heart Failure	132	Disease of the Heart, Failure (Poor Function) without Surgery
		133	Disease of the Heart, Failure (Poor Function) with Surgery
30:	Carditis, Valvular and Other Diseases	134	Disease of the Heart, Inflammation, Valve Problem without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
		135	Disease of the Heart, Inflammation, Valve Problem without Surgery with Major Secondary Diagnosis
		136	Disease of the Heart, Inflammation, Valve Problem with Cardiac Catheterization without Secondary Diagnosis or with Minor Secondary Diagnosis
		137	Disease of the Heart, Inflammation, Valve Problem with Cardiac Catheterization with Major Secondary Diagnosis

Major Diagnosis Category		Diagnosis-Related Groups
	138	Disease of the Heart, Inflammation, Valve Problem with Surgery (Valve Replacement, Other Major)
31: Cerebrovascular Diseases	139	Circulatory Disorder of the Brain, Occasional Blood Deficiency without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
	140	Circulatory Disorder of the Brain, Occasional Blood Deficiency without Surgery with Major Secondary Diagnosis
	141	Blood Clot in Brain Obstructing Circulation without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
	142	Blood Clot in Brain Obstructing Circulation without Surgery with Major Secondary Diagnosis
	143	Brain Hemorrhage (Stroke) without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
	144	Brain Hemorrhage (Stroke) without Surgery with Major Secondary Diagnosis
	145	Circulatory Dysfunction in Brain with Surgery
32: Diseases of the Vascular System	146	Disease of the Circulatory System, Inflammation of the Lymph Glands, Varicose Veins (Legs), Raynauds Disease without Surgery
	147	Disease of the Circulatory System (Hardening of Arterial Walls, Arterial Blood Clot) without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
	148	Disease of the Circulatory System (Hardening of Arterial Walls, Arterial Blood Clot) without Surgery with Major Secondary Diagnosis
	149	Disease of the Circulatory System with Surgical Procedure (Excision of Varicose Veins, Other) with Age less than 51
	150	Disease of Vascular System with Surgery (Excision of Varicose Veins, Other) with Age greater than 50
	151	Disease of Vascular System with Surgery (Excision of Nerve, Vessel) without Secondary Diagnosis
	152	Disease of Vascular System with Surgery (Excision of Nerve, Vessel) with Secondary Diagnosis
	153	Disease of Vascular System with Surgery (Arterial Reconstruction, Amputation of Extremity)
33: Pulmonary Embolism	154	Blood Clot of the Lung without Secondary Diagnosis or with Minor Secondary Diagnosis
	155	Blood Clot of the Lung with Major Secondary Diagnosis
34: Phlebitis and Thrombophlebitis	156	Inflammation of the Veins, Blood Clot without Secondary Diagnosis or with Minor Secondary Diagnosis
	157	Inflammation of the Veins, Blood Clot with Major Secondary Diagnosis
35: Hemorrhoids	158	Hemorrhoids
36: Hypertrophy of Tonsil and Adenoid	159	Enlargement of the Tonsils/Adenoids
37: Acute Upper Respiratory Tract Infection and Influenza	160	Acute Upper Respiratory Tract Infection, Influenza with Age less than 45
	161	Acute Upper Respiratory Tract Infection, Influenza with Age greater than 44
38: Other Diseases of the Upper Respiratory Tract	162	Disease of the Upper Respiratory Tract except Acute Upper Respiratory Infection and Influenza without Surgery

Major Diagnosis Category			Diagnosis-Related Groups
		163	Disease of the Upper Respiratory Tract with Surgical Procedure (Biopsy, Visualizations of the Nasal Septum)
		164	Disease of the Upper Respiratory Tract with Surgery (Nose Reconstruction, Incision and Drainage of Sinus)
39:	Pneumonia	165	Pneumonia with Age less than 31
		166	Pneumonia without Surgery without Secondary Diagnosis with Age greater than 30
		167	Pneumonia without Surgery with Secondary Diagnosis with Age greater than 30
		168	Pneumonia with Surgery
40:	Bronchitis	169	Bronchitis with Age less than 46
		170	Bronchitis without Secondary Diagnosis or with Minor Secondary Diagnosis with Age greater than 45
		171	Bronchitis with Major Secondary Diagnosis with Age greater than 45
41:	Asthma	172	Asthma with Age less than 31
		173	Asthma without Secondary Diagnosis with Age greater than 30
		174	Asthma with Secondary Diagnosis with Age greater than 30
42:	Other Lung and Pleural Diseases	175	Lung Collapse, Pleurisy, Pulmonary congestion without Surgery
		176	Emphysema, Empyema, Abscess, Acute Swelling without Surgery without Secondary Diagnosis or with Minor Secondary Diagnosis
		177	Emphysema, Empyema, Abscess, Acute Swelling without Surgery with Major Secondary Diagnosis
		178	Disease of the Lung and Pleura with Surgical Procedure (Bronchoscopy, Chest Incision, Other) without Secondary Diagnosis
		179	Disease of the Lung and Pleura with Surgical Procedure (Bronchoscopy, Chest Incision, Other) with Secondary Diagnosis
		180	Disease of the Lung and Pleura with Surgery (Removal of Lobe; Other Major)
43:	Diseases of the Oral Cavity, Salivary Glands and Jaw	181	Minor Problems of the Teeth
		182	Major Problems of the Teeth (Jaw, Salivary Glands, Other Oral Soft Tissue)
44:	Gastric and Peptic Ulcer	183	Stomach Ulcer without Surgery without Secondary Diagnosis
		184	Stomach Ulcer without Surgery with Secondary Diagnosis
		185	Stomach Ulcer with Surgical Procedure (Biopsy, Visualization, Other)
		186	Stomach Ulcer with Surgery (Removal of Portion of Stomach, Other Major) without Secondary Diagnosis
		187	Stomach Ulcer with Surgery (Removal of Portion of Stomach, Other Major) with Secondary Diagnosis
45:	Upper Gastro-Intestinal Diseases except Gastric and Peptic Ulcer	188	Upper GI Disease Except Stomach Ulcer without Surgery without Secondary Diagnosis
		189	Upper GI Disease Except Stomach Ulcer without surgery with Secondary Diagnosis
		190	Upper GI Disease Except Stomach Ulcer with Surgical Procedure (Visualization, Other Minor) without Secondary Diagnosis
		191	Upper GI Disease Except Stomach Ulcer with Surgical Procedure (Visualization, Other Minor) with Secondary Diagnosis
		192	Upper GI Disease Except Stomach Ulcer with Surgery
46:	Appendicitis	193	Appendicitis (without Peritonitis) without Secondary Diagnosis
		194	Appendicitis (without Peritonitis) with Secondary Diagnosis
		195	Appendicitis (with Peritonitis, Other) without Secondary Diagnosis

Major Diagnosis Category		Diagnosis-Related Groups
	196	Appendicitis (with Peritonitis, Other) with Secondary Diagnosis
47: Hernia of the Abdominal Cavity	197	Abdominal Hernia with Age less than 15
	198	Inguinal Hernia (without Obstruction) with Age greater than 14 and less than 65 without Secondary Diagnosis
	199	Inguinal Hernia (without Obstruction) with Age greater than 14 and less than 65 with Secondary Diagnosis
	200	Abdominal Hernia Except Simple Inguinal with Age greater than 14 and less than 65 without Surgery
	201	Abdominal Hernia Except Simple Inguinal with Age greater than 14 and less than 65 with Minor Surgery
	202	Abdominal Hernia Except Simple Inguinal with Age greater than 14 and less than 65 with Major Surgery
	203	Abdominal Hernia with Age greater than 64 without Surgery
	204	Abdominal Hernia with Age greater than 64 with Minor Surgery
	205	Abdominal Hernia with Age greater than 64 with Major Surgery
48: Enteritis, Diverticula, and Functional Disorders of the Intestine	206	Functional Disorder of the Intestine without Surgery
	207	Intestinal Pouching, Regional Enteritis, Ulcerative Colitis without Surgery
	208	Intestinal Pouching (Functional Disorder) with Minor Surgery without Secondary Diagnosis
	209	Intestinal Pouching (Functional Disorder) with Minor Surgery with Secondary Diagnosis
	210	Intestinal Pouching (Functional Disorder) with Major Surgery (Resection, Other)
49: Diseases of the Anus	211	Disease of the Anus without Secondary Diagnosis
	212	Disease of the Anus with Secondary Diagnosis
50: Miscellaneous Diseases of the Intestine and Peritoneum	213	Miscellaneous Disease of the Intestine and Abdominal Lining with Age less than 56 without Surgery
	214	Miscellaneous Disease of the Intestine and Abdominal Living with Age greater than 55 without Surgery without Secondary Diagnosis
	215	Miscellaneous Disease of the Intestine and Abdominal Lining with Age greater than 55 without Surgery with Secondary Diagnosis
	216	Miscellaneous Disease of the Intestine and Abdominal Lining with Surgical Procedure (Local Incision, Excision)
	217	Miscellaneous Disease of the Intestine and Abdominal Lining with Visualization of the Intestine without Secondary Diagnosis
	218	Miscellaneous Disease of the Intestine and Abdominal Lining with Visualization of the Intestine with Secondary Diagnosis
	219	Miscellaneous Disease of the Intestine and Abdominal Lining with Major Surgery without Secondary Diagnosis
	220	Miscellaneous Disease of the Intestine and Abdominal Lining with Major Surgery with Secondary Diagnosis
51: Diseases of the Liver	221	Hepatitis, (Infectious, Serum) Subacute Necrosis of the Liver with Age less than 41
	222	Hepatitis (Infectious, Serum) Subacute Necrosis of the Liver with Age greater than 40
	223	Liver Cirrhosis without Secondary Diagnosis or with Minor Secondary Diagnosis
	224	Liver Cirrhosis with Major Secondary Diagnosis
52: Diseases of the Gallbladder and Bile Duct	225	Disease of the Gallbladder and Bile Duct without Surgery with Age less than 51
	226	Disease of the Gallbladder and Bile Duct without Surgery with Age greater than 50

Major Diagnosis Category		Diagnosis-Related Groups	
		227	Disease of the Gallbladder and Bile Duct with Surgery without Secondary Diagnosis
		228	Disease of the Gallbladder and Bile Duct with Surgery with Secondary Diagnosis with Age less than 65
		229	Disease of the Gallbladder and Bile Duct with Surgery with Secondary Diagnosis with Age greater than 64
53:	Diseases of the Pancreas	230	Disease of the Pancreas without Surgery
		231	Disease of the Pancreas with Surgery
54:	Diseases of the Kidney and Ureter	232	Disease of the Kidney and Bladder without Surgery without Secondary Diagnosis
		233	Kidney Inflammation without Surgery with Secondary Diagnosis
		234	Nephrotic Syndrome, Nephritis (Chronic) Uremia without Surgery with Secondary Diagnosis with Age less than 65
		235	Nephrotic Syndrome, Nephritis (Chronic) Uremia without Surgery with Secondary Diagnosis with Age greater than 64
		236	Disease of the Ureter Nephrotic Syndrome, with Surgical Procedure (Cystoscopy, Biopsy, Other Minor)
		237	Kidney Inflammation and Degenerative Disease (Including Kidney Pelvis) with Surgical Procedure
		238	Disease of the Kidney and Ureter with Surgery (Kidney Removal, Kidney Transplant, Other Major)
55:	Urinary Calculus	239	Urinary Stone without Surgery without Secondary Diagnosis
		240	Urinary Stone without Surgery with Secondary Diagnosis
		241	Urinary Stone with Surgical Procedure (Visualization, Catheter to Kidney, Other)
		242	Urinary Stone with Surgery (Incision and Drainage of Kidney, Bladder, Ureter and Other Major)
56:	Cystitis and Other Urinary Diseases	243	Bladder Inflammation with Other Urinary Disease without Surgery without Secondary Diagnosis
		244	Inflammation of the Bladder and Urethra with Narrowing of the Urethra without Surgery with Secondary Diagnosis
		245	Bladder (Abnormal Passage, Pouching, Other Disease) without Surgery with Secondary Diagnosis with Age less than 46
		246	Bladder (Abnormal Passage, Pouching, Other Disease) without Surgery with Secondary Diagnosis with Age greater than 45
		247	Disease of the Bladder and Urethra with Surgical Procedure (Visualization, Opening)
		248	Disease of the Bladder and Urethra with Surgical Procedure (Visualization, Excision, Dilatation, Repair) with Age less than 15
		249	Disease of the Bladder and Urethra with Surgical Procedure (Visualization, Excision, Dilatation, Repair) with Age greater than 14
		250	Disease of the Bladder and Urethra with Surgery (Removal of Bladder, Removal of Prostate, Other Major)
57:	Disease of the Prostate	251	Disease of the Prostate without Surgery
		252	Disease of the Prostate with Surgical Procedure (Bladder Visualization, Dilatation of Urethra, Biopsy) without Secondary Diagnosis
		253	Disease of the Prostate with Surgical Procedure (Bladder Visualization, Dilatation of Urethra, Biopsy) with Secondary Diagnosis
		254	Disease of the Prostate with Surgery (Non-Incisional Removal of Prostate) without Secondary Diagnosis
		255	Disease of the Prostate with Surgery (Non-Incisional Removal of Prostate) with Secondary Diagnosis

Major Diagnosis Category			Diagnosis-Related Groups
		256	Disease of the Prostate with Surgery (Incisional removal of the Prostate)
58:	Disease of the Male Reproductive System	257	Excessive Foreskin over the Glans Penis with Surgery
		258	Disease of the Male Reproductive System Except Circumcision without Surgery
		259	Disease of the Male Reproductive System Except Circumcision with Surgery with Age less than 45
		260	Disease of the Male Reproductive System Except Circumcision with Surgery with Age greater than 44
59:	Disease of the Female Reproductive System	261	Disorder of Menstruation without Surgery
		262	Disease of the Female Reproductive System Except Disorder of Menstruation without Secondary Diagnosis
		263	Disease of the Female Reproductive System Except Disorder of Menstruation with Secondary Diagnosis
		264	Disease of the Female Reproductive System with Surgical Procedures (D&C, Visualization, Removal Fallopian Tubes) without Secondary Diagnosis
		265	Disease of the Female Reproductive System with Surgical Procedure (D&C, Visualization, Other) with Secondary Diagnosis
		266	Disease of the Female Reproductive System with Surgery (Removal of Womb, Repair of Female Reproductive Organ, Other Major)
60:	Diseases of the Breast	267	Benign Breast Tumor, Chronic Cystic Disease without Secondary Diagnosis
		268	Acute Inflammation of the Breast, Enlarged Breast without Secondary Diagnosis
		269	Disease of the Breast with Secondary Diagnosis with Age less than 56
		270	Disease of the Breast with Secondary Diagnosis with Age greater than 55
61:	Abortion	271	Abortion with Secondary Diagnosis
		272	Abortion with Secondary Diagnosis
62:	Obstetrical Diseases of the Antepartum and Puerperium	273	False Labor without Surgery
		274	Threatened Abortion Premature Separation of the Afterbirth, Other Hemorrhage During Pregnancy without Surgery
		275	Obstetrical Complications, Poisons in Blood, Excessive Vomiting, Blood Clot Vein-Extremity without Surgery
		276	Obstetrical Disease Before the After Delivery with Surgical Procedure (D&C, Repair of Neck of Womb)
		277	Obstetrical Disease Before and After Delivery with Surgery (Removal of Tubes and Ovaries, Other Major)
63:	Normal Delivery	278	Delivery without Surgery or with Surgery Assisting Delivery
		279	Delivery with Tying of Tubes, Removal of Tubes
		280	Delivery with Cesarean Section
64:	Delivery with Complications	281	Delivery with Complications without Surgery or with Surgery Assisting Delivery
		282	Delivery with Complications with Cesarean Section
65:	Diseases of the Skin and Subcutaneous Tissue	283	Excessive Scar Tissue, Excessive Pigment, Fatty Cyst, Other Minor Skin Disease without Secondary Diagnosis
		284	Excessive Scar Tissue, Excessive Pigment, Fatty Cyst, Other Minor Skin Disease with Secondary Diagnosis
		285	Skin Inflammation, Abscess, Eczema, Chronic Ulcer without Surgery with Age less than 21
		286	Skin Inflammation, Abscess, Eczema, Chronic Ulcer without Surgery with Age greater than 20

	Major Diagnosis Category		Diagnosis-Related Groups
		287	Skin Inflammation, Abscess, Eczema, Reddened Skin with Surgery without Secondary Diagnosis
		288	Skin Inflammation, Abscess, Eczema, Reddened Skin with Surgery with Secondary Diagnosis
		289	Psoriasis, Eruptive Skin Lesions, Chronic Skin Ulcer
66:	Arthritis	290	Arthritis without Surgery with Age less than 65
		291	Arthritis without Surgery with Age greater than 64
		292	Arthritis with Surgery (Excision of Bone, Joint, Membrane Surgical Joint Fixation)
		293	Arthritis with Surgery (Joint Incision, Spinal Fusions, Excision of Tissue Between Vertebrae)
		294	Arthritis with Surgery (Repair and Restoration of Joint, Removal of Membrane between Vertebrae)
67:	Derangement and Displacement of Intervertebral Disc	295	Disorder and Displacement of disc Between Vertebrae without Surgery
		296	Disorder and Displacement of Disc Between Vertebrae with Surgery
68:	Diseases of the Bone and Cartilage	297	Rheumatism and Inflammation Tissue Covering Bone, Other Minor Bone Disease without Surgery
		298	Disease of the Bone, Inflammation of Marrow (Acute, Chronic), pongy Bone, Unaided Fracture without Surgery
		299	Disease of the Bone, and Bone Tissue Lining, with Surgery (Excision Bone Lining, Repair of Other Joint)
		300	Disease of the Bone and Bone Tissue Lining with Surgery (Joint Incision, Bone Excision, Bone Fusion)
		301	Disease of the Bone and Bone Tissue Lining with Surgery (Amputation, Hip Restoration, Other Major)
69:	Other Disease of the Musculo-Skeletal System	302	Inflammation of the Component Parts of the Joints, Curvature of the Spine, Deformed Foot without Surgery
		303	Backache, Diffuse Disease of Connective Tissue, Inflammation of Muscle without Surgery without Secondary Diagnosis
		304	Backache, Diffuse Disease of Connective Tissue, Inflammation of Muscle without Surgery with Secondary Diagnosis
		305	Inflammation of the Component Parts of Joints with Deformity (Palm, Finger, Toe) with Surgery
		306	Other Disease of the Muscle and Bone (Major) with Surgical Procedure
		307	Other Disease of the Muscle and Bone (Major) with Surgery (Removal, Repair of the Small Joint, Bone)
		308	Other Disease of the Muscle and Bone (Major) with Surgery (Joining Vertebrae, Other)
70:	Congenital Anomalies	309	Birth Defect (Bone, Stomach, Testicle) without Surgery
		310	Birth Defect (Heart, Kidney, Other Major) without Surgery
		311	Birth Defect (Testicle, Skin, Stomach, Other Minor) with Surgery
		312	Birth Defect (Heart Valve, Other Unspecified Heart Site) with Surgical Procedure (Cardiac Catheterization)
		313	Birth Defect (Palate, Lip, Hip or Other Extremity) with Surgery (Repair of Mouth, Fixation of Hip)
		314	Birth Defect (Heart Valve, Other Unspecified Site) with Surgery (Heart Valve, Septal Repair)
		315	Congenital Diseases (Tetralogy of Fallot, Atrial Septal Defect, Hypospadia, Other) with Surgical Procedure (Catheterization, Repair of Urethra)
		316	Congenital Diseases (Tetralogy of Fallot, Atrial Septal Defect, Other) with Surgery (Valve, Septum, Shunt)

	Major Diagnosis Category		Diagnosis-Related Groups
		317	Birth Defect (Spine, Gullet, Large Bowel) with Surgery
71:	Normal Mature Newborn	318	Normal Full Term Newborn
72:	Certain Diseases and Conditions Peculiar to Newborn Infants	319	Well Baby Care (Pregnancy greater than 9 months), Other Minor Disease or Condition of the Newborn Infant
		320	Immaturity, Hyaline Membrane Disease, Other Major Disease or Condition of the Infant without Secondary Diagnosis
		321	Immaturity, Hyaline Membrane Disease, Other Major Disease or Condition of the Infant with Secondary Diagnosis
73:	Signs and Symptoms Pertaining to the Nervous, Respiratory, and Circulatory Systems	322	Indications of Nervous, Respiratory, Circulatory System Disease without Surgery without Secondary Diagnosis
		323	Convulsions, Fainting, Nosebleed, Chest Pain without Surgery with Secondary Diagnosis
		324	Brain Disorder of Dizziness, Shortness of Breath, Coughing up Blood without Surgery with Secondary Diagnosis
		325	Indications of Nervous, Respiratory, Circulatory System Disease with Surgical Procedure
		326	Indications of Nervous, Respiratory, Circulatory System Disease with Major Surgery
74:	Signs and Symptoms Pertaining to the Gastro-Intestinal and Urinary Systems	327	Indications of Gastro-Intestinal, Urinary System Disease without Surgery without Secondary Diagnosis
		328	Indications of Gastro-Intestinal, Urinary System Disease without Surgery with Secondary Diagnosis
		329	Indications of Gastro-Intestinal, Urinary System Disease with Surgical Procedure (Visual Inspection, Other)
		330	Indications of Gastro-Intestinal, Urinary System Disease with Surgery (Abdominal, Other Major)
75:	Miscellaneous Signs, Symptoms, and Ill-Defined Conditions	331	Sterility (Male, Female), Admission for Observation without Surgery
		332	Chemical Imbalance, Headache, Fever, Other Ill-Defined Indication of Disease without Surgery with Age less than 15
		333	Chemical Imbalance, Headache, Fever, Other Ill-Defined Indication of Disease without Surgery with Age greater than 14
		334	Miscellaneous Indication of Disease with Surgical Procedure (Visual Inspection, Other)
		335	Miscellaneous Indication of Disease with Surgery (Abdominal Surgery, Removal of Uterus, Other Major)
76:	Fractures	336	Fracture (Skull, Face, Forearm, Leg, Foot, Hand) without Surgery with Age less than 30
		337	Fracture (Skull, Face, Forearm, Leg, Foot, Hand) without Surgery with Age greater than 29
		338	Fracture (Spine, Ribs, Bone of the Upper Arm) without Surgery with Age less than 65
		339	Fracture (Spine, Ribs, Bone of the Upper Arm) without Surgery with Age greater than 64
		340	Fracture (Thigh Bone, Pelvis, Multiple) without Surgery
		341	Fracture (Nose, Forearm, Hand, Lower Leg, Foot) with Surgical Procedure (Closed Reduction) without Secondary Diagnosis
		342	Fracture (Nose, Forearm, Hand, Lower Leg, Foot) with Surgical Procedure (Closed Reduction) with Secondary Diagnosis
		343	Fracture (Lower Jaw, Upper Arm, Ankle) with Surgical Procedure (Closed Reduction, Open Reduction of Face) without Secondary Diagnosis
		344	Fracture (Lower Jaw, Upper Arm, Ankle) with Surgical Proce-

Major Diagnosis Category		Diagnosis-Related Groups	
			dure (Closed Reduction, Open Reduction of Face) with Secondary Diagnosis
		345	Fracture (Arm, Hand, Foot, Shoulder Blade) with Surgery (Open Reduction, External Fixation, Other)
		346	Fracture (Ankle, Leg Bones) with Surgery (Open Reduction, External Fixation, Other)
		347	Fracture (Thigh Bone, Pelvis) with Surgery (Open Reduction, External Fixation, Other)
		348	Fracture with Major Surgery (Amputation, Restoration of Hip Joint, Other Major)
77:	Dislocations and Other Musculo-Skeletal Injuries	349	Dislocation (Shoulder, Elbow, Wrist, Knee), Sprains (Ankle, Foot, Hand) without Surgery
		350	Dislocation (Jaw, Hip), Sprains (Knee, Sacroiliac, Other Unspecified) without Surgery
		351	Dislocation (Shoulder, Elbow, Hand), Sprains (Elbow, Wrist, Hand) with Surgery
		352	Dislocation (Knee, Ankle), Sprains (Shoulder, Knee, Ankle) with Surgery
		353	Dislocation (Hip, Multiple), Sprains (Hip, Sacroiliac, Other Unspecified) with Surgery
78:	Internal Injuries of the Cranium, Chest, and Other Organs	354	Internal Injury of the Skull, Other Organ without Surgery without Secondary Diagnosis with Age less than 41
		355	Internal Injury of the Skull, Other Organ without Surgery with Secondary Diagnosis with Age less than 41
		356	Internal Injury of the Skull, Other Organ without Surgery with Age greater than 40
		357	Internal Injury with Surgical Procedure (Suture of Skin, Nerve, Nerve Repair, Other)
		358	Internal Injury with Surgery (Removal of Spleen, Drainage of Chest Cavity, Excision of Skin)
		359	Internal Injury with Surgery (Opening of Skull, Exploration of Abdominal Cavity)
79:	Open Wounds and Superficial Injuries	360	Open Wound (Uncomplicated), Superficial Injury, Foreign Body without Surgery
		361	Open Wound (Complicated), Bruise, Multiple Injuries without Surgery without Secondary Diagnosis
		362	Open Wound (Complicated), Bruise, Multiple Injuries without Surgery with Secondary Diagnosis
		363	Open Wound (External), Foreign Body with Surgical Procedure (Visualization, Suturing, Other)
		364	Open Wound (Complicated) of the Head, Multiple Sites with Surgical Procedure (Visualization, Suturing, Other)
		365	Open Wound (External), Superficial Injury with Surgery (Excision, Other Major)
		366	Open Wound (Complicated) of the Head, Multiple Sites with Surgery (Excision, Other Major)
80:	Burns	367	Burn of the 1st Degree (Uncomplicated) Covering less than 20% of the Body
		368	Burn of the 2nd Degree (Complicated), 3rd Degree Covering more than 20% of the Body
81:	Complications of Medical and Surgical Care	369	Complications of Medical or Surgical Care without Surgery without Secondary Diagnosis
		370	Complications of Medical or Surgical Care without Surgery with Secondary Diagnosis

Major Diagnosis Category		Diagnosis-Related Groups
	371	Complications of Medical or Surgical Care with Surgical Procedure
	372	Complications of Medical or Surgical Care with Surgery (Replacement of Heart Device, Repair of Stomach)
	373	Complications of Medical or Surgical Care with Surgery (Revision of Shunt, Other Major)
22: Adverse Effects of Certain Substances	374	Adverse Effect of a Drug, Toxic Effect of Alcohol without Secondary Diagnosis
	375	Adverse Effect of a Drug, Toxic Effect of Alcohol with Secondary Diagnosis
	376	Toxic Effect (Lead, Acid, Alkali, Carbon Monoxide, Radiation) without Secondary Diagnosis
	377	Toxic Effect (Lead, Acid, Alkali, Carbon Monoxide, Radiation) with Secondary Diagnosis
23: Special Admissions and Examinations without Reported Diagnoses	378	Prenatal Care, Medical and Surgical after Care (Dialysis) without Surgery
	379	Admission for Sterilization, Chemotherapy, Radiation Therapy without Surgery
	380	Follow up (Cancer) Surgery, Medical after Care (Colostomy, Orthopedic, Other) without Surgery
	381	Special Admission with Surgery (Sterilization, D&C, Other)
	382	Special Admission with Surgical Procedure (Bladder Visualization, Removal of Fixed Internal Device)
	383	Special Admission with Surgery (Exploration of Abdominal Cavity, Removal of Uterus, Other Major)

MANAGEMENT SCIENCE
Vol. 22, No. 4, December, 1975
Printed in U.S.A.

A COMPUTER SIMULATION MODEL FOR THE CONTROL OF RABIES IN AN URBAN AREA OF COLOMBIA*†

RALPH R. FRERICHS‡ AND JUAN PRAWDA§

A simulation model is developed describing the transmission of canine rabies within and between 116 spatially distributed barrios (neighborhoods) in Cali, Colombia. The discrete time, dynamic model considers both discrete random variables (incubation and infective periods, appearance and movement of rabid dogs through the city, etc.) and deterministic variables (demographic components of barrio canine populations). Values for the input variables were acquired through field observations, other Colombian sources, and a review of the literature. Various canine vaccination strategies were tested in the model over a ten-year planning horizon for their cost-effectiveness with regard to the prevention of canine rabies. The model is recommended to the Pan American Health Organization to be used as an interactive gaming model to aid health system managers in Cali, Colombia and in other Latin American cities in scheduling the time and locations of vaccination teams in a more cost-effective manner.

1. Introduction

The problem of rabies has been with man for many years. Celsus, as far back as the first century A.D., described the infection in man and observed that the disease is transmitted by the bite of a rabid animal. Since then, it has been shown that rabies is a viral infection of the nervous system primarily of warm-blooded animals, and is invariably fatal once the virus has infected the body [20], [33].

In the last twenty years, the United States has rid its urban dog populations of this disease but the problem still lingers in wildlife, namely skunks and foxes [35]. Since man has his closest contact with urban animals, the elimination of rabies from dogs has reduced human rabies in the United States to one or two cases per year. Unfortunately, South and Central American countries have not had similar success [25]. Since other diseases compete for allocations within the health budget, rabies control programs have not always received the funds necessary to eliminate the problem in urban areas [2].

The World Health Organization feels that the vaccination of dogs is one of the most important methods for controlling rabies in cities and many immunization campaigns have proven the efficacy of this approach [41]. In addition, they recommend that stray dogs be euthanized in order to decrease the canine density in the streets to such a point that there is little opportunity for rabies transmission. The collection and subsequent killing of stray dogs, however, is highly unpopular as evidenced by the disdain shown throughout the world toward the neighborhood dog catcher. The assumption is made by the authors that in Colombia, a democracy with a history of internal violence, a canine euthanasia campaign would be politically unacceptable. Therefore, the scope of our analytical effort is limited to a control program featuring different vaccination policies.

* Processed by Professor William W. Cooper, Departmental Editor for Public Administration and Associate Editor E. S. Savas; received October 1974.

† This research study was supported by the Tulane University International Center for Medical Research and Training, Grant AI-10050 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, United States Public Health Service, Washington, D. C.

‡ Louisiana State University Medical School.

§ National Company of Basic Commodities (CONASUPO), Mexico and National Polytechnic Institute, Mexico.

When planning immunization programs for a given time period, the question arises as to what percent of a population should be vaccinated and how often the immunization campaigns are to be repeated. The probability of rabies transmission decreases when the number of immunized dogs in an area increases. However, if too few dogs are vaccinated, the risk of a rabies epidemic will remain high. Even after a successful vaccination program, the risk of a rabies epidemic will slowly start to increase, since new susceptible puppies are born into the population and old immune dogs either die or lose their immunity. Dogs in the poorer areas of a city have higher birth and death rates than dogs in wealthier sections. Consequently the rate of turnover from immune to susceptible varies from neighborhood to neighborhood, creating over time a checkerboard effect in which each area has a different probability of a rabies outbreak.

The problem facing the health system manager is to determine the frequency and intensity of vaccination campaigns which will maximize the cost-effectiveness of his control program in every section of a city throughout the planning horizon. In order to do this however, he must have some idea as to the number of rabid dogs which his policy prevents from occurring. The best policy is clearly the one which minimizes the cost per prevented rabid dog while maintaining rabies below a socially acceptable level.

In the real world it is difficult and frequently impossible to estimate the exact outcome of a vaccination campaign. The incidence of canine rabies is greatly under-reported, one reason being that extensive laboratory tests are necessary to confirm or deny the existence of the disease. In addition, many people observe dogs suspected of having rabies but destroy them or allow them to die without reporting the incident to the local health department.

Mathematical disease models are usually constructed in order to provide insight into the biological interactions necessary for the transmission of the disease. Many of the mathematical models reviewed by the authors (listed below) were found to be exceedingly complex with no derived analytical solutions. Other models, although elegant in structure, may be of very little use to health system managers who find them difficult to understand and, consequently, to implement.

Among the literature reviewed, we list the following: historical review and state of the art [3], [4], [9], [10], [19], [29]; mathematical models of communicable diseases [17], [22], [30]; recent deterministic disease models [6], [15], [21], [26], [38]–[40]; stochastic disease models [1], [3], [4], [11]-[13]; computer simulation disease models [6], [8], [12], [13], [21], [26], [32], [38]–[40]; and disease control models [5], [18], [23], [27], [31], [39], [40].

It was felt that a simulation model of urban rabies patterned after real world processes would enable the manager to experiment with different policies and observe the resultant outcomes. The knowledge derived from this approach should greatly aid him in making the quantitative and qualitative decisions necessary for controlling urban rabies.

The purpose of this paper is first to describe a simulation model for the control of urban canine rabies that was developed for the city of Cali, Colombia, and second to compare the cost-effectiveness of various simulated vaccination campaigns in order to recommend to the health system managers that preferred policy which is the most efficient and effective among all policies tested using the model.

In developing the model, we first had to define the problem and then convert the problem into mathematical form. Estimates of the parameter values were derived from a review of the literature or from prior canine surveys conducted in Cali, Colombia. Certain values, however, were unobtainable from these sources and had to be derived from field experiments in Cali or estimated from existing knowledge. The computer simulation was tested on the IBM 7044 at Tulane University.

Once the model was developed, two sets of vaccination policies were tested. The first set consisted of massive vaccination campaigns at selected intervals which resulted in the immunization of 70 percent of the dog population in all parts of the city. The second set, which we refer to as the Preferred Vaccination Policies, involved a series of experiments in which dogs were vaccinated only in those areas of the city that had the potential to contribute most to the incidence of rabies.

In the following sections we will describe the states and interactions of canine rabies as defined for the simulation model. The mathematical formulation will not be presented here but can be found in the principal author's doctoral dissertation [14]. In addition, we will discuss the different policies and the effect each policy has on canine rabies. At the end we will briefly mention the rabies control recommendations which were made to the Colombian health officials and some of the problems which arose when attempting to determine the validity of the model.

2. Definition of the Problem

Cali is located in southwestern Colombia and has a population of over 800,000 people and approximately 84,000 dogs. A sizeable portion of the population is impoverished, with only four percent of the families having a total combined income, based on a 1969 survey, of more than the equivalent of 500 dollars per month. The poorer areas have dirt streets, and children and dogs are seen everywhere.

For this model, the city was divided into a spatial grid with 116 neighborhoods or barrios, each barrio being approximately 50 hectares in size (one hectare = 10,000 square meters). The assumption is made that infectious rabid dogs migrate stochastically between barrios and set off new epidemics in the barrios to which they have migrated. The susceptible and immune dogs, as observed in Cali, are assumed to remain within their specific barrios during the entire time period. Only a certain proportion of all infected rabid dogs actually transmit the disease to others. If the infected dog develops the "furious" symptoms, he tends to wander aimlessly and bite at whatever crosses his path, while if the infection results in the "paralytic" symptoms, the dog becomes paralyzed and is ineffective at disease transmission [7]. In addition, the rabid dog must have virus present in its saliva before the disease can be transmitted through a bite [36]. In the model, the number of rabid dogs exhibiting the furious symptoms of the disease with virus in their saliva is treated as a discrete random variable resulting from the encounter between wandering rabid dogs and the susceptible dogs living in any given barrio. The system is a discrete time model with 3600 daily interations representing rabies during a ten-year planning horizon.

The epidemiology of urban canine rabies is described diagrammatically in Figure 1. Dogs in the rabies cycle can be categorized into one of the following five, mutually exclusive states: (1) Susceptible, (2) Immune, (3) Incubating, (4) Infective, or (5) Dead. In order for rabies to exist in a city, changes must occur in the various states of dogs over time.

Dogs which are in the susceptible state, X, are those which have never been vaccinated against rabies or those which have been vaccinated but have subsequently lost their immunity. Once a proportion, ξ, of the susceptible dogs are vaccinated they become part of the immune state, Z, for a variable length of time. We are assuming that only dogs over three months of age are vaccinated and that all dogs lose their immunity over time at approximately the same rate. Some vaccines maintain a level of immunity for longer than others. Therefore, the rate of conversion, γ, from immune back to susceptible is a characteristic of the vaccine. Puppies born to both susceptibles and immunes with a birth rate, μ, are themselves susceptible to rabies and are added to the susceptible state. The same nonrabies death rate, λ, is experienced by both the immunes and the susceptibles, since the vaccine offers protection against no other disease except rabies.

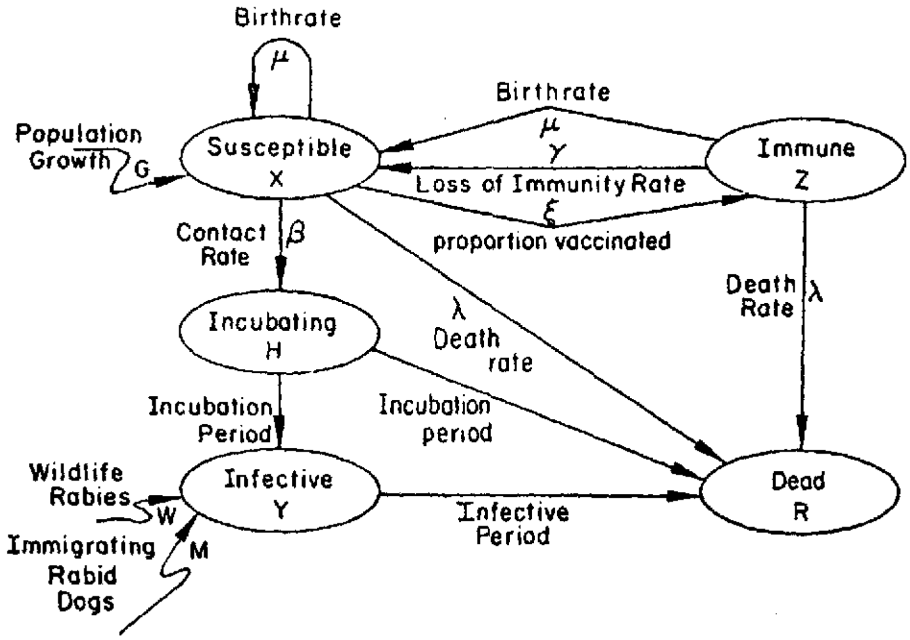


FIGURE 1. Epidemiological model of interactions and states of canine rabies.

If an infective dog has adequate contact with a susceptible dog, the susceptible becomes infected with the rabies virus and is moved to the incubating state, H. Adequate contact means that a rabid dog with virus in its saliva has bitten a susceptible dog enough times to transmit the disease. Therefore, the probability of a new rabid dog is dependent on the number in the infective state, Y, the number in the susceptible state, X, and the rate of contact, β, between them.

Once a dog is infected with rabies, it goes through a variable length incubation period. The length of stay in the incubation state is determined by a number drawn randomly from a log-normal frequency distribution which has a mean and variance characteristic of the incubation period in the real world [7], [28], [33]. Only the proportion of rabid dogs which are in the furious symptomatic stage with virus in their saliva are considered infective. The other rabid dogs, either those in the paralytic stage or those with negative saliva, are assumed not to be involved in the transmission of the disease. Instead, they move directly to the dead state, R.

All rabies infected dogs eventually die, since the recovery rate is assumed to be zero. Once in the infective state, Y, the rabid dog remains infective for a variable length of time. The length of stay in the infective state is determined by drawing randomly a value from a log-normal frequency distribution which has a mean and variance characteristic of the infective period derived from clinical observations [20], [34], [36].

While in the infective state, Y, the rabid dogs may wander throughout the city. The tendency to roam is a characteristic of the furious phase of the disease. Everyday there is a certain probability that rabid dogs will remain in the area where they are presently located or leave and move to adjacent barrios. Only one potential move is allowed per day. Figure 2 illustrates the movements of rabid dogs between barrios. The probability, g, that infectives leave a barrio, is dependent on where they originated and where they are going. For the simulation model, each day's movements are handled in a stochastic manner. Each rabid dog is given the daily option, if still alive, of remaining stationary or moving to one of five neighboring barrios. After entering a new barrio, the infective dog interacts with the susceptible dogs in that

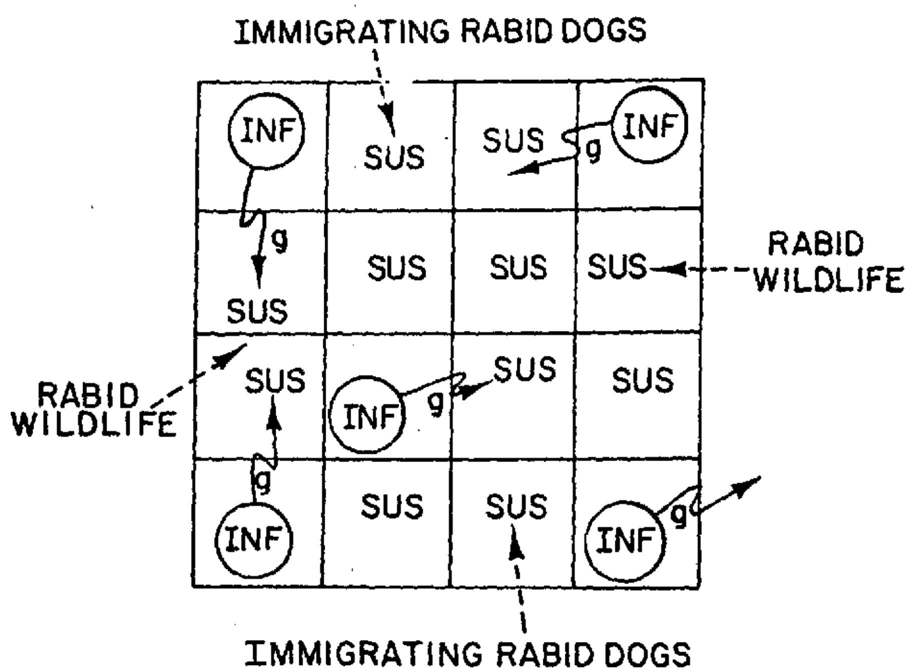


FIGURE 2. Movement of infective rabid dogs and wildlife into and through a spatial grid representing an urban area. Abbrev.: INF = infective rabid dog. SUS = susceptible dog. g = probability of movement in a given direction.

barrio and possibly, depending on the amount of contact, continues the transmission of the disease.

A certain number, W, of rabid wildlife (see Figures 1 and 2) enter the peripheral barrios of the city and interact with the susceptible dog population. The wildlife are given the opportunity to make only one move since it appears likely that a wild animal would be killed before it could roam very far in a city. Once in the urban area, the rabid wild animals are included in the infective state, Y.

Urban canine population growth, G, in every barrio (see Figure 1) is assumed to be proportional to the human population growth in that barrio [37]. The assumption is also made that these entering dogs are not vaccinated and therefore increase the susceptible canine population in a specified barrio by G, per day. Some of the newly arriving dogs may have had contact with rabies outside the system and be currently incubating the disease. A certain number, M, would become part of the infective state, Y, during the year (see Figures 1 and 2). The probability that they will appear in any specific barrio is dependent on the proportion of the city's human population presently living in that barrio. The assumption is based on the idea that the more people there are in an area, the greater is the chance that their dogs will have had some interaction with other dogs in surrounding areas or cities where rabies currently exists.

The model accepts as inputs the vaccination schedule, the number of vaccinating teams and trucks, and the number of rabid dogs and rabid wildlife which enter each year from outside the system.

3. Experiments Using the Model

Since the model is stochastic, numerous epidemic trails were run per experiment in order to determine the mean number of cumulative rabid dogs at the end of every year. All costs mentioned in the following sections are estimates, based on Colombian data, of what each vaccination campaign would cost if the experiment were done in the real world.

Rather than presenting frequency distributions derived from multiple epidemic trials per experiment, we will deal only with mean values. The policy of no vaccination and the policy of 70 percent initial vaccination were the only ones which consisted of 100 epidemic trials per experiment. The results from all other tested policies are based on 10 epidemic trials per experiment.

3a. *A Single 70 Percent Vaccination Campaign*

Looking only at the mean number of canine rabies cases per year, Figure 3 indicates that the policy of 70 percent initial vaccination was able to reduce the yearly incidence of rabies for approximately five years. Thereafter, the yearly incidence of rabies was greater than if we had followed the policy of doing nothing at all (no vaccination). With rabies absent from a barrio, the density of susceptible dogs and the concurrent likelihood of a rabies outbreak increased. All that was required was for an infectious rabid dog to wander by chance into the barrio and trigger a new epidemic. This same phenomenon of a disease being kept out of an area and then suddenly appearing in an epidemic form has been frequently observed in the real world with rabies as well as with other infectious diseases [16], [24].

Comparing after ten years the cumulative mean number of rabid dogs in the absence of any control policy and the value with an initial 70 percent vaccination campaign, we see in Table I that the immunization program prevented 2,985 rabid dogs (21,588 minus 18,603). The theoretical vaccination campaign, however, took 85 daily iterations to complete and required the operator to assign 48 vaccinators traveling in 8 trucks to all barrios of the city. During these three months, 59,323 dogs were vaccinated. Taking into account the cost of wages, the use of the trucks, the

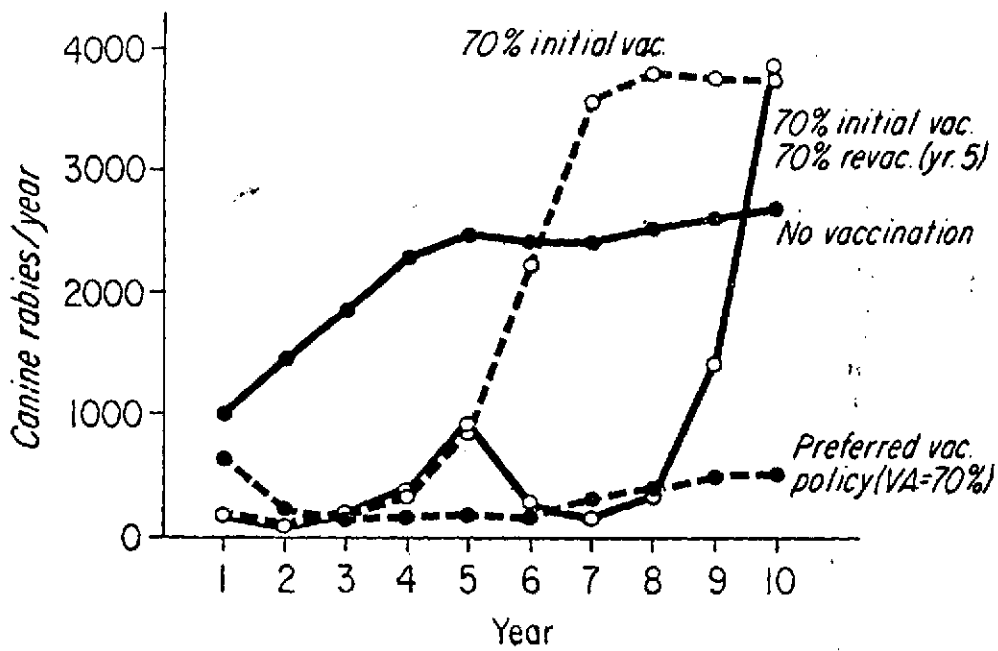


FIGURE 3. Computer Simulation Experiments: Comparison of the mean number of rabid dogs per year during a ten-year period with policies of (1) no vaccination (mean, 100 epidemic trials), (2) initial 70 percent vaccination (mean, 100 epidemic trials), (3) two 70 percent vaccination campaigns, one at the beginning and the other after 5 years (mean, 10 epidemic trials), and (4) the Preferred Vaccination Policy with VA set at 70 percent (mean, 10 epidemic trials).

TABLE 1.

Computer Simulation Experiments: Comparison of four vaccination policies, over a ten-year period, with regard to the mean cumulative number of rabid dogs, the total number of vaccinated dogs, and the cost of the individual policies.

Vaccination policy	Number of epidemic trials per experiment	Mean total rabid dogs	Total dogs vaccinated	Cost (dollars) Total*	per dog vaccinated	per rabid dog prevented
No vaccination	100	21,588	0	0	0	0
70% initial vaccination (entire city)**	100	18,603	59,323	22,157	0.37	7.42
70% initial vaccination 70% revaccination (yr. 5) (entire city)***	10	7,711	139,638	60,184	0.43	4.34
Preferred Vaccination Policy (VA = 70%)****	10	3,197	100,282	64,982	0.65	3.53

* The cost of vaccination is annually increased by an inflation factor of 5 percent

** The simulated campaign took 85 days and required the use of 8 trucks and 48 vaccinators.

*** The initial simulated campaign took 85 days and the second simulated campaign, due to the canine population growth, took 112 days. Both campaigns required the use of 8 trucks and 48 vaccinators.

**** The policy required the continuous use of 1 truck and 2 vaccinators.

Rabies Control Center, and the unit cost of the vaccine, the 70 percent vaccination campaign theoretically cost the administration 22,157 dollars. The cost per prevented rabid dog is calculated for the ten-year time period as 22,157 dollars divided by 2,985 prevented rabid dogs or 7.42 dollars per prevented rabid dog. The question then arises, "Can the health system manager do better with a different vaccination policy?"

3b. *Two 70 Percent Vaccination Campaigns*

It appears obvious that a second 70 percent vaccination campaign at year 5 would be very beneficial. Figure 3 compares a policy of one initial 70 percent campaign versus a policy of two 70 percent campaigns, one initial and one at year 5. The double campaign reduces rabies for about nine years before there is an epidemic.

The cost of the double campaign, as seen in Table 1, is almost triple the cost of the single campaign due to the increase in both the city's canine population and the cost of labor, supplies, etc. associated with a five percent annual inflation rate. The cost-per-rabid dog prevented, however, is greatly decreased due to the sizable reduction in the mean cumulative number of rabid dogs.

3c. *Preferred Vaccination Policy*

In order to reduce further the incidence of rabies, we developed an alternative policy which continuously employs throughout the planning horizon two vaccinators traveling in one truck to various selected barrios. We refer to this control policy as the Preferred Vaccination Policy since we felt it was the best among all tested policies. A formula for the Preferred Vaccination Policy was developed which ranks each barrio according to the potential contribution it would make to the incidence of canine rabies. The risk of contributing to new rabies cases is calculated for each barrio as follows:

$$R_{i,t} = C_i X_{i,t} + \tfrac{1}{5} \sum_{j=1}^{5} C_{i(j)} X_{i(j),t}$$

where

$R_{i,t}$ = the index of rabies risk for a barrio i at time t,

C_i = the relative proportion of barrio i dogs typically found in the streets,

$X_{i,t}$ = the number of susceptible dogs in barrio i at time t, and

$i(j)$ = a subscript denoting neighboring barrios j, $j = 1 \ldots 5$, which surround barrio i.

Since a rabid dog moves from barrio to barrio, the assumption is made that the sum of the values of $C_{i(j)}X_{i(j),t}$ in each of the five neighboring barrios is equally as important to the subsequent generation of rabies as the value of $C_iX_{i,t}$ in barrio i itself.

All values of $R_{i,t}$ are ranked from highest to lowest. A vaccinating team and truck are sent to the highest ranked barrio. If there are numerous trucks, each one goes to a separate high ranked barrio and remains there until the required proportion of susceptible dogs is vaccinated. Thereafter, the trucks return at different time periods to the rabies center for reassignment to the next highest ranked barrio. The most important factor is that the vaccinators go house-to-house in the selected barrios and vaccinate all susceptible dogs they encounter. Since the necessary information for the formula can be continuously updated by the vaccinators with limited data collected during their visit to the various barrios, we feel that the policy will be easy to operate and will require little supervision. Further details of this barrio ranking procedure are published elsewhere [14].

The Preferred Vaccination Policy is compared in Figure 3 with a policy of no vaccination. The term "VA" refers to the fact that 70 percent of the susceptible dogs in the selected barrios were immunized during each visit by the vaccinators. Or restated, it means that 30 percent of the barrio dog owners were either not home or were unwilling to cooperate with the rabies prevention program. As can be seen in Figure 3 the Preferred Vaccination Policy greatly reduced the incidence of canine rabies throughout the planning horizon. But at what cost?

Comparing the three vaccination policies which we have shown so far, Table 1 indicates that the Preferred Vaccination Policy with 70 percent public cooperation cost more than either of the other two policies. However, the cumulative mean number of rabid dogs over the ten-year planning horizon is also reduced to a lower level than either of the other two policies.

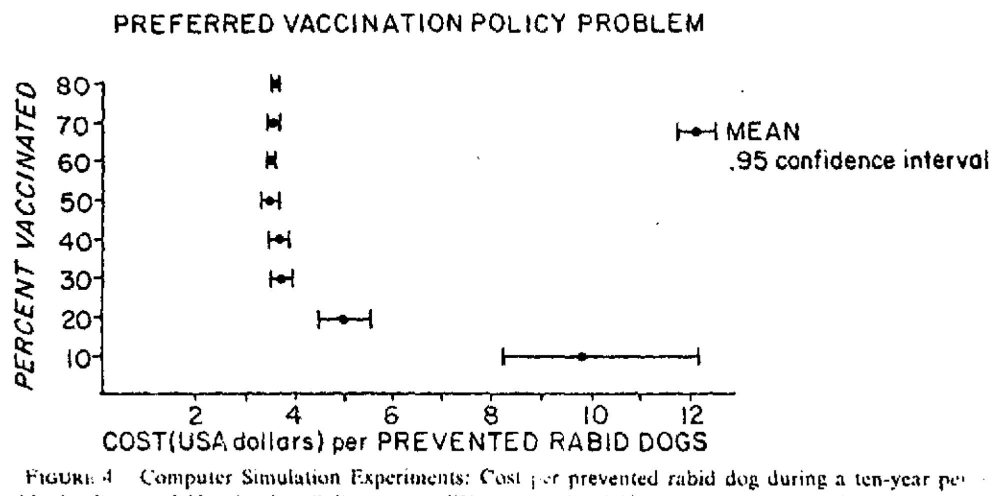


FIGURE 4 Computer Simulation Experiments: Cost per prevented rabid dog during a ten-year pe... with the Preferred Vaccination Policy set at different levels of VA (percent vaccinated in each vis... barrio) (mean, 0.95 confidence interval, 10 epidemic trials per experiment).

The usual indicator of a program's success is the cost-per-dog vaccinated. We feel that this is a poor indicator since the objective of a rabies control program is not to vaccinate dogs but rather to decrease the incidence of canine rabies. Consequently the cost-per-rabid dog prevented should be the correct indicator. As is seen in Table 1, the Preferred Vaccination Policy with 70 percent public cooperation cost the most per vaccinated dog but it cost the least per prevented rabid dog.

An additional feature of the Preferred Vaccination Policy is that it is fairly insensitive to the percentage of barrio residents being home or cooperating. We selected VA (the percent of the susceptible dogs in a barrio that are available for vaccination) to equal 70 percent based on prior knowledge of participation rates for immunization campaigns. However, even if the estimate is in error, when VA is decreased in multiple experiments from 80 to 10 percent (see Figure 4), the theoretical cost-per-prevented-rabid dog for the Preferred Policy does not go up dramatically until less than 30 percent of the susceptible dogs in a barrio are available for vaccination.

4. Validity of the Model

Before concluding, a brief mention should be made concerning the validity of the model. The problem of determining the validity can be analyzed in two ways. The first deals with internal validity. Does the simulation model perform in the manner it was intended? We feel that the internal validity of the model should be very high since the computer program was repeatedly tested and "debugged" during the initial construction of the model.

The problem is with the external validity. The question, "Does the model tell what is occurring in the real world?", is more difficult to answer. What we would like to do is be able to predict future happenings. However, since this is a new simulation model which has never been tested for predictive accuracy, all we can do is subjectively evaluate it to see if the results appear logical. Therefore, instead of external validity, we will consider only if the model is or is not reasonable.

Based on our knowledge of rabies in general and canine rabies in Cali, the model does appear to be reasonable. An opportunity to compare the computer model output with results in the real world occurred in the summer of 1971, when more than 80 percent of the known dog population in Cali was immunized by the Pan American Health Organization. The reported monthly incidence of canine rabies quickly dropped from more than 20 to less than 5 cases per month and remained at this level for at least one year thereafter in spite of the fact that there was increased rabies surveillance. This finding is in accord with the observations made using the simulation model. A further discussion of this point can be found elsewhere [14].

5. Conclusion

The results of the simulation experiments indicate that a single vaccination campaign which immunizes 70 percent of all dogs in each barrio of Cali will maintain rabies at a minimal level for five years. During the ten-year planning horizon, a policy involving two 70 percent vaccination campaigns (one every five years) is more effective per unit cost in reducing the cumulative number of simulated rabid dogs than a single campaign during the same time period. The Preferred Vaccination Policy, utilizing information about the dog population in the city in order to identify high risk barrios for selective immunization, results in the lowest cost-per-prevented rabid dog.

The computer simulation model is recommended to the Pan American Health Orgainzation either to help formulate vaccination strategies for Cali, Colombia, or to serve as an interactive teaching or gaming model to be used during their annual class on rabies control. As a result of experimentation with the model, health system managers should learn to appreciate the complexity of the system and derive greater insight into the eventual control of urban rabies.

References

1. ABBEY, HELEN, "An Examination of the Reed-Frost Theory of Epidemics," *Human Biology*, Vol. 3 (1952), pp. 201–233.
2. ALBA, M. AND ACHA, P., "The Status of Rabies in the Americas," *Second Inter-American Meeting on Foot-and-Mouth Disease and Zoonosis Control*, Scientific Pub. No. 196, Washington, D.C., Pan American Health Organization, 1970.
3. BAILEY, N. T. J., *The Mathematical Theory of Epidemics*, Charles Griffin and Co., Ltd., London, 1957.
4. BARTLETT, M. S., *Stochastic Population Models in Ecology and Epidemiology*, Methuen and Co., Ltd., London, 1960.
5. BECKER, NIELS G., "Control of a Pest Population," *Biometrics*, Vol. 26 (1970), pp. 365–375.
6. CVJETANOVIC, B., GRAB, B. AND VEMURA, K., "Epidemiological Model of Typhoid Fever and Its Use in the Planning and Evaluation of Antityphoid Immunization and Sanitation Programmes," *Bulletin World Health Organization*, Vol. 45 (1971), pp. 53–75.
7. DEAN, DONALD J., "Rabies," in *Current Veterinary Therapy III*, W. B. Saunders Co., Philadelphia, 1968.
8. DEMOOR, P. P. AND STEFFENS, F. E., "A Computer Simulated Model of an Arthropod-Borne Virus Transmission Cycle, with Special Reference to Chikungunya Virus," *Trans. Royal Soc. of Trop. Med.*, Vol. 64 (1970), pp. 927–934.
9. DIETZ, KLAUS, "Epidemics and Rumours: A Survey," *J. Royal Stat. Soc.*, A, Vol. 130 (1967), pp. 505–528.
10. ———, "A Supplementary Bibliography on Mathematical Models of Communicable Diseases," *RECS/72.5*, World Health Organization, Geneva, Switzerland, February 1972.
11. ELVERBACK, L., FOX, J. P. AND VARMA, A., "An Extension of the Reed-Frost Epidemic Model for the Study of Competition Between Viral Agents in the Presence of Interference," *Am. J. Hyg.*, Vol. 80 (1964), pp. 356–364.
12. ———, ACKERMAN, E., YOUNG, G. AND FOX, J. P., "A Stochastic Model for Competition Between Viral Agents in the Presence of Interference," *Am. J. of Epid.*, Vol. 87 (1968), pp. 373–384.
13. ———, ———, GATEWOOD, L. AND ———, "Stochastic Two-Agent Epidemic Simulation Models for a Community of Families," *Am. J. of Epid.*, Vol. 93 (1971), pp. 267–280.
14. FRERICHS, RALPH R., *Development of a Computer Simulation Model for the Control of Urban Rabies*, University Microfilms, Doctoral Dissertation, No. 73-29,878, Ann Arbor, Mich., 1973.
15. GOFFMAN, W. AND WARREN, K. S., "An Application of the Kermack-McKendrick Theory to the Epidemiology of Schistosomiasis," *Am. J. Trop. Med. and Hyg.*, Vol. 19 (1970), pp. 278–283.
16. GRAYSTON, J. T. AND ALEXANDER, E. R., "Acute Respiratory Infections," in *Preventive Medicine*, Little Brown and Co., Boston, 1967.
17. HAMER, W. H., "Epidemic Disease in England—The Evidence of Variability and of Persistency of Type," *Lancet*, Vol. 1 (1906), pp. 733–739.
18. JACQUETTE, D. L., "A Stochastic Model for the Optimal Control of Epidemics and Pest Populations," *Mathematical Biosciences*, Vol. 8 (1970), pp. 343–354.
19. ———, "Mathematical Models for Controlling Growing Biological Populations: A Survey," *Operations Research*, Vol. 20 (1972), pp. 1142–1151.
20. JOHNSON, H. M., "Rabies," in *Viral and Rickettsial Infections of Man*, J. B. Lippincott Co., Philadelphia, 1959.
21. MACDONALD, G., CUELLAR, C. B. AND FALL, C. V., "The Dynamics of Malaria," *Bulletin World Health Organization*, Vol. 38 (1968), pp. 743–755.
22. MCKENDRICK, A. G., "The Dynamics of Crowd Infection," *Edinburgh Medical Journal, New Series* IV, Vol. 47 (1940), pp. 117–136.
23. NEUTS, MARCEL F., "Controlling a Lethal Growth Process," *Mathematical Biosciences*, Vol. 2 (1968), pp. 41–55.
24. OFICINA SANITARIA PANAMERICA, *Rabies Surveillance for the Americas, English Edition, 1970-1972*, Monthly Report, Vols. 2–4, Centro Panamericano de Zoonosis, Buenos Aires, 1970-1972.
25. ———, *Rabies Surveillance for the Americas, English Edition, 1969-1971*, Semi-annual Report, Vols. 1–5, Centro Panamericano de Zoonosis, Buenos Aires, 1969-1971.

26. ReVelle, C. S., Lynn, W. R. and Feldman, F., "Mathematical Models for the Economic Allocation of Tuberculosis Control Activities in Developing Nations," *Am. Rev. Resp. Dis.*, Vol. 96 (1967), pp. 893–909.
27. ———, Feldman, F. and Lynn, W., "An Optimization Model of Tuberculosis Epidemiology," *Management Science*, Vol. 16 (1969), pp. 190–211.
28. Rhe, Charles C., "Rabies," in *Canine Medicine*, American Veterinary Publishing, Inc., Santa Barbara, Cal., 1962.
29. Serfling, Robert E., "Historical Review of Epidemic Theory," *Human Biology*, Vol. 24 (1952), pp. 145–166.
30. ———, "Methods for Current Statistical Analysis of Excess Pneumonia-Influenza Deaths," *Public Health Reports*, Vol. 78 (1963), pp. 494–506.
31. Taylor, Howard M., " Some Models in Epidemic Control," *Mathematical Biosciences*, Vol. 3 (1968), pp. 383–398.
32. Taylor, William F., "Some Monte Carlo Methods Applied to an Epidemic of Acute Respiratory Disease," *Human Biology*, Vol. 30 (1958), pp. 185–200.
33. Tierkel, Ernest S., "Rabies," in *Diseases Transmitted from Animals to Man*, Charles C. Thomas, Springfield, Ill., 1963.
34. ———, "Rabies," in *Current Veterinary Therapy*, 1966–67, W. B. Saunders Co., Philadelphia, 1966.
35. U.S. Department of Health, Education, and Welfare, *Morbidity and Mortality Annual Supplement, 1971*, Center for Disease Control, Atlanta, Ga., 1971.
36. Vaughn, J. B., Gerhardt, P. and Newell, K. W., "Excretion of Street Rabies Virus in the Saliva of Dogs," *J. Am. Med. Assoc.*, Vol. 193 (1965), pp. 363–368.
37. Villadiego, Francisco, "A Two Percent Household Survey Sample of Cali," Universidad del Valle, Cali, Colombia, 1968-1970 (unpublished data).
38. Waaler, H. T., "A Dynamic Model for the Epidemiology of Tuberculosis," *Am. Rev. Resp. Dis.*, Vol. 98 (1968), pp. 591–600.
39. ——— and Piot, M. A., "The Use of an Epidemiological Model for Estimating the Effectiveness of Tuberculosis Control Measures," *Bulletin World Health Organization*, Vol. 41 (1969), pp. 75–93.
40. ——— and ———, "Use of an Epidemiological Model for Estimating the Effectiveness of Tuberculosis Control Measures," *Bulletin World Health Organization*, Vol. 43 (1970), pp. 1–16.
41. World Health Organization, *W. H. O. Expert Committee on Rabies, Fifth Report*, Technical Report Series 321, 1966.

MEDICAL CARE
August 1981, Vol. XIX, No. 8

The Case-Control Method in Medical Care Evaluation

SANDER GREENLAND, M.S., DR.P.H.,* ERICA WATSON, M.D., M.P.H.,† AND RAYMOND R. NEUTRA, M.D., DR.P.H.

The case-control method has been applied extensively to the study of chronic diseases, largely because of its advantage in cost and statistical power over other study designs. However, problems in ensuring the validity of case-control results have led to certain reservations regarding the general utility of the method. We discuss how several of the major validity problems in case-control studies of chronic disease are limited or absent in case-control evaluations of medical care and medical technology. As a consequence, the case-control method has important potential for application in evaluation and technology research.

RESEARCH PROBLEMS in medical care and technology often require comparisons of different groups of patients with respect to the rate of occurrence of particular outcome of interest. For example, evaluation of the impact of electronic fetal monitoring on neonatal death rates requires comparison of the rates in monitored and unmonitored deliveries; an evaluation of a nosocomial infection control program would require comparison of nosocomial infection rates in units employing and not employing the program. The ideal evaluation method is the randomized clinical trial, for this type of study can easily incorporate features to prevent most types of bias in the comparison.[1] A randomized trial, however, is expensive and often impractical. In the monitoring example, it has been estimated that more than 120,000 low-risk deliveries would have to be involved in a trial in order to detect monitoring effect of a 50 per cent reduction in the neonatal death rate among such deliveries[2] (size computed using 0.90 power, 0.05 significance level); even at this size, the number of outcome events might still be too small to allow detailed analysis. Ethical issues of withholding treatment from the control group may arise. When performance of medical care providers is to be compared, it may be impossible to randomly allocate patients between the providers. Finally, when a randomized trial or other intensive investigation is planned, it may still be desirable to obtain preliminary results before extensive expenditure of resources.

When a randomized trial is impractical, a prospective nonexperimental comparison[3] ("cohort study") of patient groups is often substituted. Such studies may still require considerable expense, however, since the sample size and expected number of outcome events for a cohort study will be essentially the same as in a clinical trial.

* Division of Epidemiology, School of Public Health, University of California, Los Angeles, California.

† Department of Health Services, County of Los Angeles, California.

‡ California Department of Health Services, Berkeley, California.

Dr. Greenland's work was supported by a grant from the Milbank Memorial Foundation.

Address for reprints: Sander Greenland, M.S., Dr.P.H., Division of Epidemiology, School of Public Health, University of California, Los Angeles, CA 90024.

0025-7079/81/0800/0872/$00.85 © J. B. Lippincott Co.

If the outcome of interest defines two distinct categories (for example, death versus survival), outcome rates in different groups can, in theory, be compared by the case-control ("retrospective") method.[1-3,4] This method involves the selection of a series of cases (those persons experiencing the outcome event) and the selection of a reference series of noncases ("controls" or "referents") for comparison to the case series. Various statistical techniques allow estimation of differences in outcome occurrence between groups based on the case-control data[3-6]—provided the case-control study satisfies certain conditions required for the validity of results. A primary advantage of the case-control method is that the number of subjects necessary is usually much less than that required for a prospective comparison.[1] In the monitoring example, a case-control study would require fewer than 400 deliveries (200 cases and 200 controls) to detect a 50 per cent reduction in the neonatal death rate due to monitoring with 90 per cent probability (power), at a 0.05 significance level[7] (assuming a 50 per cent monitoring rate in the target population).

More than 20 years ago, Cornfield and Haenszel[8] summarized the two critical "representativeness" conditions that were thought necessary for the validity of a case-control study: 1) it should be "possible to enumerate all new cases of a disease, or a representative sample of them, without having to observe all the individuals in the population"; and 2) "the sample of individuals not developing the disease [the controls] [should supply] an unbiased estimate of the prevalence of the characteristic under study among the entire nondiseased population of interest." The authors then went on to lament that most retrospective studies of that day had difficulty assuring the satisfaction of the second requirement, if not the first. Other authors of that period put forth similar criteria.[4,8] More recent writers have expressed related criteria for the validity of case-control results in a sampling theory framework,[9,10] while other authors have rejected representativeness criteria and instead have emphasized the importance of comparability of cases and controls.[11] A common feature of all these points of view, however, is an emphasis on the importance of "selection validity" in case-control studies, in that subject selection should be unbiased in order to achieve valid results.

In the time since the early commentaries, general reservations concerning the utility of the case-control method have been put forth, largely on the grounds that representativeness or selection validity conditions are difficult to fulfill or assure in ordinary practice.[12] The purpose of our commentary is to offer an appreciation of the particular suitability of the case-control approach for many problems in medical care evaluation. We will discuss why the classic problems of selection validity are frequently absent in medical care evaluation settings, and how other common problems of ordinary case-control research will often be minimized as well.

Considerations Favoring the Case-Control Method in Evaluation Settings

Representativeness and Selection Considerations

As mentioned above, problems of selection bias and subject nonrepresentativeness have led to a certain degree of distrust of case-control methods among many clinical researchers. Yet it is precisely in the realm of preliminary comparisons or evaluations of medical care or medical technology that the case-control method is most often immune to such problems. This is because many evaluation settings involve small, well-defined target populations, such as a hospital inpatient population or several combined hospital populations. In such settings it is usually possible to enumerate the entire target population, select representative samples from both the case (patients experiencing outcome) and noncase (patients without outcome) groups, and thus avoid the problems of nonrepresentativeness and classical selection bias. Often, the target population will be enumerated in advance.

Analysis Considerations

Because of the frequent paucity of information regarding the absolute disease rates in the target or source population, classical (etiologic) applications of the case-control method have placed strong emphasis on the odds ratio (or functions of it) in the analysis of case-control data.[1,2,4] Many (if not most) problems in medical care or technology evaluation, however, require assessment of effects in terms of differences in exposure-specific rates, especially if risk-benefit analysis is required.[5] Furthermore, the use of ratio measure without reference to underlying rates is inadequate in most evaluation settings. For example, to say that a treatment reduces the odds of death by a factor of 20 (an odds ratio of 1/20) would be equally true if it reduced death risk from 90 per cent to 31 per cent or from 0.1 per cent to 0.005 per cent; yet a high risk of serious side effects from the treatment would yield vastly different implications in each instance. Thus, the emphasis on odds-ratio analysis may be seen as another drawback of the case-control method. Fortunately, when enumerations of the target population and the cases are available (as is frequently the case in evaluation settings), the absolute exposure-specific rates can be estimated using Bayes' theorem,[6] and classical case-control analysis methods can be deemphasized. Thus, the existence of small, well-defined target populations in evaluation settings provides analytic benefits in addition to validity assurances.

An Example: A Comparison of Neonatal Intensive Care Units

We will illustrate the above and later points with an example from our own experience, a study recently carried out to compare the performance of two different neonatal intensive care units (NICUs). This study was an investigation of preliminary statistics indicating an elevated death rate among infants admitted to one of the units (hereafter designated unit A). A second unit (hereafter designated unit B) serving a different geographic region within the same health care system was chosen as the comparison unit. Both units served low income populations in the same metropolitan area, with no special referral plan between the units. A randomized trial was clearly impractical. Of primary interest was whether the elevated death rate at unit A could be explained solely as a result of chance, or as being due to an elevated proportion of high-risk infants being admitted to unit A. If either explanation turned out to be correct, a costly investigative comparison of quality of care in the unit could be avoided.

Given various practical considerations, total sample size was limited to approximately 200 infants. Based on an initial estimate of the death rates in the units, only 25–30 cases (neonatal deaths) would have been expected in a prospective study based on a simple random sample of 200 admissions to the NICUs. This was judged to be an insufficient number of cases to accurately assess the role of various risk factors in the units. Furthermore, the admissions to the NICUs (the target population) were completely and continuously enumerated, and the status of admittees at 28 days after birth (alive/dead) was always determined. Given these considerations, it was decided that a case-control study would be conducted from medical records. Over a specified period of time, all NICU admissions that ended in a neonatal death would be entered in the study, along with a 50 per cent random sample of the neonatal survivors from the same target population (unit A + unit B). (Multiple births, infants under 800 grams birthweight or with congenital malformations incompatible with life or transferred into or out of the NICUs were excluded from the study. Very few infants from either NICU were transferred out.) There were 361 eligible admissions to units A and B over the study period (199 to unit A and 162 to unit B). Of these, 61 suffered neonatal death and thus became the study cases, while 150 of the 300 survivors were selected for the control (reference) group.

The initial results are summarized in Table 1. The crude ratio of the rates between the units, 1.9, was significant at the 0.02 level, fairly well ruling out chance alone as an explanation of the difference (Table 1). The method of rate computation is briefly described in the appendix. Using both univariate and multivariate statistical techniques,[4-6] more than 30 risk factors were examined as possible contributors to the difference. These are listed in Table 2. Few turned out to be of any importance, most because of being similarly distributed between the units, and some because of their very weak contribution to risk of death. Ethnicity (race) was distributed quite differently between the units and nearly all of the interunit difference in death rates was concentrated in the premature infants (less than 1500 grams birthweight). After adjusting for race, the rate in unit A was more than three times the rate in unit B among prematures, and the absolute rate differences were quite high in this group (more than 35 per cent). Table 3 summarizes these details. Further investigation of the elevated rates in unit A will benefit by this initial elimination of several possible explanations (such as ethnic differences) and will be able to focus on risk factors and aspects of care that primarily affect premature infants.

Other Considerations

A major class of problems in case-control studies, noted throughout the literature,[1,2,4,10,11] are difficulties in obtaining accurate (reliable and valid) risk factor information and exposure histories. While not always completely absent, these problems are often minimized in medical care evaluation settings. For example, case-control studies of diseases with long latent periods may be plagued by recall bias (bias arising from selective recall of exposure events) or simple lack of information about exposures occurring in the distant past. In contrast, most evaluation settings involve short time spans, and consequently have less difficulty in obtaining accurate recall or records regarding important exposure events. In the above example, the key risk factors (unit, ethnicity and birthweight) were accurately recorded for all NICU admissions, and no subjects had to be eliminated because of incomplete or ambiguous information on these factors.

Other classic problems of case-control research often are absent in medical care evaluation settings. Temporal ambiguity (indeterminacy of whether the exposure under investigation actually preceded the development of the outcome event) can be a problem in etiologic studies of disease with long latent periods, but would not be a problem in most evaluation settings, such as the study described above. Similarly, detection bias[12] (bias arising from differential detection of the outcome in different exposure groups) would be impossible in the large number of evaluation settings in which the outcome event is always detected, as in the neonatal mortality studies described earlier.

TABLE 1. Initial Comparison of Neonatal Intensive Care Units

	No. Cases	No. Controls	Estimated Crude Death Rate*
Unit A	43	78	21.6%
Unit B	18	72	11.1%
Total	61	150	

Fisher's exact test $p = 0.02$; crude rate ratio estimate = 1.94.

* Computed using Bayes' theorem (see appendix).

TABLE 2. Factors Examined for Contribution to Interunit Death Rate Difference

Maternal age (years)
Race (Black; Hispanic; other)
Gravidity
Number of spontaneous abortions (<20 weeks)
Number of prior stillbirths (>20 weeks) or prematures (<37 weeks)
Diabetes mellitus (no; gestational; pre-pregnant)
Chronic Hypertension (bp ≥ 140/90 pre-pregnant) (yes; no)
Drug/excessive alcohol utilization (yes; no)
Other maternal medical condition potentially affecting pregnancy (yes; no)
Pre-eclampsia/eclampsia (yes; no)
Anemia (no; mild; severe)*
Gestational bleeding (no; <20 weeks; >20 weeks)
Prenatal care (yes; no)
Hydramnios (yes; no)*
Abnormal presentation (yes; no)
Cord prolapse (yes; no)*
Placental abnormalities (abruptio, previa, etc.) (yes; no)
Oxytocin (no; augmentation; induction)*
Vaginal delivery (no; uncomplicated; complicated—breech, midforceps, vacuum, failed forceps or vacuum, etc.)
C-section (no; repeat or primary elective; nonelective)
Abnormal duration of labor (no; <3 hrs total; >20 hrs total)*
Sex
Birthweight (grams)
Gestational age (Dubowitz) (weeks)
Apgar 1 min. (by pediatrician)
Apgar 5 min. (by pediatrician)
Resuscitation at delivery (endotracheal intubation; no)
Major congenital abnormality, chromosome/genetic defect, etc. (no; yes, potentially viable)
Respiratory distress syndrome/respiratory failure (associated with immaturity and/or perinatal asphyxia) (yes; no)
Meconium/blood/amniotic fluid aspiration (yes; no)
Proven infection (no; congenital or neonatal, not hospital-acquired) (infections felt to be of nosocomial origin—e.g., bacterial sepsis with onset >3 days postdelivery—were coded "no")
Isoimmunization/other congenital hemolytic disorder (requiring exchange transfusion) (yes; no)
Significant birth trauma/injury (yes; no)
Birth out of hospital (home, in transit, etc.) (yes; no)
Number of days in hospital (0–24 hrs = 1 day)

* These factors suffered from high levels of recording inaccuracy in the records examined.

Discussion

Proper application of the case-control approach requires consideration of and concerted efforts to meet a number of criteria for study (internal) validity. These criteria have been extensively discussed and systematized in the recent epidemiologic literature.[9-11,14] Chief among the criteria have been selection validity criteria (concerning bias in selection), information validity criteria (concerning the quality of the information on the risk factors under study) and comparison validity criteria (concerning the control of confounding). Efforts should be made to meet these criteria as far as possible in any area of application of the case-control method. A researcher considering a case-control approach should carefully evaluate whether his study setting would allow fulfillment of the various validity criteria. If selection, information and comparison validity criteria cannot be met, the study will be unsound. These criteria apply to record-based cohort ("historical prospective") studies as well.

Special caution is required in the use of routine medical records for case-control and historical prospective studies, since biased recording of information on the study subjects will produce corresponding biases in the study results. Investigators must be alert to the possibility that the use of a particular treatment regimen led to a more frequent recording of the outcome under study or that the treatment under study was recorded more frequently among those experiencing the outcome. These were clearly not possibilities in our NICU study (where the "treatment" was the NICU), but such problems are frequent in nonexperimental drug evaluations.[12]

Important in any nonexperimental study of medical treatments or technology will be control of confounding.[13] Confounding warrants special attention in nonexperimental technology evaluation because of the possibility of "confounding by indication"[14]; this occurs whenever prognostic

TABLE 3. Final Comparison of Neonatal Intensive Care Units

Race	Birth Weight	Unit	No. Cases	No. Controls	Estimated Death Rate*	Fisher's Exact Test	Estimated Rate Ratio
Black	Under 1500 Grams	A	16	8	50.0%	$p < 0.05$	4.0
		B	2	7	12.5		
Black	Over 1500 Grams	A	4	31	6.1	NS	0.5
		B	2	8	11.1		
Other†	Under 1500 Grams	A	11	2	73.3	$p < 0.02$	3.4
		B	5	9	21.7		
Other†	Over 1500 Grams	A	12	37	14.0	NS	1.6
		B	9	48	8.6		

* Computed using Bayes' theorem (see appendix).
† Primarily Hispanic.
NS = not significant at 0.05 level.

factors for the outcome under study served as indications or contraindications for the application of the treatment under study. Analytic control of confounding will require that accurate information be available for all subjects on important potential confounders (e.g., in the NICU study, it was necessary to have accurate information on ethnicity, birthweight and other determinants of neonatal mortality).

Control of confounding can only be evaluated relative to subject-matter knowledge and judgment regarding the study situation. Even when "hidden" (uncontrolled) confounding is believed to exist in the study results, however, the results may still provide useful information. For example, in our NICU study, it is still possible that some unrecorded risk factors (perhaps with differential implications for admission to units A and B) rather than quality-of-care differences were responsible for the observed differences in death rates. Nevertheless, our study does provide evidence that the elevated death rate in unit A cannot be explained away by any of the risk factors listed in Table 2, and so is useful even if "hidden" confounding due to unrecorded risk factors occurred.

We have emphasized that several of the most critical problems of case-control studies can take on an entirely manageable character in many settings of medical care and medical technology evaluation. These considerations, along with the usual advantages of case-control designs, should encourage clinical investigators to consider the case-control study as a legitimate and useful method for many evaluation problems. The case-control study will be especially useful in preliminary investigations and in situations in which the outcomes of interest is too infrequent to permit prospective studies.

Acknowledgment

We are grateful to Morgan Stewart and the reviewers for their helpful comments. We would also like to thank Virginia Hansen for editing and Connie Brown for typing of the final manuscript.

References

1. MacMahon B, Pugh TF. Epidemiology—principles and methods. Boston: Little Brown & Co, 1970.

2. Anonymous. Antenatal diagnosis. Washington, D.C.: National Institutes of Health, 1979. (NIH publication no. 79-1973.)

3. Cornfield J, Haenszel WH. Some aspects of retrospective studies. J Chron Dis 1960;11:523.

4. Mantel N, Haenszel WH. Statistical aspects of the analysis of data from retrospective studies of disease. Natl Cancer Inst Monogr 1959;22:719.

5. Neutra RR, Drolette ME. Estimating exposure-specific rates from case-control studies using Bayes' theorem. Am J Epidemiol 1978;108:214.

6. Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. Biometrika 1979;66:403.

7. Schlesselman JJ. Sample size requirements in cohort and case-control studies of disease. Am J Epidemiol 1974;99:381.

8. Dorn HF. Some problems arising in prospective and retrospective studies of the etiology of disease. N Engl J Med 1959;261:571.

9. Kleinbaum DG, Morgenstern H, Kupper LL. Selection bias in epidemiologic studies. Am J Epidemiol (in press).

10. Sackett DL. Bias in analytic research. J Chron Dis 1979;32:51.

11. Cole P. The evolving case-control study. J Chron Dis 1979;32:15.

12. Feinstein AR. Clinical biostatistics. St. Louis: Mosby, 1977.

13. Greenland S, Neutra RR. Control of confounding in technology assessment. Int J Epidemiol (in press).

14. Miettinen OS. Efficacy of therapeutic practice—will epidemiology provide the answers? In: Melmon KL, ed. Drug therapeutics. Elsevier: New York, 1980.

Appendix

The rates given in Tables 1 and 2 were computed from the case-control data using Bayes' theorem. A detailed and general discussion of this method is given in reference 5. Briefly, suppose we know the overall rate of the outcome, P(D), the probability of exposure among cases, P(E/D), and the probability of exposure among noncases, $P(E \mid \overline{D})$. Letting $P(\overline{D}) = 1 - P(D)$, Bayes' theorem states that the outcome rate among exposed persons, P(D/E), is given by

$$P(D \mid E) = \frac{P(E \mid D)P(D)}{P(E \mid D)P(D) + P(E \mid \overline{D})P(\overline{D})}$$

In our NICU study, the total size of the target population (unit A + unit B) was equal to the number of cases (61) plus twice the number of controls (2 × 150) or 361; the total number of cases over the study period was 61; and so P(D) = 61/361 = 0.169 and $P(\overline{D}) = 0.831$. To illustrate Bayes' theorem, consider computation of the rate in the first row of Table 1. Here the exposure is "unit A," $P(E \mid D) = 43/61 = 0.705$, and $P(E \mid \overline{D})$ is estimated as 78/150 = 0.520. Thus, using Bayes' theorem $P(D \mid E)$, the rate in unit A, is estimated as

$$P(D \mid E) = \frac{0.705(0.169)}{0.705\,(0.169) + 0.520\,(0.831)} = .216$$

or 21.6 per cent.

Socio-Econ. Plan. Sci., Vol. 10, pp. 159-166. Pergamon Press 1976. Printed in Great Britain

PATIENT FLOW ANALYSIS AND THE DELIVERY OF RADIOLOGY SERVICE†

BENJAMIN LEV‡, GEORGE REVESZ§, FRANCIS SHEA§
and ROBERT CALTAGIRONE§
Temple University, Philadelphia, PA 19122, U.S.A.

(Received 5 February 1976)

Abstract—In recent years there has been an increased awareness regarding the cost of radiologic health care, and the patient delays encountered in the delivery to the consumer. The purpose of this paper is to demonstrate that, at least in one case in the Diagnostic Radiology Department at Temple University, the assumption that better service can be given to patients provided more technicians and orderlies are available, is not valid. The facts tend to indicate that the real problem lies in scheduling techniques, and improved utilization of available equipment. Therefore, it is safe to conclude that for improved radiologic services, the emphasis should be directed towards the design of the management systems and scheduling techniques, and not the staff and/or facilities.

INTRODUCTION

At a time, when there is a continued increase in demand for health care service,it is assumed that a reason for ineffective service to patients is the shortage of manpower and facilities. An analysis of the Diagnostic Radiology Department (DRD) at Temple University Hospital was organized, searching for reasons associated with the long waiting times of patients in the DRD and the low utilization of equipment. It has been previously assumed that better service could be presented to the consumers if more technicians, orderlies and examination rooms were available at the hospital. However, such a move would have a tendency to increase the costs of radiologic services at a time when costs are increasing at an alarming rate. In many instances, managers suspect reasons for their inefficient systems performance, but it is only after an in-depth study is made that the basic causes of the inefficiency are evident. Moreover, the suspected reasons appear to be of secondary importance.

In this particular case, the long periods of patient waiting time cannot be decreased by adding more technicians or orderlies as some had surmised, but that patient waiting time can be reduced by improving scheduling techniques. Better management methods; such as more sophisticated scheduling algorithms, automated systems and computer control can be utilized with the present staff and equipment. The effects will contribute to decreased patient waiting time, and the total time spent in the DRD. At the same time, patient service capacity should be increased in the DRD with the present number of rooms, technicians, orderlies and staff. This increased capacity will then enable administrators to accommodate the expected increase in demand for radiographic services as predicted in Knowles[1], Morgan[2] and National Advisory Committee on Radiation[3].

Recently, there have been several reports alluding to a potential shortage of radiologists in the United States in the near future. This potential shortage is predicted since the demand for radiologists' time is increasing at a more rapid rate than is the supply of radiologists. Any new national health insurance program, if enacted, would presumably cause an even greater inbalance between the demand for radiological services, and the supply of radiologists.

Improvements in DRD can be categorized in three areas. The first is increased utilization in personnel and facilities which has been reported by Lindhein[4] and Revesz *et al.*[5, 6]. The second area is improvements attempted by simulation technique. See studies by Covert *et al.*[7], Jean *et al.*[8], Kenny and Murray[9] and Lodwick[10]. The third area emphasizing Computer Scheduling and Control is reported in Donald and Waxman[11], Hansen and Snider[12] or Hsish[13]. Computerization of manual methods has demonstrated that scheduling can also be applied to large departments. To date, however, no mention has been made of an on-line scheduling system, which can be dynamically updated as the patients are processed. In addition, limited research has been conducted concerning the development of scheduling rules which have general applicability.

In view of the many possible policies that can be considered for improving the delivery of radiologic care, it is necessary to present the data quantitatively for analytical purposes. Consequently, this paper will investigate some of the effects of these policies on the DRD. In most instances DRDs at hospitals are confronted with similar problems. The methods of providing radiological services is apparently identical at all departments. The patient arrives at the department and immediately enters a sequence of service facilities, which culminates with an X-ray examination. It should be recognized, however, that departments may differ in the number of examination rooms, the procedure used and other non-major facilities. In most DRDs, however, the patient flow is basically the same. It appears to be more logical and convincing to analyze the DRD at Temple University Hospital, using its specific data and not discuss a hypothetical DRD. Moreover, the generalized model developed in this study is of sufficient generality to be associated with the DRDs in most hospitals with no more than minor modifications.

SYSTEM DESCRIPTION

The system under study is the Temple University DRD which has an annual volume of 72,000 X-ray examinations. The service is provided both for inpatients (IP) and outpatients (OP), and the ratio of IP/OP is 55/45.¶ Figure

†This work was supported in part by Grant GM 14548-06, National Institute of General Medical Sciences, United States Public Health Service. All correspondence should be directed to the first author.

‡School of Business Administration.

§Department of Diagnostic Radiology.

¶Table 1 lists and explains all symbols.

1 indicates the flow of patients through the department. Many of the patients are scheduled. Experience has shown that the scheduled patients will either arrive and enter the department too early, or will arrive too late. Scheduled inpatients are assisted by orderlies to arrive at DRD in 20 min, depending on the availability of elevators. Due to these random elements, it is assumed that the patients enter the DRD with an interarrival distribution time of $f(t)$; with probability P_1 of being an outpatient, and with probability $1-P_1$ of being an inpatient. The inpatients will wait for one of the ten orderlies (OR = 10) for assistance to be taken to the control desk by one of three means of transportation; walking, wheelchair and stretcher with probabilities P_2, P_3, P_4, respectively. An outpatient reports to the reception desk at which time the individual is serviced for time τ_1. A fraction P_5 of the outpatients will change to a hospital gown which takes time τ_2 if one of the dressing rooms (DR = 16) is available. The outpatient then reports to the control desk and spends time τ_3. Each patient using the dressing rooms will lock the room for the entire period of the examination and therefore prevent others from utilizing the room until the individual is released.

At a typical DRD, there are over 200 different types of examinations which can be grouped into 13 major categories with minor variations within a category.

For the most part, however, the demand for the different radiographic examination will vary over the day. GI, BE and IVP† are performed during the morning hours before the patient has eaten; while a Myelogram is usually scheduled for the afternoon.

Table 2 presents a list of the 13 examinations with the corresponding mean examination time τ_4, and the range for each category. $P_i(j)$ is the probability of having examination type j at period $i = 1$ (8–10 a.m.), $i = 2$ (10 a.m.–12 p.m.), $i = 3$ (12–2 p.m.), $i = 4$ (2–4.30 p.m.).

The patient is then assigned an Examination Room (ER = 14), at the control desk. This assignment is an important one since it determines the length of stay for that patient in the queue. The DRD at Temple is one where not all examination rooms are equipped with the

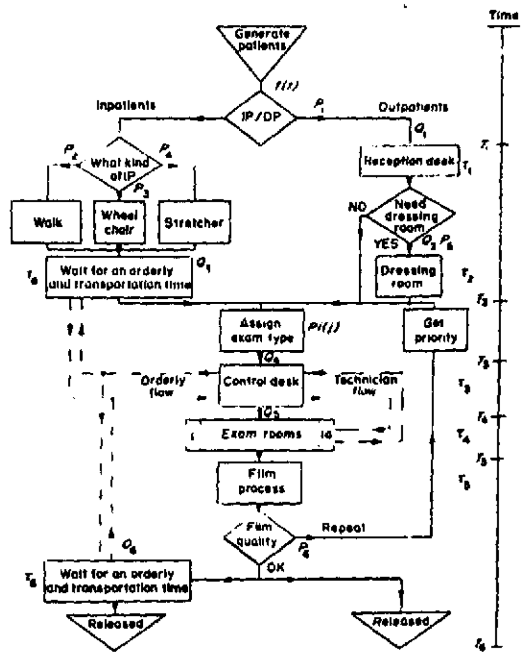


Fig. 1. Macro flow chart of patient flow.

same facilities, and only certain rooms can perform specific examinations. Most of the examinations require one of the Technicians (TE = 10), while other rooms require radiologists. When the examination is completed, the patient waits time τ_5 for film processing with the probability P_6 that the individual may need additional films to complete the study. If supplementary films are necessary, the patient then returns to the control desk with high priority. If the study is adequate, the outpatient is released, and the inpatient will wait for an orderly to be returned to the room which takes time τ_6.

EVALUATION OF SUCH A SYSTEM

It is important to recognize that in most large systems, there is more than one measure of performance. Thus, it is not surprising that in this case there are several, some

†Explanations of category names are given in Table 2.

Table 1. List and explanation of variables

Variable	Explanation	Distribution
$f(t)$	Inter Arrival Time Between Two Patients	Uniform
τ_1	Service Time At Reception Desk	Uniform
τ_2	Dressing Time	Uniform
τ_3	Service Time At Control Desk	Constant (1 Min.)
τ_4	Examination Time	See Table 2
τ_5	Film Process Time	Constant (10 Min.)
τ_6	Transportation Time To or From {Walk / Wheelchair / Stretcher} The Department	Uniform / Uniform / Uniform
P_1	Percentage Of Outpatients	45% (55% Inpatient)
P_2 P_3 P_4	Percentage Of Walk/ Wheelchair/ Stretcher	5/20/75%
P_5	Percentage Of Outpatients Needing Dressing Room	80%
P_6	Percentage Of Patients Needing Additional Films	10%
ER	Examination Rooms	14
OR	Orderlies	10
TE	Technicians	10
DR	Dressing Rooms	16

Table 2. Data concerning various examinations: probability of having the examination (as function of time in the day), times (mean, range), room equipped to perform the examinations (before 11 a.m. and after 11 a.m.)

	1	2	3	4	5	6	7	8	9	10	11	12	13
Name of Examination	Skull	Spine	Chest	Bone	Myelo	Abd.	GI	BE	IVP	Tomo	Fluoro	PEG	GB
P_1 (j) 8-10 A.M.	0	0	32	0	0	0	20	20	20	0	0	0	8
P_2 (j) 10-12 A.M.	13	17	42	20	0	0	0	0	0	2	2	2	2
P_3 (j) 12-2 P.M.	13	17	42	20	4	0	0	0	0	1	1	1	1
P_4 (j) 2-4:30 P.M.	13	17	38	20	0	12	0	0	0	0	0	0	0
Mean Exam. Time (Min.)	12	18	6	10	60	10	100	45	60	40	20	90	45
Range Exam. Time (± Min.)	7	12	4	8	15	8	75	15	30	20	10	30	15
Rooms Equipped to Perform the Examination													
Before 11 A.M.	3,5	3,7	4,7	2,3,6,7,8	11	7	1,2, 13,14	11,12 13,14	5,6 9	8,10	1,12	3	8
After 11 A.M.	3,5,6,7	3,5, 6,7	4,6,7 8,9	2,3,6,7,8	11	2,7,9	1,2 13,14	11,12 13,14	5,6 9	8,10	1,12	3	8

Examinations 7 and 8 require pairs of rooms 1 and 2, 11 and 12, or 13 and 14.

$P_i(j)$ is probability of examination j in period i.

Abbreviations: Myelo—Myelogram, Abd.—Abdomen, GI—Upper Gastrointestinal Study, BE—Barium enema, IVP—Intravenous Pylogram, Tomo—Tomograph, Fluor—Fluoroscopy, PEG—Pneumoencephalogram, GB—Oral Gall bladder Study.

being more important than the others. The following is a list of 5 measures of performances used in this study.

A. $T_6 - T_1$ *total time*†

This measure is most important to the referring physician who may be waiting for the patient. Consequently, a delay in the DRD may possibly mean another day in the hospital with additional cost to the consumer. When total time becomes excessive, it is then obvious that the department will receive numerous complaints from the referring physicians.

B. *Waiting time before study*

$T_4 - T_2$ for inpatients;

$T_4 - T_1$ for outpatients.

Except for a few minutes allocated to registration and a change of clothing, the patients are idle while waiting for their examination.

C. $T_6 - T_5$ *time the patient spends in the DRD after the examination is completed*

This time includes the film processing time and the time until the patient is released.

D. *Utilization of various resources*

Efficient utilization of the following facilities is essential: rooms, technicians, orderlies, dressing rooms, reception desk and control desk.

For example, some of the examination rooms are equipped with expensive equipment costing as much as $200,000. A low utilization of such a room is inefficient and increases the operation cost of DRD.

E. *Number of patients in the system at* 4:30 *p.m.*

In many instances, the work is not completed at 4:30 p.m. and two of the technicians are required to stay after this time. In this respect, it would seem reasonable to reduce the number of patients that are in the system after 4:30 p.m.

†Times $T_1 \ldots T_6$ are indicated in Fig. 1.

The objective of this analysis is to improve the operating efficiency of the radiology department. Needless to say, efficiency is a difficult term to define. The concept may be easily used when a weighted combination of the above five measures of performances is taken with the weights differing for inpatients and outpatients. However, it is apparent that these weights are subjective unknowns. Another fact to consider is that one measure may be more important than another, with various weights assigned for inpatients and outpatients. And besides, different weights may be given to the different examinations (there are 13 major categories). For example, emphasis may be placed on trying to reduce the waiting times of patients requiring short examinations (10 min or less), rather than those who need lengthy studies (greater than 60 min). Moreover, different weights may be assigned to individuals based on the cost of the waiting time. For example, an unemployed person may not object to waiting as compared to an employed individual that has a demanding time schedule.

In this study, the System Performance is evaluated on the basis of a combination of all the measures previously mentioned. Generally speaking, utilization figures cannot directly indicate performance. However, a reduction in waiting time, due to improved scheduling procedures, will result in additional time available for processing patients. Since it is apparent that patient load is expected to increase, the ability to handle the additional patients increases the utilization of staff and equipment.

In evaluating the various scheduling procedures, each measure of performance is first analyzed separately in order to determine which procedure produces the best results. After each is examined separately, a combined analysis is then conducted to determine a choice procedure for all measures of performance.

If a suggested change has a tendency to yield a better performance over all measure, it is then safe to conclude its superiority. On the other hand, if there is a discrepancy in the change of measures—better for some and unsatisfactory for others—then weights must be assigned to determine overall performance.

CONGESTION POINTS AND CONTROLLABLE VARIABLES

It should be recognized however, that such a complex system has several congestion points. Specifically, some of the more important ones are:

- Q_1 In front of the reception desk;
- Q_2 In front of the dressing room;
- Q_3 Inpatients waiting for an orderly to be assisted to the department;
- Q_4 In front of the control desk;
- Q_5 In front of each examination room;
- Q_6 Inpatients waiting for an orderly to be returned to the room.

In addition to the queues, there is the problem of technician flow and orderly flow. In many instances, technician and orderly flows are delayed because of the congestions in the system which are beyond their personal control.

Before evaluating the behavior of the system as results from changes in the input, it is necessary to determine what parameters can be changed. The capacity of the facility is greater than the volume of patients and, therefore, it does not seem reasonable to reject patients. There is a limit to the number of patients in some categories, but, in general, more patients can be seen than at the present level. The scheduling procedure determines the time of arrivals but the number of patients is under limited control.

There is some control regarding the patient's arrival time, but in many cases this situation is limited. As noted previously, some of the examinations may require an empty stomach, and therefore must be completed early in the morning. Another problematic area is that patients will arrive two hours before or after the scheduled time, and must be accepted for treatment. In other words, there is some control on the arrival but it is often limited by extraneous circumstances.

Certainly, the parameters of the system can be changed; i.e. number of technicians, the number o' orderlies and the increase rate of service at the reception desk. But most important, the decision that is made at the control desk regarding the room assignment appears to be one of the key factors for the efficiency of such systems. In this sense, one way to improve the performance of the system would be through changes in the room assignment decision. This decision is not easy since it must take into consideration the rooms that have the quipment to perform the examination, as well as the lines in front of each room. At present, this decision is performed by a clerk. An example of the decision tree for the room assignment of a patient requiring a chest examination is presented in Fig. 2. It should also be pointed out that Rooms 4 and 7 are available for chest X-rays before 11 a.m., and Room 4 does not have the equipment to handle stretcher cases. Decision trees for other examinations are even more complicated when requiring multiple phase examinations, and more than one room as for cases concerning upper gastrointestinal studies (GI), and barium enemas (BE). Table 2 clearly indicates the rooms that can perform the various examinations before 11 a.m. an ... 11 a.m. (before 11 a.m. some of the rooms are used for special studies).

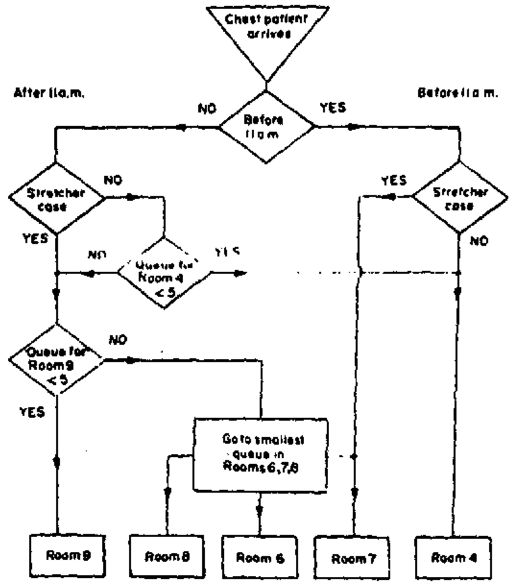


Fig. 2. Chest patient decision tree for room assignment.

METHODOLOGY

The department described above is a large and complex facility. The behavior of one part of the system is dependent on the output of another. Therefore, the arrivals to one service station represents the departures from another service station. Consequently, variables are interrelated and there are numerous system parameters. Since the matter is complex, it will be impossible to analytically optimize the operation of the department, and to develop compact results in closed forms. Therefore, a decision was made to utilize simulation techniques which have been proven useful in similar applications. A simulation program was then written for the patient flow in DRD. More specifically, GPSS (General Purpose Simulation System V) is combined with Fortrar IV, and the program has been implemented on an IBM 0/165.

For each experiment, the following sample siz. vas taken. Four vectors were generated: $P(i) = (P_1, P_2, \ldots,$ ... $i = 1, 2, 3, 4$. All four vectors have a similar patient mix (... me percentage of the 13 categories). For each one of t... ...tors $P(i)$, 10 repetitions were simulated with rando... ...r the stochastic variables. In certain cases, a larger nu... ...r of runs were performed in order to test convergence. T... randomness was in the service time, method of pati... transportation (walking, wheelchair, stretcher) and oth... stochas... parameters. Thus, each experiment included 4 days × 160 patient/day × 10 repetitions, which resulted in 6400 patients. To observe changes over various number of technicians, say 4, 6, 8, 10, the sample size was 4 × 6400 = 25,600 patients.

VALIDATION

One of the important phases of any simulation is to verify its performance as to whether or not it replicates the actual system. In ord... to satisfy the requirements of the study, statistical information was derived from DRD over a two year period, including data on performance, and compared with the results from the simulation model. The results which include a χ^2 test for goodness of fit at a level of significance, $\alpha = 0.05$, c...rly indicates a good agreement between ... ctu... DRD and the simulation model. The res... presented in Table 3.

Tests wer... ...pleted in the simulation model to measure ... performances; under a typical daytime load of 16... ...ents, under a heavy load of patients (220 pat...s) and under a light load of patients (135 patients).

Table 3. Comparison between simulation and actual system

Waiting Time Before Study (Min)	135 Patients/Day Measured	Simulation	$\chi^2(\chi^2)$*	160 Patients/Day Measured	Simulation	$\chi^2(\chi^2)$*
All Bones	17 (34)	18 (35)	6.78 (.93)	21 (61)	17 (46)	2.03 (3.35)
Chest	12 (58)	16 (62)		21 (69)	16 (70)	
Major Fluoros	56 (28)	41 (24)		25 (18)	20 (24)	
Miscellaneous	21 (15)	18 (14)		19 (18)	12 (20)	
All Patients	22 (135)	23 (135)		21 (166)	19 (160)	
Total Time in Department (Min)						
All Bones	48 (34)	55 (35)	7.6 (.93)	58 (61)	50 (46)	2.24 (3.35)
Chest	30 (58)	42 (62)		50 (69)	40 (70)	
Major Fluoros	130 (28)	137 (24)		121 (18)	122 (24)	
Miscellaneous	102 (15)	90 (14)		81 (18)	87 (20)	
All Patients	63 (135)	70 (135)		69 (166)	65 (160)	
Technician Utilization %	36	31-36 range		50	51-55 range	
Orderly Utilization %	42	38-43 range		54	56-71 range	
Room Utilization %	31	34-41 range		38	38-44 range	

Note : 135 patients, 10 technicians, 10 orderlies, 12 examination rooms
160 patients, 10 technicians, 10 orderlies, 14 examination rooms

(·) The numbers in parentheses are the number of patients that the times are based upon

* The measured data and the simulated data are from the same population at the level of $\alpha = .05$

The results disclosed that the model behaved as expected. (See Shea *et al.*[14].) Having ascertained the accuracy of the model with respect to the real life situation, sensitivity analysis would then be performed with respect to the parameters, and the measure of performances at various conditions. The experiments concerning the simulation can be described as follows.

A. *The effect of changing the number of technicians*

Presently, there are 10 technicians in the core department consisting of 14 rooms. Several "runs" were made with the simulation model, varying the number of technicians from 4 to 10. As anticipated, the system is overloaded with 4 technicians, and the waiting times and total times are very high—50 min and 110 min—as seen in Fig. 3. When the number of technicians increases, the measure decreases but only to a certain level when the times reach a threshold. One implication is then clear; waiting time and total time cannot be decreased below a certain level by adding more technicians. This fact can be explained by noting that the problem is not with the technicians, but with a combination of the number of orderlies, rooms and the scheduling procedure. The DRD operates efficiently with 8 technicians, but an additional technician is necessary for absenteeism.

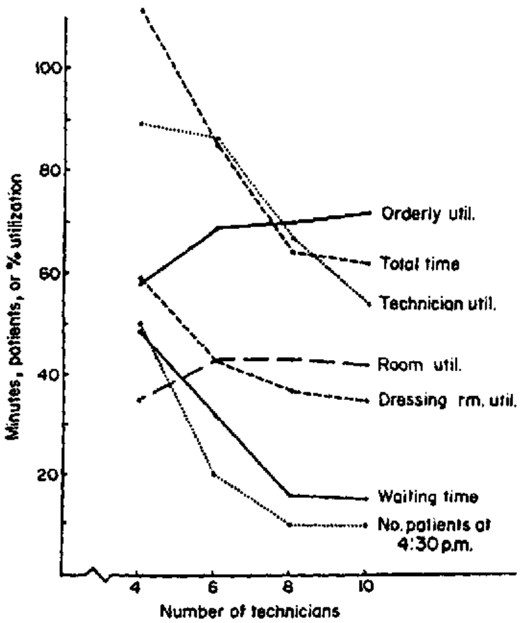


Fig. 3. The effect of changing the number of technicians.

Similarly, the number of patients in the department at 4:30 p.m. decreases with the number of technicians, but the residual number stabilizes with 8 technicians, and does not decrease further with additional technicians.

Technician utilization decreases as the number of technicians increases due to the fact that the amount of work for all technicians remains constant. Therefore, increasing the number of technicians results in a lower utilization for each technician.

The utilization of other service stations is stabilized and constant when the number of technicians rises above 8. Orderly utilization is about 0.70, reception desk utilization is about 0.55, room utilization 0.43, and control desk utilization is about 0.30. It is important to point out that none is affected by the change in the number of technicians.

B. *The effect of changing the number of orderlies*

The next parameter to be varied is the number of orderlies in the system. At present, there are 10 orderlies and Fig. 4 demonstrates the effect of changing the number of orderlies from 6 to 12. Figure 4 clearly indicates that the number of orderlies has little affect on the plotted measures of performance. The most affected is the utilization of orderlies which decreases as the number of orderlies increases since the same amount of work is shared among more orderlies. It also has an effect on the number of patients in the department at 4:30 p.m. This number decreases from about 25 for 6 orderlies to 10 patients for 10 orderlies. Other than the two changes, the measures of performance are almost unaffected and reach some constant value for 8 or more orderlies.

C. *The effect of decreasing the number of examination rooms*

The succeeding experiment involves the possibility of reducing the number of examination rooms from 14 at the present time to 12 rooms. Table 4 indicates the changes in

Table 4. The effect of changing the number of examination rooms (at 135 patients/day, 10 technicians, 10 orderlies)

		12 Rooms	14 Rooms (At Present)
Total Time			
in the System	(All patients)	79 Minutes	58 Minutes
"	(Inpatients)	59	51
"	(Outpatients)	99	69
Waiting Time			
in the System	(All patients)	28	12
"	(Inpatients)	17	8
"	(Outpatients)	40	20
Number of Patient at 4:30 PM		9	7
Technician Utilization		.91	.98
Orderly	"	.40	.61
Room	"	.38	.37
Dressing Room	"	.58	.31

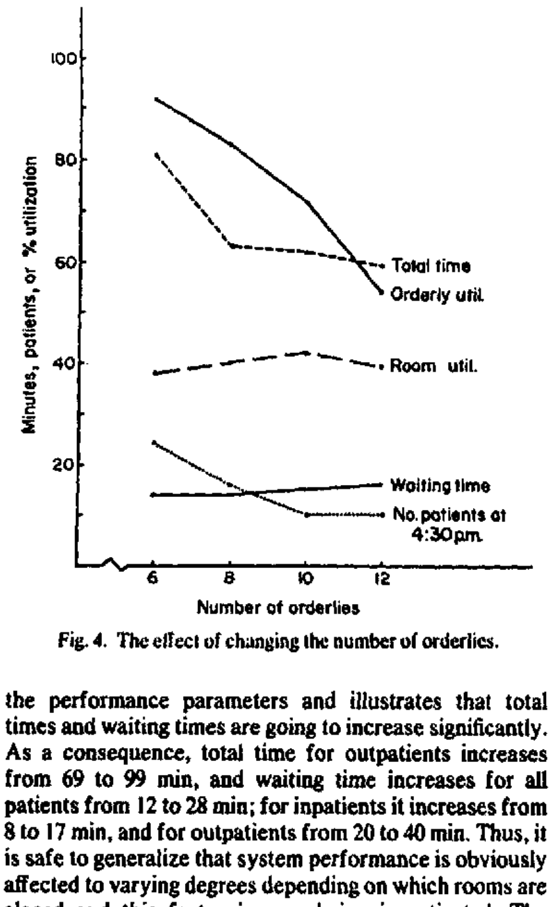


Fig. 4. The effect of changing the number of orderlies.

the performance parameters and illustrates that total times and waiting times are going to increase significantly. As a consequence, total time for outpatients increases from 69 to 99 min, and waiting time increases for all patients from 12 to 28 min; for inpatients it increases from 8 to 17 min, and for outpatients from 20 to 40 min. Thus, it is safe to generalize that system performance is obviously affected to varying degrees depending on which rooms are closed and this factor is now being investigated. The addition of examination rooms was not considered as a measure to decrease waiting time because of the high cost of each examination room ($100,000–$200,000). h surplus existing in the departments' capacity to handle more patients, the hospital administrators will not invest vast sums to decrease patient waiting time and total time.

D. *Changing the scheduling method*

An experiment was designed to observe the system performance for a change in scheduling methods. Sev... different scheduling procedures were considered:

1. *PS—Present Scheduling.* At present, the room assignment is based on the smallest number of patients in front of each examination room, regardless of the amount of work each patient requires. Once a patient is assigned to a room, the individual will not be changed to another room and no switching is permitted.

2. *MWL—Minimal Work Load.* This scheduling procedure assigns the patient to the examination room according to the minimal work load in front of each examination room. The work load is defined as the total expected time that the patient will use the examination room, including the one in process.

3. *MMWL—Modified Minimal Work Load.* This process is a modification of MWL which considers the time a patient enters an examination room and : time a decision of the present room assignment is ... de. It is evident that this concept is more accurate and bet...r than MWL. In this technique, an examination room is assigned ... on the smallest load in front of each examination roo..., ... ubtracting the process time that the patient has already ... in the room.

4. *CQ—Co... on Queue.* This algorithm assigns the patients in one large queue. As one of the examination rooms is availabl... e first patient in the common queue from those who can ta...e the exami... ...s in that room, is assigned to the room.

5. *MCQ—Modified Common Queue.* This process is a modification of CQ. While in CQ, the patients were ordered according to the time available to take the examination, MCQ orders the patients according to the arrival time at the radiology department.

6. *SPT—Shortest Process Time.* As pointed out earlier, the decision by which patients enter the assigned room in PS is "First Come, First Served". In SPT, the decision is made according to the shortest examination of all patients waiting for this particular examination room. Consequently, patients with short examination times have priority over those with long examination times.

7. *TSPT—Truncated SPT.* T...T is a modification of SPT scheduling by assignin... h...gh priority to patients who have been wai... ...ger beyond a certain time. The disadvanta... ... ...T is that patients with long examination tim... ... get low priority and stay in the system for a lon... ...e. Truncation was done to guarantee that on... a patient waited over a certain amount of time, the individual would enter the room regardless of the length of the examination time.

Four different simulation days were performed consisting of the same patient mix and arrival times. Each day was repeated 10 times for random service time, with the same expectations concening random personnel characteristics (dressing time, transportation time, etc.). Thus, for each one of the seven scheduling procedures, there were:

$$4 \text{ days} \times 160 \frac{\text{patients}}{\text{day}} \times 10 \text{ repetitions} = 6400 \text{ patients}.$$

Table 5 summarizes the data in relation to this experiment. It is clear to see that MMWL is superior for all measures of performance with the exception of two cases. One case is where the number of patients in the system at 4:30 p.m. is 15, while the best (MWL) is 14. The second case is when the total time for which 95% of patients complete their service is accomplished in 165 min, while the best (PS) is 163 min.

There is a significant difference between CQ, MCQ and the other five scheduling methods. In all five procedures, the examination room is assigned to the patient at the time the person leaves the control desk. The patient moves physically, and waits at the front of the examination room. However, in regard to CQ and MCQ, this is not the case. The patient has to wait in a common waiting room until an examination room becomes available. At that time, the technician returns to the control desk and determines which patient is next in line and can be examined in this room. The technician has to identify the patient and direct the individual to the examination room. In some instances, the distance may be as great as 50 yards, and the patient may be on a stretcher. It is then safe to assume that extra activities may take about 5 min which was incorporated into the model. This fact may explain the reason that CQ and MCQ did not perform as well as theoretical results previously indicated. Once again, this study demonstrates the possible conflict between queueing theory and scheduling theory. In scheduling theory, there is more control on the patients (jobs) in terms of the arrivals, and service times. In queueing theory, as in the department described above, there is more randomness.

SPT and TSPT were clearly the most unsuccessful techniques among the seven procedures. SPT in theory should produce the best mean waiting time, however, this result tends to diminish as the flexibility of machine (rooms) selection increases (see Wayson[15]). This appears to be the case in the radiology department where examinations can be performed in several different examination rooms.

When evaluating the various procedures, attention has to be given to the feasibility and to the cost of implementation. The CQ and MCQ procedures can be implemented manually but would require a considerable amount of bookkeeping, and a means of scanning the entire queue to determine which types of examinations are waiting. In addition, the SPT and TSPT procedures could also be implemented manually. However, it is anticipated that certain difficulties would arise in attempting to implement TSPT manually. Moreover, the MWL and MMWL procedures will require a computer to constantly update the expected load.

SUMMARY AND CONCLUSION

This culmination of observations had desirable results and a model of the DRD has been developed. The model has been tested and verified for replication of the performance of the department at a satisfactory level of accuracy. In addition, any changes in the department are tested and verified first on the simulation model.

The model used for the study indicates that increasing the number of technicians or orderlies will not decrease patient waiting times, and total time in the department. Increasing examination rooms, which may be promising, was considered and rejected because of high costs. It should be pointed out that decreasing the number of examination rooms is costly in terms of higher waiting time and total time. A recommendation was then made that there is no need for additional technician and orderly staff to improve service, and that a reduction of the number of technicians by one will not affect the department's performance. This is based on the fact that there is no

Table 5. Comparison of 7 scheduling procedures for various measures of performances

	PS	MWL	MMWL	CQ	MCO	SPT	TPST
Waiting time before study							
OP	33	28	26	32	30	35	31
IP	18	12*	12*	19	19	19	19
Both	23	18	17*	23	23	24	29
Total time in the system							
OP	79	75	72*	78	77	81	77
IP	62	57	56*	67	65	63	66
Both	68	64	62*	71	69	70	70
No. of patients in system at 4:30	20	14*	15[1]	18	18	18	18
No. of patients waiting over 60 min for service	16	12	11*	25	22	20	20
Waiting time within which 95% of patients receive service	69	64	62*	99	93	101	88
Total time within which 95% of patients left the system	163*	172	165[2]	177	174	188	183

*Best value.
[1]Off by 1 patient.
[2]Off by 2 min.

significant difference between the performance when there are 9 or 10 technicians and 10 orderlies (see Figs. 3 and 4).

As a result of the study, a decision was then issued to change the scheduling procedure to MMWL. In order to implement the procedure, a PDP-11 Mini-Computer was purchased and the operation of the DRD is now computerized. In particular, the room assignments which were performed manually are now completed by MMWL, taking into consideration the anticipated load on the rooms, and updated feedback from the rooms concerning the amount of remaining work. This is accomplished by the technicians who report to the control desk after completing the last patient, and the control desk then notifies the technician of the next patient. However, it is too early to report on the efficiency of this dynamic on-line scheduling system. Moreover, the staff is currently involved in the phase of adding options and functions. In any event, the first reaction indicates promising results which are at least as good as the previous performance.

REFERENCES

1. J. H. Knowles, Radiology—a case study in technology and manpower—II, *New Eng. J. Med.* **280**, 1326 (1969).
2. R. H. Morgan, The emergence of radiology as a major influence in American medicine, *Am. J. Roentgenology* **111**, 449 (1971).
3. National Advisory Committee on Radiation, *Protecting and Improving Health through the Radiological Sciences—Report to the Surgeon General, Public Health Service*, p. 9. U.S. Department of Health, Education, and Welfare, Washington, D.C. (1966).
4. R. Lindheim, *Uncoupling the Radiology System*, pp. 55, 150. University of California Center for Planning and Development Research, Berkeley, Calif. (1967).
5. G. Revesz, F. J. Shea and M. C. Ziskin, Patient flow and utilization of resources in a diagnostic radiology department, *Radiology* **104**, 21–26 (1972).
6. G. Revesz, F. J. Shea, B. Lev and M. S. Lapayowker, Computer aided design for improving the delivery of radiological services, In *Planning of Radiological Departments* (Edited by M. Kormano and F. E. Stieve), pp. 101–106. George Thieme, Stuttgart (1974).
7. R. P. Covert, G. S. Lodwick and E. W. Wilkinson, Simulation modeling of a diagnostic radiology department, *Proc. Conf. on the Use of Computers in Radiology.* University of Chicago (1966).
8. W. D. Jeans, S. R. Berger and R. Gill, Computer simulation model of an X-ray department, *Br. Med. J.* **1**, 674–679 (1972).
9. J. J. Kenny and G. R. Murray, Computer models evaluate alternative method, *Indust. Engng* (Dec. 1971).
10. G. S. Lodwick, *A Systems Model for an Operating Radiology Department*, Final Progress Report on Grant #HM 00477 (March 1971).
11. E. C. DeLand and B. D. Waxman, *Review of Hospital Information Systems, A State-of-the-Art Report.* Health Care Technology Division, NCHSRD.
12. K. E. Hansen and R. L. Snider, Scheduling system shows way to expand X-ray productivity 30%, *Modern Hospital* **102**, 110 (1964).
13. R. K. C. Hsieh, Scheduling of patients, technologists, and facilities in a hospital radiology department, *Proc. Conf. on the Use of Computers in Radiology.* University of Chic (1966).
14. F. J. Shea, G. Revesz, R. J. Caltagirone and M. S. apayowker, Improving the efficiency of the diagnostic radiology department using computer simulation. In *Planning of Radiological Departments* (Edited by M. Kormano and F. E. Stieve), pp. 107–111. George Thieme, Stuttgart (1974).
15. R. D. Wayson, The effects of alternate machines on two priority dispatching disciplines in the general job shop. Master's Thesis, Cornell University (1965)

FURTHER READING

Barnett, G. O., Computers in patient care, *New Eng. J. Med.* **279**, 1321 (1968).

Barnhard, H. J. and Long, J. M., Computer autocoding selecting and correlating of radiologic diagnostic cases, *Am. J. Roentg.* April 1966, p. 854.

Conway, R., Maxwell, W. and Miller, L., *Theory of Scheduling.* Addison-Wesley, Reading, Mass. (1967).

Ferrucci, J. T., Jr., Bauman, R. A., Pendergrass, H. P., *et al.*, Application of systems analysis in a diagnostic radiology department, *Radiology* **97**, 17 (1970).

Geoffrion, A. M., Dyer, J. S. and Feinberg, A., An interactive approach for multi-criterion optimization with an application to the operation of an academic department. *Mgt. Sci.* **19**, 357–368 (1972).

Harvey, R. O., Radiology scheduling system, *22nd Ann. Conf. Am. Instit. Indust. Engng*, Boston (May 1971).

Harvey, R. O., Radiology department survey report and radiology scheduling system, Medical College of Virginia, Personal communications (1971).

Lev, B., Caltagirone, R. J. and Shea, F. J., Patient flow and utilization of resources in a diagnostic radiology department: analysis by simulation techniques. *Investigative Radiology*, **7**, 517–525 (1972).

Lev, B. and Caltagirone, R. J., Operational analysis of a radiology department by computer simulation. In *Advances in Cybernetics and Systems* (Edited by J. Rose), Vol. 2, pp. 580–590. Gordon and Breach, London (1974).

Lev, B. and Shea, F. J., Operations management in radiology. In *Current Concept in Radiology* (Edited by J. Potchen), pp. 1–18. Mosby (1975).

Lodin, H. and McLander, K., The planning and organization the diagnostic X-ray department in a large hospital. *Sjukhuset* [illegible]) 39, 211. (Abstracted in *Hospital Abstracts* 3640) (Nov. [illegible])

McCreadie, [illegible], Radiology department planning. *Hospital Health and [illegible]ment (London)* 26, 26 (July 1963).

Nuffield Provincial Hosp[illegible] Trust, *Towards a Clearer View: The Organization of Di[illegible]tic X-Ray Departments.* Oxford University Press, Londo[illegible] 1962).

Schaefer, J. E., Temple [illegible]versity Hospital Administration. Personal communication (1971)

Shuman, L. T. and Wolfe, H., *A Gen[illegible] Model for Hospital Micro-Costing.* University of Pittsburg Press (1972).

Theusen, G. J. and Sullivan, W. G., Cost sensitivity analysis for a radiology planning problem. *Proc. 39th Nat'l. ORSA Meeting*, Dallas, Texas, May 1971.

Trecate, A. and Zoboli, C., Elements in planning a radiology department for a general hospital, *Cliniche Moderne (Milan)* 1, 117 (July/Aug. 1961).

Warner, T. N., Use of simulation to generate scheduling heuristics in a diagnostic radiology department. Master's Thesis. MIT (1968).

Welch, J. D. and Noble, D., Functional studies in hospitals and particularly in X-ray departments, *Radiology (London)* 27, 131 (May 1961).

W[illegible]stman, C. E., Planning of new diagnostic de[illegible]artments, *Brit. J. Radiol.* **35**, 265 (April 1962).

Widegren, G. B., Georgia [illegible] of Technology, Personal communication (1971)

#5

Program Evaluation Techniques in the Health Services

JACK MEREDITH, PHD

Abstract: This article addresses the issue of program evaluation in the area of health services; examples are drawn from the field of mental health. Current arguments concerning the goals, characteristics, and methodologies of program evaluation are discussed and two generally useful quantitative evaluation models are presented. The models are compared and their advantages for clinicians and administrators are detailed. (Am. J. Public Health 66:1069–1073, 1976)

Introduction

Program evaluation is currently a subject of major attention in the health services field. Much of this attention has centered around the definition of "evaluation" itself. Simply stated, the purpose of evaluation is to find out what worked, what did not work, and why.[1] However, the proper use of the evaluation results is frequently a source of disagreement. Levey and Loomba[2] distinguish between on-going evaluation and retrospective evaluation as follows: The purpose of on-going evaluation is to measure progress toward program goals so as to direct *control* over a program. In contrast, retrospective evaluation is conducted to determine the effect of a program so as to facilitate program *planning*.

If, as is generally accepted, program evaluation is the determination of the degree of progress in achieving program objectives, then evaluation is a tool for *control*. However, there also exists the opposite view—that the consideration and evaluation of predetermined goals is not only unnecessary but also possibly contaminating.[3] The concern here is that the evaluator will develop tunnel vision by accepting the validity of the program goals and thus overlook other, perhaps more important, effects of the program. This viewpoint of evaluation as an unbiased (in terms of the program's goals) analysis thus appears to place evaluation as a *planning* tool.

There is no need to make a choice between these two viewpoints. Evaluation can, and should, be used for both planning and control and thus serve as an all-important feedback link between these crucial functions.

This paper compares two general-purpose program evaluation models that can be used for both planning and control. The models are illustrated with applications in the complex, ill-defined area of mental health. To illustrate the models it is assumed that psychiatric cases can be specified objectively, outcome criteria can be designated and agreed upon, treatment programs remain stable, and patient selection, treatment selection, and outside influences are controlled or accounted for. Clearly, few situations will satisfy all these assumptions and thus the models' outputs must be tempered with experience, judgment, intuition, and other qualitative factors before a final decision is reached. If so tempered, such models can provide important information for the clinician or administrator.

An Index Model

The first model, adapted from Halpern and Binner,[4] is an index model that examines a program's effectiveness (how well it achieves its goals) and its efficiency (results per unit cost). Summary ratios for these two measures are developed from program value and program cost data. Mental health programs may result in changed individual functioning, intervention, and protection of the individual and/or so-

Address reprint requests to Dr. Jack Meredith, Associate Professor of Management, Florida International University, Tamiami Trail, Miami, FL 33199. This paper was submitted to the Journal November 17, 1975, revised, and accepted for publication June 1, 1976.

ciety. However, change in the functioning of the individual is perhaps the most important result.

A surrogate measure of the value of individual functioning is suggested by Halpern and Binner, based on the individual's primary means of contributing to society: his economic productivity. Since the patient's most recent economic productivity may not accurately reflect his potential due to his current mental handicap, his economic productivity is based instead on the average of his previous 12 months' earnings and the expected annual earnings of an average member of his educational and occupational peer group. Annual earnings is used, rather than total expected future earnings, to acknowledge the general impermanence of the improved level of functioning due to the program. A lower bound on earnings, the minimum annual wage, is imposed to reflect society's implied minimal worth of all individuals. This obviates the difficulties of inferring a wage for housewives, students, the severely retarded, children, the retired, and other such groups.

The thought of using an individual's personal economic productivity as a surrogate for program value may be anathema to some because this suggests accepting only the rich or the working male into treatment programs in order to obtain the most cost-effective programs. Thus, a more desirable alternative may be to utilize the mean national (or regional) income so as to equally value all patients. Another alternative would be to develop monetary values based on broader concepts than wages. For simplicity's sake, however, we will illustrate the example here with individual economic productivity.

The surrogate measure of improved functioning is then the product of the patient's average annual productivity and an *index of change*.* The index of change is a rough measure of the effect of the program on the individual. Table 1 lists some examples of indices of change, on a gross scale, as a function of impairment at admission and discharge.

TABLE 1—Specification of Indices of Change (A Model Input)

Admission Level of Impairment	Discharge Level of Impairment: None	Slight	Moderate	Severe
Slight	+40%	0	−30%	−70%
Moderate	+70%	+40%	0	−30%
Severe	+100%	+70%	+40%	0

NOTE: 100 per cent or 1.0 represents progression from severe impairment to no impairment, 70 per cent or .7 from moderate to none, etc. Level of impairment categorization is preferably determined by place on a standardized scale whose reliability and validity have been established.

Table 2 presents hypothetical data for an Incarceration program serving three individuals to illustrate the determination of program value. The index of change is based on columns (4) and (5). The program value is then based on this index of actual change in column (6) and the average productivity in column (3). The overall program value index computed in the table, 477, is computed on a *per patient* basis so as to be comparable with other programs. The maximum program value possible in column (9) is found in the same manner as the program value in column (8) but the index of possible change is used instead of the actual. Thus, the index of possible change for Individual 1 whose admission level was moderate is, from Table 1, 70 per cent (discharge level of none). For Individual 2, the maximum is 100 per cent, and for Individual 3, it is the same value as before, 40 per cent.

Program costs are more straightforward. The only difficulty here is the decision of whether to charge the program with only direct costs or whether to include indirect costs as well. If the patient spent different amounts of time in different cost statuses (hospital, family care, outpatient) within a program, a total cost figure may be obtained by multiplying the days spent in each status by the daily cost of the status and then summing over all statuses.

Table 3 presents the hypothetical summary program evaluation indices for an agency with four drug treatment programs. The incarceration program's values, columns (2) & (3), were derived from Table 2; the values for the remaining programs would be derived in the same manner. These numbers are interpreted in relation to annual wages—the treatment home saves three times the per patient wages as incarceration (1,510 vs 477). The maximum possible value per program is shown in column (3) for comparison with column (2). Thus, incarceration saved 477 out of a possible 4,480 while methadone maintenance saved les[illegible] 450, but out of a much smaller possible, 1,753. Thus, metha[illegible]e maintenance is less effective than incarceration in terms of absolute value [illegible] more effective in relation to its potential, as shown by the eff[illegible]ess index in column (6).

C[illegible] (4) gives the program cost per patient and is the basis for the [illegible]iency index in column (5). Of course, by utilizing the maximum possible program value, column (3), a maximum poss[illegible] efficiency index could have been determined—in some cas[illegible] it still ma[illegible] [illegible]t exceed unity, a fact of significant interest.

Comparing the indices in columns (5) and (6), it is seen that not all programs have an efficiency index exceeding unity. This is a reflection of the fact that not all programs are self-sustaining, in terms of investment—some are successful, some are not. Those least successful tend to be the pure custodial programs. Similarly, the four programs vary considerably in terms of their effectiveness with incarceration making a very poor showing and the treatment home looking very good. Note, however, that the most effective program is not the most efficient, nor is the most efficient (methadone detoxification) necessarily very effective. For that matter, the least effective (incarcer[illegible]ion) here is not the most inefficient either.

If the[illegible] [illegible]e real data we would conclude that the treatment h[illegible] [illegible] not only unusually effective but also quite effic[illegible] [illegible]arceration is neither effective nor exceptionally effic[illegible]. Methadone maintenance is weak while methadone de-

*This index may be derived through the use of scales of impairment in functioning whose reliability and validity have been established. Rehabilitation frequently uses such types of scales, for example.

TABLE 2—Example of the Determination of Program Value for Incarceration

	(1) Previous 12 Month's Earn.	(2) Group Ave. Annual Earn.	(3) Avg. of (1) & (2)	(4) Admission Level	(5) Discharge Level	(6) Index of Change	(7) Index of Poss. Change (Discharge Level of None)	(8) Pgm. Value (3) × (6)	(9) Pgm. Value Possible (3) × (7)
Individual 1.	7,400	11,000	9,200	Mod.	Severe	-30%	+70%	-2,760	6,440
Individual 2	4,680*	4,680	4,680	Sev.	Mod.	+40%	+100%	1,872	4,680
Individual 3	5,100	6,500	5,800	Slight	None	+40%	+ 40%	2,320	2,320
Total (T)								1,432	13,440
Avg. = T/3								477	4,480

*Based on a minimum wage of $2.00 per hour.

toxification, although not extremely effective in terms of its potential, is an extremely efficient program. We might therefore move to expand treatment homes at the expense of incarceration and try to improve the effectiveness of methadone detoxification.

Table 3 can also give overall efficiency and effectiveness indices for the agency's four programs. These are found by multiplying each program index by the number of patients in that program (column 1) and then dividing by the total group size:

$$\text{Overall Agency Efficiency} = \frac{.82\times 3 + 1.16\times 57 + 1.28\times 81 + .76\times 163}{3 + 57 + 81 + 163} = .97$$

$$\text{Overall Agency Effectiveness} = \frac{.11\times 3 + .68\times 57 + .34\times 81 + .26\times 163}{3 + 57 + 81 + 163} = .36$$

The poor performance of incarceration is minimized in the overall indices due to the small number of patients in that program. In terms of efficiency, the agency is saving almost as much, in terms of patient productivity, as the program cost. The effectiveness of .36 indicates that they are achieving about one-third of the maximum possible achievable—not necessarily a poor showing.

This example has compared different programs but the same model can also be applied to the same program at different time periods or to groups of patients categorized by means other than treatment programs; e.g., by sex, age, or diagnosis.

Halpern and Binner[1] point out one of the inevitable arguments against the index evaluation model when they state that ". . . the administrator of a mental health program may not be able to maximize his return on investment. In fact, he may have to follow strategies that lower his return, if he is to serve those who need his help most." This, of course, raises a fundamental question: Who should receive the benefit of the program's limited resources? For example, should the patient who can utilize them the most receive the resources or the patient who needs them the most? There are no methodological or technical answers to such moral and ethical questions.

The Markov Chain Model

Eyman's[5, 6] statistical evaluation model, the Markov Chain, differs significantly from that of Halpern and Binner in that it focuses exclusively on the functional level of the patient, using it as the measure of change and the basis for evaluation. The Markov model delineates, via contingency tables, the movement of the patient (or groups of patients) along a scale in terms of the patient's initial position on the scale.

Eyman used the Markov model to evaluate the effectiveness of a school program[5] and an intensive treatment program[6] in a hospital for the mentally retarded. However, the Markov model is much more powerful than these limited applications suggest. For example, as will be shown, program costs can be included in the model so that it has the potential of being combined with the index model advanced by Halpern and Binner. In addition, under very general conditions the model has been used[7, 8] to predict the probable time-

TABLE 3—Drug Treatment Program Evaluation Indices (Hypothetical)

Program	(1) No. of Patients	(2) Pgm. Value	(3) Program Value Poss.	(4) Pgm. Cost	(5) Efficiency Index = (2)/(4)	(6) Effectiveness Index = (2)/(3)
Incarceration	3	477	4,480	583	.82	.11
Treatment Home	57	1,510	2,205	1,300	1.16	.68
Methadone Detroxification	81	975	2,870	760	1.28	.34
Methadone Maintenance	163	450	1,753	590	.76	.26

varying outcome of a treatment program on a patient. Furthermore, the model can handle more than one treatment program at a time so a sequential variation of treatments can also be analyzed.

The basis of the Markov model is a one period, (e.g., one year) matrix which tabulates the probabilities of a patient changing ability or status levels over the duration of the period. These probabilities are typically obtained from historical cohort data concerning patient's changes in self-help abilities such as arm-hand use, toilet training, communication, etc. An example from a hospital for the mentally retarded is given in Table 4.

TABLE 4—Markov Matrix of Probabilities of Moving between Ability Levels in One Year in "Standard Care" Program

1970 Ability Group	1971 Ability Group (percentage distribution) I	II	III	IV	V	VI
I	**72**	20	06	0	0	0
II	14	**49**	31	04	0	0
III	01	20	**51**	24	02	0
IV	0	02	28	**50**	17	01
V	0	0	05	35	**47**	11
VI	0	0	0	0	04	**94**

The horizontal rows in Table 4 correspond to groups of patients with successive levels of general ability, Group VI being the highest level. The time period covered in this example was July 1970 to July 1971 and the treatment program being evaluated was "standard" (custodial) care. The table indicates, for example, that 49 per cent of the patients in Group II in 1970 neither progressed nor regressed in their general ability level by 1971. However, 14 per cent of the group regressed and of the 35 per cent who progressed, 4 per cent advanced all the way to Group IV. Note that the largest numbers in each row (set in bold face type) are those along the diagonal, thus indicating the high probability of remaining in the starting group at the end of the period. Larger numbers above the main diagonal than below it (e.g., 1970 Group I) indicate improvement in group functional ability. Conversely, larger numbers below the diagonal than above it (e.g., 1970 Group V) indicate deterioration. By adding the probabilities across each row it will be noticed that 2 per cent of each starting group is unaccounted for; this is the annual death rate for the particular group of patients used in this study.*

To compare treatment programs, the probability matrices can be statistically compared to determine if significant differences exist; this was Eyman's[5] objective. However, as mentioned earlier, the matrices may also be used, under certain general conditions,* to predict the probable outcome and cost-effectiveness of a series of treatment programs on a patient. Consider the set of programs and corresponding matrix shown in Table 5 for example.

Table 5 shows the probabilities for an assumed set of treatment programs and their costs as a function of the ability level of the patients. Note that death is now explicitly considered so that all enumerations of the prognosis of the initial cohort (or, equivalently, the probabilities of movement for the individual patient) will be exhaustive. The result of using the set of programs in Table 5 is shown in Tables 6 and 7. The comparison of tables such as 6 and 7 for different sets of treatment programs (such as in Table 5) thus shows the administrator which program to use for each ability level. The mathematics involved in obtaining Tables 6 and 7, although not complex, will not be presented here since they are not relevant to the purpose of this paper. A simple summary of the methods may be found in the references 7, 8 or 9.

Table 6 presents an example of the first set of supplementary information. This table shows the expected change in the distribution of groups due to the programs listed in Table 5 between 1970 and some future year; 1974 was used [illegible] as an example. As can be seen by comparing the column and row of total patients, the distribution is expected to shift considerably toward the higher groups in the four-year period. The number from each group expected to regress, progress, and [illegible] are detailed in the cells of the matrix. This information then tells the administrator what patient distribu-

*These conditions are discussed in Kemeny and Snell.[9]

TABLE 5—Varying Treatment Program Matrix with Costs

1970 Ability Group	Program	1971 Ability Group (percentage distribution) I	II	III	IV	V	VI	Death	Cost ($/Pat./Yr.)
I	Std. Care	72	20	06	0	0	0	02	7675
II	Behav. Mod	11	[illegible]	37	07	0	0	02	8600
III	School	0	[illegible]	51	30	03	0	02	[illegible]030
IV	Intensive Training	0	0	01	48	48	01	[illegible]	12220
V	School	0	0	03	32	[illegible]	[illegible]	02	10030
VI	Placement	0	0	0	0		94	02	2016
Death		0	0	0	0	0	0	100	0

Note: Each ability group is placed in the specific program indicated.

TABLE 6—Expected Change in Cohort Distribution from 1970 to 1974 with Varying Treatment Programs

1970 Ability Group	Total Patients	1974 Ability Group I	II	III	IV	V	VI	D
		89	72	95	133	118	58	50
I	221	73	46	46	27	11	0	18
II	112	12	16	26	27	19	3	9
III	114	4	9	18	34	32	8	9
IV	96	0	1	3	30	36	18	8
V	52	0	0	2	14	18	14	4
VI	20	0	0	0	1	2	15	2
D	0	0	0	0	0	0	0	0

tion he will have in any future year he desires from which he may determine what resources he will need to service those patients; of course, distant projections have less reliability than near projections.

Table 6 was generated, specifically for the 1970 cohort shown, from a four-year probability matrix similar to that of Table 5. The next three sets of supplementary information are all shown in Table 7. This table lists the number of years a patient is expected to spend in each of the groups before either dying or being placed in a community foster home (limited to Group VI as per Table 5) for the first time. The total cost the patient is expected to incur up to this event is also shown and was derived by multiplying the years spent in each group by the annual cost of the program selected for that group (from Table 5). Lastly, the table gives the probability of each of the outcomes occurring first.

Table 7 shows, for example, that a patient in the lowest ability group (I) is expected to spend 4.4 years in that group, including time spent there due to regressing from higher groups, about two and one-half years in Groups II and III each, and about four and one-half years in Groups IV and V each before he either dies or is placed in a foster home, for a total of 18.3 years spent in the hospital at a total cost of almost $180,000. The likelihood is almost twice as great that he will be placed in a foster home rather than dying first, a very positive prognosis for a patient in this group.

TABLE 7—Expected Stay Times in Years with Varying Treatment Programs Until Either First Placement or Death

1970 Ability Group	Ability Group I	II	III	IV	V	Total Years	Total Cost ($/Pat)	Outcome Probability Placement	Outcome Probability Death
I	4.4	2.2	2.8	4.6	4.5	18.3	179,500	.64	.36
II	1.1	2.8	2.6	4.9	4.8	16.2	166,400	.68	.32
III	0.4	0.9	3.2	5.1	5.1	14.6	155,400	.71	.29
IV	0.1	0.2	0.6	5.6	5.4	11.9	130,900	.76	.24
V	0.1	0.2	0.6	3.9	5.8	10.5	113,000	.79	.21

Lastly, the Markov model is not limited to evaluating only those programs with which an agency has experience. For instance, if a new mode of operant conditioning becomes available or a modification to an existing program is contemplated, subjective estimates of the transition probabilities could equally well be used in the transition probability matrix. The effect of the new or modified program could then be determined in a manner similar to Tables 6 and 7. This would also give the effect on the total cost to the hospital and thus ascribe an equivalent worth to the program.

Discussion

Although the Markov model and the Halpern and Binner model require a considerable amount of effort to use, even more effort is required to provide accurate, meaningful input data for the models. Clearly, the statistical manipulation of inaccurate data is worse than useless—it may well be misleading. Thus, to use these models for the purpose they were intended requires an enormous amount of careful work ascertaining the reliability and validity of the raw data. In addition, an appropriate scoring and recording system must be designed, monitored, and properly utilized. And finally, the results of the model analyses must be interpreted and very carefully explained lest incorrect inferences be drawn.

Furthermore, any evaluation model is only one tool in the total program evaluation process. The Markov model and the Halpern and Binner model appear to be especially useful in this regard. Undoubtedly, more evaluation models for the health services will be developed and become available as time progresses but during the interim the two models described here appear to possess the most general applicability for quantitative program evaluation. When used in conjunction with additional qualitative information and real world constraints these models should prove extremely helpful to the clinicians and health administrators faced with the task of program evaluation.

REFERENCES

1. Weikel, K. Evaluation of national health programs. Am. J. Public Health 61:1801, 1971.
2. Levey, S. and Loomba, N. P. Health Care Administration, p. 421. Lippincott Co., 1973.
3. Scriven, M. Goal-free evaluation. Evaluation 1:62, 1973.
4. Halpern, J. and Binner, P. R. A model for an output value analysis of mental health programs. Administration in Mental Health 1:40–51, 1972.
5. Eyman, R. K., Tarjan, G., and McGunigle, D. The Markov Chain as a method of evaluating schools for the mentally retarded. Am. J. Mental Deficiency 72:435–444, 1967.
6. Eyman, R. K., Tarjan, G., and Cassady, M. Natural history of acquisition of basic skills by hospitalized retarded patients. Am. J. Mental Deficiency 75:120–129, 1970.
7. Meredith, J. A Markovian analysis of a geriatric ward. Management Science 19:604–612, 1973.
8. Meredith, J. Program evaluation in a hospital for the mentally retarded. Am. J. Mental Deficiency 78:471–481, 1974.
9. Kemeny, J. C. and Snell, J. L. Finite Markov Chains. D. Van Nostrand, 1960.

MEDICAL CARE
March 1981, Vol. XIX, No. 3

Assessing the Performance of Medical Care Systems: A Method and Its Application

PAUL A. NUTTING, M.D.,* GREGORY I. SHORR, M.D.,† AND BARTON R. BURKHALTER, PH.D.‡

As health care becomes more differentiated, fewer people receive the majority of their care from a single source. Yet, most methods for assessing health care focus on the care provided by a single facility or group of practitioners. A method is described which tracks individuals through the diffuse medical care "system" and examines the process of care received for complete episodes of care. Through the use of tracer conditions the individual's pathway through the system is followed and the contribution of the various system components (e.g., facilities and providers) is assessed for various functions of care (e.g., screening, diagnosis, treatment), thus pinpointing deficiencies in the process of care. The method is designed to sample systematically from the entire provider and consumer system. Use of this methodology in a variety of settings, including American Indian communities, has proved to be feasible and has uncovered deficiencies in the delivery of health services which might have been overlooked by other approaches. This article describes the assessment method and presents selected results which demonstrate the assessment outputs.

AS TECHNOLOGY and specialization increase, consumers of health services are faced with a bewildering array of different individual and institutional providers with ever fewer receiving their care from a single source. Solutions to the problem are offered from at least two philosophical viewpoints. Some argue that technology and specialization should be decreased and the goals of the health care system revised to give more control to consumers and more attention to the environment and styles of living,[1-5] while others find evidence that the high degree of specialization should be maintained and better management added to it.[6]

Methods of quality assessment and assurance may eventually contribute to better management of health care systems. However, quality assessment and assurance have yet to demonstrate their worth. Most assessments of the quality of care address only narrow segments of the complex array of services. Some focus only on the care provided by single facilities or groups of providers. Others focus only on care provided to those patients who have

* Associate Director for Research, Office of Research and Development, Indian Health Service, Department of Health and Human Services.

† Director, San Xavier Ambulatory Care Project, Office of Research and Development, Indian Health Service.

‡ Senior Scientist, Community Systems Foundation and Professor (Adjunct), Department of Family and Community Medicine, University of Arizona.

All opinions expressed herein are the authors' and do not necessarily represent the policy of either the Indian Health Service or the Department of Health and Human Services.

Address for reprints: Paul A. Nutting, M.D., Associate Director for Research, Office of Research and Development, Indian Health Service, P.O. Box 11340, Tucson, AZ 85734.

0025-7079/81/0300/0281/$01.30 © J. B. Lippincott Co.

sought care for specific health problems. Most emphasize only diagnostic and treatment functions.

More comprehensive methods are needed. This article presents an approach to quality assessment which tracks members of the community into and through episodes of care provided by various parts of the medical care system.

Requirements for Quality Assessment in the Indian Health Service

The Indian Health Service (Department of Health and Human Services) has evolved a complex system of health services delivery. Charged by Congress to assure comprehensive health services to more than 600,000 American Indians and Alaskan Natives, the Indian Health Service (IHS) has developed and is operating more than 85 local comprehensive health care systems, called Service Units. The typical Service Unit serves a dispersed population, often scattered over several thousand square miles. It usually consists of a 30–50 bed hospital and outpatient department staffed by a variety of health professionals and administrative staff. The main Service Unit facility serves as a referral center and administrative base for one or more full or part-time field clinics, public health nurses, environmental engineers, health educators, and a variety of problem-specific health programs (nutrition, mental health, maternal and child health, alcoholism, etc.) operated by the tribal government. Each Service Unit can also refer patients to secondary and tertiary care centers, either operated by the Indian Health Service or through contract within the private sector.

As Service Units have grown more complex and Indian communities have become more mobile and more active in health delivery programs, the methods traditionally used by IHS to assess and assure the quality of care have become inadequate. Therefore, the Indian Health Service has undertaken a long-term research and development effort in the assessment and assurance of quality.

In an earlier study in one Service Unit, patients were tracked through episodes of ambulatory care for several prevalent health problems.[7] The study demonstrated that 1) explicit criteria for minimal care could be defined by the Service Unit physicians; 2) reliable data could be collected retrospectively from medical records by nonphysicians; 3) failures in the process of care could be identified; 4) the failures tended to occur at the same places in the process of care for different health problems; and 5) this information caused the Service Unit to take action aimed at correcting the problems.

As a result of this experience, eight requirements were defined which, it was felt, would make the assessment method applicable throughout Indian Health Service.

First, *the assessment must examine the performance of the total health system.*§ All facilities, organizational subunits, and programs providing health services within the community should be included whether or not the individual components consider themselves as a part of a "system." The contribution of physician extenders, pharmacists, public health nurses and community health personnel as well as physicians should be included. The combined effect of consumer and provider on system performance should be measured.

Second, *the assessment must examine the care received by all members of the community, both patients and nonpa-*

§ The *total health system* contains the *consumer subsystem* and the *provider subsystem*. Much of our attention focuses on the *medical care system*, which we define as that part of the provider subsystem providing medical care. *Components* refer to the various parts of the two subsystems. For example, the assessment examines three types of components in the medical system, including facilities, organizational subunits and professional disciplines, although other components could also be identified.

tients. It should determine if certain groups, especially those at high risk, receive lower quality of care than others.

Third, *the assessment must examine the performance of the system across a broad range of functions.*‖ In addition to diagnosis and treatment, system performance for other functions such as prevention and screening should be examined also. Although practicality generally dictates that several specific health conditions be reviewed, the assessment should focus on the performance of the functions rather than on the specific health conditions, because corrective action can be aimed more readily at functions than at specific health conditions.

Fourth, *the assessment must examine the progress of consumers through the many stages of care.* For each episode, it should determine if the consumer received the appropriate elements of care in the proper sequence and within the appropriate time. Interactions between functions which occur at different stages in the process of care should be taken into account as well as the cumulative effect of deficiencies on outcomes.

Fifth, *the assessment should examine particular health conditions which are common within the community, which meet the criteria established by Kessner,* # *and which as a group are representative of the health conditions in the community and span the health care system.* Common health conditions have the advantages that they usually do not require sophisticated care from a specialized component of the system, are understood by consumers and are prevalent.

Sixth, *the assessment should employ standards of care that are minimal and agreed upon by local practitioners.* In comparison with stringent standards, minimal standards require less professional time to reach agreement, and more nearly reflect a level of care which providers expect of themselves.

Seventh, *the assessment must identify the major deficiencies in system performance that can be remedied.* It should point to the causes of the deficiencies; at a minimum it should distinguish among consumer behavior, provider behavior and system characteristics as causes of the deficiencies.

Eighth, *the cost of the assessment must be reasonable and within the limitations of Indian Health Service.*

While all of these requirements have been stated by other authors,[8,9,12-14] there is relatively more emphasis placed on examining the total health system and ... ing on functions in the method presen... e.

A... ssment Strategy

The assessm... method derives in large part from the trace... appro... h proposed by Kessner.[8,9] Like Kessner's approach it examines a set of health conditions¶ that can trace through the process of care to outcome, that as a group span the part of the health care system under study, and that are selected according to Kessner's six criteria noted previously. Like Kessner's approach it assumes that the care provided

‖ We use the term *function* to refer to primary prevention, screening, health status monitoring, diagnostic evaluation, treatment planning, treatment, follow-up, and ongoing management in order to emphasize that their objectives are well-defined and independent of any particular system components or structure. Other authors have used "major activities,"[4,6] "clinical elements,"[7] "levels of care"[10] and "process categories"[11] in reference to what we have called functions.

Kessner proposes six criteria for selecting health conditions: 1) have functional impact; 2) be ... defined and easy to diagnose; 3) be prevalent; 4) be influencable by medical care; 5) have a well-defined technology for care; and 6) have a relatively well-understood epidemiology.

¶ The term *tracer* refers to th... health conditions examined by the assessment. This follows the terminology of Kess... ... implies that the health conditions meet ... ...ditions established in the fifth requireme...

for the tracers as a group is similar to all care, and assumes that efforts that improve deficient care in the tracers will also improve care for other similar conditions, although these assumptions are not yet supported by experimental evidence.

Consistent with Kessner's use of tracers in sets, rather than singly, the assessment aggregates data by function across the tracers, in order to focus corrective action on function rather than disease.** The functions we have used—prevention, screening, health status monitoring, diagnostic evaluation, treatment planning, follow-up and ongoing management—can be recombined to fit the needs of a particular analysis. Other functions, such as patient education or rehabilitation, can be defined if desired.

Most functions can be separated into three *sequential events*, namely, *contact* between a consumer and a provider, *recognition* of the need for service once contact is made and *provision* of service after contact and recognition. This particular classification of functions and sequential events was chosen because it is generally familiar, because the adaptive processes required to correct deficiencies would appear to differ by function and event, and because the Indian Health Service may be in a position to influence these categories of activities. This classification may be altered as more experience is gained with patterns of homogeneity across tracers and with their usefulness in improving care.

When carrying out a particular assessment, we define an assessment space†† that is delineated by the functions, tracers, system components and populations to be studied. By defining the assessment space and measures of performance first, the assessment focuses only on issues of interest.

Table 1 shows the functions and tracers that we used in early applications of the assessment method. Every function is examined using at least two tracers, by aggregating the data across the tracers for each function. This particular assessment space was constructed to examine common issues in ambulatory care, and did not address care for mental health problems, medical or surgical problems requiring specialized care, or rehabilitation. Since many of the functions are further broken down into the three sequential events of contact, recognition and provision of required service, the assessment space permits analysis of patient utilization, system outreach and problem recognition. However, this assessment space does not allow for analysis of inappropriate utilization of inpatient or outpatient services, surgical procedures or inappropriate drug therapy.

For each tracer, minimal criteria for the process of care are established for each function which the tracer examines. In no case was a tracer used to examine a function for which "valid" criteria were unavailable or for which the criteria originally proposed were questioned by the local providers of care. Only criteria considered essential for basic health care are included. Criteria that may apply to a relatively small percentage of instances are useful in de-

** The data for a particular function are aggregated by 1) summing the number of consumers found to be in need of care in the study cohort of each one of the tracers; 2) summing the number found to be receiving adequate care in all tracer cohorts; and 3) dividing the second sum by the first to obtain the fraction of consumers for all tracers who are receiving adequate care for that function. This procedure weights the tracers in proportion to the size of their cohorts. Usually the size of the cohorts have been in proportion to the prevalence of the tracers, that is, the number of consumers in need of care for each function of a tracer, although in some cases cohort sizes for all tracers have been equal. The aggregating procedure assumes that when the tracers are weighted in proportion to cohort size, the care received for the tracers is representative of the care received for all health conditions. However, our experience indicates that this assumption is not always justified and that care should be taken in interpreting aggregated results.

†† The *assessment space* for a particular assessment is defined by specifying 1) the functions to be studied; 2) a set of tracers for each function; 3) the components of the health system for each function; and 4) a population of consumers (from which an appropriate sample is selected) for each function of each tracer.

tailed examination of specific issues in health care, but are of less value in attempts to examine generic system performance. Since the assessment involves abstracting data from the medical record, criteria that are likely to be documented in the record are more often incorporated into the assessment design. In general, prescription data, measurements, lab results and diagnoses are reliably documented, while historical data, physical findings (especially negative findings) and education treatment plans are not.

The assessment method produces three different types of *indicators* of health system performance: population-based indicators, encounter-based indicators, and health-status indicators. *Population-based indicators*‡‡ are computed from a sample of the community or a patient subset of the community and express the *percentage of individuals* in need of a specific health service who receive that service within a specified period of time. They track specific cohorts of the consumer popul[illegible] tion through the system of health care and examine the adequacy of the process of care and how it is distributed in the population. ***Encounter-based*** indicators are computed from consumer contacts with a particular component of the provider subsystem and express the ***percentage of consumer encounters*** in which a specific need for service is satisfied. These indi-

‡‡ For many issues in health care it is not practical to employ a true population-based indicator, however desirable this may be. For example, an instructive indicator might describe the percentage c[illegible] [illegible]eople in the community with a urinary tract infection who received an appropriate antibiotic within an appropriate time frame. However, short of a special survey to identify all people in the community with a urinary tract infection, this subset of the population i[illegible] [illegible]arly impossible to define. Consequently, the i[illegible] would more likely express the percentage of people in the community screened positive for a urinary tract infection who received an appropriate antibiotic within three weeks. In this case the indicator more accurately might be called patient-based rather than population-based, although for simplicity we use population-based to refer to all indicators with number of individuals in the denominator.

TABLE 1. Relationship Between Five Fun[illegible]ns of Care and the Tracers Used to Examine Them

	[illegible]eizure Disorder	Hypertension	Prenatal Care	Ir[illegible] C[illegible]	Urinary Tract Infection	Nutritional Anemia	Lacerations	Streptococcal Pharyngitis	Gonorrhea
Preven[illegible]			X	X					
Screenin[illegible]		X	X	X			X		X
Health sta[illegible]us monitorin[illegible]		X	X	X					
Diagnostic evaluation							X	X	X
Treatment					X	X	X	X	X
Follow-up					X	X	X	X	X
Ongoing management	X	X							

Each function is examined by at least two tracers, and most of the function[illegible] [illegible]e further broken down into consumer-provider contact, recognition of need for service, and provision of the required service.

cators focus on the components of the provider system such as discipline or facility in order that their contribution to total system performance can be appraised.

Finally, ***health status indicators*** express the *percentage of patients* for whom a change in health status has been documented. Health status indicators should not be equated with measures of incidence or prevalance since the latter requires a random sampling of the population. Health status indicators, on the other hand, often reflect change in health status of selected patient groups, e.g., only those who were followed up. Since health status indicators do not distinguish changes in health status that are the result of health care from changes related to behavioral or environmental variables, they should not be considered measures of the outcomes of medical care unless the appropriate causal relationships linking process measures to health status have been demonstrated.

Population-based indicators can be constructed in a sequence in order to examine the continuity of the process of care,§§ which reflects the extent to which consumers pass successfully through sequential steps in a defined process of care. Figure 1 illustrates a sequence of population-based indicators for urinary tract infection from which the continuity-of-process can be obtained for the treatment and follow-up functions. The functions are divided into contact, recognition and provision of service, and indicators at each of these sequential events show the flow of patients through the treatment and follow-up functions. Patient flow from one event to the next can be expressed as the transition rate P_{ij}, where P is the proportion of patients at event *i* moving to event *j* during a specified period of time. Likewise, the transition rate of patients through multiple successive events in the process of care can be expressed as the product of the intervening rates, and the dropout rate between any two events, *i* and *j*, is $1 - P_{ij}$.‖‖ By examining care in this way, the assessment can identify deficiencies in health systems performance and distinguish between problems related to provider performance, those related to patient utilization of service and those related to the system itself.

Selection of the study cohorts and their health records largely determines the extent to which the assessment examines the entire community of consumers rather than those receiving care. As nearly as is practical, the study cohorts are generatec randomly from the entire community of po tential consumers. This is especially impor

§§ The *continuity* of the process of care is defined as the likelihood that consumers will receive needed health services, in a proper sequence and within an appropriate interval of time, and is expressed as a sequence of conditional probabilities based on empirical data. This definition derives from Shortell,[15] who conceptualized continuity as the extent to which medical services are received as a coordinated and uninterrupted succession of events consistent with the medical needs of the patients. The term "continuity" has been used differently. Some authors prefer a focus on continuity of process such we have done,[7,15,16] while many others focus on continuity of provider by measuring the extent to which care is received from a single source or by referral.[17-20] Although it appears that greater continuity of provider results in greater continuity of process,[18] Indian Health Service has often found it difficult to achieve continuity of provider and therefore has focused its attention on continuity of process. Here, the term "continuity-of-process" is used to distinguish our usage from other usages of the term "continuity."

‖‖ Care must be taken in multiplying together P_{ij}' obtained from different tracer conditions and differei cohorts, because patients receiving adequate cai early in the process of care may not be independent (those receiving adequate care later on. Specificall patients receiving adequate care at one stage may al: be more likely to receive adequate care at another. I fact, such nonindependence is illustrated in the da reported by Novick,[19] although it was not explicit pointed out by him. We have avoided this problem l calculating continuity-of-process scores only for fun tions or sequences of functions where we were able observe the passage of the cohort throughout the s quence. Thus, our continuity-of-process scores ai observed, not calculated. When greater knowledge obtained about the interdependence of functions w may be able to combine continuity-of-process score obtained from different cohorts.

tant for prevention, screening and health status monitoring. As an example, a cohort of women may be selected for examination of prenatal care by generating a master list from delivery room logs, birth certificates, operating room logs and laboratory logs (searching for prenatal lab work ordered). When the redundancies are removed from the master list, a standard sampling technique is used to select the study cohort. For many tracers a single document may be used to generate an age- and/or sex-specific study cohort, e.g., birth certificates alone may suffice for the generation of a cohort for examination of infant care.

It also is important to sample from the most basic source document available. For example, in generating a cohort of patients with urinary tract infections, it is better to sample the laboratory log for patients with a positive urine culture than to generate a sample of medical records for patients who were diagnosed with a urinary tract infection. The latter technique biases the sample in favor of patients who have made contact and for whom the system has recognized the problem.

Patients are eliminated from the study cohort when they do not contribute to the objectives of the study. For example, when examining uncomplicated urinary tract infections, it might be preferable to eliminate patients with chronic urinary tract infections, chronic renal disease, urinary tract anomalies, etc. These characteristics become apparent in the record review and patients thus eliminated from the study cohort may be replaced from the master list.

Selection of the health records to be examined largely defines the scope of the medical care system to be examined. Since medical care systems usually are not clearly delineated, judgment is required to define the system in a way that will produce useful assessment results. Most assessments of IHS Service Units must examine medical records at the main

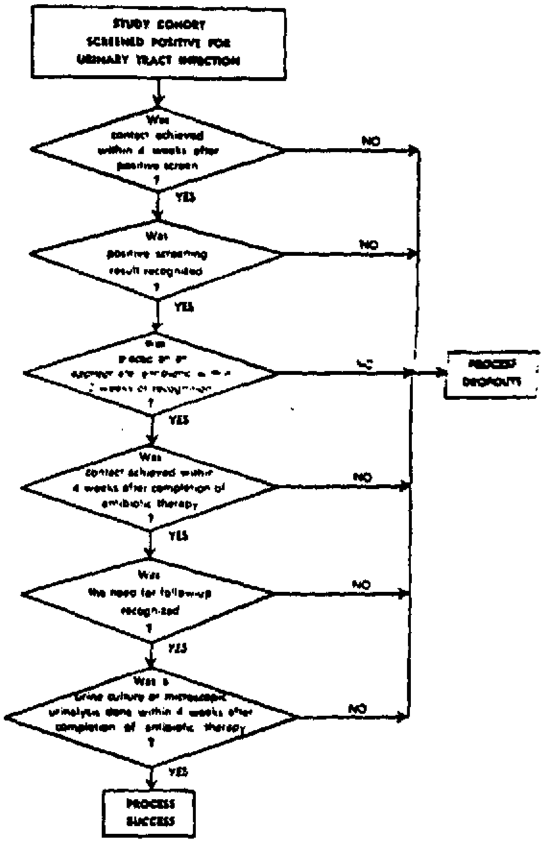


FIG. 1. A sequence of population-based indicators for urinary tract infections constructed to examine the continuity of the process of care.

hospital–outpatient facility and at one or more field clinics, public health nursing records, medical records at one or more referral centers and sometimes health records from tribal health programs (e.g., nutrition, mental health or alcoholism). In applications of the method outside the Indian Health Service, a similar set of record types have been included in the assessment procedure. In an assessment of one rural private practice, records were examined from two private practices, the community hospital, the county health clinic and the county public health nursing program.

Audit instruments are designed for each tracer to extract data from each record for each individual in the study cohort. The audit instrument documents each contact between the consumer and any component of the provider system and captures the date, location of contact, provider of services and any of a predetermined list of services required to compute the indicator results for that tracer. When completed, the audit instrument contains a complete profile for that consumer of each contact with every component of the provider system and each of the relevant services provided. The results for each of the indicators and continuity-of-process scores for each function or sequence of functions can be computed from the completed audit.

Application of the Assessment Method

The assessment method has been applied in 20 Service Units of the Indian Health Service, three rural private practices and two closed-panel health maintenance organizations. Selected data is presented from these applications to demonstrate the major characteristics of the method and some of the results achieved.

The first step in an analysis is to examine the performance of each function. In one Service Unit of the Indian Health Service, both population-based and encounter-based indicators were employed to examine the follow-up function, using as tracers iron-deficiency anemia, urinary tract infection, laceration of the scalp and extremities and prenatal care. Table 2 presents the follow-up criteria for each tracer and the results aggregated across the tracers. The consumer–provider contact rate is population based and expresses the *percentage of patients* due for follow-up who made contact with the medical care system during the time interval appropriate for follow-up. The encounter-based follow-up rate expresses the *percentage of patient encounters* due for follow-up, in which the follow-up criteria were met. Finally, the population-based follow-up rate expresses the *percentage of patients* due for follow-up who contacted the system and received the follow-up services within the appropriate time. The aggregate data indicate that 71 per cent of patients make contact with some component of the medical care system when due, resulting in only 44 per cent of patients receiving follow-up services. These results suggest that the relative weakness in follow-up care is the sys-

TABLE 2. Examination of the Follow-up Function With Aggregate and Tracer-Specific Data

Follow-Up Function	Aggregate Results	Iron Deficiency Anemia	Urinary Tract Infection	Lacerations	Prenatal Care
Contact rate (population-based)	71 (92/129)	58% (15/26)	63% (19/30)	65% (15/23)	86% (43/50)
Provision of service rate (encounter-based)	51% (57/112)	44% (7/16)	52% (12/23)	80% (12/15)	45% (26/58
Provision of service rate (population-based)	44% (57/129)	27% (7/26)	40% (12/30)	52% (12/23)	52% (26/50

Follow-up criteria include:

Anemia—All patients placed on therapy should have a hematocrit or hemoglobin or reticulocyte count between three and six weeks of the initiation of therapy.

Urinary Tract Infection—All patients placed on antibiotic therapy should have a urine culture or microscopic urinalysis within four weeks after therapy is completed.

Lacerations—All patients with a laceration requiring sutures should have an examination and documented statement of wound healing between 5 and 21 days after the sutures were applied.

Prenatal Care—All women delivering should have their blood pressure documented and an examination of the uterus between 2 and 10 weeks after delivery.

tem's recognition of and response to patients presenting when due for follow-up services, rather than the utilization behavior of the patient population. Table 2 illustrates how data for several tracers can be aggregated for a single function, and for sequential events within that function. It also illustrates that the results are not always similar across tracers. For example, the fact that lacerations, with sutures acting as an apparent clinical sign calling for attention, produced higher recognition and a very different pattern than the other tracers, suggests that it may be possible to devise more efficient ways of aggregating across tracers or to categorize health conditions in a way that allows for more representative sets of tracers to be selected.

In order to examine the performance of different system components by function, the assessment often employs a population-based indicator in conjunction with an encounter-based indicator that is disaggregated by relevant system components as well as by function. Table 3 shows the data examining infant immunization in a rural private practice setting, two Service Units of the IHS and a large closed-panel health maintenance organization. The population-based immunization rate expresses the percentage of infants who had received three DPT and two polio immunizations by 12 months of age. The encounter-based immunization rate expresses the percentage of visits by infants due for an immunization in which the immunization was provided. The DPT immunization was considered to be due at 2 months of age and to be repeated monthly until three doses had been given. If, at the time of a visit, the infant had a rectal temperature greater than 100.5°, then an immunization was not considered due on that visit.

Among the infant population served by the private practice, only 32 per cent had received three DPT and two OPV immunizations by age 12 months and the encounter-based indicator revealed that immunizations were provided on only 22 per cent of the visits for which they were due. This private practice had assumed that infants were receiving their immunizations from the nearby county health clinic. But when the encounter-based indicator was sorted by the physician's office

TABLE 3. Data for Infant Immunization From a Rural Private Practice, Two IHS Service Units, and One Health Maintenance Organization (HMO) Illustrating the Performance Patterns Resulting From the Assessment Method

	Private Practice	IHS-A	IHS-B	HMO
Immunization rate (population-based)	32% (26/50)	86% (43/50)	56% (28/50)	58% (29/50)
Immunization rate (encounter-based)	22% (63/285)	46% (179/387)	38% (119/316)	86% (127/147)
Sorted by facility				
Medical officer	19% (21/112)			
County clinic	24% (42/173)			
MCH clinic		85% (103/121)		
General clinic		34% (70/208)		
2 fild clinics		11% (6/53)		
Inpatient service		0% (0/5)		
Sorted by provider discipline				
Physician			34% (64/189)	
Physician extender			75% (36/48)	
Clinic nurse			50% (6/12)	
Public health nurse			87% (13/15)	
Pharmacist			0% (0/52)	

and the county clinic, it revealed that neither location was taking advantage of its opportunities to provide immunizations. In discussing the results, both sites agreed that it had assumed the other was responsible for immunization, and both agreed to begin immunizing more vigorously. An informal follow-up study by one of the private practitioners several months later indicated that the encounter-based immunization rate had increased three fold at both locations.

One Service Unit (IHS-A) had a population-based immunization rate of 86 per cent and an encounter-based rate of 46 per cent. When the encounter-based rate was sorted by clinic, it was noted that the MCH Clinic was performing well at 85 per cent, while the general clinic (34 per cent) and the two field clinics (11 per cent) were missing many opportunities to provide immunization. Since most of the missed opportunities occurred at the general clinic, the Service Unit instituted a standing order for immunizations in the general clinic.

In the second Service Unit (IHS-B), 56 per cent of the infant population was immunized by 1 year of age. When the encounter-based rate of 38 per cent was sorted by provider discipline, it was noted that the physicians were providing immunizations only 34 per cent of the time and were commonly referring infants to the physician extender for well baby care. Also, 52 of the 316 infant visits made when an immunization was due had been to the pharmacist, who had recently begun a program providing nonprescription medication directly from the clinic pharmacy. This result and the pattern apparent for other functions led to the development of a checklist of potential service needs for prevention and chronic disease surveillance for use by the pharmacist while dispensing over-the-counter medications.

A contrasting pattern was seen in the immunization indicators for the health maintenance organization. Although the providers were immunizing infants on 86 per cent of the visits when an immunization was due, only 58 per cent of the population was being immunized. This pattern suggested that patient contact was the limiting factor in achieving higher immunization rates in the infant community. This was later confirmed by a study of the utilization pattern of the infants, which indicated that many of the infants contacted the provider system only when they were ill.

The above experiences with immunization illustrate several benefits of the assessment method. It can focus attention on a particular site or discipline; it can point to areas where deficiencies are greatest; and feedback of the results may, in some cases, generate action. Several limitations are also illustrated. The method does not identify the causes of the deficiencies; it does not uncover and analyze the many possible remedial actions; nor does it evaluate remedial actions during and after implementation unless a separate study is undertaken.

In several applications, treatment and follow-up were examined by obtaining the study cohorts from the laboratory logs of each laboratory at every study site. For iron-deficiency anemia the study cohort included individuals with laboratory records showing a hematocrit less than 33 vol % and a hemoglobin less than 11 gm %. For urinary tract infections the study cohort included individuals with a urine culture resulting in greater than 10^5 colonies per milliliter. In order to examine the care for routine and uncomplicated conditions only, patients whose charts showed a nonnutritional cause for their anemia, or that had a chronic urinary tract infection, urinary tract anomaly or chronic pyelonephritis, were eliminated from the study cohort.

Table 4 illustrates varying patterns of performance in treatment and follow-up in three IHS service units. The numbers in the table are population-based indicators that express the probability based on empirical data that a patient at a given point in the process of care will pass successfully to the next. Likewise, the probability that a consumer will pass successfully through multiple successive elements of care can be obtained directly from the number of individuals in the study cohort who successfully completed the sequence, but also can be calculated as the product of the intervening probabilities.[##] Thus the "continuity-of-process index" of Table 4 expresses the probability that a patient screened positive (in this case for either anemia or a urinary tract infection) will achieve contact with the system, have the problem recognized, receive treatment, make contact for follow-up, have the need for follow-up recognized, and receive the follow-up service. As shown in Table 4, this probability ranges from 0.55 in Service Unit C to 0.09 in Service Unit E. The aggregated results suggest that the three Service Units have three distinct patterns of care. Service Unit C, with a continuity-of-process index of 0.55 for treatment and follow-up, appears to have no particular step in the process of care which stands out as a relative deficiency. Service Unit D, with a continuity-of-process index of 0.36, is similar to Service Unit C except for recognition (indicators 2 and 5), which appears as the relative impediment. Service Unit E, with a continuity-of-process index of only 0.09, appears to have substantial deficiencies in both contact (indicators 1 and 4) and recognition (indicators 2 and 5). The next step in the analysis would be to determine if these distinctions continue when the results are disaggregated and examined by tracer.

The assessment method can also examine the distribution of care among various types of consumers, with one of the most useful distinctions being consumers at different risk to particular health problems. In a study of prenatal care at one IHS service unit, 22 of 50 pregnant women in the study cohort were classified as high risk because they were under the age of 18 years, over the age of 35 years, primigravida, with parity equal to or greater than 5, or with a history of miscarriage or spontaneous abortion, while the other 28 women were classified as average risk. Table 5 shows results for three indicators (of the 25 employed in the study) which constitute a simple sequence to examine the continuity of process of gonorrhea screening. Respectively, the indicators examine the proportion of women achieving contact with the system by the 20th gestational week, the proportion of those with pregnancy recognized by the 20th week, and the proportion of those having a cervical culture by the 20th gestational week.

As is apparent from the indicator results and continuity-of-process index of Table 5, system performance favors the average risk group at each step of the process. The reasons for the disparity are suggested by the encounter-based indicators for pregnancy recognition and gonorrhea screening disaggregated by site of contact as shown in Table 6. Unlike population-based indicators, encounter-based indicators are computed in units of patient encounters with the system. Thus, only 43 per cent of encounters by high-risk patients compared with 67 per cent of encounters by average-risk patients due for recognition of pregnancy resulted in pregnancy recognition. Similarly, only 33 per cent of encounters by high-risk patients compared with 71 per cent of encounters by average-risk patients with pregnancy recognized and due for gonorrhea screening received a cervical culture. When disaggregated by site of contact, the results suggest that the prenatal clinic performs well in both the recognition and screening function, the public

[##] Care must be taken in multiplying together P_{ij}'s obtained from different tracer conditions and different cohorts, because patients receiving adequate care early in the process of care may not be independent of those receiving adequate care later on. Specifically, patients receiving adequate care at one stage may also be more likely to receive adequate care at another. In fact, such nonindependence is illustrated in the data reported by Novick,[10] although it was not explicitly pointed out by him. We have avoided this problem by calculating continuity-of-process scores only for functions or sequences of functions where we were able to observe the passage of the cohort throughout the sequence. Thus, our continuity-of-process scores are observed, not calculated. When greater knowledge is obtained about the interdependence of functions we may be able to combine continuity-of-process scores obtained from different cohorts.

TABLE 4. Results From Three Service Units of the Indicator Sequence Designed to Examine the Continuity of Process for Treatment and Follow-Up

		Service Unit C	Service Unit D	Service Unit E
1. Contact for evaluation	The proportion of patients with a positive screening result who made contact with the health care system within an appropriate time frame (within 3 weeks for anemia and within 2 weeks for urinary tract infection).	0.92 (62/67)	0.93 (93/100)	0.73 (73/100)
2. Recognition of problem	The proportion of patients making contact for whom there was any evidence that the problem was recognized.	0.90 (56/62)	0.76 (71/93)	0.62 (45/73)
3. Provision of treatment	The proportion of patients with the problem recognized who received appropriate treatment within 2 weeks after recognition.	0.95 (53/56)	0.99 (70/71)	0.93 (42/45)
4. Contact for follow-up	The proportion of patients achieving contact with the health care system within a time frame appropriate for follow-up (within 3–6 weeks after treatment initiated for anemia and within 4 weeks after antibiotic course completed for urinary tract infection).	0.83 (44/53)	0.89 (62/70)	0.52 (22/42)
5. Recognition of need for follow-up	The proportion of patients making contact for whom there is any evidence that the problem and/or the need for follow-up was recognized.	0.89 (39/44)	0.61 (38/62)	0.41 (9/22)
6. Provision of follow-up	The proportion of patients with recognition of problem and/or the need for follow-up, who received an appropriate follow-up (hemotocrit, hemoglobin, or reticulocyte count for anemia; urine culture or microscopic urinalysis for urinary tract infection).	0.95 (37/39)	0.95 (36/38)	1.0 (9/9)
Continuity-of-process index	The proportion of patients with a positive screening result who completed the entire health care sequence. (The continuity-of-process index may also be computed as the product of the proportions for each of the indicators 1 through 6).	0.55 (37/67)	0.36 (36/100)	0.09 (9/100)

The results shown are an aggregate of data derived from nutritional anemia and urinary tract infections. The study cohorts for Service Units D and E are random samples of 50 patients for each tracer. The study cohorts from Service Unit C consist of a total sample of 36 patients with anemia and 31 patients with urinary tract infection.

health nurses perform well in recognition, but the hospital outpatient department and the field clinics contribute substantially less to the performance.

The data in Table 6 show that the average-risk women who were due for care made a higher proportion of their encounters with the prenatal clinic than did the

TABLE 5. Results From a Service Unit Showing Population-Based Indicators for Prenatal Care

		Total Cohort (n = 50)	High Risk (N = 22)	Average Risk (N = 28)
Contact	The proportion of pregnant women who made contact with the health care system by the 20th week of gestation.	0.64 (32/50)	0.54 (12/22)	0.71 (20/28)
Recognition	The proportion of pregnant women making contact who had their pregnancy recognized by the 20th week of gestation.	0.69 (22/32)	0.50 (6/12)	0.80 (16/20)
Screening	The proportion of pregnant women with pregnancy recognized who had a cervical culture by the 20th week of gestation.	0.82 (18/22)	0.50 (3/6)	0.94 (15/16)
Continuity-of-process index	The proportion of pregnant women who had a cervical culture by the 20th week of gestation.	0.36 (18/50)	0.14 (3/22)	0.54 (15/28)

A patient was considered high risk if she was less than 18 years of age, over 35 years of age, primigravida, had parity equal to or greater than 5, or had a history of miscarriage or spontaneous abortion.

high-risk women. Specifically, for average-risk women 42 per cent (10 of 24) of all encounters when due for recognition were with the prenatal clinic and 67 per cent (14 of 21) of all encounters when due for screening were with the prenatal clinic, while for high-risk women only 7 per cent (1 of 14) and 33 per cent (3 of 9) of all encounters when due for recognition and screening, respectively, were with the prenatal clinic. Consequently, the superior performance of the prenatal clinic favors the average-risk group by virtue of the different utilization patterns of the two risk groups. A similar pattern at another Service Unit has been studied in detail and is reported elsewhere.[21]

The above examples illustrate how the assessment method operates, and how it can point to areas where there are de-

TABLE 6. Results for Encounter-Based Indicators of Pregnancy Recognition and Screening for Gonorrhea, Disaggregated by System Component

	Cohort Total	High Risk Patients	Average Risk Patients
Recognition rate: The proportion of visits by the 20th week of gestation by patients not previously recognized as pregnant, in which recognition of pregnancy occurred.			
Hospital outpatient department	0.35 (6/17)	0.25 (2/8)	0.44 (4/9)
Prenatal clinic	1.00 (11/11)	1.00 (1/1)	1.00 (10/10)
Field clinic	0.00 (0/5)	0.00 (0/2)	0.00 (0/3)
Home visit by public health nurse	1.00 (5/5)	1.00 (3/3)	1.00 (2/2)
Total visits	0.58 (22/38)	0.43 (6/14)	0.67 (16/24)
Screening rate: The proportion of visits by the 20th week of gestation by patients with pregnancy recognized and due for a cervical culture, in which a cervical culture was taken.			
Hospital outpatient department	0.11 (1/9)	0.00 (0/3)	0.17 (1/6)
Prenatal clinic	1.00 (17/17)	1.00 (3/3)	1.00 (14/14)
Field clinic	0.00 (0/1)	0.00 (1)	— (0)
Home visits by public health nurse	0.00 (0/3)	0.00 (0/2)	0.00 (0/1)
Total visits	0.60 (18/30)	0.33 (3/9)	0.71 (15/21)

ficiencies. The anecdotal data presented on corrective actions suggests that in some cases the method may enhance the assurance function, although we do not yet have systematic evidence.

Discussion

Requirements for a method to assess the quality of care in Indian Health Service service units were defined. A method was developed to try to fulfill these requirements and was applied in 25 health systems to test whether the requirements had been met. Some of the requirements were fulfilled while others were not.

The first requirement calls for examining the performance of the entire health system and is based on the assumption that the performance of the total health system cannot be inferred from the performance of individual components, which is especially important in Indian Health Service, where consumers are often scattered geographically and highly mobile. Rather than being limited to specific facilities or provider groups, the assessment begins with a sample of the community with specific needs for health service and tracks them through their encounters with all system components, including components that do not necessarily view themselves as part of the system, thus fulfilling the requirement.

The second requirement calls for the examination of care for all members of the community, not just patients, and for differentiation among subgroups. By selecting study cohorts from the entire known community of consumers, as demonstrated in the immunization and prenatal care tracers, and using original source documents, as demonstrated in the anemia and urinary tract infection tracers, the second requirement is largely met. However, more work must be done to determine the extent to which the entire community of consumers has been identified in IHS service units.

The distribution of care by risk group was successfully examined, with the finding that two maternal and child health clinics provided better care to average-risk mothers than to high-risk mothers because of different utilization patterns. Although it appears that the method could also examine the distribution of care across other subgroups of the population, this has not yet been demonstrated. **

The assessment focuses on the performance of a broad range of functions, including especially those most important for ambulatory care, rather than on particular health conditions, as called for in the third requirement. However, the usefulness of the resulting information depends on the assumption that the care received for the tracers is representative of all care. The original application of this method found that the functions had similar relative performances for all the tracers used,[7] which was supportive of this assumption. More recent applications of the method also found similar patterns for most tracers, but with some notable exceptions, for example the high recognition rate for suture follow-up care compared with much lower recognition rates in follow-up care for other tracers. Such variation across tracers indicates that a thorough investigation is needed on whether and how to aggregate across tracers. For example, classes of tracers might be found that have similar patterns of care within classes but different patterns across classes. Although current aggregates are possibly useful starting points, they are not based on sound theoretical or empirical foundations and therefore should be accompained by the tracer-specific results.

** In concept, any attribute of medical care can be distributed across different subgroups of the population. The distribution of accessibility[32] and utilization[33] across various types of population subgroups has been widely studied, but the process of care as we have defined it has not. Shortell[34] has attempted a more precise definition of distribution of care by applying concepts of economics.

The fourth requirement is largely met in that the assessment method traces the care received throughout an episode and relates it temporally to health status. However, causal linkages between process and health status are not established by the assessment and so outcome of care is not established. The number of health conditions available to use as tracers is expanded greatly by examining only segments of the process of care rather than the entire process for every tracer, although this limits the conclusions that can be drawn about the continuity of the entire process of care because the functions are not independent of one another. More work is needed to understand clearly how changes in one function influence the performance of others.

The fifth requirement established criteria for the selection of tracers. These criteria have been met with the exception that the representativeness of the tracers as a group has not been demonstrated, as noted above.

The sixth requirement establishes criteria for defining standards of care for the tracers. The standards established were both minimal and agreed upon by the local practitioners. The work of Brook[24] suggests that because the assessment method uses explicit standards for the process of care rather than implicit judgments or indicators of outcome, the absolute performances will be relatively low. However, this is not important so long as the performance indicators are consistent over time and across components, functions and consumer groups and their statistical attributes are not relatively worse than methods using implicit judgments or outcome indicators. More investigation of these assumptions is needed.

Although the assessment method identifies areas of major deficiencies, it has not demonstrated that it systematically identifies *remedial* deficiencies, as called for in the seventh requirement. In order to determine if the deficiencies uncovered are able to be corrected, the assessment must be imbedded in an effective assurance process. Alternative possible remedies must be identified and evaluated and then implemented successfully. This is the point where other assessment methods have encountered the most difficulty, and it is widely recognized that assessments have not reliably led to improvements in quality.[13,16,24,25]

The eighth requirement deals with cost. Our experience indicates that the cost of employing this method in IHS service units is modest. Nonprofessionals were used successfully to collect and tabulate data from the nonstandardized health records currently found in Indian Health Service.¶¶ In 10 service units it took 13 data collectors and two supervisors a total of 291 working days to identify the cohorts and abstract data from the health records, plus another 69 working days to tabulate the data. Thus an average of 36 nonprofessional days was required to collect and tabulate the data per service unit. Each worker also spent eight days in training. On the average, data were abstracted from 3.4 different health records for each consumer, with study cohorts of approximately 50 consumers for each of nine tracers. This produces an average time for data collection and tabulation of 38.4 minutes per episode of care studied and 11.3 minutes per record abstracted.

Others have also reported on the cost of data collection by nonprofessionals for episodes of care using explicit standards. Some agree with our experience that the cost is modest[7,10] while others have found it expensive.[14,26] Recent work by Albrecht and Kessner[27] suggests that while the cost of abstracting may be modest, the fixed

¶¶ Although not reflected in the above cost estimate, IHS is implementing a computerized health information system which integrates data from all the health records maintained for a given consumer. Once in operation, this system should significantly increase the efficiency of the assessment through its use of standardized encounter forms, a well-defined list of consumers from which to draw cohorts, and identification of the location of all relevent records.

costs for operating such an assessment method may be high.

This assessment method is not offered as a replacement for other methods currently available that provide an in-depth examination of specific system components. Rather, it is offered as a broad-brush assessment which provides a mechanism for detecting problems in system performance which may then be examined in more detail using the other methods. Use of this approach in a variety of health care settings has proved to be feasible and has uncovered areas of deficiency in total system performance. Work on improving and applying the method continues by Indian Health Service, with particular effort being given to testing the reliability of its measurements, examining how to select more representative sets of tracers, and in imbedding it in an assurance process that improves the quality of care.

Acknowledgment

The authors wish to express their gratitude to Robert H. Brook, M.D., for his many helpful comments and encouragement, and to Louise Burbank, Claudina Noriega and Sandra J. Woolbright for their diligent efforts in the preparation of the manuscript.

References

1. Carlson R. The end of medicine. New York: John Wiley and Sons, 1975.

2. Illich I. Medical nemesis: the expropriation of health. London: Calder and Bayars, 1975.

3. Navarro V. The underdevelopment of health of working America: causes, consequences, and possible solutions. Am J Public Health 1976;66:538.

4. Fuchs V. Who shall live? New York: Basic Books, 1974.

5. Moss GE. Illness, immunity, and social interaction. New York: John Wiley and Sons, 1973.

6. Lawrence PR, Lorsch JW. Differentiation and integration in complex organizations. Administrative Science Quarterly 1967;12:1.

7. Shorr GI, Nutting PA. A population-based assessment of the continuity of ambulatory care. Med Care 1977;15:455.

8. Kessner DM, Kalk CE, Singer J. Assessing health quality—the case for tracers, N Engl J Med 1973;288(i):189.

9. Kessner DM, Kalk CE. Contrasts in health status. Vol 2: a strategy for evaluating health services. Washington, D.C.: National Academy of Sciences, 1973.

10. Novick LF, Dickinson K, Asnes R, May Lan SP, Lowenstein R. Assessment of ambulatory care: application of the tracer methodology. Med Care 1976;14:1.

11. Nobrega FT, Morrow GW Jr, Smoldt RK, Offord KP. Quality assessment in hypertension: analysis of process and outcome methods. N Engl J Med 1977;296:145.

12. Gonnella JS, Daniel ZL, McCord JJ. The staging concept—an approach to the assessment of outcome of ambulatory care. Med Care 1976;14:13.

13. Hirschhorn NH, Lamstein J, Klein SF, McCormack J, Warner TN. Quality by objectives: a model of quality of care assessment and assurance for ambulatory health centers. Journal of Ambulatory Care Management 1978;1:55.

14. Howell JR, Osterweis M, Huntley RR. Curing and caring—a proposed method for self-assessment in primary care organizations. J Community Health 1976;1:256.

15. Shortell SM. Continuity of medical care: conceptualization and measurement. Med Care 1976;14:377.

16. Donabedian A. Needed research in the assessment and monitoring of the quality of medical care. Washington, D.C.: Department of Health Education and Welfare, 1978 (DHEW publication no. PHS-78-3219.)

17. Beck MH, Drachman RH, Kirscht JP. A field experiment to evaluate various outcomes of continuity of physician care. Am J Public Health 1974;64:1062.

18. Starfield BH, Simborg DW, Horn SD, Yourtee SA. Continuity and coordination in primary care: their achievement and utility. Med Care 1976;14:625.

19. Breslau N, Reeb RG. Continuity of care in a university-based practice. J Med Educ 1975;50:965.

20. Steinwachs DM. Measuring provider continuity in ambulatory care: an assessment of alternative approaches. Med Care 1979;17:551.

21. Nutting PA, Barrick JE, Logue SC. The impact of a maternal and child health care program on the quality of prenatal care: an analysis by risk group. J Community Health 1979;4:267.

22. Aday L, Anderson R. Development of indices of access to medical care. Ann Arbor: Health Administration Press, 1975.

23. Anderson R. Health service use: national trends and variations. Washington, D.C.: Department of Health Education and Welfare, 1972.

24. Brook RH. Quality of care assessment: a comparison of five methods of peer review. Washington, D.C.: National Center for Health Services Research and Development, 1973. (Publication no. HRA 74-3100.)

25. Williamson JW. Assessing and improving health care outcomes. Cambridge, Mass.: Ballinger Publishing Company, 1978.

26. Morehead MA. Ambulatory care review: a neglected priority, Bull NY Acad Med 1976;52:60.

27. Albrecht JA, Kessner DM. Assessing ambulatory care: a comparative analysis of three methodologies. (Unpublished memo.)

Systems and Procedures of Patients and Information Flow

Arnold Reisman, Ph.D., Joao Mello da Silva, Ph.D., and Joseph B. Mantell, Ph.D.

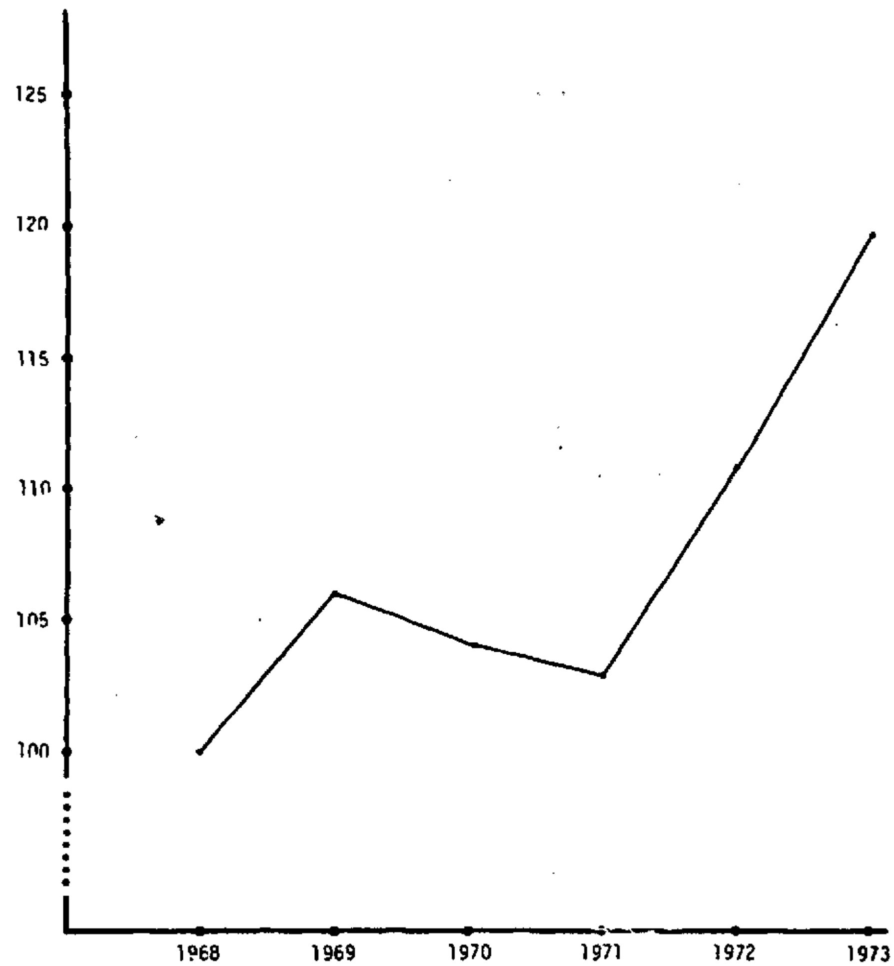
THE TASK of properly scheduling an annual load of several hundred thousand patient visits to more than 100 doctors in 28 different departments, and supporting these visits with timely, accurate, and complete flows of information is very complex.

This task is further complicated by the fact that patients generally require timely sequencing of laboratory tests, X-rays, and consultative appointments. From 1968 to 1973 in the clinic of one large health center, the number of patient visits per year increased approximately 20% (Figure I) and the institution's physical plant was greatly expanded, tying up much of the administrative talent and resources. More importantly, however, the staff capability had become more diversified. It had disproportionately increased in number to serve the rapidly-expanding inpatient population and to perform additional research and teaching activities. The systems and procedures for scheduling and processing the patients through the outpatient clinic degraded to the consternation of patients, doctors, and administrators alike.

The objective of this study was to investigate the systems and procedures for outpatient flow and to recommend improvements. The specific objectives were to:

1. Understand and describe, in operationally meaningful terms, outpatient flow systems and procedures in effect at the time of the study;

Figure I. Total Patient Visits



THE PATIENT VISITS IN 1968 ARE USED AS THE BASE FIGURE (=100)

2. Identify problem areas in any and all patient flow related operations;
3. Collect and analyze meaningful hard data to test various hypotheses regarding the status of the patient flow systems and procedures and the reasons for the evolution of these systems to their current state;
4. Delineate the most meaningful and reasonable goals and objectives for the patient flow system;
5. Recommend both long-range and short-range improvements to the patient flow related operations.

Historical background

The "old appointment system" was comprised of a central appointment desk (CAD) and several partially decentralized non-uniform procedures. Within the Medical Division, the CAD scheduled physical examinations only. All types of appointments of the surgical division as well as "former" appointments for all departments were made either by the desk receptionists or by the medical secretaries. Routing Section personnel were responsible for consults, X-rays, laboratory tests, special exams, and reports. Each doctor kept an appointment sheet specifying the time of day during which different types of patients should be scheduled.

Because of the rapid growth of the outpatient clinic and a lack of adequate support, the CAD became less effective in performing its job. As matters worsened, complaints about patient flow were expressed by physicians throughout the clinic. The major criticisms were incompetence in the scheduling function and favoritism (more patients were directed to certain physicians or departments than to others.) Consequently, some departments transferred control of their appointments (except reports and consults which were kept by the Routing Section) to appointment secretaries stationed on the department floors. Further movement toward decentralization followed. As each department devised its own rules for scheduling patients, a gradual degeneration of the clinic's overall systems and procedures for making appointments resulted. This degeneration was further aggravated by the rapid turnover of clerical personnel which was spurred on by conflicting directives, lack of systematic training and enforcement, and a system which was difficult to master or comprehend. Figure II depicts the dynamic feedback process which led to the degradation of systems and procedures.

Productivity of the clerical staff, measured in terms of patient visits/clerical position ratio, is shown in Figure III. The ratio dropped sharply for the appointment-making personnel. In contrast, the ratio for the other clerical staff categories remained fairly constant. It is interesting to note that while the "actual" system effectiveness was dropping, the *perceived* effectiveness, as seen by the physicians, had discontinued its downward trend and appeared to be improving because of the ability of physicians to exercise more control over clerical staff as the clinic's systems became more decentralized.

A study of the problem was made in early 1972 by a Subcommittee on Registration and Routing which generated a report. Another attempt to study the problem was made in late 1973, with the establishment of a Patient Flow Committee. In January, 1974, this committee established the task force which was led by the authors of this paper.

Method of study

The aim of this study was to report on the system description of the clinic and to provide a set of recommendations. It was recognized that these were preliminary recommendations and that more detailed investigation of the problem areas identified by this study should be undertaken. Specifically, it was felt that subsequent studies should repeat the activities of this study, utilize mathematical models, and develop an implementation scheme (Figure IV).

As shown in Figure IV, this study phase was devoted to learning, describing and analyzing the clinic as a system. As part of this phase, a graphical description was obtained by flow charting the various patient flow related processes. Also, hard data were collected to test various hypotheses which emerged from discussions with members of the clinic's medical, administrative, and clerical staffs on patient flow related operations. Future work

Figure II. Historical Background

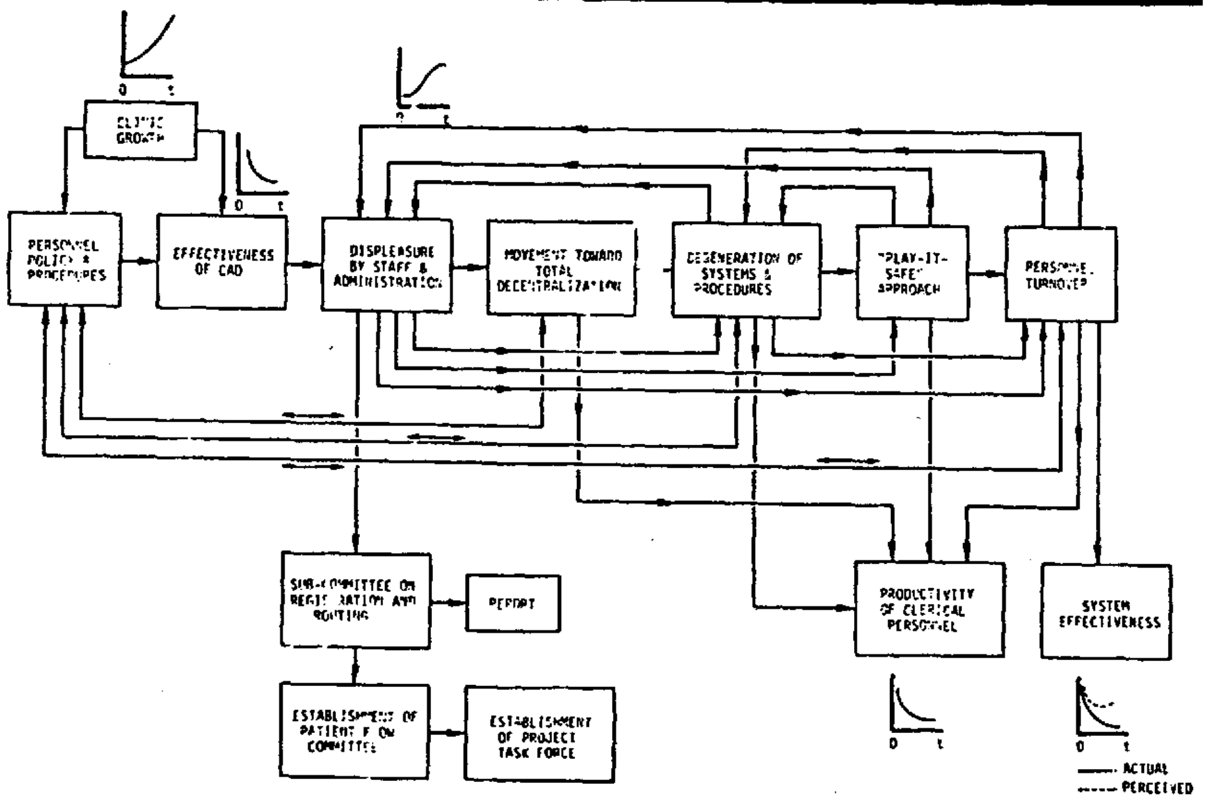


Figure III. Patient Visits/Clerical Staff Ratio

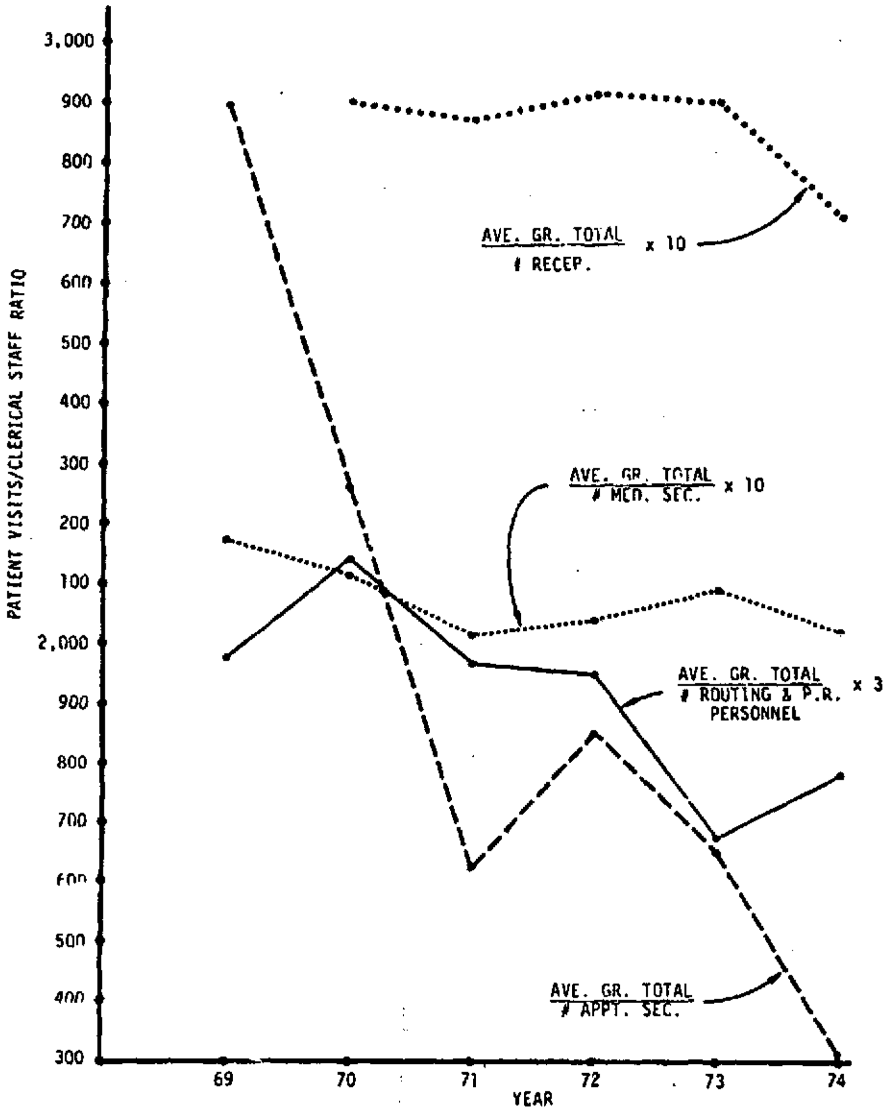
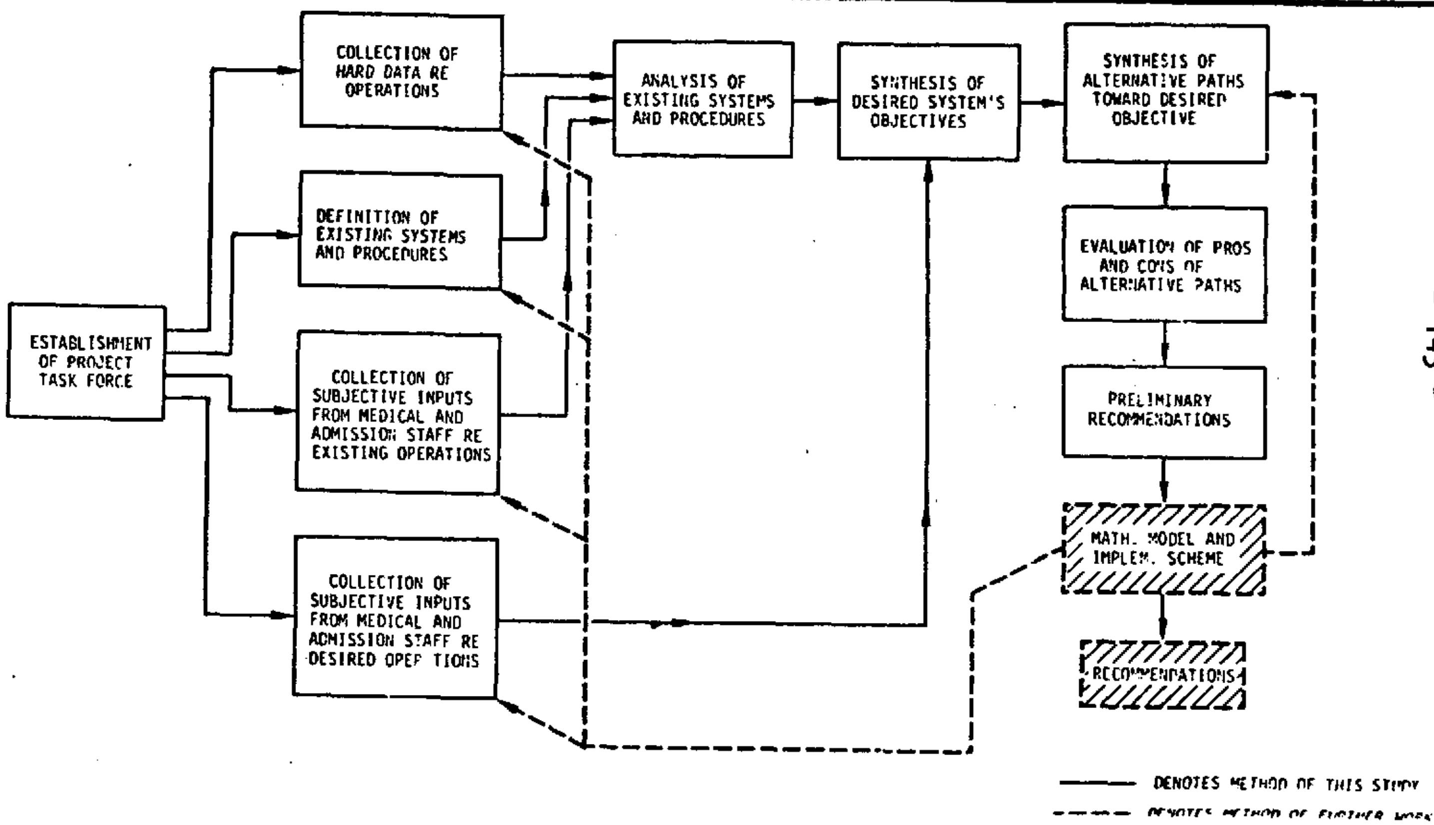


Figure IV. Method of Study



should be devoted to the development, testing, and evaluation of each of the preliminary recommendations through mathematical analysis and/or computer simulation.

Patient, appointment and physician classification

The patients at the clinic are classified into three categories: "new," "new-old," and "former." A new patient is one that is coming to the clinic for the first time. In most medical departments, a former patient is one who has been at that department previously and has had a complete physical examination within a reasonable time. A "new-old" patient is one who, regardless of having been at that department previously, has not had a complete physical examination recently. This length of time varies from one to two years depending on the department. Some medical and surgical departments do not require physical examinations. In these departments, a former patient is one who has been at that department within a specified length of time, while a new-old patient is one who is coming to that department for the first time or returning after the specified length of time.

Patient appointments are classified into five categories: new, new-old, former, consult or referral, and report. A consult is necessary when the patient's doctor feels that the patient should be seen by specialists or subspecialists. Normally this takes place in another department. A report is an appointment (which may or may not involve an examination) during which the doctor informs the patient of test results and diagnoses.

Some of the clinic trainees, called "fellows," are allowed to perform work-ups of patients for the staff doctors in certain departments. Depending on the department, the fellows either form a pool and work up patients for any staff doctor or each one works up patients for only one staff doctor.

Senior fellows may be given the privilege of assuming patient responsibility under staff supervision. The number of appointments available by a department depends, therefore, not only on the staff's available time to see patients but also on the number of fellows assigned to the department and on the functions they are allowed to perform.

The current appointment-making system

In general, there are two basic appointment-making systems. One is the centralized system, in which all appointments for all departments are made in one location. The other is the decentralized system, in which all the appointments of each department are made by a secretary or equivalent personnel stationed on the department floor. Every request regarding appointments is directed to a central area for the centralized system and to the corresponding secretary of a department for the decentralized system. All other appointment systems are variants and/or combinations of these two basic systems. Both systems have distinct advantages.

In the centralized system, since there is only one location responsible for all the scheduling:

1. Calls for appointments will always be correctly directed or connected. In contrast, calls to the decentralized system may be transferred several times before reaching the correct location.
2. Persons making the appointments know the available times for *all* doctors. Functions such as coordination of multiple appointments, and modifications of appointment schedules to keep up with the demands of the various appointment types can therefore be efficiently performed.
3. Patients will have only one appointment card. In the case of a multiple appointment, paper work will be kept to a minimum and the Record Room will receive only one request for the patient's chart.
4. Economy of scale may result.

In the decentralized system, since each department has its own appointment-making person(s):

1. Appointments are made for but a few doctors in a single specialty. Thus it is easier for the person to know the doctors' characteristics (e.g., preferences for certain cases and

the lengths of time usually required for the various appointment types) than it is for an appointment secretary in a central location who must know the characteristics of all clinic doctors.

2. The learning period is shorter and so replacement personnel are more easily trained.
3. Follow-up appointments are made at the floor just after the examinations.
4. The doctors can easily check their schedules for availabilities and patient load and any changes in schedule can be easily communicated to the appointment secretary.

The clinic studied had a combined centralized-decentralized system. The centralization was characterized by the work of the Routing Section. This section made most of the consult and report appointments, and scheduled special exams, e.g., EKG, EEG, basal metabolic rate test. The decentralization was characterized by the fact that the other appointment types (and also some reports and consults) were made in the departments by several clerical people. A study was conducted to determine who was scheduling patients for each department.

Five types of appointment-making personnel were identified: appointment secretaries, receptionists, medical secretaries, Routing Department personnel, and personnel from the Appointment Service Desk.[1] For nearly all departments, different types of personnel were responsible for different appointment types and different appointment-making procedures were being used. In general, the study clearly showed the lack of uniformity of systems and procedures throughout the clinic.

Current appointment practices

Upon making an appointment, a patient is assigned to a doctor or for a special exam in a specific department at a specified time.

Appointment systems may be classified into three categories: pure block, individual, or mixed block-individual.

[1]The Appointment Service Desk makes some of the new and new-old appointments. Its main function is to screen calls which have not specified a doctor or a department.

A pure block appointment system assigns the same appointment time at the beginning of the clinic session to all patients. An individual system assigns a different appointment time to each one of the patients. A mixed block-individual assigns a group of patients to certain times (generally at the beginning of a clinic session), but also assigns individual appointment times to each one of the other patients.

At the time of the study only one department of the clinic used the pure block system. The individual appointment system was used by most doctors; the mixed block-individual appointment system was used by the others.

There are two basic types of prescheduling. In "prior prescheduling" the patient is scheduled for certain laboratory tests, X-rays and/or special exams *before* an appointment with the primary doctor. In "posterior pre-scheduling," tenative appointment times are saved for the patient, in addition to the patient's primary doctor appointment. *After* the primary examination, the doctor decides whether the tentative appointments should be used or cancelled. Combinations of the two types of prescheduling can and have been used. There are advantages and disadvantages to both kinds of prescheduling.

One advantage of prior prescheduling is that the primary doctor will have laboratory, X-ray and special studies results available at the time of the appointment. This may help in the initial examination. Another advantage is that when no further tests and/or consults are necessary for the patient, the doctor may be able to give a full diagnosis at the initial appointment. One disadvantage is that the patient may sometimes take unnecessary tests. Another is that as a result of the doctor's examination, the patient may be required to return to the laboratory[2] or to the Radiology Department.

The advantage of posterior prescheduling is that the time span for the complete routing may be less than the time span which could be obtained on the day of the appointment. One of the disadvantages is that when the doctor decides to cancel some of the prescheduled appointments, these appointments may go un-

[2]In some cases, it may be possible to eliminate this inconvenience to the patient by saving specimens and blood samples for a certain period of time in the lab.

filled due to the short lead-time of the cancellation. In addition, the advantage of this type of prescheduling may be negated when the doctor finds it necessary to order additional tests and/or consults.

At the time the study was conducted, neither kind of prescheduling was extensively used at the clinic.

Systems-related problems

Long lead-times for initial appointments were of great concern to the clinic as well as to the patients. If their appointment is too far into the future, patients may look for medical care elsewhere. Consequently, appointment times are not used because such patients often do not cancel (they become no-shows on the day of the appointment) or they cancel too late for the clinic to react and fill the appointment slot.

A two-week appointment distribution study[3] was run to determine the number of no-shows (including unfilled cancellations) and appointments, both unfilled and kept, for the different appointment types for each department. The study revealed that a significant percentage of appointment times, especially those reserved for consult and report appointments, were going unfilled in many departments. Also revealed was the fact that no-shows are a significant problem for the clinic. For the clinic as a whole the percentage of no-shows for new, new/old, and former appointments was more than 12%.

Proper balancing of patient appointment types on the physicians' day sheets was a problem. This can be seen quite vividly by examining the results of the appointment distribution study mentioned above and comparing these with the results of another study which examined the appointment availabilities for new, new-old, and consult appointment types, as they were one week after the beginning of the project. In this study it was found that in some departments the next available new, new-old, or consult appointment was often as long as one to two months into

[3]Only patients who actually made appointments were considered. Those who did not make appointments due to the long lead-time or any other reason were not considered.

the future. It was discovered that in some departments certain appointment times, e.g., reports, were unfilled while at the same time the demands for new, new/old and consult times exceeded their availability.

Figures V and VI identify the problems:

1. The number of ways that a patient can be transferred before making an appointment or cancellation.
2. Patients wishing to make cancellations may not reach the correct locations. Consequently, if a patient has multiple appointments, some appointment times will remain reserved, even though the patient has contacted the clinic to cancel them.[4]

Both problems resulted because the various types of appointments are made in many different locations and, in general, the persons making appointments (and cancellations) do not know exactly who make the various types of appointments for other departments.

Especially for the out-of-town patients, the time span of routing should be kept as short as possible.[5] The major obstacles to a short time span are the difficulty in obtaining consults, time availability constraints (i.e., certain tests are only available at certain times of the day), and procedural constraints (i.e., some tests which require special preparation may not be followed by certain other tests.)

It was found that missing or incomplete charts were a major problem for many people in the clinic. A missing or incomplete chart may cause delay in the patient examination, a delay in the report, and often results in a great loss of time in locating the missing material.

When an appointment is made, the person who makes it should generate an appointment slip and send a copy to the chart processing area. This is called preregistration. If a patient is not preregistered, the paper work is started when the patient arrives

[4]A patient with multiple appointments may assume that if he cancels one of the appointments, the remaining ones are also cancelled. Usually this is not true.

[5]It was the clinic's policy to try to schedule out-of-town patients for the shortest possible routing time span.

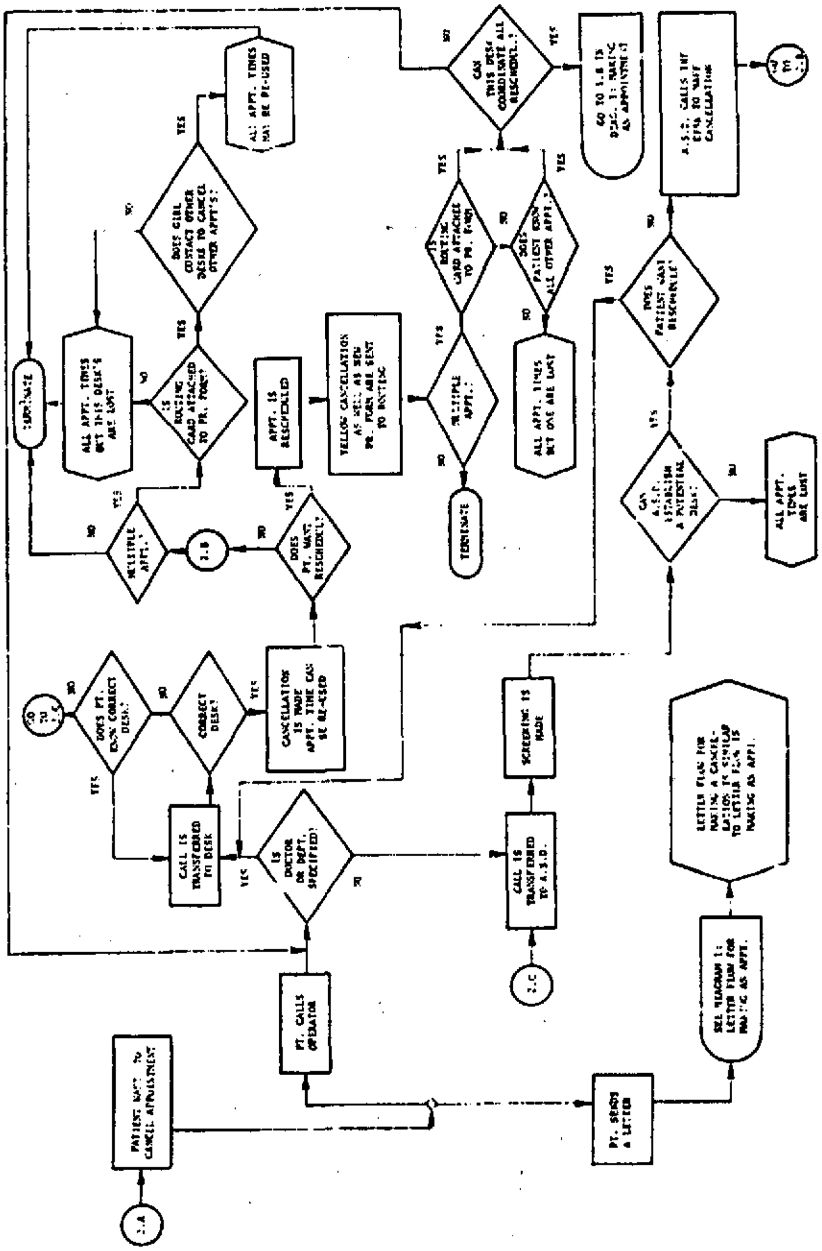
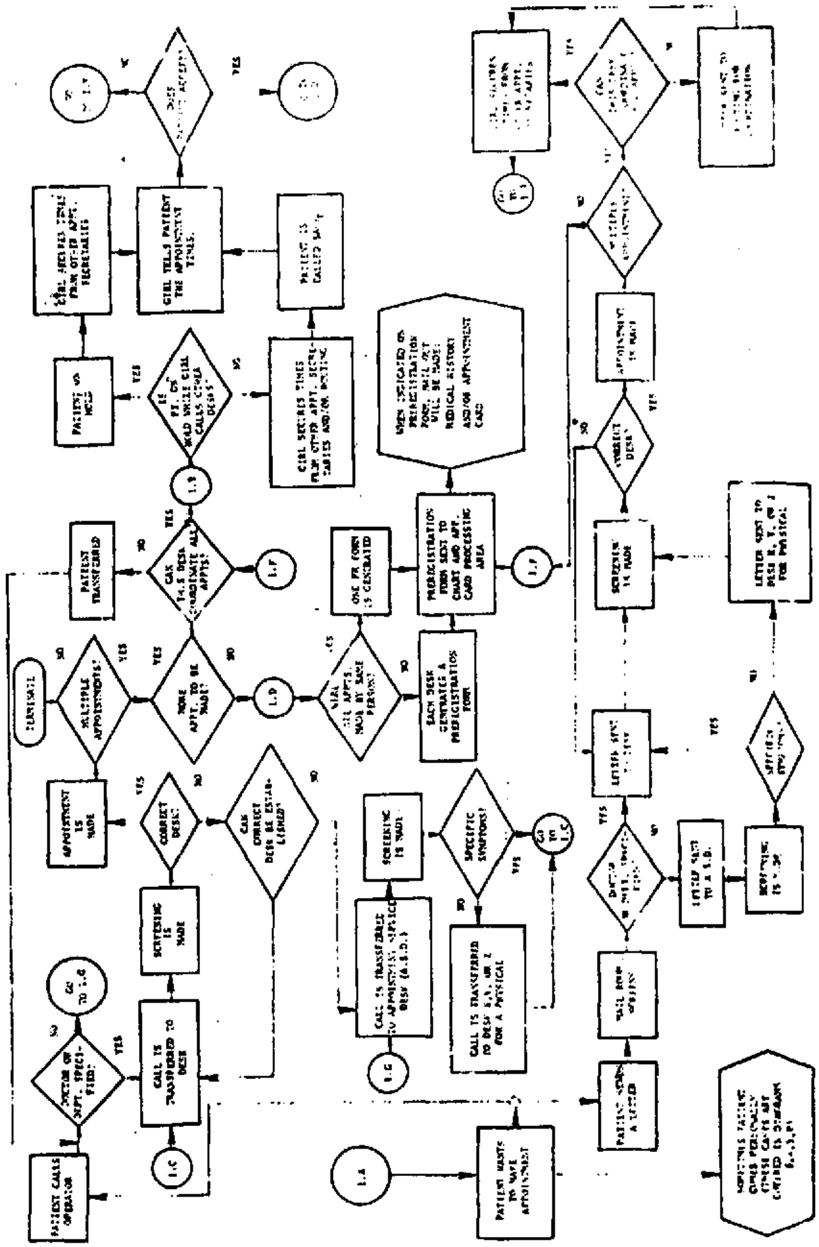


Figure V. Making a Cancellation

Figure VI. Making an Appointment



at the clinic. As a consequence, the patient may be late for the appointment or may be missing necessary paper work.

In the case of multiple appointments, it is possible for the patient to be preregistered more than once. This occurs when all the appointments are not made by the same person. As a result, there is duplication of paper work, the chart may not be sent to the place of the first appointment (it will be sent to the first place requesting it), and if the patient decides to cancel, appointment times are usually lost (Figures V and VI).

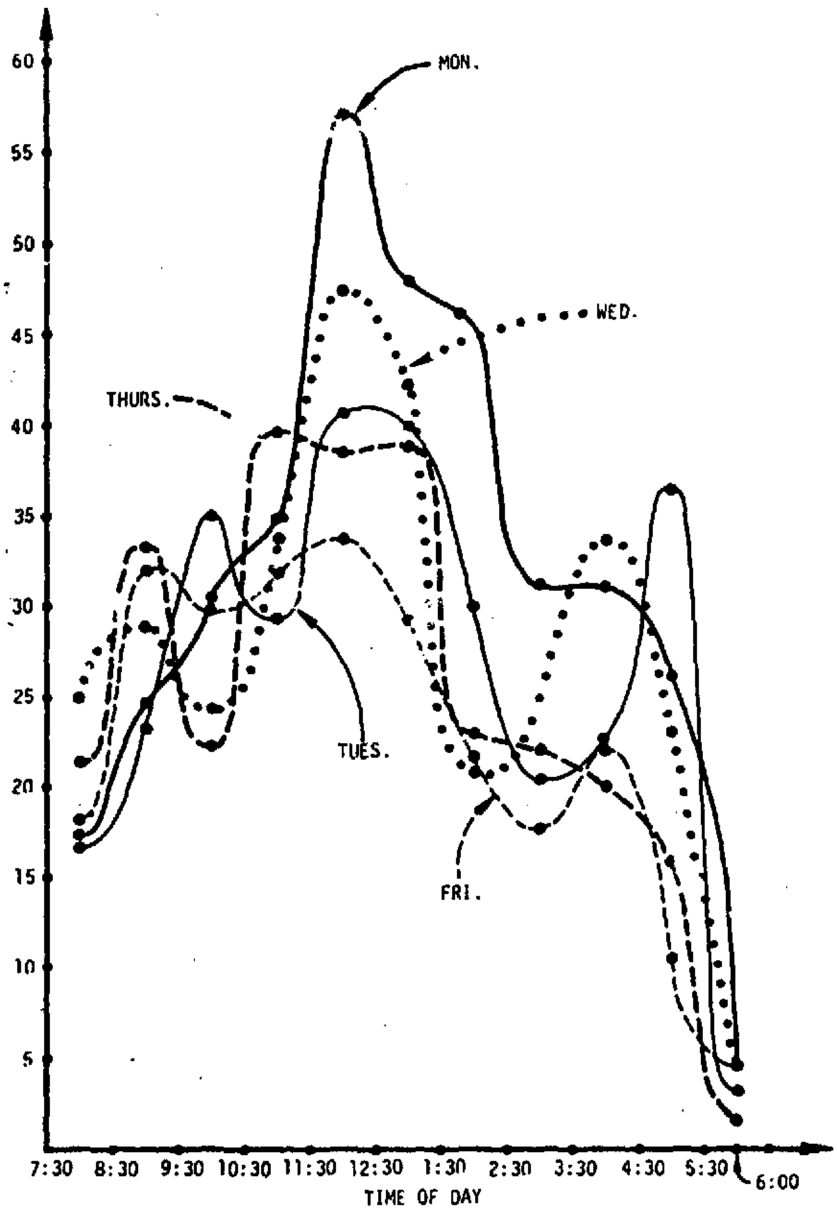
Physical examinations are usually scheduled early for both morning and afternoon clinic sessions, mostly for the morning. Often routing, laboratory tests, and X-rays are required after these examinations. Since a physical, on the average, takes one to two hours, the bulk of the physicals are finished after 10 a.m. and before noon. As a consequence, bottleneck situations develop in routing, laboratory, and radiology between 10 a.m. and 1:30 p.m. A smaller bottleneck occurs around 3 p.m. as a result of the completion of the early afternoon physicals. Figure VII shows these peak-load situations at the laboratory. The situation is worsened by the fact that the heaviest patient load occurs during the lunch hours when these three areas (routing, radiology and laboratory) are often short of personnel.

Room availability is another source of problems. These problems tend to be more noticeable in the medical departments, especially those performing complete physical examinations. This is less of a problem for the surgical division because surgeons, in general, spend more time in the hospital than the medical division physicians. As a result of room constraints both the physician and the patient may be required to wait for a vacant examination room.

A study of this problem was conducted. In recognition of the difference in needs for rooms between medical and surgical divisions, the data on the examination room availability were separated by division. The number of examination rooms in a department was compared with the number of staff physicians on duty. It was apparent from this study that some departments had more than enough rooms while others had too few.

Physicians' schedules also have an impact on the problem. For

Figure VII. Patient Arrival Distribution for Blood Drawing in Laboratory—Day of Week/Time of Day



example, a physician may schedule two physicals at 8 a.m. These patients will be worked up by fellows while the staff doctor may be seeing former patients. In this situation, the staff physician's patients will be using three or more examination rooms at the same time. As was seen in the study, few departments can afford this many examination rooms for each staff physician.

Personnel-related problems

Obviously, the clinic's patient flow system is very complex. For the entire existing system to work properly, it was necessary for the many patient flow related processes to interact smoothly. Thus, it was imperative that *all* personnel understand how their work affected not only their particular process but also all the others with which it interacts. At the time of the study, there were no structured training programs in which the new employees were given an overall perspective of their job as it related to other clinic operations. The only training which took place was on-the-job because a new employee was hired only when the position was already vacant. Therefore, adequate training was very difficult because of the urgency with which the employee was needed. Poor job performance and lack of coordination resulted since a new employee did not know how his or her work interacted with the other employees' functions.

In some areas there were certain inequities in job classifications. Some jobs, where specific skills were a definite requirement (such as typing), were classified on the same level as jobs requiring no specific skills, no patient contact, and no machinery to operate. This was due, in part, to incorrect original classifications but more often it was the result of lack of reclassification as job duties changed and became more complex. There was no procedure apparent for regular updating of job descriptions and classifications. The job classification inequity seriously delayed recruitment of qualified personnel. To compensate for this problem supervisors and department heads juggled personnel within their departments to insure efficient operation. They, at times, had to shift employees from one classification to another without changing salary or job classification. This often resulted in personnel working in positions outside their classification and in dissension among employees if variations in salary scales became known.

Depending on the department, the work-load varied considerably for the same job classification. For example, an appointment secretary of a department attending many patients had a heavier work load than one stationed in a department attending fewer patients.

A lack of upward mobility in the job often resulted in lack of interest and motivation to perform well since there was no incentive or additional compensation for doing so. There were only two job levels for medical secretaries; medical secretary and senior medical secretary. Appointment secretaries had but one job category and no opportunity for advancement.

Conflicting directives caused personnel problems. In many instances, clerical personnel received instructions from physicians as well as from a supervisor. Sometimes these instructions conflicted with one another and it therefore became difficult for the supervisors to enforce those systems and procedures which should be uniform throughout the clinic. This situation presented a dilemma for clerical personnel and resulted in a decline of morale.

Rapid personnel turnover further aggravated the problems. This lack of training resulted in poor job performance and criticism from both physicians and supervisors (sometimes conflicting directives), which in turn resulted in low morale. Low morale causes personnel turnover. As one can see, the situation was a vicious circle in which turnover led to more turnover, to further degeneration of systems and procedures, which led to staff dissatisfaction, etc.

Ultimate objectives

The objectives reflect a synthesis of hard data and subjective inputs from medical, administrative and clerical staff regarding existing and desired operations. They are to:

1. Tailor make to the needs, wishes and constraints of every

doctor, the best possible methods of scheduling.

2. Effectively, quickly and uniformly process all appointment-making functions, e.g., making, cancelling, and rescheduling appointments, in a manner which is consistent with good medical care, and with human and material resource constraints in mind.
3. Give each patient the advantages of a large group practice, e.g., laboratory and radiology facilities, in-house expert consultative capabilities, attached hospital facilities, etc.
4. Effectively triage patients.
5. Enable each doctor and/or department to pretest any and all changes of scheduling procedures.
6. Effectively requisition, deliver in a timely fashion, and control patient-related information flow, e.g., charts, X-ray and laboratory reports, and financial information.
7. Balance patient flow to avoid overloading clinic units.
8. Have a patient flow system which can easily adapt to changes.
9. Have available, in an easily accessible and updated form, all relevant data regarding patient flow, e.g., appointment availabilities.
10. Keep both the patient waiting time and physician idle time low.
11. Make appointments available in a manner which is consistent with the patient's needs.
12. Efficiently handle multiple appointments and the time span necessary to complete a multi-appointment diagnostic and/or treatment plan.
13. Reduce the number of no-shows and of unfilled time slots.
14. Provide for flexibility in doctor's schedules within a reasonable planning horizon so that changes in availability times and policies can be made easily.
15. Have the advantages of both centralized and decentralized scheduling systems.
16. Make available to physicians innovative means of scheduling patients and an ability to pretest such in a "laboratory" environment, e.g., prior and posterior prescheduling.

17. Keep the time span of the patient routing as short as possible.
18. Improve receivables handling, e.g., reducing number of lost billings.
19. Identify all patient flow associated administrative costs to design a system which will optimize the same.
20. Improve personnel morale and effectiveness.
21. Improve hiring practices.
22. Have the administrative support necessary to implement the clinic's patient flow objectives.

Personnel related recommendations

1. Using inputs from this project, a patient flow procedures manual should be generated and kept current.
2. A systematic training of existing clerical staff in systems and procedures should be established. This training should consist of lecture sessions on the various patient flow processes. This could be done with the help of the process diagrams discussed. The essence of each job (goals and objectives as well as systems and procedures) and its role in the clinic operations must be discussed in detail with the employees performing the job.
3. A systematic training of new medical and clerical personnel in systems and procedures should also be established.
4. A systematic enforcement of systems and procedures is imperative.
5. An improvement in hiring practices is necessary. This can be done by proper testing procedures and by maintaining a small pool of trained personnel to move into both temporary and permanent vacancies.
6. A reclassification of jobs, allowing more upward mobility, should be made. For the same job positions in different departments with different work loads, some kind of incentive should be established.
7. Commendations from doctors, when work is done properly, should be encouraged and would help the morale of clerical personnel.

8. Consider relieving medical secretaries of appointment-making duties, where feasible.
9. Define more clearly lines of responsibility and control of clerical staff to avoid conflicting directives.

Scheduling-related recommendations

1. Consider providing, at each heavy consultative load desk, scheduled "standby" time on a rotating basis, to take up urgent consults without overloading any one doctor. This could be accomplished by having at least one physician at all times performing some clinic functions which could be easily interrupted.
2. Based on historical data, consider devising a battery of lab and/or X-ray studies which result from some percentage (say 80% or 90%) of physicals for a given doctor or department.
3. At the doctor's discretion, consider prescheduling the above lab and/or X-ray studies prior to the patient's appointment for a physical with sufficient lead-time to allow for processing, reporting and delivery of results to the doctor.[6]
4. When historical data indicates that certain appointment types with a given doctor, inevitably or to a large extent, are followed by some other appointment, either with the same doctor or elsewhere in the clinic, then consider prescheduling the latter as well.
5. Consider storage of specimens for some prescribed period of time in the laboratory. The inconveniences caused by additional visits to the laboratory will be avoided. This would be especially useful for prescheduled patients.
6. Consider a patient standby list. The patients in this list should be easily reached and would help to alleviate the problem of no-show and late cancellation appointment slots going unfilled.

[6]As a first attempt to identify which further appointments result from physical appointments, studies were run in each department to determine how many of each type of consult, X-ray tests, laboratory tests, and special studies were ordered over two-week periods. The results of these studies are displayed in Table I.

TABLE I: Primary Desk/Consultant Desk Input/Output

CONSULT DESK / PRIMARY DESK	ALLERGY	CARDIOLOGY	CARD. FUNC. LAB		. . . .				CMI SCAN
1. ALLERGY					. . . .				
2. CARDIOLOGY					. . . .				
3. DENTAL					. . . .				
					. . . .				
					. . . .				
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮
					. . . .				
28. UROLOGY					. . . .				

7. Consider overbooking policies based on no-show and late cancellation statistics, by department or by physician.
8. Consider calling patients to ascertain arrival. By doing this, many of the predictable no-shows could be determined. Personnel requirements for performing this job need to be considered.
9. Consider alternative scheduling methods to optimize patient waiting time and physician idle time. The following factors are among those which should be considered: staff-fellow interaction, room constraint, patient load, and appointment sheet time increments. Computer Simulation can be used to pretest the efficacy of various alternatives.
10. Consider altering the availability of each appointment type, based on both physician determined criteria and patient needs. Patient needs can be determined to a certain extent by examining appointment availability data and data on unfilled appointments.
11. Study room availability constraints and the effect of same on doctors' schedules, fellows' schedules, staff-fellow interaction, and hospital rounds. Again, Computer Simulation is a good study vehicle.
12. Define a patient priority classification scheme, based on patient's urgency and importance, and doctor's professional interest. Institute this classification scheme where feasible and acceptable to the individual's practice.
13. Consider providing booking options which will have, for example, 100% filled appointment sheets for the most immediate time period, 80% filled for the next time period, 60% filled . . . , etc. Such a booking procedure will allow more urgent patients (as determined by the patient priority scheme, described above in 12) to be scheduled sooner and without overloading the doctors.
14. Define the optimal time periods for the above, for each desk or each doctor and each appointment type. As an example, a department could have new appointment times fully booked for a two week period while the consult times could be obtained on the day of the request.

Recommended system

Many of the clinic's problems, identified earlier, are concerned with information flow. Problems arise when information is not readily available or is not transmitted properly.

In response to these problems, a patient flow information data bank was recommended. This information bank should contain all information relevant to the patient flow processes and should be easily and quickly accessible. This information would include appointment availabilities for all doctors and central file of scheduled appointments. In addition to receiving, the data bank would be responsible for distribution of all information regarding patient flow, such as billing, scheduled appointments, and demographic information.

In the recommended system all doctors' appointments would be made at the departments, with the exception of consults, reports, and special exams which would be held by the routing area. This is, in fact, what was presently being done at the clinic.

After an appointment is made, all information regarding it should be sent to an appointment processing area. This area would be responsible for all paperwork associated with appointments and updating the central file of appointments as well as appointment availabilities. Further, by checking the central file of scheduled appointments, this area would easily locate appointments, make cancellations, catch duplicate appointments, and coordinate multiple appointments, multiple cancellations, and rescheduling.

In the recommended system, appointment information would be updated both at the departments and in the appointment processing area. When a doctor or a department is not specified, the call will be directed to the Appointment Service Desk where a triage is made. Some patient triage situations are too complex to be properly handled by a person with non-medical training. In response to this problem, a Physician Triage Officer should be established. It could possibly be implemented by providing physicians' consultative services (on a rotating basis) to handle these triaging of non-routine patients. With such a service available, a non-medical person who does triaging of patients would have the opportunity to consult a physician about difficult situations.

For both routine and non-routine triaging situations, a questionnaire could be developed to be completed by the patient, which would provide a means for systematic evaluation of triaging accuracy. On the basis of this questionnaire, an appointment could be changed at the discretion of the Physician Triage Officer or any other medical staff.

For multiple appointments, the appointment processing area will make a triage (with the help of the Physician Triage Officer when necessary). Then the area will call the proper desks and coordinate the appointments. This coordination will be made easily because all appointment availabilities will be known at the appointment processing area. In the new system, the problem of excessive transferring of calls from department to department will be avoided.

A major problem for the clinic was that of no-shows.[7] However, many of these no-shows are predictable. For example, a patient with multiple appointments may call his primary department to say that he will not be able to make this appointment. The primary department appointment will be cancelled; however, often the person making the cancellation does not know about the other appointments for that day. These appointments will not be cancelled and will result in no-shows. This situation can be avoided in the recommended system since, whenever a patient cancels an appointment, all other appointments can easily be checked in the scheduled appointment file. Part of the updating procedure in the appointment processing area could be to check in this file (under the patient's name) for additional appointments for the same day or associated, in some way, with the cancelled appointment.

A predictable no-show can also occur when a patient with an appointment scheduled sometime in the future has an appointment with the same doctor at an earlier date. This often happens due to a change in medical urgency for the patient. In the present

[7]The billings foregone due to patient no-shows are estimated to be in the $1,000,000 to $2,000,000 range per year.

system, there is no systematic procedure for cancelling the original appointment and often the result is a no-show. This situation can also be avoided in the recommended system by making it part of the updating procedure to check for these duplicate appointments.

Another feature of the recommended system is that data collection for studies such as the ones described in this paper would be much more easily collected since all patient flow related data could be collected at one location, i.e., the information data bank.

Thus, such an appointment system would have most of the advantages of both centralization and decentralization. Further, it can be easily implemented since no major changes in clinic operations are necessary.

Strategy of going toward the ultimate system

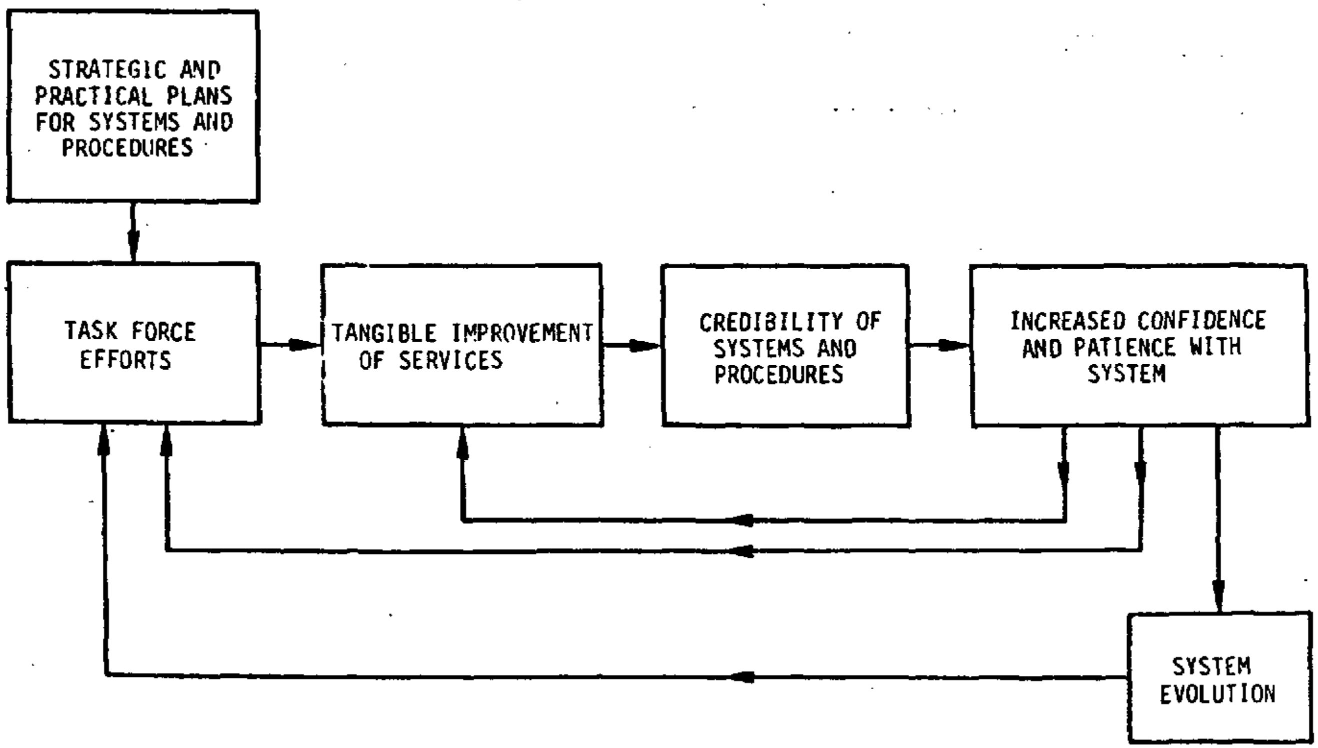
There are two ways in which improvements in systems and procedures can be implemented. One is to schedule a complete changeover on a certain day. The other is to do it gradually, step-by-step.

The first method would be too risky and too disruptive to the clinic operations and not readily accepted by the clinic personnel.

On the other hand, gradual improvement will not be so disruptive and will be more easily implemented. In this implementation scheme, as tangible improvements are realized, credibility and increased confidence and patience with the system will result. Consequently, further changes will be more readily accepted. Figure VIII shows the dynamics of this implementation process.

As an initial step in the implementation process, a manual system could be devised to handle appointment-related information. As soon as an appointment is made, information regarding appointment availability will be immediately updated at the department (on the appointment sheet). At the same time, some mechanism for transferring this information to the appointment processing area would be initiated. In this manual system, the means of transferring this information would be by telephone, lift or runners. Such a manual system would not be disruptive to present clinic operations.

Figure VIII. Implementation Process



A major problem with such a system would be the lead-time for transferring information from the desks to the appointment processing. During this time, the appointment availability file will not be up to date. This would cause some inefficiency in the appointment processing area's ability to reschedule and coordinate multiple appointments.

The ideal situation would be to have the appointment availability file updated immediately from the desk. This feature necessitates some kind of mechanized information transmission.

This mechanized transmission could be accomplished by using teletypewriters or some other telecommunication device at the desk to transfer information to the appointment processing area. In addition to the speed with which information would be transmitted, this system would have the advantage of being able to generate typewritten copies of the appointment information at the appointment processing area directly from the desks immediately after the appointment is made. This capability would eliminate the need for filling out and sending a preregistration form, since all the necessary information could be sent via the teletype. Further study could be made to determine the feasibility and desirability of sending other types of information, such as requests for charts and X-rays in the same manner to other clinic areas.

Once a mechanized information transmission device is in operation, the major system bottleneck would be in the time necessary to manually update the files. This problem could be solved by mechanizing the updating procedures using some computer configuration.

Once the appointment information system is mechanized, the next step would be to incorporate other patient-related information, such as billing, which could be accessed easily from the proper places. Thus, this appointment information system could be the foundation of a complete communications-based information system.

An advantage of this implementation scheme is that, at each step, improvements are made in the patient flow system. Also, each of these steps serves as a transition and learning period for the next. Further, if, at any step in the implementation scheme, the feasibility and desirability do not justify the implementation of the next step, all the advantages of the current stage will be retained. Moreover, with proper forethought and planning, each of the implementation steps could serve as a backup mode for the more advanced system stages.

It is important to note that the successful implementation of each step is, to a large extent, dependent on personnel support. Therefore, improvements in the personnel situation must accompany any and all changes in systems and procedures.

In future work, detailed study of each of these steps should be made. Cost-benefit analyses of different means by which each step could be implemented should be performed. Also, the scheduling recommendations presented earlier should be extensively investigated.

More than a year has passed since this study was completed. The clinic personnel are currently methodically implementing most if not all of the recommendations delineated, using the directed step-by-step approach of achieving the ultimate goals and objectives.

The Role of Operations Research in Regional Health Planning

Larry J. Shuman, Harvey Wolfe

University of Pittsburgh, Pittsburgh, Pennsylvania

and

R. Dixon Speas, Jr.

Stanford University, Stanford, California

(Received August 1, 1972)

Operations-research workers have not met with much success in being accepted as integral members of regional-health-planning teams, owing in part to a lack of understanding by health planners of the skills the operations researcher has to offer and in part the analyst's inability to demonstrate that he can close the gap between theoretical modeling and the implementation of his results. This paper explores the growth of regional health planning in the United States and highlights its important problem areas. The literature of operations-research applications to health planning is reviewed critically with respect to the feasibility of models and the appropriateness of assumptions. Specific problems with the types of studies currently in the literature are identified and recommendations are made for improved coordination between operations-research workers and health planners.

HEALTH PLANNERS have inherited a delivery system that developed almost randomly in accord with local political, economic, and/or parochial interests. Community-wide planning typically played a minor role. As a result, the uneven distribution of health resources and services throughout a region has been a persistent problem complicating planning. Further, as long as hospitals were required to provide relatively simple procedures, no great strain was placed on manpower supplies, nor was the economic burden caused by duplication of services unduly heavy. Specialization and advancing technology have changed this situation. Today's complex medical procedures require expensive, sophisticated equipment and facilities, as well as teams of highly skilled personnel.

MAY[36] has described the history of health planning in the United States as highlighted by three landmark events: The passage of the Hill-Burton Act (PL 79-725) in 1946; the 1961 joint publication by the American Hospital Association and the Public Health Service of the report, "Area-wide Planning for Hospitals and Related Health Facilities;"[28] and the passage of the Comprehensive Health Planning Act of 1966 (PL 89-749). To these might be added the Regional Medical Program (PL 89-239) enacted in 1965, and the Experimental Health Services Delivery Systems program adopted in 1971.

Health planning in the United States most likely originated with a voluntary group that formed the Cincinnati Public Health Council in 1917.[49] The growth of community health councils mushroomed; by 1949 there were 1222 such local and state organizations. Many of these councils served as the main health-planning bodies for their localities until the early 1960's when 'area-wide hospital-planning' agencies came into existence. Early accomplishments by these agencies were primarily of a political nature, supporting bond issues for facility construction, tax levies for program operations, changes in laws, development of new services, and public education on health problems. It was not until the Hill-Burton Act, when quantitative justification for new facilities became necessary, that planning agencies began to recognize the need for policies and guidelines to aid in decisionmaking.

Prior to the Hill-Burton Act, health planning consisted of trying to satisfy arbitrary ratios established by medical-care investigators; e.g., five beds per 1000 population[41] or 135 physicians per 100,000 population.[32] Planning based on such national norms or standard ratios ignored many important variables that affect utilization and, in fact, assumed that these norms were generally appropriate for local conditions.

The purpose of the Hill-Burton Act was to assist in developing and carrying out comprehensive state-wide and local plans for the orderly development of hospital and health-related facilities.[68] The Act required states to create a planning program that would survey existing facilities and promote hospital construction plans based on community needs as a prerequisite to receiving federal funds. In developing these plans, states were instructed to establish planning regions in which the health-care providers could cooperate in delivering care. Although the general characteristics of a health 'region' were specified in the Hill-Burton Act, a model for determining an effective region for health planning has not been developed.[7]

The American-Hospital-Association/Public-Health-Service Committee report on Health Facility Planning developed a formula for approximating the gross number of general hospital beds needed in a planning region for a target year.[28] This report provided guidelines that were generally adopted for the organization and operation of hospital-planning agencies. Today many agencies have modified these guidelines in order to concentrate on the planning process within each health facility. A consequence of the report was an increase in the number of area-wide planning agencies, particularly in metropolitan areas. The federal government added impetus to health planning by making seed money and operating funds available to most of these agencies.

The report of the President's Commission on Heart Disease, Cancer, and Stroke further dramatized the need for coordinated regional planning of medical services and led to the passage of legislation creating the Regional Medical Programs (RMP).[62] Congress directed local RMP's to plan ways to improve the delivery of services, primarily through the achievement of cooperative arrangements on a regional basis.[47] RMP was later broadened to consider preventive and ambulatory care (PL 91-515). The promotion of a regionalized system of comprehensive health-care services was considered to be a practical means of implementing the legislative intent of the Regional Medical Program. In actuality, local RMP's have had difficulty in translating goals and objectives into programs that can be realistically implemented. Ann Somers points out that the RMP has ignored the specific issue of planning that it was directed to tackle.[59] Indeed, she feels that only lip service has been given to the concept of regionalization or "systematization" among the various levels of health-care institutions. Sasuly and Ward concur; in their experience the RMP has discovered how not to plan.[47]

One of the reasons why RMP has had limited impact was its failure to utilize effectively available technical skills on a local level. This absence of certain technical skills within RMP was highlighted by a federal survey that indicated almost no utilization of operations researchers or systems analysts in local programs.[26] Moreover, individuals whose job titles indicated high levels of technical competence were frequently computer programmers rather than experienced analysts. Although one objective of RMP was to provide specialized care with increased effectiveness in a region, there is little evidence that sophisticated interdisciplinary planning approaches have been used to achieve this end. As a result of such shortcomings, federal support of RMP appears likely to be terminated.

In the mid-60's, Congress also passed the "Partnership for Health" legislation (PL 89–749), which has funded comprehensive health planning at state and areawide levels. In issuing a mandate for planning, this Act declared health a national priority and the basic right of every citizen. JACOBS AND FROH conclude that the CHP strategy for social change meant that health planning had finally "come of age in America." [27] Traditional techniques, such as program budgeting, and new procedures, such as systems analysis and operations research, were being recognized as necessary tools for making better decisions in areas of social concern. The Comprehensive Health Planning legislation clearly implied a need for multidisciplinary, broad-based approaches to planning and, consequently, definite opportunities to draw upon operations research in problem-solving and decisionmaking situations. In fact, Jacobs and Froh suggest that CHP is based on the premise that for too long those interested in particular areas of health care have not looked beyond their own sphere of interest.

POLK has reported some of the problems that one regional agency experienced in implementing the guidelines of PL 59–749.[41] Certain impediments appeared to be attributable to the constraints of innumerable local, state, and federal licensing standards, operating codes, and funding programs, which limit a hospital's flexibility in decisionmaking. To these, SEIVERTS adds quasipublic accreditation and approval programs.[66]

As still another possible way to approach the problems that have hampered the health-planning process, the Health Services and Mental Health Administration (HSMHA) funded Experimental Health Services Delivery Systems (EHSDS) programs in 1971. The EHSDS programs represent government's first major attempt to develop health-service delivery systems rationally at the community level. Grants were awarded to communities in which payers, providers, the public, and politicians have indicated a desire to establish formal mechanisms by which comprehensive personal health services could be provided.

The EHSDS program has three primary objectives: (1) to test whether the community management structure can improve the avenues of access into the delivery system and its efficiency of providing services without decreasing the level of quality; (2) to determine whether the EHSDS organizational form can bring about better integration and coordination of Federal funds; and, (3) to test new forms of providing services within the community.[13]

QUANTITATIVE STUDIES FOR HEALTH PLANNING

QUANTITATIVE HEALTH-PLANNING studies first appeared in the early 1960's. Most of these studies can be classified into six general areas:

very least, a qualitative improvement. Included among those activities for which planning decisions have resulted in shared enterprises are maternity and emergency care, data processing, inservice education, laboratories, and laundries. Decisions to form a joint venture for sharing services are not difficult to implement, aside from the obvious logistical problems that result.

Implementation is not so simple when one or more institutions give up a service entirely, while selected hospitals upgrade and emphasize this service. Nevertheless, in a number of cases, through negotiation and the investigation of the economies of eliminating duplicate services, a successful redistribution of resources has been accomplished. The awareness that not all institutions need provide every service has led to the concept of regionalization.

The last two general areas of quantitative planning activity are studies relative to the location of hospitals and distribution/location of specialized services. The rising health costs of the past decade have made these areas critical with respect to the need for a rational approach to decisionmaking. Unfortunately, no quantitative technique has been produced that can be utilized practically by planning agencies to make locational decisions. Much of the work directed toward the location of hospitals has naively minimized distance or travel time without considering behavioral variables and cost factors. It is little wonder that planning agencies have not been able to integrate these suboptimizations into their decision processes.

The distribution of specialized services is becoming increasingly important as new high-cost techniques are being developed for low-volume, critically-ill cases. Such services as renal dialysis, open-heart surgery, burn centers, emergency coronary care, or cobalt therapy have such high costs that improper planning decisions result in unnecessary expenditures due to wasteful duplication. In addition to the costs involved, many of these services must be performed at some minimum-volume level if the quality of the service is to be acceptable. Given the high costs involved and the critical need for methods to aid in making these decisions, the use of operations research (with emphasis on defining a realistic and achievable goal) offers considerable potential in creating guidelines for locating specialized services.

OPERATIONS RESEARCH APPROACHES IN REGIONAL HEALTH PLANNING

ALTHOUGH OPERATIONS RESEARCHERS have been studying health-related problems for almost fifteen years, it is only recently that their attention has been drawn away from the individual hospital and directed toward regional applications. This late recognition of area-wide problems may be attributed to several factors, including (1) the difficulty in obtaining valid, reliable regional data, (2) recent federal emphasis on regional planning, and (3) the fact that early health operations researchers developed their skills and were probably more comfortable in the single-facility setting.

Initial operations-research health-planning efforts concentrated on the allocation of individual hospital resources within institutions. Over 400 in-hospital studies have been catalogued by the STIMSONS,[60] who have presented a critical analysis of hospital operations research from the point of view of the hospital administrator. They note that almost any hospital operations-research model can be viewed as a planning model, yet they are disenchanted with the success these models have had with respect to implementation.

Two in-hospital studies that utilized some of the operations-research techniques now being applied to regional-planning problems are by WOLFE[65] and FETTER AND THOMPSON.[16] In 1964, Wolfe developed an integer-programming model for allocating nursing services within a hospital. An important attribute of Wolfe's model was the adaptation of psychometric-scaling techniques to form quantitative measures of qualitative factors. Techniques that use subjective estimates must be incorporated with traditional measures of effectiveness (quality, cost, time) if realistic regional models are to be constructed. Fetter and Thompson first used simulation models to describe certain hospital subsystems: maternity suite, outpatient clinic, and surgical pavilion. These models were general in nature and were developed for use in designing and evaluating future facilities.

The recognition of operations research as a useful technique for studying regional planning problems can be attributed to FLAGLE AND YOUNG.[18] They noted in 1966 that the greatest challenges then facing operations-research analysts lay in problems of long-term planning of new health-care systems and the integration of the hospital within a comprehensive health-care framework. Flagle proposed a simulation model AESOPH (All Encompassing Simulation Of Public Health) for predicting health needs based on socioeconomic groupings and classification of health status within a region.[17] Although necessary data for implementing this model were never available in a usable form, it nevertheless represented an important theoretical formulation.

TORRANCE, THOMAS, AND SACKETT,[63] and CHEN AND BUSH[9] have presented models for selecting which health programs should be funded for a particular region. These models are important, not so much because of their formulation, but rather owing to the measures of effectiveness used. Both papers represent variations of the knapsack problem in which costs are minimized subject to a budget constraint. Torrance et al. have utilized the VON NEUMANN-MORGENSTERN Standard Gamble[63] in cooperation with a group of physicians to develop a health-utility index. They used this index to rank health programs with respect to cost effectiveness, assuming that only budget is constrained. When other resource restrictions are included, an integer-programming formulation is proposed for solution.

Chen and Bush form the objective function for their mathematical-programming models by scaling subjectively weights obtained from 60 raters. These weights reflect the 'quality adjusted years of life' that would result if an individual participated in a particular health program. The constraints on the objective function were formulated to reflect available budget, resources, and interrelations among projects. In addition, Chen and Bush cited several political constraints that might be added to such planning models.

Even though the models developed by these two teams are not mathematically sophisticated, both groups' contributions lie in the development of potentially viable measures of effectiveness marked by the active involvment of physicians (Sackett and Bush) in each study. Other approaches for constructing measures of effectiveness relative to health status are discussed by BARNOON AND WOLFE.[3] They propose measures based on expected life span and place values on each year of life, percent disability, and weighted combinations of disability and expected life span.

Simulation planning models have ranged from the in-hospital work of Fetter and Thompson to the abstract model postulated by Flagle. SMALLWOOD, SONDIK, AND OFFENSEND[36] propose a regional simulation model that describes the patient's movement through the health system, detailing resources required for various stages of illness and recovery. The model's purpose is to calculate potential demand for services given specific regional populations and to evaluate the effects of particular programs, e.g., preventive care. An important conceptualization in the model is to incorporate patient behavior in determining the utilization of various regional facilities based on incidence of disease.

In a later paper, Smallwood[37] has proposed a formal framework for regional health-planning models. Smallwood's formulation is intended to account for postulated causal relations between the health-care system and the health of the population. In actuality, the framework consists of five interrelated dynamic models, which describe a complete health system. This series of models would be used to evaluate the eventual consequences of various alternative decisions in a regional health system. While Smallwood's formulations are conceptually interesting and detailed, the problem of obtaining data greatly limits any real potential for application.

Most operations-research planning studies that have utilized real data in developing simulation models were concerned with a subsystem rather than an entire health-delivery system. Two examples of this level of simulation are a planning model for emergency medical transportation constructed by GIGLIO, KAMINSKY, AND WATTS,[19] and a model for studying the utilization of prenatal care that was developed by KENNEDY AND WOODSIDE.[29]

Simulation was among the techniques proposed by NAVARRO AND PARKER[40] in developing several simplistic regional planning models. Hypothetical states describing patient condition were postulated, along with a specification of resources (beds and physicians) appropriate for each state. Given the state population and the transition probabilities, the requirements for physician services or hospital beds could be calculated. Another of their models could be used to determine the optimal transition probabilities in order to obtain a total population with a specified mix of patient conditions. Subjective estimates of the cost of altering transition probabilities are used as the measure of effectiveness for determining the optimal solution. Although this model represents a mathematically clever formulation, the subsequent problem of regional allocation of services is ignored and costs are considered only hypothetically.

Navarro and Parker's models represent an early attempt to apply optimization techniques to planning models. Other operations-research optimization models directed toward health planning within a specified region have been proposed by Fetter and Thompson,[15] REVELLE, FELDMAN, AND LYNN,[48] FELDSTEIN,[14] KRYSTYNAK,[31] LOVE, MATHIAS, AND TREBBI,[33] and SHUMAN.[52]

Using data generated by a simulator, the Fetter and Thompson paper develops an optimization model for planning services. Specifically, a prediction of occupancy and service requirements are determined based on a specification of the patient mix. Given this prediction, the optimal arrangement of services, measured in terms of beds, is found with respect to subjective planning weights. The model includes both the investment costs and the operating costs as budgetary constraints. The final form of the model is a mixed-integer linear program.

Three different optimization approaches for planning disease-control programs are proposed in the Revelle et al., Feldstein, and Krystynak papers. The first two are concerned with tuberculosis and the last with rheumatic fever. Revelle et al. first postulated a descriptive model of tuberculosis epidemiology for a developing nation in order to evaluate the cost effectiveness of various methods of control. The descriptive model can be formulated as a linear program. The solution of the linear-programming model yields the minimum-cost method for attaining a given reduction in tubercular cases during a specified time period.

Feldstein proposes a tuberculosis planning model for use in developing countries. This model selects the set of activities that would maximize the total economic benefit for given resource constraints. It considers the interrelations of public-policy preference, population structure, disease incidence and prevalence, limited resources, and technology.

Krystynak constructs a probabilistic network model for the regional control of rheumatic fever. It is converted to a linear program in which actions are selected that minimize the annual incidence of disease subject to a budgetary constraint on the total community cost for rheumatic fever prevention.

The Love, Mathias, and Trebbi model is concerned with planning facility construction over a given time period. Patient demand is assumed to be Markovian, and a dynamic-programming formulation results. Costs considered are those associated with the acquisition and operation of the facility. If requirements for services are greater than the facility can provide, an additional cost is imposed as a penalty. The model was originally designed for planning the development of a single facility. In order to use this model for regional problems, planners must first determine where facilities should be located. Moreover, the cost and availability of specialized services are not considered in the original model and would have to be incorporated in any extension.

Shuman's model provides a framework for allocating health services optimally throughout a region by controlling the quantity and type of manpower, the level of technology available at each facility, the amount of construction necessary, and the amount of regional funds allocated to the facility. The effectiveness of this allocation is measured by the total cost to society for providing medical services, including operational, construction, manpower-development, and patient costs, as well as the economic loss that would result when needed services are not provided. Shuman, Young, and Naddor[51] have shown how this model could be used to plan for a neighborhood health center. Unfortunately, these models cannot be implemented without better data to estimate the measures of effectiveness.

While OR workers were developing health-planning models, several groups of geographers began formulating theories for regional and locational analysis using qualitative techniques to study health-facility-location problems. The general theory of regional analysis as a decisionmaking tool was developed by Isard,[26] whose method synthesized elements of sociology, geography, political science, and planning. One aspect of this theory, locational analysis, was applied by both Coughlin and Schneider to evaluate urban hospital locations using patient travel time as a measure of effectiveness. Coughlin[10] has presented a framework and method based on travel time for planning a metropolitan system of facilities to provide adequate community health services. Schneider[49] believes that it is the task of planning agencies to develop hospital systems in a manner that minimizes the patient and provider costs attributable to the locations of hospitals within a region.

CHERNIACK AND SCHNEIDER[9] proposed a statistical technique known as sectorgrams for describing a hospital service area where natural barriers exist. In actual application, the overlapping of hospital service areas make defining uniques boundaries for a particular region extremely difficult. In a later paper, Schneider[49] noted that the service-area approach to estimating bed needs is untenable in an urban region. He suggested that the problem of allocating services be treated apart from that of locating facilities. He felt that primary attention should be directed towards locational problems using the sectorgram approach and concentrating on the spatial correspondence between facilities and patients.

Earickson and Morrill provide another approach to the locational analysis of regional hospital systems. EARICKSON[12] used a transportation model to investigate the effect on Chicago hospitals if indigent blacks utilized the facilities nearest their residences. This was accomplished by minimizing personal travel time with utilization constrained by facility capacity. MORRILL AND EARICKSON[37] state that the model failed to account for patient behavior, but was valuable in identifying poorly served patient groups as well as poorly located health facilities. They propose a more complex model which, as a first step, simulates patient behavior with respect to existing physician and facility location. Physician locations are changed and patients are reallocated using a modified gravity model. Hospital locations are then altered and the allocation process repeated. MORRILL AND KELLEY[38] indicated that the simulation model was able to provide a good replication of the Chicago hospital system and that the gravity model was a good predictor of patient behavior with respect to facility choice.

Another locational analysis was performed by LUBIN, DROSNESS, AND WYLIE,[34] who evaluated hospital locations by employing techniques previously used by transportation-planning agencies. They calculated the shortest estimated travel time from census tracts to hospitals and physician office clusters. The utilization of facilities with respect to distance was then evaluated for present populations and projected population shifts and transportation improvements. Lubin et al. were among the first to apply urban-land-use, market-concept, and central-place theory to health-facility planning. There are two basic operations-research approaches to facility location, location on a network, and location on a plane. REVELLE, SWAIN, AND TOREGAS[46,61] have used the first of these approaches in developing a series of models that select an optimal set of facility locations as measured by total travel time. Constraints are introduced to limit the number of facilities, allocate patients to the nearest facility, and limit the maximum travel time. Their models are integer-programming formulations, but modifications are made to allow using linear programming to obtain solutions. Although these models are more complex than Earickson's transportation approach, they likewise ignore behavioral and political aspects of the location problem that would cause a mathematically optimal allocation to be infeasible or suboptimal when all factors are considered.

SHUMAN, HARDWICK, AND HUBER[54] have used a variation of this first approach in developing a model for the location of Health Maintenance Organization clinics within a metropolitean area. Unlike the Revelle et al. models, in this formulation not every member of a census tract must elect to utilize the facility. The model finds the locations from a set of potential sites that maximize the number of people who would choose the prepaid group practice mode of the receiving care. Utilization is considered a function of travel time as estimated from existing hospital-usage patterns. Constraints are included that place minimum capacities on each clinic, limit the number of clinics, assign people to the nearest clinic, and restrict expenditures to the available budget.

ABERNATHY AND HERSHEY[1] propose to solve the health-facility-location problem by using the second method previously described. Four objective functions are postulated: (1) maximize utilization, (2) minimize distance per capita to the nearest facility, (3) minimize distance per visit, and (4) minimize the decline in utilization. Population is divided into census tracts and then subdivided into socioeconomic classes. Utilization is a function of distance, social class, and personal preference. The Abernathy/Hershey model considers all points in a region as possible sites rather than evaluating predetermined potential locations. A modified search algorithm is used to select the locations. In considering all possible points as locations, there is a possibility that sites that are infeasible, e.g., rivers, parks, or very expensive property, may be chosen as locations.

Many of the aspects of locational models proposed by geographers and operations researchers have been incorporated by BAUM[4] into a simulator for evaluating primary- and ambulatory-care delivery systems. Baum's simulator is general in that it can consider any configuration of hospital outpatient activities, satellite clinics, and neighborhood health centers. In addition to patient travel time to facilities, the model treats some behavioral aspects by considering the probability of facility choice, keeping appointments, and remaining within a particular delivery system.

Locational-analysis models provide valuable methods for evaluating the spatial relations between a health delivery system and the community. However, they do not consider a complete spectrum of services, nor do they treat the problem of distributing specific services throughout the health region nor the consequent effects of removing or adding services. The measures of effectiveness used in these models have been overly simplistic and inadequate. Except for travel costs, minimal consideration is given to the cost components of the system. There have been no investigations of the economic costs involved in closing units, relocating physicians and beds, or constructing new facilities, even though in many cases these costs may dominate patient travel costs. A comprehensive review of location/allocation models is presented by SCOTT.[56] A critique of locational models applied to health planning is given by GROSS.[22]

The operations-research approach has evolved from in-hospital to multihealth facility applications. Yet, many people, including those actively involved in health operations research, have questioned the value and impact of most studies. Both GUE[23] AND YOUNG[64] have noted that very few, if any, operations-research projects have had a profound effect on the delivery of health care. Gross[21] has pointed out that, with few exceptions, health operations-research studies have resulted in erudite models that either have little meaning or are given low priority by health professionals. He cites the serious lack of mutual respect, understanding, and communication between physician, medical researcher, and systems analyst as another reason for the lack of success in implementing quantitative studies. "Before health planning moves from the rhetoric to the action stage, physicians will have to be convinced, through education and through demonstration, that planning is worthwhile, . . . and that the health-planning process is not an extravagant whim of academics, intellectuals, and bureaucrats sitting in splendid isolation from the real world."

McLaughlin,[35] on the other hand, is not as pessimistic. He notes that in the health environment it is extremely attractive to select first the problems that are most tractable, even though they are often the most trivial. McLaughlin believes that one intent of health operations-research studies has been to move decisionmakers away from an action-document approach to one that considers all alternatives, including those not recommended. Special emphasis is placed on having the decisionmaker understand the assumptions that underlie each alternative. Kissick[30] also has indicated optimism that systems analysis can assist in the precise identification of health-care goals and in determining the most effective and economic means of attaining these goals.

A NEED FOR MORE PRACTICAL OPERATIONS RESEARCH IN FUTURE PLANNING

This synopsis of quantitative approaches to health planning has reviewed numerous techniques and philosophies. The question still persists: Why have the applications of quantitative methods and, in particular, operations research, not resulted in useful models for aiding planners? The answer is complex. It is easy to blame planning agencies for what the Comprehensive Health Planning Legislation sought to overcome, an inability to look beyond its own traditional methods. Only a few comprehensive-health-planning agencies have had the courage or sophistication to support meaningful operations-research activity.

Even so, a great many studies have been carried out, and it seems apparent that much of the failure of these studies to be implemented lies with those doing the quantitative work. They have not integrated themselves into the planning system. They operate as outsiders viewing the planning system as an abstract structure, ignoring the fact that it is political, social, and highly dynamic. DeVise[11] has noted that a chasm exists between researchers and health planners to the point that they do not speak the same language. Consequently, he has recommended that interdisciplinary exchanges take place among concerned parties. According to DeVise, prescriptive models must be capable of specifying alternative future plans in the system, predicting the consequences of choosing each alternative, and evaluating these consequences according to an acceptable measure of planning goals.

Hilleboe and Schaefer[24] have stated that, in the health field more than almost any other area, the professionals have usually asserted that the subject matter is too technical and the consumer too inexpert to play an important role in planning. Meaningful health planning, they philosophise, will be controversial if it is to meet society's needs. Health planners cannot be content with attempting a better coordination of the existing system, but will have to change this system. While it is true that planners have not extensively embraced quantitative approaches, it is also true that quantitatively-oriented analysts have not produced any work that is meaningful in the immediate context of planners' needs.

Operations research has not been useful in regional planning because no one has used the operations-research approach to aid in the kinds of decisions that current planners are making. A large portion of the quantitative work that has been done has ignored the basic tenets of operations research and has degenerated into an exercise in mathematical modeling: the problems that have been attempted have not been well formulated; the objective functions have been naively constructed and the measures of effectiveness yield suboptimizations; quantitative variables are considered of prime importance, but behavioral variables are neglected; alternatives that are politically and socially infeasible are still considered in the feasible set.

The failure of planners to view operations research as a viable tool is partially attributable to a misunderstanding of the terminology and power of the operations-research process. Model building and the developing mathematical techniques cannot be construed as synonymous with operations research. Yet, many people confuse these terms, expecting model building by itself to lead to implementable solutions. This confusion is further accentuated by operations researchers who first conceive a mathematical model and then attempt to fit a health-related problem to it under the guise of an application. Although this technique quite often leads to journal articles, it rarely results in implementable decisions. The planner is left with a theoretical contribution that he can neither use nor appreciate, while the analyst qualifies his model by claiming that no data were available for validation. In effect, the health-planning problems that have been defined and solved do not resemble the problems for which rational decisions must be made.

Developing and conceptualizing mathematical models and techniques for solution are important and should be supported, but this kind of work must be recognized for what it is, theoretical rather than applied. Real-world problems must be approached comprehensively and systematically. This requires analysts to reorder their priorities relative to work on optimizing regional location of services. Prior to building and solving models, emphasis should be placed on developing measures of effectiveness and corresponding objective functions to reflect the many variables—quantitative, social, and political—that planners must consider.

Rising costs have made the regionalization of certain services a necessity. Studies have been performed relative to social and behavioral variables, but these have not been comprehensive. The many locational studies that minimize distance or time have been useful in establishing a theoretical base for the quantitative modeling of planning systems. Now, a true operations-research approach is needed to take the best of these partitioned areas of study and merge them into comprehensive planning models that are meaningful and, in turn, useful.

In using operations research, one must recognize the special nature of regional-health-planning problems, particularly with respect to the types of decision that must be made and the types of constraints imposed upon planning systems. Regional health planning deals with policy decisions rather than traditional operational considerations. The usual output measures such as quantity, time, or dollars are inadequate for decisions. Additional indices such as health status must be postulated and quantified. Typical constraints, e.g., plant capacity or available capital, which can be represented by well known relations, do not necessarily result in solutions that are politically or socially feasible. The thrust of comprehensive health-planning legislation has been to provide adequate funds for the major multidisciplinary investigation of planning problems. It is time to reduce the support of piecemeal, uncoordinated studies of planning problems and to emphasize comprehensive and sophisticated analyses that coincide with planning-agency objectives and can be implemented in current planning practice.

REFERENCES

1. W. J. ABERNATHY AND J. C. HERSHEY, "A Spatial Allocation Model for Regional Health Services Planning," *Opns. Res.* **20**, 629–642(1972).
2. ABSTRACTS OF HOSPITAL MANAGEMENT STUDIES, The Cooperative Information Center, The School of Public Health, University of Michigan, Ann Arbor, Michigan.
3. S. BARNOON AND H. WOLFE, *Measuring the Effectiveness of Medical Decisions—An Operations Research Approach*, Charles C Thomas, Springfield, Illinois, 1972.
4. M. A. BAUM, "A Model for the Examination of Urban Primary Care Health Delivery Systems with Special Emphasis on the Poor," Ph.D. Dissertation, School of Government and Public Administration, American University, 1971.
5. M. S. BLUMBERG, "Shared Services for U.S. Short-Term Hospital," a report of a preliminary survey, Stanford Research Institute, Menlo Park, California, 1964.
6. CALIFORNIA DEPARTMENT OF PUBLIC HEALTH, "Hospital Utilization Research Project: 1961–1965," Berkeley, California, 1965.
7. THE CARNEGIE COMMISSION ON HIGHER EDUCATION, *Higher Education and the Nation's Health—Policies for Medical and Dental Education*," McGraw-Hill, New York, N. Y., 1970.
8. M. M. CHEN AND J. W. BUSH, "A Mathematical Programming Approach for Selecting an Optimum Health Program Case Mix," paper presented at the 40th National Meeting of the OPERATIONS RESEARCH SOCIETY OF AMERICA, Anaheim, California, 1971.
9. R. D. CHERNIACK AND J. B. SCHNEIDER, "A New Approach to the Delineation of Hospital Services Areas," Discussion Paper Series No. 16, Regional Science Research Institute, Philadelphia, Pennsylvania, 1967.
10. R. D. COUGHLIN, "Hospital Complex Analysis: An Approach to Analysis for Planning a Metropolitan System of Service Facilities," Unpublished Ph.D. Dissertation, University of Pennsylvania, 1965.
11. P. DEVISE, "Methods and Concepts of an Interdisciplinary Regional Hospital Study," *Health Services Res.* **3**, 166–173 (Fall, 1968).
12. R. EARICKSON, "The Case for Decentralizing Cook County Hospital: Some Applications of Linear Optimization in Hospital Planning," Hospital Planning Council for Metropolitan Chicago, Working Paper, Vol. III, No. 4, Chicago, Illinois, 1968.
13. "Experimental Health Services Delivery Systems: Information for Applicants," Health Services and Mental Health Administration, 1972.
14. M. S. FELDSTEIN, "Cost Benefit Analysis and Health Program Planning in Developing Countries," paper presented at the Second Conference on the Economics of Health, Baltimore, Maryland, 1968.
15. R. B. FETTER AND J. D. THOMPSON, "A Decision Model for Design and Operation of a Progressive Patient Care Hospital," *Medical Care* **7**, 450–462 (1969).
16. ——— AND ———, "The Simulation of Hospital Systems," *Opns. Res.* **13**, 689–711 (1965).
17. C. D. FLAGLE, "Simulation Techniques Applicable to Public Health Administration," paper presented at the First Annual Conference of American Statistical Association and Public Health Association, Proceedings on Simulation in Business and Public Health, New York, 1966.
18. ——— AND J. P. YOUNG, "Application of Operations Research and Industrial Engineering to Problems of Health Services, Hospitals and Public Health," *J. Indust. Eng.* **17**, 609–614 (1966).
19. R. H. GIGLIO, F. C. KAMINSKY, AND J. WATTS, "A Mathematical/Computer Planning Model for Emergency Medical Transportation Systems," paper presented at the Conference of Areawide Comprehensive Health Care Planners, September, 1969.
20. GLASGOW HOSPITALS SURVEY, *Report of Working Party*, Western Regional Hospital Board, Scotland, Western Regional Hospital Board 351, March, 1965.

21. P. F. GROSS, *Some Determinants of Development and Implementation of Health Care Technology and Some Lessons of the U.S. Experience*, University of Saskatchewan, Regina, Canada, 1971.
22. ———, "Urban Health Disorders, Spatial Analysis and the Economics of Health Facility Location," paper presented at the 40th National Meeting of the OPERATIONS RESEARCH SOCIETY OF AMERICA, Anaheim, California, 1971.
23. R. GUE, "Operations Research in Health and Hospital Administration," *Hospital Administration* 10, 6–25 (Fall, 1965).
24. H. HILLEBOE AND M. SCHAEFER, "Evaluation in Community Health: Relating Results to Goals," *Bull. New York Academy of Medicine* 44, 140–158 (1968).
25. F. ICHNIOWSKI, "Operations Research and Regional Medical Programs," paper presented at the National Conference Hospital and Health Services Division, New Orleans, Louisianna, February, 1970.
26. W. ISARD, *General Theory: Social, Political, Economic and Regional*, MIT Press, Cambridge, Mass., 1969.
27. A. R. JACOBS AND R. B. FROH, "Significance of Public Law 89-749—Comprehensive Health Planning," *New England J. of Medicine* 279, 1314–1318 (1968).
28. JOINT COMMITTEE OF THE AMERICAN HOSPITAL ASSOCIATION AND PUBLIC HEALTH SERVICE, "Areawide Planning for Hospitals and Related Health Facilities," PHS-855, Washington, D.C., 1961.
29. F. D. KENNEDY AND M. B. WOODSIDE, *Development of a Simulation Model of a Community Health Service System*, The Health Service System—Research Triangle Institute, Vol. III, 1968.
30. W. L. KISSICK, "Planning Programming and Budgeting in Health," *Medical Care* 5, 201–220 (1967).
31. L. F. KRYSTYNAK, "A Model for the Regional Control of Rheumatic Fever," paper presented at the 38th National Meeting of the OPERATIONS RESEARCH SOCIETY OF AMERICA, October, 1970.
32. R. I. LEE AND L. W. JONES, *The Fundamentals of Good Medical Care*, University of Chicago Press, Chicago, Illinois, 1933.
33. I. G. LOVE, R. A. MATHIAS, AND G. TREDBI, "Dynamic Planning of Health Care Systems," paper presented at the 37th National Meeting of the OPERATIONS RESEARCH SOCIETY OF AMERICA, Washington, D.C., April, 1970.
34. J. W. LUBIN, D. L. DROSNESS, AND L. G. WYLIE, "Highway Network Minimum Path Selection Applied to Facility Planning," *Public Health Reports* 80, 771–778 (1965).
35. C. MCLAUGHLIN, "Health Operations Research and Systems Analysis Literature," *Systems and Medical Care*, ALAN SHELDON, FRANK BAKER AND CURTIS MCLAUGHLIN (eds.) MIT Press, Boston, Massachusetts, 1971.
36. J. J. MAY, "Health Planning: Its Past and Potential," Health Administration Perspective A5, Center for Health Administration Studies, University of Chicago, Chicago, Illinois, 1967.
37. R. L. MORRILL AND R. EARICKSON, "Locational Efficiency of Chicago Hospitals: An Experimental Model," *Health Services Res.* 4, 128–141 (1969).
38. ——— AND M. B. KELLEY, "The Simulation of Hospital Use and the Estimation of Location Efficiency," *Geographic Anal.* 2, 283–300 (1970).
39. NATIONAL COMMISSION ON COMMUNITY HEALTH SERVICES, *Action-Planning for Community Health Services*, Public Affairs Press, Washington, D.C., 1967.
40. V. NAVARRO AND R. PARKER, "A Planning Model for Personal Health Services," Working Paper, Department of Medical Care and Hospitals, The Johns Hopkins University, Baltimore, Maryland, 1967.
41. L. D. POLK, "Areawide Comprehensive Health Planning: The Philadelphia Story," *American J. Public Health* 59, 760–764 (1969).
42. THE PRESIDENT'S COMMISSION ON HEART DISEASE, CANCER AND STROKE, "Report to

the President—A National Program to Conquer Heart Disease, Cancer and Stroke," February, 1965.

43. Public Health Federation of the Cincinnati Area, 40th Annual Report 1957, Cincinnati: The Federation, 1957.

44. L. S. Reed and H. Hollingsworth, "How Many General Hospital Beds are Needed? A Reappraisal of Bed Needs in Relationship to Population," PHS-309, HEW, 1953.

45. C. Revelle, F. Feldman, and W. Lynn, "An Optimization Model of Tuberculosis Epidemiology," *Management Sci.*, Application Series, 16, B-190-B211 (1969).

46. ——— and R. Swain, "Central Facility Location," *Geographical Anal.* 2, 30–42 (1970).

47. R. Sasuly and P. Ward, "Two Approaches to Health Planning, The Ideal vs. the Pragmatic," *Medical Care* 7, 235–241 (1969).

48. J. Schneider, "A New Approach to Planning the Growth of a Metropolitan Hospital System," Regional Science Research Institute, Philadelphia, Pennsylvania, 1965.

49. ———, "Measuring the Locational Efficiency of the Urban Hospital," *Health Services Res.* 2, 154–169 (1967).

50. A. J. Scott, "Location-Allocation Systems: A Review," *Geographical Anal.* 2, 95–119 (1970).

51. Sedgwick County Health Facilities Planning Council Survey and Recommendations, Sedgwick County Health Facility Planning Council, Kansas Health Facilities Information Service, Publication No. 494, Topeka, Kansas.

52. L. J. Shuman, "Mathematical Models for Health Manpower Planning," Department of Operations Research, The Johns Hopkins University, Baltimore, Maryland, 1969.

53. ———, J. P. Young, and E. Naddor, "A Planning Model for Regional Health Services," *Health Services Res.* 6, 103–119 (1971).

54. ———, C. P. Hardwick, and G. Huber, "A Model for the Allocation of Ambulatory Health Care Centers in a Metropolitan Area," *Health Services Res.* 8, 121–138 (1973).

55. S. Sieverts, "Influences of Areawide Planning," *Hospitals* 44, 63–65 (1970).

56. R. D. Smallwood, E. J. Sondik, and F. L. Offensend, "Towards an Integrated Methodology for the Analysis of Health Care Systems," *Opns. Res.* 19, 1300–1322 (1971).

57. ———, "A Quantitative Approach to Analyzing Regional Health Care Systems," paper presented at 39th National Meeting of the Operations Research Society of America, Dallas, Texas, 1971.

58. J. D. Snyder and M. J. Enright, "Federal Aid," *Hospital Management* 102, 37–56 (1966).

59. A. R. Somers, "Hospitals Costs and Payment: Suggestions for Stabilizing the Uneasy Balance," *Medical Care* 7, 348–360 (1969).

60. D. H. Stimson and R. H. Stimson, "Operations Research and Systems Analysis in Hospital Administration," Chicago: Hospital Research and Educational Trust, 1972.

61. C. Toregas, R. Swain, and C. Revelle, "The Location of Emergency Service Facilities," *Opns. Res.* 19, 1363–1373 (1971).

62. G. W. Torrance, W. H. Thomas, and D. L. Sackett, "A Generalized Cost-Effectiveness Model for Evaluation of Health Programs," paper presented at 39th National Meeting of the Operations Research Society of America, Dallas, Texas, 1971.

63. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1947.

64. J. P. Young, "No Easy Solutions," George K. Chacko (ed.), "The Recognition of Systems in Health Services," paper presented at the 35th National Meeting of the Operations Research Society of America, Arlington, Virginia, 1969.

65. H. Wolfe, "A Multiple Assignment Model for Staffing the Nursing Units," Department of Operations Research, Johns Hopkins University, Baltimore, Maryland, 1964.

Programming, Budgeting, and Control in Health Care Organizations: The State of the Art

Robert A. Vraciu

The planning, budgeting, and controlling processes (PBCP) largely subsume all of the planning and controlling activities of an organization. This paper discusses these activities within the context of a single management control system, focusing on three topics. First, a brief historical perspective of management concerns which relate to PBCP is presented and several important external pressures currently imposed on the health care industry are discussed. Second, normative models of the processes—programming, budgeting, and controlling—are presented. The discussion focuses on the elements and relationships of these processes, and numerous references to the literature are provided. Third, several issues related to the gap between the state of the art in PBCP for hospitals and the current state of practice are discussed.

> Much of the management control process involves informal communication and interactions. Informal communication occurs by means of memoranda, meetings, conversations, and even by such signals as facial expressions. Although these informal activities are of great importance, ... most companies also have a *formal* management control system. [1, p. 19]

PROGRAMMING, budgeting, and controlling processes (PBCP) represent the three major phases of such a formal management control system. A simple schematic diagram (Figure 1) illustrates the relationships between these activities as they occur in a regular cycle and the role of strategic planning as a starting point.

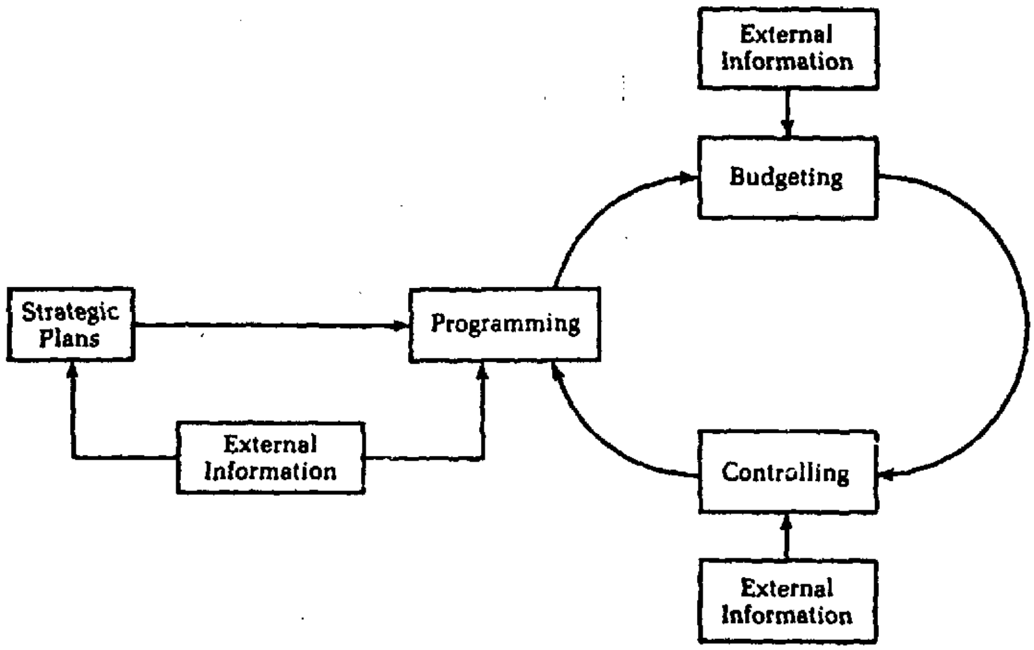
"Programming is the process of deciding on the programs that the company will undertake and the appropriate amount of resources that are to be allocated to each program" [2, p. 670]. This aspect of the planning process includes the formulation of corporate objectives and strategies based on marketing, legal, regulatory, and social concerns, as well as on the organization's own perception of its role. During this stage, health care organizations would identify the types of medical services to be provided, the types of teaching and research activities to be conducted, the service populations to be served, and the general mode of operations for both a long-run and a short-run planning horizon.

The second phase is the budgeting process which expresses the programming decisions in monetary terms, and covers a specific time period, generally one year. The budget represents the best plan for allocating resources to achieve the objectives and implement

Address communications and requests for reprints to Robert A. Vraciu, Assistant Professor, Program in Hospital Administration, School of Public Health, University of Michigan, Ann Arbor, MI 48109.

0017–9124/79/1402–0126/$2.50/0

Figure 1:
Management Control Process*



*Adapted from R.N. Anthony and J.S. Reece, [2, p. 670].

the programs approved in the programming phase. The agreed-upon budget is often considered a bilateral commitment in which responsibility center managers commit themselves to produce the planned output with the agreed-upon amount of resources, and their superiors commit themselves to a definition of satisfactory performance [3, p. 31].

. The third phase of the management control process represents a feedback mechanism in which plans and expectations for subsequent years can be modified based on experience from previous years. The process involves the reporting and analysis of performance. Such reports and analyses provide the basis for control over expenditures and useful data for identifying the causes of operating problems.

This model of management control systems is applicable to any organization and is used as a *framework* for discussing programming, budgeting, and control in health care institutions. Superimposed over the provider organizations are several layers of planners, regulators, etc., each trying to manage the system. As the focus of the discussion is on provider institutions, these interests are viewed as constraints and external factors which impinge upon the PBCP. It is recognized that these constraints can be influenced by providers, but this type of interaction is not discussed here.

The remainder of this paper presents a discussion and evaluation of the state of the art of PBCP, with a distinction being made between state of the art, (i.e., that which is technically possible) and the state of practice (i.e., that which is currently being employed in the health care industry. While much of the discussion is focused on the nonprofit, nongovernment hospital industry, many of the processes and issues are relevant to other health care provider groups. Because the topic area of PBCP subsumes other areas of financial management, and because it overlaps with many other areas of management activities not typically associated with financial management, discussion here must be narrowly focused. The paper presents a normative model of the three processes (i.e., PBCP) and discusses several issues related to the gap between the state of the art and the state of practice. To accomplish this, the paper is divided into three parts. The first section provides a brief historical perspective to PBCP and summarizes several important external pressures currently imposed on the health care industry. The second section presents a generic model of each of the three processes (PBCP). The elements of the models and the relationships between their elements are highlighted. The final section discusses several issues associated with orientation, design and implementation of PBCP currently practiced in the hospital field.

Background—Legacy and Inertia

As the hospital industry developed from a cottage industry to being the nucleus of a technologically sophisticated health care system in the early 1970s, it did so with few constraints. Each institution had cash-flow concerns like any other economic entity; however, major programming decisions were made largely independent of any external planning or regulatory bodies. By 1970, the hospital industry was undergoing rapid growth in the hope of fulfilling the promise that "high quality health care was a right." The inertia generated by this period is still influencing managerial decisions and has left the industry with a legacy of high and rapidly increasing costs. Enthoven [4, p. 1229] summarizes the management philosophies that have resulted:

> In the system of fee-for-service, cost reimbursement and third-party intermediaries that dominate health care financing today, the question of efficient use of resources does not even arise. The problem of how best to spend a given amount of money for the health care of a population is not posed. Providers are not required to set priorities, look at alternatives and make hard choices. From the point of view of the provider, there is an apparently unlimited amount of money. The system rewards cost-increasing behavior with more revenue; it punishes cost-reducing behavior with less revenue. Such a system must produce inflation in prices and waste in the use of resources.

To better understand the elements of this philosophy as they relate to PBCP, four general areas of planning concern are discussed: service mix, production efficiency, production effectiveness, and capacity.

Service Mix. The mix of services available in a given hospital was influenced by the rapid technological change in the medical field [5], the high degree of medical specialization [6], and the ability of hospitals to rapidly adopt new technologies. The rapid technological change was stimulated by research funded through the federal government, as well as universities, foundations, and industry. The resultant new clinical technology required expensive and highly specialized hospital facilities. The clinical efficacy of these technologies were often not questioned or tested [7] and the cost-based reimbursement adopted by Medicare and Medicaid programs, " . . created an unprecedented opportunity for physicians and hospital administrators to do what they always wanted to do—improve the quality of care as they see it. This means more equipment, more personnel, more tests, more x-rays, and so on" [8, p. 94]. This natural tendency of health care providers and administrators received both moral and fiscal support from the federally financed Regional Medical Program created in 1965. As a result, no hospital wanted to be technologically inferior and sought to provide a wide range of services regardless of the impact (or lack thereof) on the health status of its service population or the resultant duplication of services.

Production Efficiency. The hospital industry is often characterized as having few inherent incentives for efficiency. The profit motive is generally lacking, contributions have historically allowed many hospitals to operate at a loss, health insurance programs have insulated the consumer from the provider, and the guaranteed payment of cost-based reimbursement have created an environment in which managers have had little need or incentive to strive for efficient production of services [8]. In addition, cost-based reimbursement even provides disincentives to reduce costs through improved efficiency [9, pp. 165–167]. As a result, hospital management has generally failed to introduce efficiency-minded styles of management and similarly failed to develop information systems necessary for proper control.

Production Effectiveness. Historically, hospitals have had little concern about providing the right number and right types of health services to their service populations. Until recently, few people have questioned medical judgments regarding admissions, length of stay, and surgery rates. As a result, the incentives inherent in the fee-for-service, litigation-prone health care system have dictated the modus operandi, and increasing evidence of excessive utilization is surfacing [10,11].

Capacity. While the medical profession operated on the "more is better" philosophy described above, hospital administrators developed a "bigger is better" philosophy toward facility planning. The availability of donations, the guaranteed reimbursement of depreciation and interest expense by Medicare, Medicaid, and many Blue Cross programs, and the availability of debt financing provided hospitals with sufficient means to add to their bed complement and replace existing facilities. Even though regional planning was an intended feature of the Hill-Burton Program, the Regional Medical Program, and later the Comprehensive Health Planning Act, large numbers of unused hospital beds are found in most regions of the country. The depreciation, interest, and operating expenses associated with these facilities and subsequent replacement costs are all considered legitimate expenses and reimbursed under cost-based reimbursement. As a result, there have been few financial constraints or penalties linked to excessive capacity.

While many other forces and activities were at play prior to 1970, those cited above were instrumental in shaping the hospital industry as it is currently operating. Programming decisions were made based on the "more is better" and "bigger is better" philosophies adopted by the medical staff, administration, and trustees. There was apparently little concern for the efficient allocation of resources since there was apparently no scarcity. The net result is the often cited increase in hospital costs.

Given the trends toward rapidly increasing costs of health care, the early 1970s saw the federal government, state governments, and private industry exert pressure on the hospital industry. These major purchasers of health care services began to recognize that they would soon be unable to afford the health care system which was evolving. The attempts to control systems costs include rate regulation [12], controls on capital investments, budgeting requirements, and utilization review. The associated regulations act as constraints to the PBCP.

Attempts to control capital investments include the early comprehensive health planning structure, state certificate of need requirements, and certificate of need review mandated by P.L. 93–641. Under P.L. 93–641, states are required to designate an agency which performs a reviewing and comment process for capital expenditures which exceed $150,000, change the bed supply, or substantially change the services of the facility. P.L. 93–641 also authorizes health systems agencies to recommend to this state agency whether new institutional health services should be improved or not. While the experience of such programs is mixed [13,14], the process of requiring hospitals to justify proposed expenditures does force a level of planning not previously observed. Moreover, the federal government and some states have made inroads in defining and implementing quantitative criteria based on conservative estimates of justifiable demands [15,16].

Annual operating budgets and capital budgets with a three-year planning horizon are mandated of all hospitals under Section 234 of P.L. 92–603 [17, p. 369]. While the law can make this requirement of a process, it cannot insure that the process is effective in the context of the management control process outlined in Figure 1.

Professional standards review is another effort to affect costs by monitoring the appropriate use of hospital services. "But since the mechanism is that of peer review—one physician vis-à-vis another—the conflict of quality versus cost is immediately established, and with the emphasis on the maximizing of care, the net result may be increased rather than decreased services" [18, p. 6].

The four categories of regulatory programs represent direct attempts to lower the rate of increase in hospital costs by 1) providing services which are medically necessary rather than simply demanded, 2) reducing the capacity of the system to correspond to some justifiable demand for medical services, and 3) providing incentives for production efficiency and effectiveness.

The models of the programming, budgeting, and controlling processes presented in the next section adopt these three objectives for institutional planning and control. (Cost-effective planning is the term used by Dowling to refer to "determining the least costly arrangement of facilities and services consonant with community needs" [19, p. 26]. The two definitions are seen to be identical in meaning.) It should be noted, however, that cost-effective planning as defined here does not necessarily coincide with the private interests of health care providers. It is this inherent conflict of interests which spawns many of the issues discussed in the final section of this paper.

Programming, Budgeting, and Control—A Model of Processes

This section presents an overview of the state of the art of programming, budgeting, and controlling processes (PBCP) for health care institutions. Schematic diagrams of each of the three processes are presented (Figures 2, 3, and 5) for the purposes of highlighting the important data elements, relationships, and decision points. These diagrams are not intended to preclude additional considerations or steps leading toward decisions, and the ensuing discussions must, because of space constraints, be restricted to general characteristics of the diagram elements. (For more detailed discussions of methodological options, evaluations, etc., see [20].)

Programming Phase

In the management control model described above, the initial planning step is that of deciding on the strategic plans for the organization. Once decided, the programs to be undertaken and the approximate amount of resources to be allocated to each program are determined. The outputs of this programmng process are decisions regarding the revenues, expenses, and capital requirements of patient care, teaching, research, and overhead programs; facility size; and objectives in the area of improving or ensuring production efficiency and effectiveness. The programming decisions are based on four major categories of information (see Figure 2).

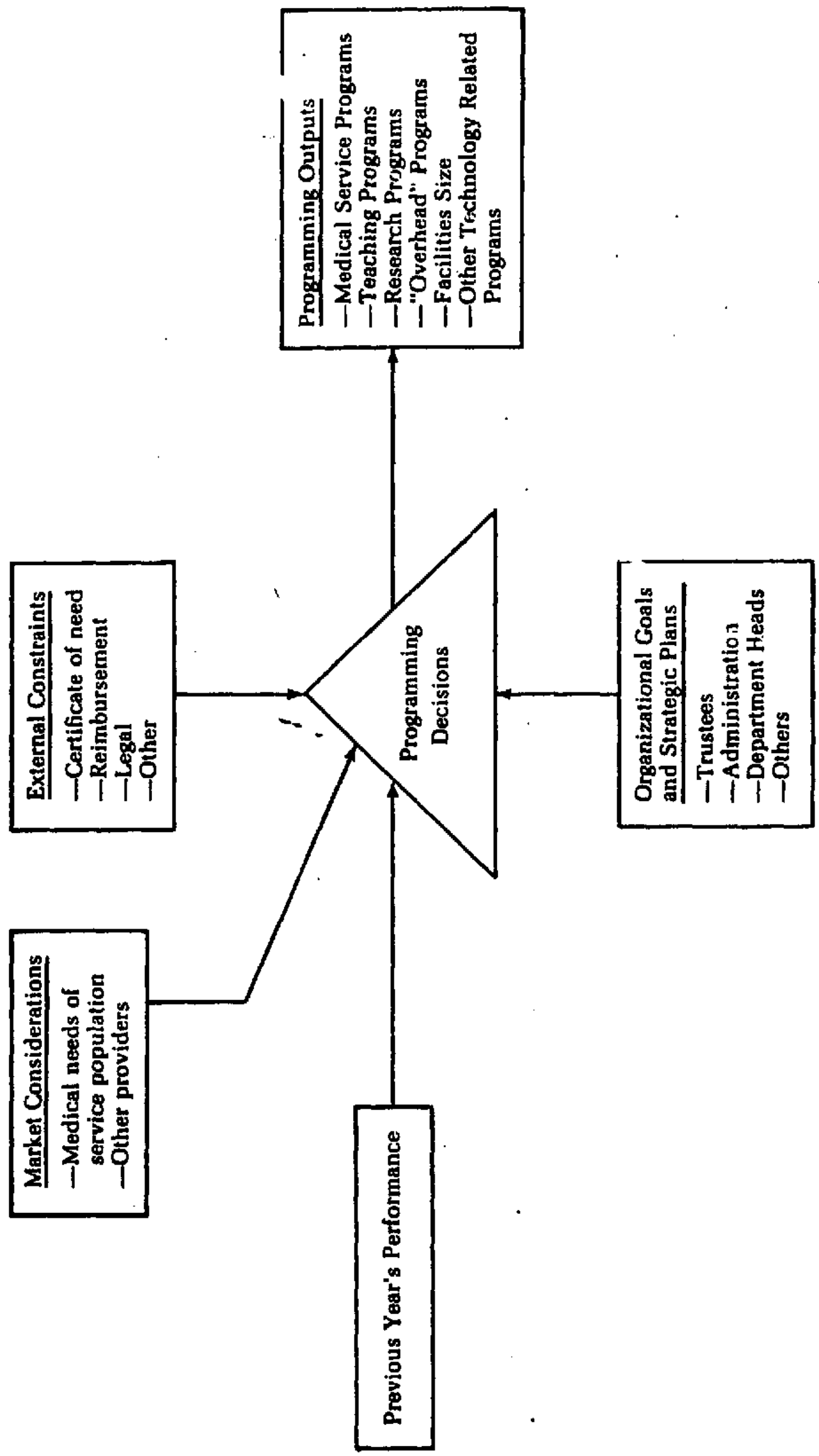
Marketing Information. The market served by the hospital should be understood well enough to permit cost-effective planning, i.e., to determine the least costly arrangement of facilities and services, consonant with community needs. This should include three important types of information: 1) determination of service population of the hospital and its characteristics [21], 2) assessment of medical needs of the service population [22], and 3) assessment of potential areas for cooperation with other hospitals and providers [23]. It is recognized that difficulties arise in trying to define and measure need for medical services [24]. As a result, many of the contemporary planning models and the marketing information are demand based. It should be noted that increasingly hospitals are trying to influence the amount and configuration of their demand through various marketing efforts.

Organizational Goals and Strategic Plans. The development, statement, and updating of organizational goals and strategic plans is the second major category of information necessary for the programming process. These goals initially provide direction for the decision process and, later, become yardsticks along which various programs can be evaluated. It is important that the formulation of these goals reflect inputs from a variety of individuals within the organization, e.g., trustees, administrators, department heads, etc. [17]. Such an exchange of ideas between individuals at the different levels not only provides useful suggestions in the planning process, but helps foster a commitment to the organization and its plans [25].

Previous Year's Performance. The feedback loop in the general management control process can suggest areas where adjustments are necessary and areas where expectations should be changed. The primary use of this information is the evaluation of ongoing programs [3, Chap. 9], and the data is generated during the Controlling Phase which is discussed below.

External Constraints. The hospital industry operates within a larger legal and economic system and is regulated by a host of statutes and agencies.

Figure 2:
Schematic Diagram of Programming Phase



Consequently, hospital programming decisions must be made with a full understanding of the reimbursement restrictions, legal constraints, tax considerations, and certificate of need restrictions associated with each program.

Information from these four categories are necessary for hospital managers to make cost-effective programming decisions. A purely algorithmic approach to programming decisions is not available, and it is doubtful whether one could be of much benefit given the complex nature of organizational decision making. Such decisions generally result from consideration of a large number of political, competitive, social, economic, and personal concerns. The application of algorithms would undoubtedly be plagued with the same types of conceptual and measurement problems identified with cost-benefit analysis [26].

Nevertheless, useful decision models have been developed for several types of programming decisions that are conceptually consistent with the objective of cost-effective decision making. Where they can differ, and where application is weakened, is often in the specification of key parameters. A number of these models are cited below.

Facility Requirements. Dowling [27] identifies five steps for translating demand forecasts into facility requirements.

1. Identify the facility units (e.g., medical/surgical beds, obstetrical beds, operating rooms, x-ray units, etc.) associated with each major medical service (e.g., deliveries, surgical procedures, etc.).
2. Determine the service producing capacity of a single facility unit.
3. Determine the number of facility units required to accomodate the average demand for service.
4. Determine the number of additional facility units that must be available to accommodate emergency demands or peak loads based on the desired level of protection sought (the reserve or standby requirement).
5. Calculate the total number of facility units required and the load factor (e.g., occupancy level) from the results of Steps 3 and 4.

The calculations associated with each step are necessary to match the capacity of a facility unit to its expected demand in an efficient way. The demand estimates are derived from service population estimates [21,28]. Reductions in use rates based on normative standards can be made to demand estimates (use rates for fixed service population) to provide activity levels that are closer to need than the observed demand [16]. The validity of such a reduction is based on an assumption of excessive utilization due to excessive supply and inefficient management of medical services [6, pp. 96–100; 10,16,29]. Once the demand is characterized, four general approaches can be used to determine the total facility requirements [27]:

1. Use of normative occupancy levels (e.g., build and/or staff enough medical/surgical beds to provide an 85 percent occupancy level).
2. Analyze the past census data to identify the number of beds necessary to accommodate the expected census, x percent of the time (e.g., 95 percent of the time).
3. Apply a formula based on an assumption of a simple (generally Poisson) distribution of arrivals. Such models range from the simple square root model used by the Hill-Burton Program [30] to more complex mathematical formulations (see analysis and references in [31]).
4. Simulation models for more com-

plex stochastic distribution of arrivals and census data [32, Chap. 7].

These models should be quite useful for determining target bed sizes of hospitals. When applied to institutional planning, the need for cooperation between health facilities is apparent. The technology has not been developed to the same degree for sizing ancillary departments. (Useful studies include those by Thomas and Stokes [33], and Conrad et al. [34].)

Support Services. All organizations incur systems' costs for "overhead programs." The mix, scope, and level of such programs must be addressed in the programming phase. Neumann [35] discusses why such programs often go unevaluated and how organizations have a natural tendency to foster their growth. Overhead value analysis [35] and functional value analysis [36] are pragmatic methods for evaluating the worth of such overhead programs to the organization.

For almost every overhead service, there is an implicit "make versus buy" decision. Increasingly, hospitals are buying many services (e.g., management contracts, food services, data processing, etc.) and increasingly, hospitals are looking to shared service arrangements with other hospitals for fiscal services, purchasing, laundry, etc.

Programs to Improve Production Efficiency and Effectiveness. Significant cost savings can be achieved by implementing contemporary admissions scheduling systems, preadmission testing, concurrent review, and outpatient surgery [37]. Successful implementation of these methodologies requires considerable advanced planning/programming to develop the requisite data systems [38] and to develop the necessary support from critical actors in the system [39]. The involvement in this planning process of all affected parties and the "top-down" support is essential to the successful implementation of these efficiency methodologies.

The importance of these methodologies in the efficient management of patient services should be noted. A hospital which for example does not have a preadmission testing program and outpatient surgery will experience a higher inpatient census than if the hospital had the programs. Using the approaches described above, this higher demand will be translated into more hospital beds. Thus, for the same medical services, a larger inpatient facility would be justified. In addition, if admission and operating room scheduling systems were not used, the variance of admissions would likely be higher, thus justifying more beds for the same target occupancy. In both cases, implementation of these efficiency technologies will tend to reduce the number of necessary beds for the service population. If the facility resizes accordingly, it will probably increase outpatient expenses, but there should be net savings.

It is through the programming decisions identified above and in Figure 2 that hospitals can plan in advance how to allocate resources and provide facilities and services which meet the needs of the population served, while minimizing capital and operating costs. These decisions provide the context within which budgeting and control decisions are made and, to a large extent, set the stage for a hospital's level of production efficiency and effectiveness. The major cost-related decisions are program decisions; budgeting and control can be viewed as fine tuning.

Budgeting Phase

The programming decisions provide the fiscal officer and budget committee with a programmatic statement of organizational objectives. During the budget process, the approved programs are translated into a detailed statement of monetary requirements and financial consequences, i.e., the budget package is prepared.

The process of budgeting is useful to management because it formalizes communication between the hospital's governance, administration, department heads, and others. This dialogue serves a useful plannng function and can be extremely important if the budget is to be used effectively in the control process. "Management by objectives" is a formal program to facilitate and improve the effectiveness of this process [40].

The end result of the budgeting process, the budget package, can serve three major functions. First, it describes the implementation and consequences of the programming decisions in the preceding step. Thus, it may be viewed and used as a simulator of financial consequences of programming decisions. Second, the approved budget is used as a standard in the control process. Thus, the measures of performance (often costs) must be stated in terms of responsibility (discussed below). The budget's third function is in the area of reimbursement. In some states, rate setting programs of third party participating contracts require hospitals to negotiate an expense budget. Subsequent reimbursement is then based on the results of this negotiation process. When a budget is used in a rate setting context, it may reflect ploys to increase reimbursement rather than accurately reflecting management intentions. If this is the case, the budget would have questionable value in the controlling phase.

There are several ways of structuring the budget schedules depending upon management needs. Four dimensions to this structure are identified and discussed below.

Basis of Accumulating Costs. The expense budget is a key operating budget and the basis of accumulating costs represents the basic unit for planning and control. Management faces a choice of bases, and four commonly cited options are:

1. Line items. A budget developed on this basis would show expenses grouped by major items of expenditures for the entire organization, e.g., salaries broken into several categories, medical supplies, laboratory supplies, etc. With expenses accumulated this way, management can control items of expenditure and plan for their use. There is, however, no clear link to output and such a classification fails to identify responsibility for expenditures.
2. Program. A budget prepared on this basis would augment the programs identified in the programming phase by showing more detailed budgets for each program. Using such a base, organizations must decide whether they are planning for or intend to control only the direct program costs or the full program costs (direct plus indirect). In either case, this method of preparation is useful in planning, but often fails to accumulate costs along lines of responsibility. Thus, the control function is subverted.
3. Medical Services. A budget prepared using this basis identifies the costs associated with each category of medical services. Costs can be defined in total, on a per-unit basis, and by direct and indirect costs. Bundles of medical services associated with specific diagnoses are the base proposed by Thompson [41].
4. Responsibility Center. Costs are ac-

cumulated according to any organizational unit headed by a responsible manager. In many cases these responsibility centers correspond to departments, although not necessarily. Budgets prepared using this basis are quite useful for controlling expenditures, but often lose their direct usefulness in planning based upon medical needs of a service population. In addition, the control function can be hampered in cases where areas of responsibility overlap.

Fixed versus Flexible Budgets. A fixed budget is one developed using a single estimate of activities. The usefulness of such a budget is contingent upon the accuracy of the activity forecasts. A flexible budget is essentially a set of fixed budgets covering a specified range of activity levels [17,42,43].

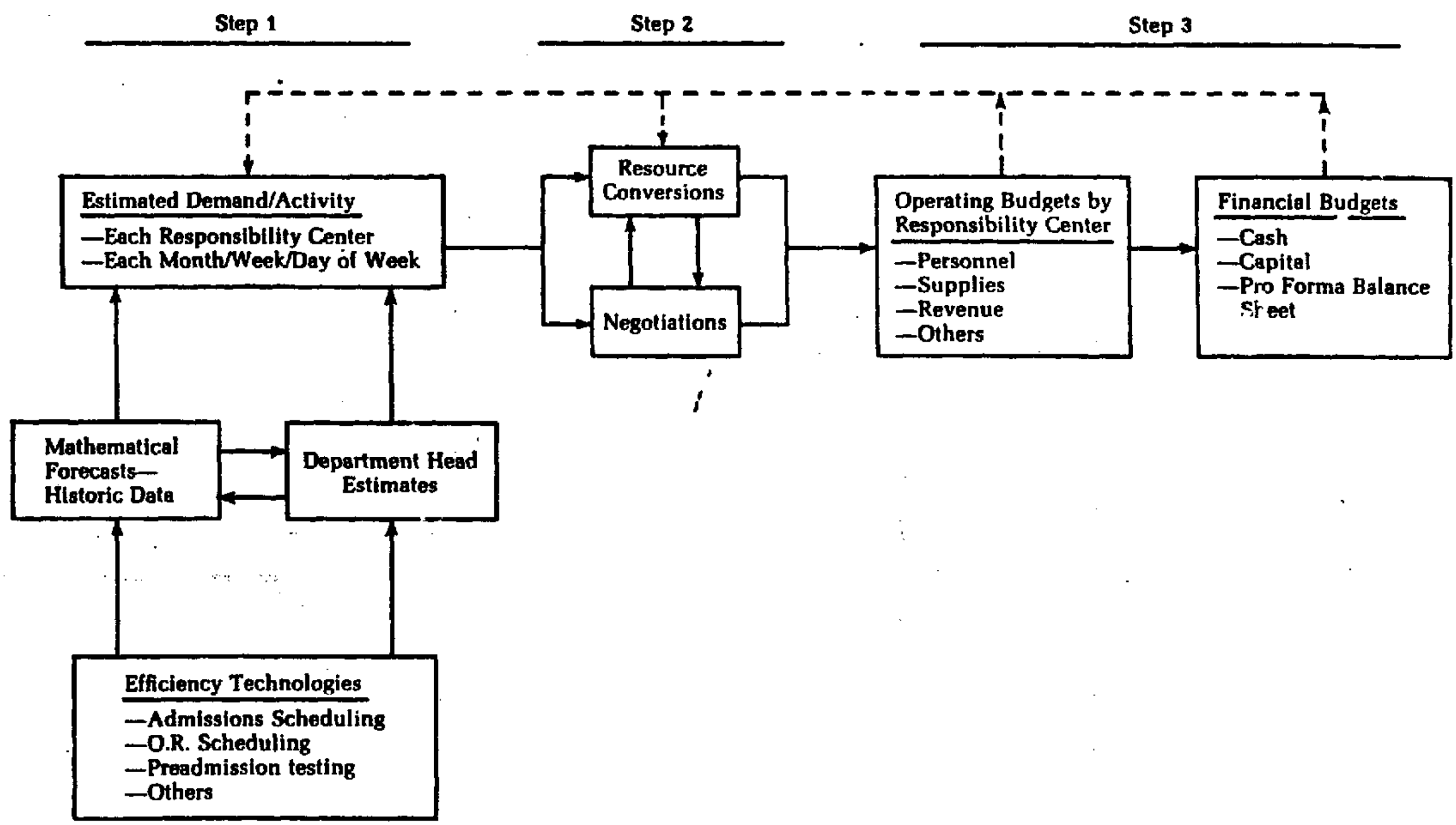
Time Periods. The expense forecast can correspond to the budget year, in total, each of the four quarters, each of the 12 months, or 13 four-week periods. The smaller the unit of time, the more tedious the planning process. However, the ability to use variable budgeting (i.e., coordinating staffing levels with seasonal variations in demand) can only be accomplished efficiently with reasonably small units of time. In addition, the specified period of time is directly related to the time lag between the occurrence of unfavorable performance and its detection using budget reports.

Continuous or Discrete. A budget can be prepared on a discrete basis, i.e., a fixed 12-month period, or on a continuous basis, i.e., the budget is updated on a monthly or quarterly basis, always extending 12 months in advance.

Management can select any combination of these different types of budgets, each with their own strengths and weaknesses. For our purposes, it is assumed that the budget is being prepared for a 13-period year [44] on a discrete basis, accumulating costs by responsibility center [2] and on a fixed basis. (The benefits of flexible budgets in the planning process will depend on the uncertainty of activity levels. In cases of high uncertainty, a flexible budget should be used to specify the range of options and consequences. However, at some point, a "likely" activity level will have to be assumed. For control purposes, other means can be used to analyze performance.) Having made these assumptions, we can discuss the three principal steps in the budgeting phase (see Figure 3).

Step 1 generates forecasts of the activity levels for each responsibility center, for each of the 13 periods. In some cases, mathematical models employing the historical data (time series, leading indicators, causal factors, and/or some combination) can be quite useful for predicting activity levels [32, 45–47]. In other cases, historical data has little bearing on the future. For example, when a new medical service is implemented, there is no experience on which to base forecasts. In all cases, there will be some risk associated with each forecast. Any biases associated with the forecasts of activity levels should be identified to the extent possible, since forecasts of expenses and revenues will generally reflect the same bias. In practice, the forecasting of activity levels is seen as a combined quantitative and qualitative approach. That is, historical data and mathematical models can be useful for identifying trends and underlying relationships; however, department heads' understanding of exogenous and endogenous factors which bear upon the future must be subjectively weighed in the forecasts.

Figure 3:
Schematic Diagram of Budgeting Phase



Step 2 represents the heart of an efficient budget preparation process. In this step, expected levels of output are translated into resource requirements, i.e., personnel at various levels, supplies, etc. The intent is to match an efficient level of inputs with the expected level of output, given a desired level of quality. For two major areas of requirements—manpower and supplies—major elements of the process can be highlighted.

Manpower Requirements. Salary expenses represent 55 to 70 percent of a hospital's total expenses. Thus manpower budgets represent a major area of budgetary concern. Warner [48] discusses the current research on nurse staffing, scheduling and reallocation activities. McNally [49] evaluates six techniques for evaluating manpower levels/needs for many hospital departments. Kaplan [50] uses regression analysis to staff different nursing units. These formula-based approaches to staffing decisions can be effective if there is significant department head input into staffing standards [51]. Lipson and Hensel [52] describe a manpower budget process which is based on negotiation between responsibility center heads and administration. Such participation in which department heads not only formulate but commit themselves to performance levels is considered important for an effective control system [53]. Through practice, careful feedback and commitment, activity indexes can be developed for specific nursing units, and production functions developed for ancillary departments which represent efficient and agreed upon technologies.

Supply Requirements. Supplies vary from approximately 5 percent of nursing unit expenses to 75 percent for pharmacy. For many departments, the amount of supplies used can be correlated to some level of hospital output, e.g., patient-days, surgical procedures, etc. Estimated relationships (with an adjustment for inflation) can be used to forecast the amount of supplies for each department. An effective materials management program can reduce supply costs, including distribution and inventory costs [54].

The important elements of this resource identification step are to 1) attempt to develop an input-output model for each department and 2) allow department heads to participate in developing the model to be used for their department. The models can be developed using historical data [55] or can be constructed using subjective estimates of the parameters [56]. Caution must be exercised that past inefficiencies are not perpetuated, but the actual process of trying to understand the use of inputs for different levels of output will have educational benefits to administrators and department heads. By making current patterns of resource use explicit, the status quo can be challenged and better ways sought. (Zero-based budgeting is a systematized approach to categorizing and judging current expenditures [57].) The field testing of these better ways and feedback through the budget reporting system provide the potential for period by period improvements in performance.

The operating and financial budgets are prepared in Step 3. It is important that the full consequences of the hospital's set of plans are identified. Following the completion of Step 2, financial officers must predict or set salary levels, fringe benefits, supply prices, utility expenses, interdepartment charges, etc., in order to complete the expense budget. With the expense budget completed, revenues from cost-based reimbursers can be estimated, and charges set at a level to cover total financial requirements [58]. Total financial requirements are the sum of operating expenses (patient care, teaching, research, etc.) and capital needs (working capital, purchases of plant and equipment, etc.). The setting of charges involves decisions about cross-subsidization of departments and input from responsibility heads. Integrating these two budgets yields the pro forma income statement which can be used in preparing the cash budget, finalizing the capital budget, and preparing the pro forma balance sheet.

This package of budgets should be viewed as a model of interrelated decisions where one financial decision influences a number of others (see Figure 4). For example, decisions regarding long-term financing directly influence the cash budget, the expense budget, and the revenue budget (in particular the setting of charges). Understanding these interrelationships enables the budget package to be used as a simulation device whereby a set of programming decisions can be used to estimate activity levels, and resource requirement assumptions can be examined in terms of the financial consequences displayed in the budget package. Computerizing such a model can enhance its usefulness [59].

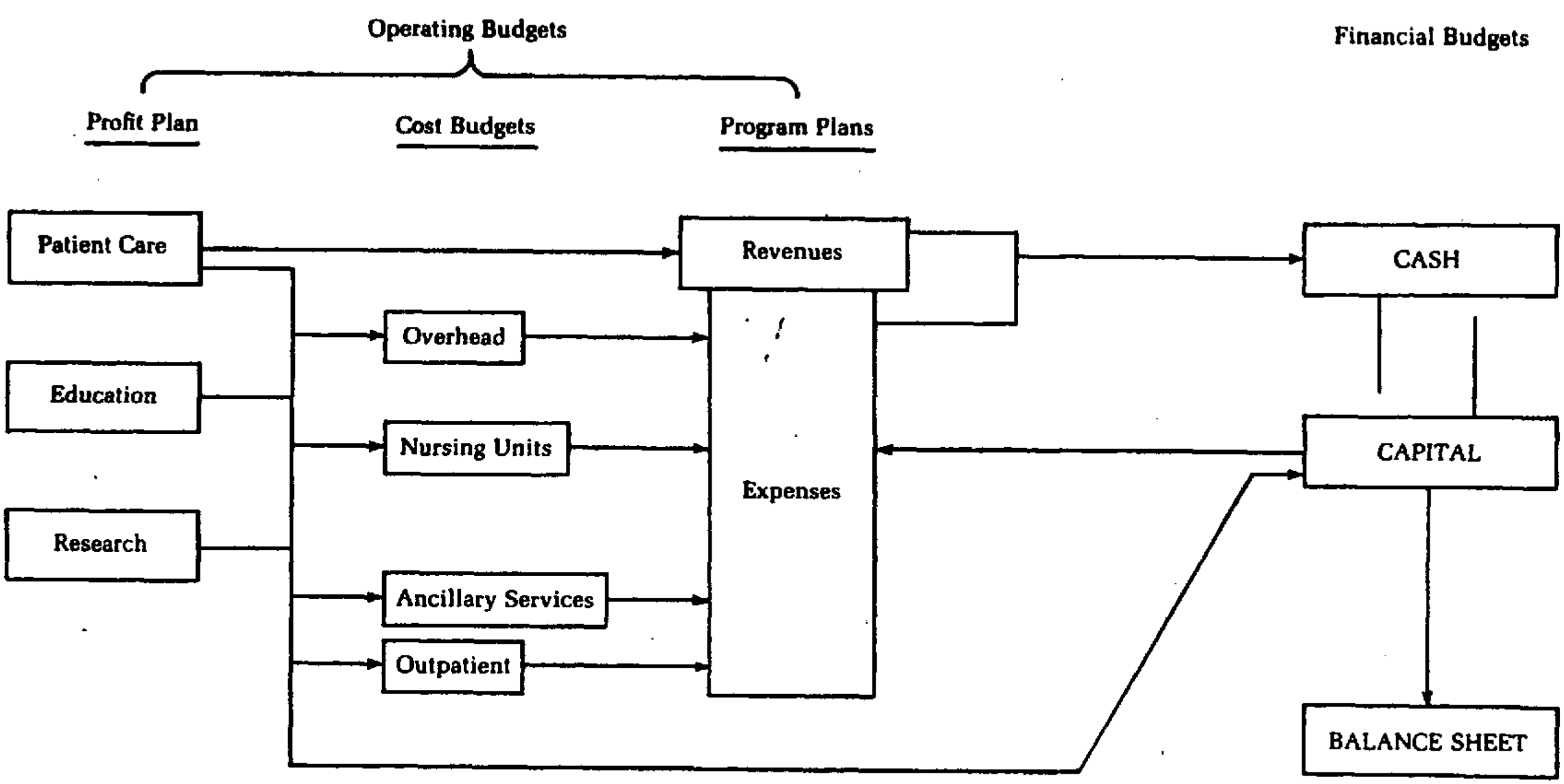
Controlling Phase

The concern in this section is with the control of performance at the responsibility center level rather than the more general management control process shown in Figure 1. The focus is on evaluation of operational performance where the results of the planning phases are considered "givens" [1, Chap. 1]. It is important that both financial and nonfinancial performance be evaluated in this retrospective analysis. The evaluation of financial performance generally begins with an analysis of operating and financial budgets relative to actual performance. Typically, this so-called Budgetary Control process stops with the financial review. However, for nonprofit health care organizations, nonfinancial objectives are often key elements of strategic plans and programming decisions. Consequently, a valid analysis of performance must include an evaluation of nonfinancial characteristics of operations.

Management must develop the structure and processes of its control systems, taking into consideration the ability to detect deviant behavior as well as the systems' effect on employee behavior. The detection concerns are discussed below. The behavioral concerns include the following: 1) Does the control system induce behavioral patterns that are conducive to achieving organizational objectives? 2)What unintended behavioral effects does the control system have? and 3) What latitude does the system provide for people to implement ploys? These behavioral concerns have been explored in the literature [9, p. 75; 25,53,60]. These studies, and others, emphasize the importance of such elements as participation, accurate information systems, communication, and incentives. This last consideration is receiving considerable attention in the hospital industry and a number of hospitals have explored profit sharing or incentive payment programs with their employees [61–63].

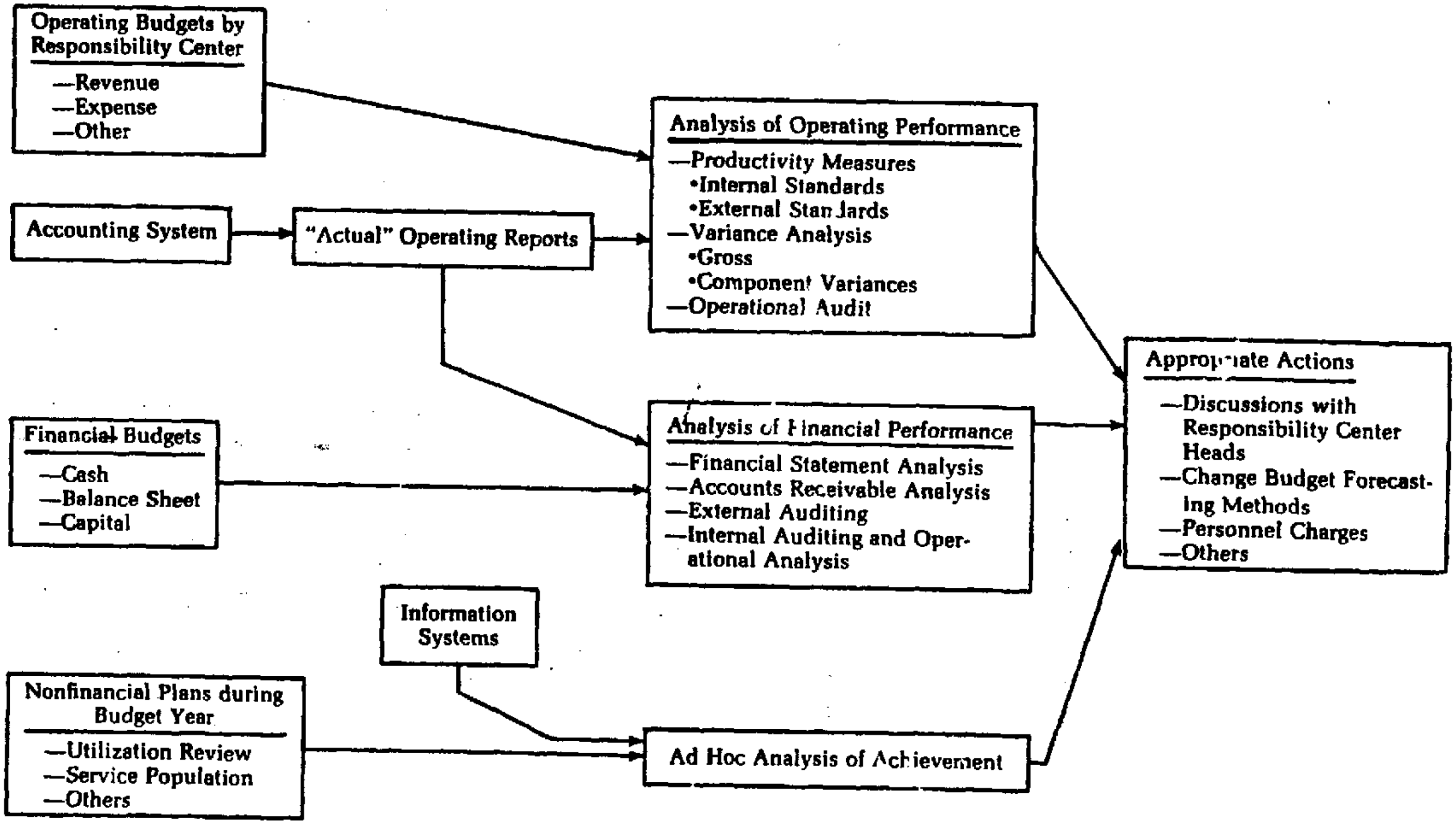
The major elements of the controlling phase are summarized in Figure 5. At the heart of this phase is an analysis of performance in which expected outcomes are compared with actual outcomes. For the operation and financial aspects of the organization, the operating and financial budgets generally provide the expected outcomes. They represent the best plans resulting from

Figure 4:
Budget Components*



*Adapted from Shillinglaw, G. *Managerial Cost Accounting*, 4th ed., p. 137. Homewood, IL: Richard D. Irwin, Inc., 1974.

Figure 5:
Schematic Diagram of Controlling Phase



analysis, forecasting, and negotiation activities in the programming and budgeting phases. For the nonfinancial objectives, explicit statements of expectations should be available from the planning phase.

Information about actual outcomes should be available from the organization's management information system. The data should be accurate, provided on a timely basis, and must be comparable with data from the budgeting system. This generally means that the hospital's financial accounting system must have the ability to 1) accumulate expense items by responsibility center, 2) provide cost accounting systems for calculating per-unit costs, and 3) collect operating statistics as an integral part of the systematic management information system. Davis and Freeman [64] offer some suggestions on how to evaluate management information systems, and DATAPRO [65] specifies some guidelines for judging commercial software packages.

The assessment of performance should be as timely as data permits. The first cut at the analysis should be standardized management reports of four types:

1. *Productivity Monitoring*. Reports in which the productivity (output ÷ input) of responsibility centers and/or individual providers is measured. These reports compare actual measures with internal and/or external standards [17, Chap. 18; 66–68].
2. *Variance Analysis*. Methodologies for estimating the difference between budgeted and actual accounts. Such analyses can be performed at the total or gross variance level for specific responsibility center or line items [69] or at the component variance levels [9, pp. 82–85; 70]. Component variance analysis attempts to partition the total experience into parts which correspond to identifiable causes, e.g., price, volume, supply quantity, etc.
3. *Operational Audit*. Process of examining the operations of departments and organizational control systems to assess their effectiveness. The process is concerned with problem detection rather than problem solving [71].
4. *Financial Statement Analysis*. Methodologies for analyzing the organization's financial performance as a whole and assessing its financial position [72,73].

The purpose of this first cut analysis is to systematically identify problems and, to the extent possible, their apparent causes. When problems are indicated, further analysis and discussions with responsibility managers are generally necessary to identify the true causes and the appropriate courses of action.

The reporting and analysis of performance serve several purposes. First, the reports are used to keep management informed of what is happening in the different responsibility centers. Management can, when necessary, take appropriate action on the basis of these reports, or can initiate additional analyses. Second, they can be used to evaluate the performance of responsibility center heads. Third, such reports may indicate that the budgeting process should be altered. Fourth, the reports can be useful in educating department heads to the financial consequences of their decisions. Without such reports, department heads may be unable to make staffing and scheduling decisions which are consistent with the overall goals of efficiency and effectiveness.

The controlling phase should be viewed as a feedback loop to the programming and budgeting phase. Analyzing performance will have little useful effect if it is done as an independent management activity. However, if conducted for the purposes of improving the planning and budgeting processes, gaining a better understanding of production processes, and improving management skills, performance evaluation can be an essential link in a hospital's attempt to improve its cost effectiveness.

State of the Art—Issues and Failings

The models presented in the preceding section outline the scope and elements of contemporary planning, budgeting, and controlling processes (PBCP). Their focus is on accurate cost-effective planning and effective management control. This section presents a number of issues associated with the orientation, design, and implementation of PBCP in the health field by highlighting the major reasons for observed differences between the state of practice and the state of the art.

Planning Attitudes

There remains a significant gap between the public interest and the private interest of health care providers. The public interest was characterized in the first section as being served by cost-effective programming decisions, i.e., based on the health care needs of a service population and using criteria which maximize system efficiency and effectiveness. The private interest is, for the most part, still geared toward increasing technological capabilities, facility size, and the volume of patients, and not toward reducing the levels of expenditures. Exceptions do exist and, in some cases, this growth is consistent with development of rational networks and multiple hospital systems. In many cases, however, this growth in assets and operating expenses is motivated by fear of becoming technologically obsolete, fear of being labeled inferior to other hospitals, and fear of losing medical staff, patients, and status.

As yet, the external constraints are still too loose to prevent hospitals from reacting to these fears out of pure self-interest. Enthoven [4, p. 1229] summarizes the attitude:

> The problem of how to best spend a given amount of money for the health care of a population is not posed. Providers are not required to set priorities, look at alternatives and make hard choices

While the debate continues on how best to improve institutional planning attitudes (i.e., by increased direct economic regulation or restructuring the financial incentives), efforts are being made to improve the means of making population-based decisions [74], and prominent students of the industry are raising the visibility of problems and methods of increasing cost-effective decision making, e.g., curtailing the "flat-of-the-curve medicine," stimulating regionalization of health care, introducing cost considerations into the physician's decision making process, and controlling the introduction of new technology [4].

It should be noted that in the current environment, cost-effective planning decisions may be impractical and ineffective unless there is strong community planning. A hospital which is very discriminating in the types of medical equipment purchased, aggressive in reducing unnecessary utilization, and effective in managing the admissions to a properly sized facility runs the risk of losing its medical staff to competing hospitals that are less restrictive. Such a danger would, however, not be as great if similar efforts were being made by all hospitals in a community.

Budgeting Attitudes

The planning attitudes extend to the budgeting process; in particular, there is no indication that hospitals are motivated to develop technically efficient budgets. Many believe this results from the fee-for-service, cost-based reimbursement system which fails to provide real incentives for efficiency. Under this system, there is little incentive to move away from peak-load staffing and to contend with the problems of part-time and temporary staff; there is little incentive to search for better modes of operation which might lower costs; and there is little incentive to say "no" to requests for additional staff which are intended to increase the quality of care without regrad to effectiveness. The negative effects of hospitals' ability to attract patients are greatly reduced by the widespread absence of these incentives and the general lack of competitive pressures in the industry. The net result has been a widespread increase in costs, quantity and perhaps quality without any type of competitive or effective regulatory mechanisms for forcing the system toward an equilibrium position.

There appears to be concensus that this system will change through either more restrictive means of calculating reimbursement rates, or a restructuring of financial incentives. While this future scenario has not altered current budgeting decisions, it has increased hospital managers' desire for improved budgeting capabilities. In some cases, hospitals are preparing for this future environment by increasing their expense base with the hope that future restrictions will be less painful.

In addition, many financial managers have sought to develop formulistic approaches to budgeting, e.g., developing departmental cost functions, and/or rigid productivity indexes. While these are important elements in the models described above, there is a danger that mathematical models will be used to develop the budget without department head input, review or understanding. Such a unilateral preparation process will reduce the effectiveness of the entire management control process and not be in the long-run interests of the hospital.

Attitudes Toward Control

The power structure in hospitals is such that management often finds its hands tied. Many decisions geared toward improving efficiency or insuring compliance with the budget often involve a perceived sacrifice by physicians and other professionals. Confrontations between administrators and the medical staff are often one-sided in that administrators are easier to replace than medical staffs. Since efficient behavior is not rewarded financially, nor does the system as yet demand efficient management, the route of choice is often the route of minimum resistance. Without clear directives and support by hospital boards, this cannot and will not change.

Information Systems

Accurate, valid, and timely data are essential for the management decisions described in PBCP. In an abstract sense, there are four key relationships which should be generated from the information systems: costs per responsibility center, costs per inputs, inputs per output, and cost per output. The information system must provide this data for the hospital's own operations and should include comparable data from other institutions. Drebin [75, p. 88] criticizes the capabilities of many (not all) hospital information systems.

> [The hospital industry] is unable to define its product or quantify the value of its product, cannot specify

> its production costs or explain why those costs are increasing at a rate higher than the average rate of costs increase in the economy. . . .

This current state of practice results more from a historical belief that managers did not need good information systems rather than from a set of unsolvable problems. There has simply been a general avoidance of implementation issues, largely in two areas: augmenting financial accounting systems for management purposes, and developing accurate cost-accounting systems.

The financial accounting system of a hospital is frequently used to provide the basic financial information regarding operations. Because the financial accounting and managerial accounting functions are quite different [2, Chap. 14], two major problems can arise. First, the data can be invalid for management decisions. The financial accounting system is based on externally determined standard accounting practices which are geared toward disclosure of financial information to the outside world. In developing these standard accounting practices, criteria are used which may not be consistent with management purposes, e.g., the criterion of objectivity often conflicts with the criterion of usefulness. This is illustrated in the accounting principles relating to the measurement of fixed assets. Financial accounting initially records an asset at cost (an objectively determined amount) and systematically reduces that value using depreciation methods (objective estimates). For the purposes of planning, it has been argued that the replacement cost of an asset is the more appropriate valuation base. The second problem arises in the definition of cost centers. There is considerable pressure for hospitals to adopt a uniform accounting system [76], and there are fears that such a system will define cost centers on the basis of function rather than responsibility [77]. If this occurs, using the financial accounting system for budgeting and control purposes will necessarily interfere with the effectiveness of the management control system [77]. Attempts are being made to develop systems which are compatible with both functional and responsibility accounting [78].

Historically, there has been little need to perform cost accounting beyond what was required by third party reimbursement formulas. As a result, definitions of output have become synonymous with patient days and fee codes, cost allocation is equated to the step-down method, and statistical bases for allocation are selected if they maximize revenue rather than if they accurately reflect the use of overhead services. As a result, most hospitals find themselves unable to measure the cost of programs, responsibility centers and specific outputs, and must make planning and control decisions based on faulty information. Berman and Maloney [79] describe these problems and their consequences for the outpatient departments of teaching hospitals, and Thompson and Cannon [80] discuss the general benefits and elements of an improved cost-accounting system.

Conclusions

The future fiscal environment will provide hospital managers with stronger incentives to develop cost-effective PBCP. The elements of such processes have been outlined in this paper and to a large extent represent a straightforward application of the techniques, activities, and philosophies practiced for decades in other sectors of the economy. The observation that most hospitals need to improve significantly their management control capabilities along these lines appears to have resulted from a perceived lack of need, followed by a failure to develop and implement—not from significant technical barriers. As the fiscal environment becomes more hostile, and as more multiple hospital systems emerge, the state of practice should approach the current state of the art. This change in management capabilities cannot occur overnight, and prudent hospital managers will recognize the time lags associated with such changes and begin implementation before the health care crisis becomes a personal problem.

REFERENCES

1. Anthony, R.N. and J. Dearden. *Management Control Systems: Text and Cases.* Homewood, IL: Richard D. Irwin, Inc., 1976.
2. Anthony, R.N. and J.S. Reece. *Management Accounting: Text and Cases,* 5th ed. Homewood, IL: Richard D. Irwin, Inc., 1975.
3. Anthony, R.N. and R. Herzlinger. *Management Control in Non-Profit Organizations.* Homewood, IL: Richard D. Irwin, Inc., 1975.
4. Enthoven, A.C. Cutting costs without cutting the quality of care. *New England Journal of Medicine* 298(22):1229, June 1, 1978.
5. Fuchs, V.R. and M.J. Kramer. *Determinants of Expenditures for Physicians' Services in the United States, 1948–1968.* New York: National Bureau of Economic Research, 1973.
6. Stevens, R. *American Medicine and the Public Interest.* New Haven, CT: Yale University Press, 1971.
7. Cochrane, A.L. *Effectiveness and Efficiency: Random Reflections on Health Services.* London: The Nuffield Provincial Hospital Trust, 1972.
8. Fuchs, V.R. *Who Shall Live?* New York: Basic Books, Inc., 1974.
9. Silvers, J.B. and C.K. Prahalad. *Financial Management of Health Care Institutions.* Flushing, NY: Spectrum Publications, Inc., 1974.
10. Wennberg, J.E. Testimony and statement presented at the Hearings before the House of Representatives, Subcommittee on Oversight and Investigations of the Committee on Interstate and Foreign Commerce, 94th Congress, 1st Session, July 15, 1975. In *Getting Ready for National Health Insurance: Unnecessary Surgery,* pp. 73–85. Washington, DC: U.S. Government Printing Office, 1975.
11. McCarthy, E.G. Testimony presented at the Hearings before the House of Representatives, Subcommittee on Oversight and Investigations of the Committee on Interstate and Foreign Commerce, 94th Congress, 1st Session, July 17, 1975. In *Getting Ready for National Health Insurance: Unnecessary Surgery,* pp. 104–116. Washington, DC: U.S. Government Printing Office, 1975.
12. Moore, R. Payments to health care institutions: State-of-the-art. Paper presented at the AUPHA Faculty Institute on Financial Management, Cincinnati, OH, Nov. 7–8, 1978.
13. Salkever, D.S. and T. Bice. The impact of certificate-of-need controls on hospital investment. *Milbank Memorial Fund Quarterly* 54(2):185, 1976.
14. Hellinger, F.J. The effect of certificate-of-need legislation on hospital investment. *Inquiry* 13(2):187, June 1976.
15. Department of Health, Education and Welfare. *National Health Planning Guidelines.* Hearing before the U.S. Congress, Subcommittee on Health and the Environment of the Committee on Interstate and Foreign Commerce, 95th Congress, 1st Session. Washington, DC: U.S. Government Printing Office, 1978.
16. Griffith, J.R. and R. Chernow. Cost effective acute care facilities planning in Michigan. *Inquiry* 14(3):229, Sept. 1977.

17. Berman, H.J. and L.E. Weeks. *The Financial Management of Hospitals*, 3rd ed. Ann Arbor: Health Administration Press, 1976.
18. Hanft, R., I. Raskin and M. Zubkoff. Introduction. In R. Hanft, I. Raskin and M. Zubkoff (eds.), *Hospital Cost Containment: Selected Notes for Future Policy*, pp. 1–30. New York: Prodist, 1978.
19. Dowling, W.L. A Procedure for Rational Planning. In J.R. Griffith, W.M. Hancock and F.C. Munson (eds.), *Cost Control in Hospitals*, pp. 25–35. Ann Arbor: Health Administration Press, 1976.
20. Association of University Programs in Health Administration. *Financial Management of Health Care Organizations: A Reference Outline and Annotated Bibliography*. Washington, DC: Association of University Programs in Hospital Administration, 1978.
21. Griffith, J.R. Measuring Service Areas and Forecasting Demand. In J.R. Griffith, W.M. Hancock and F.C. Munson, (eds.), *Cost Control in Hospitals*, pp. 36–39. Ann Arbor: Health Administration Press, 1976.
22. MacStravic, R.E. *Determining Health Needs*. Ann Arbor: Health Administration Press, 1978.
23. Brown, M. and M. Enright (eds.) Shared services. *Topics in Health Care Financing* 2(4), Summer 1976.
24. Donabedian, A. *Aspects of Medical Care Administration: Specifying the Requirements for Health Care*. Cambridge, MA: Harvard University Press, 1973.
25. Dunbar, R.L.M. Budgeting for control. *Administrative Science Quarterly* 16(1):88, Mar. 1971.
26. Klarman, H.E. Application of cost-benefit analysis to the health services and the special case of technologic innovation. *International Journal of Health Services* 4(2):325, Spring 1974.
27. Dowling, W.L. Converting Demand Forecasts Into Facility Requirements. In. J.R. Griffith, W.M. Hancock and F.C. Munson (eds.), *Cost Control in Hospitals*, pp. 70–89. Ann Arbor: Health Administration Press, 1976.
28. Boardman, J.J. Utilization Data and the Planning Process. In Ann Somers (ed.), *The Kaiser-Permanente Medical Care Program: A Symposium*, pp. 61–70. New York: The Commonwealth Fund, 1971.
29. Griffith, J.R., W.M. Hancock and F.C. Munson. Practical ways to contain hospital costs. *Harvard Business Review* 51(6):131, Nov.–Dec. 173.
30. Blue Cross Association. *Comprehensive Health Planning: Analytic Techniques*. Chicago: Blue Cross Association, 1974.
31. McClain, J.O. Bed planning using queuing theory models of hospital occupancy: A sensitivity analysis. *Inquiry* 13(2):167, June 1976.
32. Griffith, J.R. *Quantitative Techniques for Hospital Planning and Control*. Lexington, MA: Lexington Books, 1972.
33. Thomas, D. and H.H. Stokes. How many beds should a hospital department serve? *Health Services Research* 11(3):241, Fall 1976.
34. Conrad, R.B., D.A. Knee, J.M. Meade and L.M. Parrish. Utilization study saved hospitals from needless expansion of radiology facility. *Hospital Financial Management* 27(9):40, Sept. 1973.
35. Neumann, J.L. Make overhead cuts that last. *Harvard Business Review* 53(3):116, May–June 1975.
36. Bennett, J.E. and J. Krasny. Functional value analysis: A technique for reducing hospital overhead costs. *Topics in Health Care Financing* 3(4):35, Summer 1977.
37. Magerlein, D.B., W.M. Hancock, F.W. Butler, G.W. Mallett and D.R. Young. New systems can mean real savings. Parts 1 and 2. *Hospital Financial Management* 32(4):10, Apr. 1978 and 32(5):18, May 1978.
38. Hancock, W.M. Dynamics of Hospital Operational Control Systems. In J.R. Griffith, W.M. Hancock and F.C. Munson (eds.), *Cost Control in Hospitals*, pp. 129–149. Ann Arbor: Health Administration Press, 1976.

39. Munson, F.C. and W.M. Hancock. Implementation of Hospital Control Systems. In J.R. Griffith, W.M. Hancock and F.C. Munson (eds.), *Cost Control in Hospitals*, pp. 297–316. Ann Arbor: Health Administration Press, 1976.
40. Deegan, A.X. *Management By Objectives for Hospitals*. Germantown, MD: Aspen Systems Corporation, 1977.
41. Thompson, J. Planning, budgeting, and controlling—One look at the future: Case-mix cost accounting. *Health Services Research* 14(2):111, Summer 1979.
42. Houser, R. How to build and use a flexible budget. *Hospital Financial Management* 28(8):12, Aug. 1974.
43. Cleverley, W.O. One step further—The multi-variable flexible budget system. *Hospital Financial Management* 30(4):34, Apr. 1976.
44. Heda, S.S. and S.H. Lipson. Dealing with a 13-period year. *Hospital Financial Management* 28(12):54, Dec. 1974.
45. Chambers, J.C., S.K. Mullick and D.D. Smith. How to choose the right forecasting technique. *Harvard Business Review* 49(4):45, July-Aug. 1971.
46. Donlon, V.W. Statistical methods to forecast volumes of service for the revenue budget. *Hospital Financial Management* 29(4):38, 1975.
47. Koza, R.C. Time-series data is essential to health care planning. *Hospital Financial Management* 28(1):32, Jan. 1974.
48. Warner, D.M. Nurse staffing, scheduling, and reallocation in the hospital. *Hospital and Health Services Administration* 21(3):77, Summer 1976.
49. McNally, J.K. How to find manpower levels/needs before budgeting. *Hospital Financial Management* 30(12):34, Dec. 1976.
50. Kaplan, R.S. Analysis and control of nurse staffing. *Health Services Research* 10(3):278, Fall 1975.
51. Flynn, J.W. Budgeting salary savings with department head input. *Hospital Financial Management* 32(6):38, June 1978.
52. Lipson, S.H. and M.D. Hensel. *Hospital Manpower Budget Preparation Manual*. Ann Arbor: Health Administration Press, 1975.
53. Buckley, A. and E. McKenna. Budgetary control and business behavior. *Accounting and Business Research* 2(6):137, Spring 1972.
54. Holmgren, J.H. *Purchasing for the Health Care Facility*. Springfield, IL: Charles C. Thomas, 1975.
55. Cleverley, W.O. Input-output analysis and the hospital budgeting process. *Health Services Research* 10(1):36, Spring 1975.
56. Davis, R.J. *Testing the linearity assumption in hospital budgeting and control*. Unpublished Master's thesis, Program in Hospital Administration, The University of Michigan, 1977.
57. Pyhrr, P.A. Zero-based budgeting. *Harvard Business Review* 48(6):111, Nov.–Dec. 1970.
58. Lemer, R.S. and D.E. Willman. Rate setting. *Topics in Health Care Financing* 1(2), Winter 1974.
59. Benson, F.S. and M.A. Santullano. Modeling makes budgeting more efficient. *Hospitals* 50(1):93, Jan. 1, 1976.
60. Lawler, E.E. and J.G. Rhode. *Information and Control in Organizations*. Pacific Palisades, CA: Goodyear Publishing Company, Inc., 1975.
61. Groner, P.N. Employee incentives. *Topics in Health Care Financing* 3(3):63, Spring 1977.
62. Austin, C.J. Wage incentive systems: A review. *Hospital Progress* 51(4):36, Apr. 1970.
63. Williams, F.G. And D.C. Anderson. Cost control incentive programs: Appropriate for non-profit? *Hospital Financial Management* 32(5):14, May 1978.
64. Davis, S. and J.R. Freeman. Hospital managers need management information systems. *Health Care Management Review* 1(4):65, Fall 1976.

Programing, Budgeting, and Control

65. DATAPRO. How to buy software packages. *DATAPRO.* DATAPRO Research Corp., Jan. 1972.
66. Herkimer, A.G. Improving productivity. *Topics in Health Care Financing* 4(2), Winter 1977.
67. American Hospital Association. *The Management of Hospital Employee Productivity: An Introductory Handbook.* Chicago: American Hospital Association, 1973.
68. Moriuchi, M., J.D. Durham and R.D. Roberts. Productivity monitoring: First step to effective manpower allocation. *Hospital Financial Management* 32(5):30, June 1978.
69. American Hospital Association. *Budgeting Procedures for Hospitals*, Chap. 9. Chicago: American Hospital Association, 1971.
70. Holder, W.W. and J. Williams. Better cost control with flexible budgets and variance analysis. *Hospital Financial Management* 30(1):12, Jan. 1976.
71. Flesher, D.L. *Operations Auditing in Hospitals.* Lexington, MA: Lexington Books, Inc., 1976.
72. Choate, G.M. Financial ratio analysis. Hospital Progress 55(1):49, Jan. 1974.
73. Bernstein, L.A. *Financial Statement Analysis: Theory, Application and Interpretation.* Homewood, IL: Richard D. Irwin, Inc., 1974.
74. Griffith, J.R. *Measuring Hospital Performance.* Chicago: Blue Cross Association, 1978.
75. Drebin, M.E. Financial information systems: Key to hospitals' survival. *Hospitals* 52(12):88, June 16, 1978.
76. Kovener, R.R. Uniform accounting, the background. *Hospital Financial Management* 31(6):20, June 1977.
77. Ingram, J.D. Uniform accounting, how it will affect hospitals. *Hospital Financial Management* 31(6):10, June 1977.
78. American Hospital Association. *Chart of Accounts for Hospitals.* Chicago: American Hospital Association, 1976.
79. Berman, R.A. and T.W. Maloney. Are outpatient departments responsible for the fiscal crisis facing teaching hospitals? *The Journal of Ambulatory Care Management* 1(1):37, Jan. 1978.
80. Thompson, G.B. and W.G. Cannon. Hospitals, like industry, must apply cost accounting techniques. *Hospitals* 52(12):129, June 16, 1978.

DESIGN OF ALTERNATIVE PROVIDER TEAM CONFIGURATIONS: EXPERIENCE IN BOTH DEVELOPED AND DEVELOPING COUNTRIES

by

Dr. Lilia Duran* and Dr. Arnold Reisman**

Technical Memo 497

ABSTRACT

Job Evaluation, a time-honored industrial engineering technique developed for manual labor rate-setting was used to design new personnel resource teams for providing anesthesiology services in well-known U.S. hospitals on the one hand, and for delivering primary care in regions of Latin America on the other. This paper discusses the use of this technique in designing job descriptions for new allied professionals, and for designing the curricula necessary to train same. Experience to date in integrating the products of these curricula into health provider teams in U.S. tertiary care institutions and in primary care within several countries of Latin America indicates the soundness of this approach.

---

*Senior Researcher, Centro Universitario de Tecnologia Educacional para la Salud (CEUTES), National University of Mexico, Mexico City.

**Professor of Operations Research, Case Western Reserve University, Cleveland, Ohio.

BACKGROUND; THE UNITED STATES

During the last two decades or more, the United States has been experiencing a shortage of physician-anesthetists relative to the need for their service [1,2,3]. Many anesthesiology residencies over the years have gone unfilled as have positions within hospitals across the country. As late as 1976 it was reported that 50% of the anesthetics in the U.S. were given by nurse anesthetists* [4]. Parenthetically, this paper will not take a position as to whether administration of anesthesia ought to be performed by physician-anesthesiologists or nurse-anesthetists, both are in short supply. The American Board of Anesthesiology has reduced the number of residency training programs significantly, during the last five years. The programs are now concentrated in major medical centers. Residencies in smaller peripheral hospitals are few. Yet, the available residency positions go unfilled.

The shortage of anesthesiologists will probably be influenced in the future by the following:

----

*These figures have been questioned because there are no data to establish what proportion of these anesthetics were given by nurses functioning alone as opposed to those supervised by anesthesiologists [5].

1. A change in the immigration policy for foreign medical graduates has led to the reduction of the number of foreign medical graduates available to enter this specialty. In 1972 for example, 58% of anesthesiologists in formal training programs were graduates of foreign medical schools [6].

2. The percentage of U. S. medical school graduates entering anesthesia is decreasing.

3. The anesthesiologists who entered the specialty during the immediate post-World War II period - a time of the highest recruitment into the specialty - are now approaching retirement.

Much has been said about physician maldistribution. Anesthesiologists reflect the same maldistribution. The pronounced migration out of the snow belt into the southern and western coastal areas and into the sun belt of the population in general is reflected also in the movement of anesthesiologists. This outer-migration further reinforces the problem of anesthesia coverage in small peripheral hospitals of the North. The small and peripheral community hospitals have historically experienced the most difficulty in acquiring anesthesiologists. Of the 7000 hospitals in the U.S., one half have 100 or fewer beds. Of this latter 3500 hospitals, only 9% have anesthesiologist (MD) coverage [7]. There is a maldistribution of coverage for obstetrical needs as opposed to surgical needs [8,9]. Lastly, in situations where anesthesiologists work alone and without support personnel, it is fair to assume that they cannot spend the time outside the operating room to devote their attention to work in the recovery room, intensive care unit or as a person who responds to cardiac arrests or functions as consultant for pulmonary management. The best use of anesthesiologists would be to have them be involved in decision-making skills that are needed from a physician and delegate the more technical skills to people with lesser training [1].

Time-sharing of anesthesiologists: the Solution*

The use of physician extenders is a rationale reflecting a national trend to making the time of physicians available to more patients by extending their ability to see patients. Telecommunications technology in medicine or Telemedicine has been proposed as a solution to problems in medical care during the 1960s [10,11]. Health care professionals in this mode of practice use the telecommunication channels to communicate with each other or with patients in order to improve the cost-effectiveness and/or cost benefit of the delivery of health care services. Systems analysis studies have shown that using either or both the new communication technologies and different ways to organize the practice of medicine can result in improving both the quality and

----

*These studies were funded by grants from the U. S. Public Health Service and from the National Institute of Health.

accessibility of health care delivered while at the same time reducing the costs of same [12,13].

Two programs to train Physician Assistants in Anesthesia were developed[1] based on the perceived need for more non-physician anesthesia personnel possessing training that is different in kind from the Nurse Anesthetist. This training, it was thought, should make the graduates be on a par with a Nurse Anesthetist yet allow them increased upward mobility with access to professional station in or outside of medicine. Further, it was hoped that some of the inter-professional political strains which have traditionally existed between MD's and RN's would thus be avoided.

It was thought the people drawn into this kind of education would not commonly be those from the nursing pool. Nursing at the time of development of the above curricula was experiencing its own shortage problems. These shortages persist. It was also thought that the use of a non-physician to extend the contact time physicians could have with patients would extend their decision-making qualities to more patients while reducing the costs of service.

---

[1]Case Western Reserve University in Cleveland, Ohio and Emory University in Atlanta, Georgia.

Prior to developing the new curriculum, a task analysis was undertaken in anesthesiology [14]. The study identified all tasks performed and established which of them could reasonably be delegated to non-physicians in a manner that would conserve the anesthesiologist's time yet result in quality patient care and wider coverage. The determination of which tasks can be delegated, under what conditions and to whom was made by a group of qualified anesthesiologists.

The design of the curriculum was obviously an extension of the results of the task analysis and of the Job Evaluation Point-Rating System [15,16,17] to be discussed later in this paper. An assessment of the types of activities that non-physicians could reasonably engage in. resulted in the curriculum for Anesthesiologists Associates [18].

Lastly, a methodology was developed to design optimal anesthesia team involving Board-qualified MD's and the Anesthesia Associates. The mix of provider categories was considered in this methodology to be constrained by budget, personnel availabilities and the acceptable levels of worker "overloads". The overloads were used as surrogate measures of the quality of care.

The computer simulation developed to study the efficacy of alternative team configurations generates a set of daily overload indicators for different team configurations given some patient load and mix distribution. The results are plotted on a two dimensional graph to illustrate the trade-offs between tolerable overload levels and manpower requirements. These simulation results provide a basis for manpower mix tradeoffs. An overload was defined as the unavailability of a given worker within a 5-minute time period to respond to demands for his or her services because of scheduled or unscheduled commitments to another case. The three overload categories were further defined as:

(1) Emergency Overload: Number of periods of unsatisfied emergency calls per day.

(2) Urgent Overload: Number of periods of unsatisfied induction, early maintenance and surgical procedure tasks per day.

(3) Routine Overload: Number of periods of unsatisfied tasks per day.

Experience with the curriculum in both programs at the Baccalaureate (CWRU) and the Masters level (Emory) indicates that the content is adequate to train people to function in anesthesiology at the level of a Nurse Anesthetist. Yet the new professionals bring more technical or specialized expertise, into the system. This additional expertise is primarily in respiratory monitoring and care and in the utilization of more sophisticated equipment. Acceptance of graduates from both programs has been quite good. Although many graduates have gone on to other kinds of activities and to graduate study, those who chose to remain in the profession and work in anesthesia are indeed gainfully employed.

BACKGROUND; LATIN AMERICA

One of the main problems in primary health care (PHC) delivery systems in Latin-American countries is the lack of coordination between the providers of PHC services and the educational institutions engaged in training health personnel [19-26]. This lack of coordination translates itself into ill-defined statements about the quantity, quality and mix of professional resources needed to be trained. The result is an excess and maldistribution of health personnel in some areas and almost total absence of personnel in other areas. These distribution problems are further complicated by an accumulation of tasks at some professional levels while other levels are left with few tasks to perform.

Task analysis, job and team design, much like in the Anesthesiology studies, were used as the methodological tools to generate a data base for use by planners of health services and by educational institutions. The study* developed a methodology which identifies the professional profiles needed to satisfy the special conditions of several regions of Mexico.

----

*This study was and continues to be funded by the National University of Mexico and the National Council of Science and Technology (MEXICO).

Primary Care Health Service System: The Solution

The main objectives of the study were a) to develop a technique which will allow the definition of the number and kind of health personnel that ought to be trained; b) to provide a communication flow, between the training units and the providers of health care; and c) to establish a feasible way to integrate curricula for each of the categories of health personnel, thereby avoiding the proliferation of disorganized actions in providing health care.

As a point of departure, the study defined a model of the health service system approach to providing care. It was decided to work under a model in which the service was more preventive than curative oriented [19,20,24,25], even though some curative tasks were included. The model focuses on the delegation of tasks to primary care and preventive medicine personnel, and on the design of provider team configurations.

In order to do this, it was necessary to establish the different sequences of health care requiring different levels of professional skills. These sequences start with self-care, all tasks performed by the individual to preserve his or his family's health and end up with a sequence of tasks requiring the highest professional level. Lastly, the most significant health problems of Mexico were identified and prioritized. From the above a matrix (Figure 1) was constructed relating the sequences of care to the health problems found to be important.

Job Design

Following the methodology delineated in the Anesthesiology studies [14] modifications of job evaluation techniques were used to design new job descriptions for health personnel. Using the downward delegation of tasks principle, the needed health personnel profiles were thus clearly defined. All tasks that were actually performed by personnel were identified and delegated to the least expensive and/or least trained professional capable of performing the task adequately.

According to the scheme shown in Table 1, each box described the functions that each level of professional has to perform with respect to each problem in order to maintain and/or preserve a health outcome. As in anesthesia, this resulted in job profiles not in existence at the time. This, an additional benefit of the analysis, lead to a more integrated approach to health care delivery.

Rather than going to the field where the tasks were actually performed as called for by the traditional job evaluation technique, and as done in the Anesthesia curriculum study, panels from each of the professional levels in practice at different health and educational institutions in Mexico were identified and used in this study. Each panel had a high degree of expertise in each of the selected health problem areas. The task assigned to these panels was to fill each of the boxes in the table. They were asked to write the ideal professional functions that each level will be performing within the next 10 to 15 years, and to consider what would be the ideal mix of team configurations.

As in Anesthesia, this study used the four basic criteria of job evaluation, namely 1) SKILL, 2) EFFORT, 3) RESPONSIBILITY, 4) JOB CONDITIONS [15,16,17]. Each of the above factors was in turn broken down to as many levels as necessary to rate the tasks necessary for handling each of the selected health problems. The panels were instrumental in delineating/refining all the factors and subfactors of this plan. Table 2 - shows the degrees assigned to each of the subfactors and the description given to each degree. The panels were able to establish a clear definition for each of the break points on the scale, avoiding uncertainty and favoring a more precise rating. The tasks thus delineated and their ratings with respect to malnutrition, a significant health problem in Mexico, are presented in Table 3. Currently work is proceeding in the mental health area, and the same information will be developed for 25 other major health problems. With all this information in hand, it will be possible to mix and interrelate all same-level tasks to each of the health problems. This procedure will further provide descriptions for new allied health personnel.

This technique has proven useful to assign more adequate functions to already existing personnel. In Mexico, it has already defined provider functions for the family planning program. A continuing education workshop was organized in family planning for social assistants, social workers, nurse auxiliaries, registered nurses, general practitioners and gynecologists. The coverage of this workshop was nationwide, and it included all major Mexican health institutions.

The first step was to establish what task ought to be performed by the personnel, irrespective of what they were actually doing in their respective institutions. Thus, the procedure described earlier was again applied.

With the help of this panel, the tasks that each personnel category should realize were defined and rated according to a number of criteria. Data were then gathered regarding who was currently performing some of the tasks. A questionnaire was administered to solicit the panel opinion with respect to the importance of each task in achieving the objectives of the family planning program and the frequency of task performance.

The tasks that were rated "very important" and performed on a daily basis were thus identified and given greater priority. Based on this study, a course was developed and manuals for several health personnel categories were developed, including among others, an instruction manual, student manual, and actual-practice manual.

The course was developed mainly to teach "multipliers" all around the country to handle these materials, by performing the double function of instruction while providing content and method for the training of other personnel as instructors.

In August 1980, the first workshop including 150 employees of the Secretary of Health and Welfare from all over the country was conducted. It has been reported that this workshop had an impact in improving the performance of the personnel assigned to family planning programs.

Using this approach, mid-career PHC workers have been trained for Colombia, Peru, Ecuador, Honduras, Brazil, Bolivia and other areas of Latin-America. Among these were PHC physicians, nurses, social workers, etc. Feedback indicates that these procedures are already being implemented in a number of agencies of family planning, mental health and other programs within Mexico, Colombia and Nicaragua. Lastly, the task analysis/job design approach outlined here has already been incorporated in redesigning curricula at several universities in Mexico.

SUMMARY

This paper indicated how a technique developed for and used in industrial settings to evaluate manual-job compensation levels [16,17] was used to design new professional level job categories for providing more cost-effective health care. Moreover, the same basic approach was shown to apply in acute care U.S. settings [14,18] as well as in primary care within developing regions of Latin America [29,30,31,32].

TABLE 1. DISTRIBUTION OF CARE FUNCTIONS (PUBLIC HEALTH AND MEDICAL)

SEQUENCES OF CARE / HEALTH PROBLEMS	MALNUTRITION*	DIARRHEA CAUSE NOT YET DETERMINATED	ACUTE RESPIRATORY INFECTIONS	MALARIA	TUBERCULOSIS
5th sequence 3rd. professional level skills	Clinical and lab. diagnosis Treatment of malnutrition of 3rd. degree Basic research Planning of teaching programs Publish and disseminate information	Professional functions	Professional functions	Professional functions	Professional functions
4th sequence 2nd. professional level skills	Clinically and etiologically diagnose of malnutrition of 1st. and 2nd. degree Treatment of 1st. and 2nd. degree malnutrition Record data	↓	↓	↓	↓
3rd sequence 1st. professional level skills	-Organization of monitors -Training of monitors -Tabulation of health data -Realization of specific orientation campaigns	↓	↓	↓	↓
2nd sequence Monitor	-Orientation -Detection -Gather data on a population sample	Non-professional assigned functions	Non-professional assigned functions	Non-professional assigned functions	Non-professional assigned functions
1st sequence Self care	Observation Detection Gives information Elaboration of nutritive menus	↓	↓	↓	↓

*We are presenting only a few examples in each box of the functions assigned by the panel to each level.

TABLE 2. HEALTH PROBLEMS RATING PLAN

I. SKILL

(A) EDUCATION	(B) EXPERIENCE	(C) INITIATIVE AND INGENUITY	(D) MANUAL DEXTERITY	(E) INTERNATIONAL RELATIONSHIPS	(F) VERBAL COMMUNICATION	(G) WRITTEN COMMUNICATION	(H) MANAGEMENT	(I) TIME OF SPECIFIC TRAINING
0 None	0 Empirical	0 Little judgement (comply with simple specific instruction)	0 None	0 Limited	0 Moderated	0 None	0 None	0 None
1 Elementary School	1 Up to one year	1 Little judgement (comply with complex specific instruction)	1 Little manual dexterity (does not need training)	1 Moderated	1 High	1 Moderate	1 Little	1 Less than six months (general)
2 High School	2 More than one year	2 Moderate (Plans sequences of operations within set standars)	2 Moderate (needs general training)	2 High		2 High	2 Moderate	2 Up to one year (specific)
3 Preparat. College		3 High degree of judgement. (only general methods or no methods available)	3 High (needs special training)				3 High (professional level)	3 Continued Education
4 Technical								
5 Bachelor (degree)								
6 Specialization Degree								
7 Master Degree								
8 Doctoral Degree								

TABLE 2. (CONTINUED)

II. RESPONSABILITY FOR

(A) PATIENT	(B) EQUIPMENT AND MATERIAL	(C) WORK OF OTHERS
0 No responsability	0 None	0 0 Persons
1 Delegated responsability	1 Low cost material	1 Up to 15 persons
2 Ultimate responsability	2 low cost general equipment	2 15 or more persons
	3 High cost specialized equipment	

III. EFFORT

(A) MENTAL	(B) PHYSICAL
0 Requires little concentration (observing ECG)	0 Little
1 Requieres moderate concentration (Reading flow-meter indicator)	1 Moderate
2 Requires intense concentration (Administering exact dosage in IV)	2 High

TABLE 2. (CONTINUED)

IV. JOB CONDITIONS

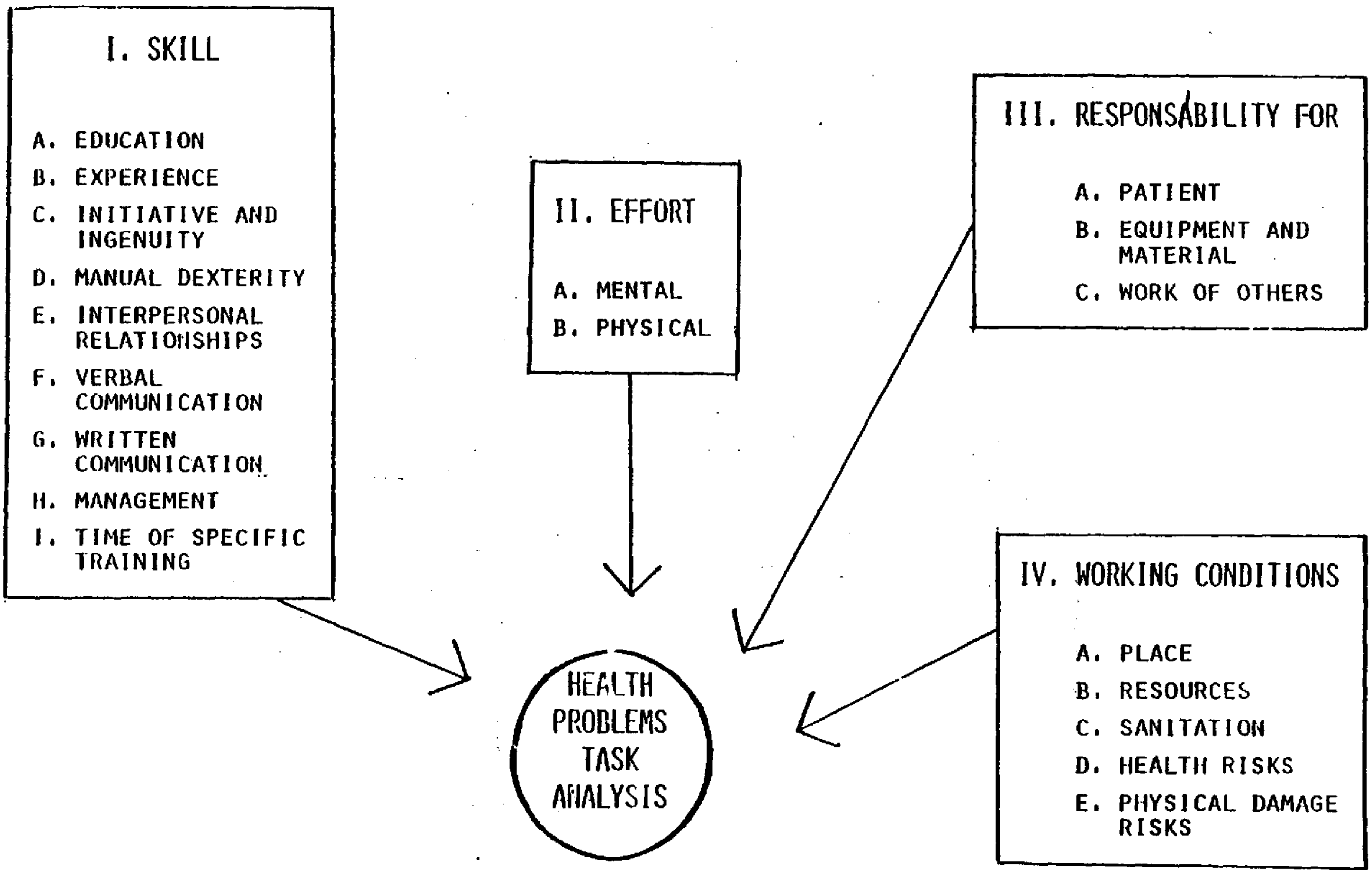
(A) PLACE	(B) RESOURCES	(C) SANITATION	(D) HEALTH RISK	(E) PHYSICAL DAMAGE RISK
0 Field - home - community	0 None	0 None	0 Without risk	0 Absence of risk
Health home	1 Scarse	1 Low	1 With high risk	1 Presence of risk
Health Center	2 Moderate	2 Moderate		
General Hospital	3 High	3 Good		
Hospital of specialities				

TABLE 3

TASK ANALYSIS OF MALNUTRITION

TASK	I									II			III		IV				
	A	B	C	D	E	F	G	H	I	A	B	C	A	B	A	B	C	D	E
SELF-CARE																			
1. Weight and measure the child periodically.	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	1
2. Watch over the child in his daily activities and educational progress.	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1
3. Assist to the service of well baby care.	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	1
4. To provide to the child a balanced diet prepared hygienically according to the family's economic conditions.	1	1	0	1	1	0	0	0	1	2	0	0	0	0	0	3	3	0	0
5. To promote that the child be engaged in frequent recreative activities in open places.	1	0	0	0	0	0	0	0	1	1	1	0	0	1	0	1	2	1	1
MONITOR																			
6. Transmite information to organized groups.	1	1	1	2	1	1	1	1	1	2	2	1	1	1	1	1	0	0	0
7. Realize surveys and distribute printed material with information on malnutrition.	1	1	1	2	1	1	1	1	1	2	2	1	1	2	0	1	0	1	1
8. Control the ralization of problem's detection campaigns and their prevention.	2	1	2	2	1	1	2	1	2	2	2	1	1	2	0	1	0	1	1
9. To differentiate presumtive degrees of malnutrition.	4	1	2	2	1	1	2	2	2	2	1	0	1	0	1	2	1	1	1
10. To canalize to specialized treatment the required cases.	4	1	2	2	1	1	2	2	2	2	1	0	1	0	1	1	1	0	0
1ST. PROFESSIONAL LEVEL																			
11. To carry out lab. examinations.	4	1	3	2	0	1	2	2	2	2	3	0	1	1	2	3	3	1	1
12. To capacitate monitors.	4	1	3	2	1	1	2	2	2	2	2	2	1	2	1	1	1	0	0
13. To supervise monitors.	4	1	3	2	1	1	2	2	2	2	0	2	1	2	0	1	1	0	0
2ND. PROFESSIONAL LEVEL																			
14. To establish clinic diagnostic of certainty of malnutrition and degrees of it.	5	2	3	2	2	1	2	2	3	2	3	0	2	0	2	2	2	0	0
15. To indicate and to interpret lab. exams.	5	2	3	2	2	1	2	2	3	2	3	0	2	1	2	2	2	0	0
16. Specific treatment of the cases.	5	2	3	2	2	1	2	2	3	2	1	0	1	0	2	3	2	0	0
17. Follow directly the evolution of acute cases.	5	2	3	2	2	1	2	2	3	2	1	0	1	1	0	3	0	1	1
18. Register in total form the cases treated.	5	2	3	2	2	1	2	2	3	2	2	1	2	1	3	3	3	0	0
19. To interput treatment results.	5	2	3	2	2	1	2	3	3	2	0	0	2	1	3	0	3	0	0
20. Send to subordinated levels the cases in resolution phase in order to control them.	6	2	3	2	2	1	2	3	3	2	1	0	1	1	3	3	3	1	1

FIG. 1. FACTORS AND SUBFACTORS FOR HEALTH PROBLEMS TASK ANALYSIS



TASK ANALYSIS OF MALNUTRITION

TASK	I									II			III		IV				
	A	B	C	D	E	F	G	H	I	A	B	C	A	B	A	B	C	D	E
21. To maintain interchanges of experiences of the case with health units related to the problem.	5	2	4	2	2	1	2	3	3	2	2	1	2	1	4	3	3	0	0
22. Plan and develop training courses for auxiliar personnel.	7	2	4	3	2	1	2	3	3	2	2	2	2	2	3	3	3	0	0
3RD. PROFESSIONAL LEVEL																			
23. Planning of detection campaigns.	7	2	4	3	2	1	2	3	3	2	2	2	2	1	4	3	3	0	0
24. To diagnose nutritional deficiences due to specific factors.	6	2	4	3	2	1	2	3	3	2	3	0	2	1	4	3	3	0	0
25. To prescribe specific treatment.	6	2	4	3	2	1	2	3	3	2	3	0	2	1	4	3	3	0	0
26. To do research on malnutrition.	6	2	4	3	2	1	2	3	3	2	3	0	2	1	4	3	3	0	0
27. To publish papers with respect to research outcomes.	6	2	4	3	2	1	2	3	3	2	3	1	2	1	4	3	3	0	0
28. To disseminate research's results.	6	2	4	3	2	1	2	3	3	2	3	1	2	1	4	3	3	0	0
29. Elaborate, coordinate and administer teaching and service plans.	7	2	4	3	2	1	2	3	3	2	3	2	2	1	4	3	3	0	0
30. To interchange interinstitutional experiences.	6	2	4	3	2	1	2	3	3	2	3	1	2	1	4	3	3	0	0

REFERENCES

1. Gravenstein, J.S., J.E. Steinhaus, P.P. Volpitto. "Analysis of Manpower in Anesthesiology," Anesthesiology, 33(3), September 1970.

2. Gravenstein, J.S., A.Esogbue, A. Reisman, B.V. Dean, V.V. Aggarwal, V. Kaujalgi, and P. Lewy. "Physician Supply and Surgical Demand Forecasting: A Regional Manpower Study," Management Science, August 1973.

3. Reisman, A., B.V. Dean, A.O. Esogbue, V. Aggarwal, V. Kaujalgi, P. Lewy, C. deKluyver, J.S. Gravenstein. "Supply and Demand of Anesthesiologists in Cuyahoga County, Ohio," The Ohio State Medical Journal, 69(10), pp. 760-763, October 1973.

4. Moore, Francis, "Ravenstine Lecture," Anesthesiology 48:125-138, 1978.

5. Ament, Richard, "Anesthesia and Surgical Care: Manpower Needs and Ultilization," Anesthesiology, 48:88-90, 1978.

6. Orkin, Fred, American Society of Anesthesiologists, 1981 annual meeting, unpublished report.

7. Professional Activities Survey, P.A.S. Ann Arbor Michigan 1978-1980.

8. "National Study of Maternity Care: Survey of Obstetric Practice and Associated Services in Hospitals in the United States - 1970," Committee of Maternal Health, American College of Obstetricians and Gynecologists, 1970.

9. Rasmussen, J., T. George, A. Reisman, W. Cull, J.S. Gravenstein, "Obstetrical Anesthesiology Practice and Attitudes in Cuyahoga County, Ohio Hospitals: A Survey," The Ohio State Medical Journal, Vol. 73, pp. 139-144.

10. Bashshur, R.L., Armstrong, P.A., Youssef, Z.I. Telemedicine: Explorations in the Use of Telecommunications in Health Care. Springfield, Charles C. Thomas, 1975.

11. Gravenstein, J.S., B. Grundy, Y. Pao, A. Reisman, F. Staub, "The Cleveland Telemedicine Project," Telecommunications For Civic And Social Services, No. 20, 1976 Wescon Professional Program Proceedings, Los Angeles, pp. 1-2, September 1976.

12. Reisman, A., F.Staub, B.V. Dean, J.S. Gravenstein, V. Aggarwal, V. Kaujalgi, C. deKluyver. "Anesthesiology Manpower Planning Study Phase III - Progress Report I," Technical Memorandum No. 280, Department of Operations Research, Case Western Reserve University, December 1972.

13. Reisman, A., W. Cull, H. Emmons, B. Dean, C. Lin, J. Rasmussen, P. Darukhanavala, T. George, "On the Design of Alternative Obstetric Anesthesia Team Configurations," Management Science, Vol 23, No. 6, February, 1977.

14. Reisman, A., B.V. Dean, V. Kaujalgi, V. Aggarwal, P. Lewy, J.S. Gravenstein. "A Task Analysis in a Clinical Specialty Providing Data for a New Curriculum for Anesthesia Personnel," Socio-Economic Planning Sciences, Vol. 7, pp. 371-379, 1973.

15. Job Evaluation Plan, American Institute of Industrial Management.

16. Lanham, E., Job Evaluation. McGraw-Hill, New York, 1955.

17. Lytle, C.W., Job Evaluation Methods. Ronald Press, New York, 1954.

18. Gravenstein, J.S., and M. F. Rhoton. "Teaching Anesthesia to Undergraduate College Students," Anesthesiology, 37(6), 641-646, 1972.

19. Arango, Jaime "Planeacion Curricular y Evaluacion en Funcion de las Necesidades de un Pais," Educacion Medica y Salud, 6, 3 y 4, 1972.

20. Butler, Willis, "The Undergraduate Education of Physicians in Cuba," Journal of Medical Education, Vol. 48, September, 1973.

21. Galvao Lobo, Luiz Carlos, "O Uso de Nova Tecnologia Educacional no Formacao de Recursos Humanos," Educacion Medica y Salud, 8, 2, 1974.

22. Gutierrez, Rodrigo, "Participacion de la Universidad en la Planificacion de la Salud," Educacion Medica y Salud, 7,3-4, 1978.

23 Hutchins, Edwin B. y Hoack, Horst, "Evaluacion de los Programas de las Escuelas de Medicina," Educacion Medica y Salud, 4. 4. 1970.

24. Duran, Lilia, "An Alternative Way of Curriculum Planning in the Health Sciences." Conference presented at the International Coloquium of Education, National University of Mexico, October, 1979.

25. Duran, Lilia, "Desarrollo de una Metodologia para Establecer Prioridades de Atencion en el Sector Salud." Internal working paper, Centro Universitario de Tecnologia Educacional para la Salud (CEUTES), June, 1981.25.

26. LaSalle, Gerald, "Nuevos Modelos en la Formacion Profesional del Personal de la Salud," Universidad de Sherbrooke, Educacion Medica y Salud, 4, 4, 1970.

27. Maddison, D.C. "What's Wrong with Medical Education?" Medical Education. 1978, 12, pp. 97-102.

28. Rexed, Bror, "The Role of Medical Education in Planning. The Development of a National Health Care System," Journal of Medical Education, Vol. 49, January 1974.

29. Duran, Lilia, "Una Aplicacion de la Tecnica de Analisis de Tareas a Partir de un Modelo de Servicio en el Area de la Salud y su Derivacion en Objetivos Generals de Carrera." Technical Report. Centro Latinoamericano de Tecnologia Educacional para la Salud (CLATES), 1979.

30. Duran, Lilia y Torres, A., "Planificacion Familiar: Una Aplicacion de la Tecnica de Analisis de Tareas." Technical Report. Centro Latinoamericano de Tecnologia Educacional para la Salud (CLATES), 1979.

31. Duran, Lilia, "Algunas Sugerencias de Apoyo al Catalogo Nacional de Ocupaciones en el Area de Planeamiento Curricular." Documento Presentado a la Direccion General de Planeacion de la S.E.P. (Secretary of Public Education) Marzo, 1980.

32. Duran, Lilia, "Technology Development for a Quantitative and Qualitative Definition of Human Resourcres in the Health Field." Research proposal submitted to the National Council of Science and Technology, June, 1979.

#11

METHODS

Parameters Affecting Hospital Occupancy and Implications for Facility Sizing

By Walton M. Hancock, David B. Magerlein, Robert H. Storer, and James B. Martin

Simulation is used to investigate the effects on hospital occupancy of the number of beds in the facility, the percentage of patients who are emergencies, the percentage of elective patients who are scheduled, and the average lengths of stay of emergency and elective patients. A practical method is presented for estimating the optimum size of a short-term hospital on the basis of expected demand, and use of the results in planning is discussed.

To operate at minimum cost it is imperative that a hospital not contain more beds than necessary to meet demand. The nontrivial costs of building, staffing, and maintaining unused beds are unnecessary and are ultimately borne by the health care consumer. The variable cost of staffing and maintaining a hospital bed was found to be $10,130 per year in one study [1] and $16,201 per year in another [2], and the fixed cost of building a bed has been quoted as $50,000 [3]. Thus, considerable savings are possible by eliminating beds in an overbedded hospital (saving variable costs) and by preventing the construction of unneeded beds (saving both fixed and variable costs).

A simulation-based analysis of the effects of several parameters on hospital occupancy is presented. The parameters investigated are number of beds in the facility, percentage of patients who are emergencies, percentage of elective patients whose admission date is set in advance (scheduled patients), and mean lengths of stay of emergency and elective patients. A practical method is developed for predicting the correct size of hospital facilities, given the expected demand. Current planning methods such as the Hill-Burton formula, the Poisson assumption, and that of Shonick [4] are shown to be inappropriate

This study was supported by PHS grant no. HS00228 from the National Center for Health Services Research (DHEW) and by the W.K. Kellogg Foundation through the Greater Detroit Area Hospital Council, Inc.

Address communications and requests for reprints to Walton M. Hancock, Professor of Industrial and Operations Engineering and Hospital Administration, Program and Bureau of Hospital Administration, University of Michigan School of Public Health, 1420 Washington Heights, Ann Arbor, MI 48109. David P. Magerlein is a graduate student and Robert H. Storer is an undergraduate student, both in industrial and operations engineering; and James B. Martin is assistant professor of hospital administration. All of the authors are at the University of Michigan.

0017-9124/78/03027614/$02.00/0

since they substantially overestimate the necessary number of beds in most cases.

Simulation Model

The Admissions Scheduling and Control System

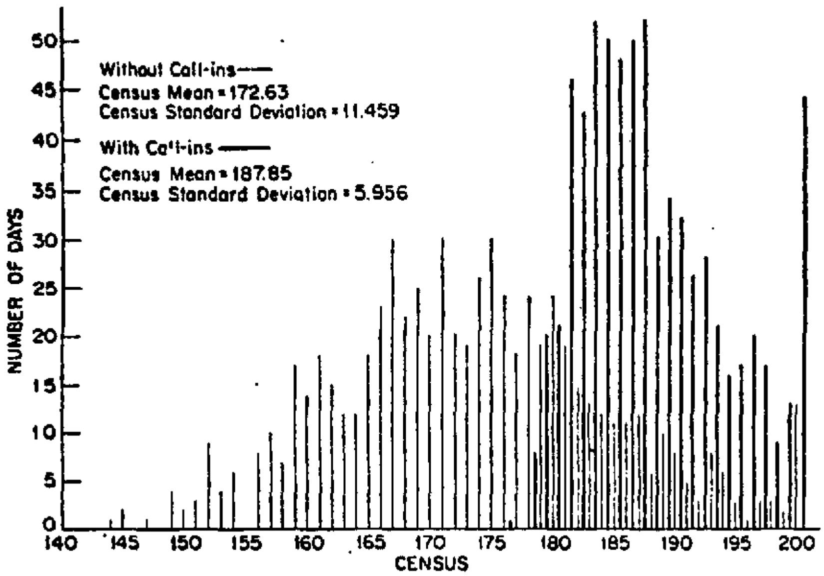
In order to model and simulate a hospital, specific rules, policies, and priorities regarding patient admissions must be defined. The admissions guidelines of the Admissions Scheduling and Control System (ASCS) developed by Hancock et al. [5] are used in this model. Simulation studies have predicted that hospitals using this system would operate at occupancies in excess of those found in most hospitals, and implementation of the ASCS in several hospitals has shown these predictions to be realistic [6,7]. A facility using the ASCS should require fewer beds to meet demand and thus should operate at a lower cost. A model using the ASCS should predict the optimum number of beds needed to operate a hospital.

In the ASCS, admissions to the hospital are classified in three categories: emergency, scheduled, and call-in. An emergency patient is defined as one requiring immediate admission to the hospital. Thus an emergency admission is uncontrollable and may be considered a random event. Scheduled and call-in patients are elective and thus do not require immediate admission; they may be put on a waiting list or scheduled for admission at some future date. The admission of a scheduled patient is planned for a specific time in the future, and a call-in patient is called in for admission at the hospital's convenience. Thus, if at some time during the day the hospital has beds available, patients are called in.

A turnaway is defined as an emergency patient who cannot be accommodated in the normal manner because all hospital beds are full. A cancellation is defined as a scheduled admission that must be cancelled in order to save room for emergency patients who may arrive before the next day's discharges. Thus a scheduled admission is cancelled in order to prevent the possibility of an emergency turnaway later in the day. In practice, cancelled patients are rescheduled at the next open date or are called in with the highest priority. In this study, turnaways were constrained to be between 1 and 3 percent of all emergency arrivals and cancellations were constrained to be between 1 and 3 percent of all scheduled admissions. These percentages were chosen because they appeared to be acceptable to the hospitals that had become aware of the ASCS.

At some point during the day when all discharges are known, the hospital must decide if it is necessary to call in patients or to cancel any scheduled admissions. This decision is based on the census reduction allowance (CRA) and the cancellation allowance (CA). The CRA and the CA represent the upper and lower bounds on the number of beds left empty at the decision point and may be different for each day of the week. The number of filled beds includes those to be occupied by patients scheduled for admission later in the day. If the number of empty beds is greater than the CRA, patients are called in until

Fig. 1. Daily census with and without call-ins.



the number of empty beds equals the CRA. If the number of empty beds is less than the CA, scheduled admissions are cancelled until the number of empty beds equals the CA. Thus the census at the decision point is always between the CRA and the CA. The overall effect of these allowances is to reduce significantly the variance in the hospital census and thus increase the attainable occupancy while maintaining a given turnaway level.

Reduction of census variance through use of the call-in algorithm is the mechanism that allows the ASCS to achieve high average occupancies. This census variance reduction is illustrated in Fig. 1, which shows two separate simulation runs in which the numbers of scheduled and emergency patients admitted are equal. It is apparent that a facility using the call-in algorithm will operate at a higher average occupancy and with lower census variance than a unit without call-ins. In addition, a facility using the call-in algorithm will be at its bed capacity (200 beds in Fig. 1) much more often than a facility without call-ins.

Probability Distributions of Emergency Arrivals and Patient Lengths of Stay

In order to simulate the randomness of a hospital, it is necessary to assign a probability density function to emergency arrivals and to patient length of stay. The Poisson distribution is used here to model emergency arrivals. This has been done often and has been found to be a good fit to empirical daily emergency arrival distributions [8,9] as well as being theoretically appealing. In the simulation model, a day is divided into two periods. The first period extends from the time when discharges begin to the time when all discharges are known (the decision point). The second period extends from the decision point to the time when the next day's discharges begin. For simplicity, it is assumed that the distributions of emergency arrivals in each period are identical. The assumption should have little effect on the results. The percentage of emergency arrivals, a variable closely allied to the overall mean emergency arrival rate, is one of the parameters whose effects are investigated in this study.

Assigning a probability distribution to inpatient length of stay (LOS) is difficult. The nature of these distributions varies greatly among hospitals, services, and days. Some researchers have used negative exponential, lognormal, and gamma distributions to model LOS [7,8,9]. Others report that no distribution fits their LOS data [7]. One possibility is to use an empirical distribution from a specific hospital, but that idea is discarded here in an effort to keep the model as generally applicable as possible. The use of a distribution described by a mathematical function makes it easy to produce probability density functions with given means and variances. This in turn allows LOS to be varied as a parameter.

To model LOS, the lognormal distribution, which has been found to be a good approximation to empirical LOS distributions, is used [10]. To justify its use, 140 LOS distributions from five different hospitals were examined. The use of the lognormal distribution appeared to be reasonable, and it was concluded that a variance of about seven times the mean is a good approximation. In this study, mean LOS is assumed to be somewhat higher for emergency patients than for elective patients.

Because of computer capacity restrictions, the LOS distributions were truncated at 50 days within the simulation, but this had little or no effect on the results because the LOS distributions had low frequencies at 50 days. (All means and variances refer to the truncated distributions.)

Simulation Runs

As stated previously, the effects of several parameters on the maximum average occupancy of a facility were examined. Maximum average occupancy is the highest average occupancy that can be achieved by searching for the best choices of the CRA and CA for each day of the week. Initially, the following paramaters were investigated: number of beds, percent of emergency admissions (percent EMG), and percent of elective patients scheduled.

During the first phase of the study, patients were scheduled five days a week (to reflect the scheduling policy of most hospitals) and emergency and elective LOS means were 10.38 and 8.39 days, respectively. These lengths of stay are typical of those found in midwestern general hospitals.

The number of beds was varied over 40, 80, 120, 160, and 200. At each level the percent emergencies and percent scheduled were varied as shown in Table 1 (p. 280).

Table 1. Parameters Varied for Simulation of Hospital with Five-Day Scheduling to Determine Maximum Average Occupancy at 40, 80, 120, 160, and 200 Beds

Phase 1		Phase 2		
% EMG	% scheduled	Mean emergency LOS	Mean elective LOS	% scheduled
30	40, 70, 90	5.1	4.3	40, 70, 90
50	40, 70, 90	10.4	8.4	40, 70, 90
66	40, 70, 90	14.9	13.1	40, 70, 90
95	0			

In the second phase of the study, the effects of mean LOS on maximum average occupancy were investigated. Percent EMG was held constant at 50 percent and patients were scheduled five days a week. Again, the number of beds was varied over the same five levels. Within each level, mean LOS and percent scheduled were also varied as shown in Table 1.

In the third part of the study, the effect of scheduling seven days a week rather than five was investigated. Percent EMG was held constant at 50 percent and the emergency and elective LOS means were 10.38 and 8.39 days, respectively. The number of beds was again varied over 40, 80, 120, 160, and 200, and percent scheduled was varied over 40, 70, and 90 percent.

A Fortran simulation program developed by Hamilton, Hancock, and Hawley [11] was used. This program embodies the rules of the ASCS and allows much flexibility in modeling different hospital and facility settings. The simulator has been documented and validated [5,11], but an attempt was made here to estimate the magnitude of possible errors involved in the simulation process.

The parameters percent emergencies and percent scheduled are not directly controllable through the input but are dependent on the number of call-ins that occur. Thus a search procedure is involved in which the inputs are varied until the proper levels of the parameters are approximated and the turnaway and cancellation constraints are satisfied. Once this is accomplished, the allowances are varied until the maximum average occupancy at the given parameter levels is determined. This procedure causes two possible sources of error.

It is impossible to set all of the parameters exactly at the specified levels. Percent emergencies and percent scheduled are generally within ±2 percent of the assigned level, and the error that this induces in maximum average occupancy may be estimated from the results using linear interpolation. For example, Fig. 2b shows that in a facility with 40 beds, a change of 30 percent in percent scheduled results in a change of 6.5 percent in maximum average occupancy. Thus an error of 2 percent in percent scheduled would cause an error

Fig 2. Maximum average occupancy vs. number of beds for different percentages of emergency arrivals and with 40, 70, and 90 percent of electives scheduled.

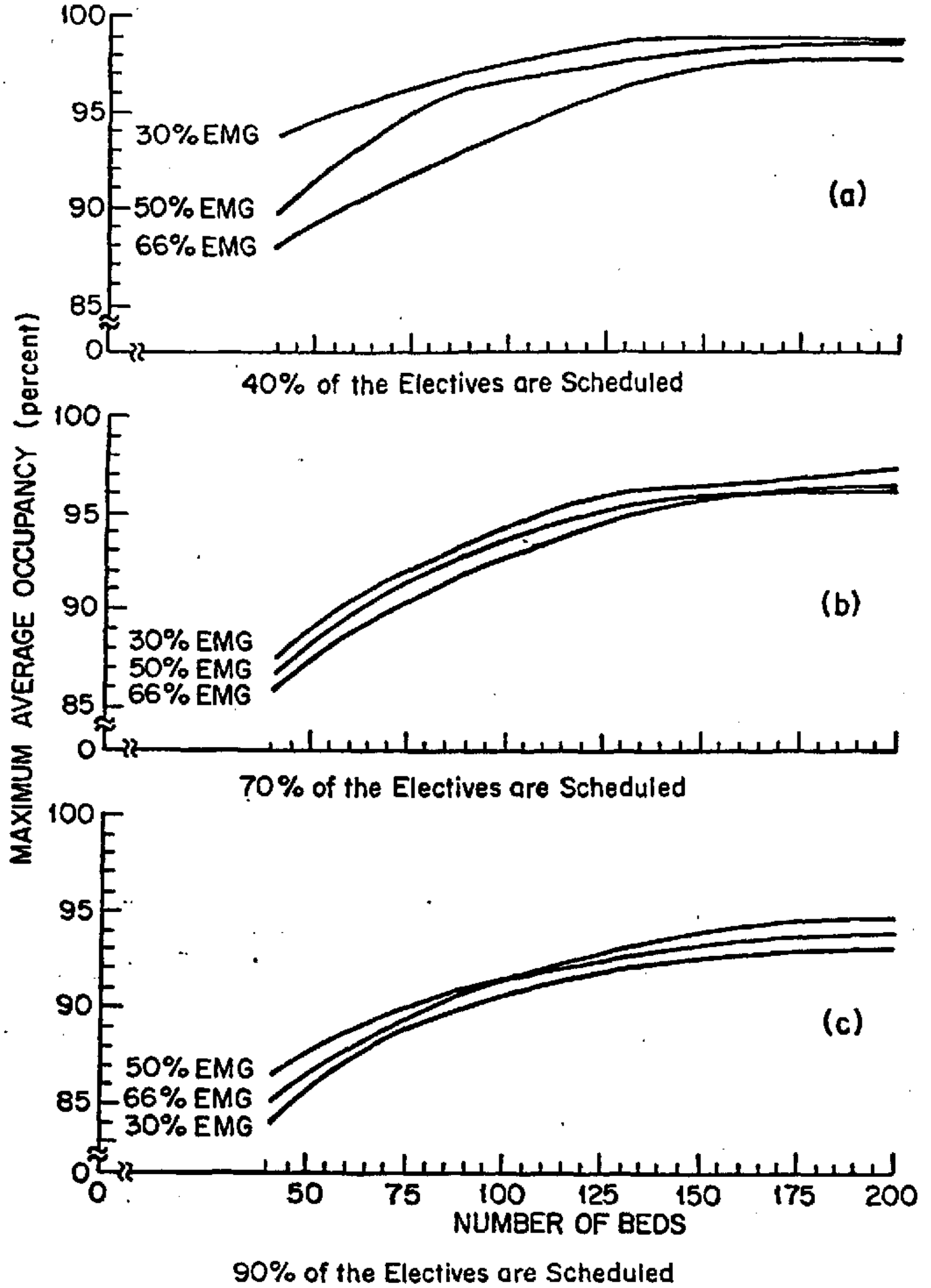
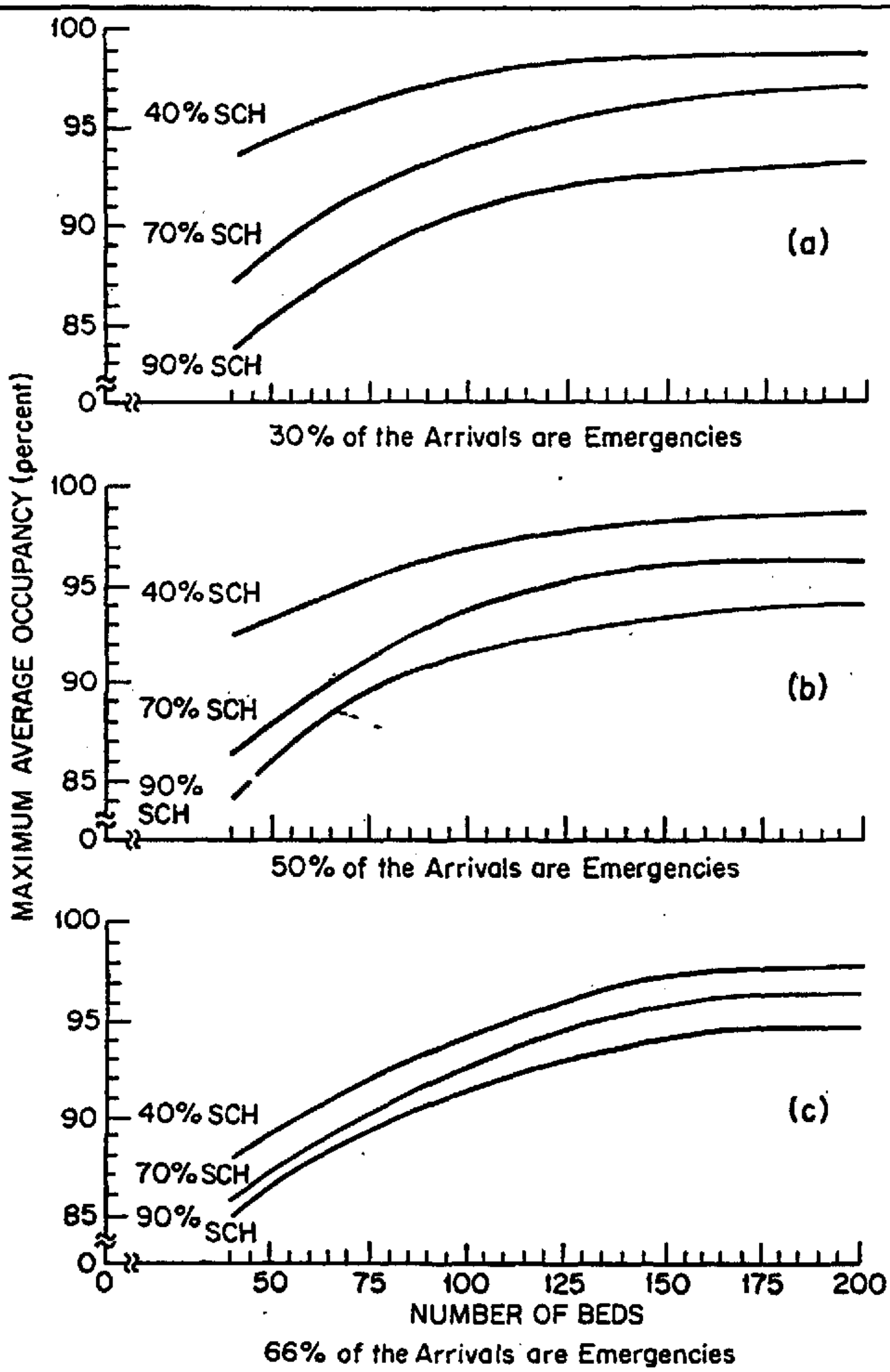


Fig. 3. Maximum average occupancy vs. number of beds for different percentages of scheduled arrivals and with 30, 50, and 66 percent emergency arrivals.



of $(0.02 \times 0.065/0.30) \times 100 = 0.43\%$ in maximum average occupancy. Similarly, for percent EMG (Fig. 3b, 40 beds), an error of 2 percent in percent emergency would cause the following error in maximum average occupancy for the 30 and 50 percent emergency levels:

$$(0.02 \times (0.937 - 0.897)/0.20) \times 100 = 0.40\%$$

These figures represent the maximum errors in occupancy. In general (at different parameter levels) the error will be much smaller because the differences in the numerator are smaller. The possible errors are easily calculated from the graphs for an individual case. (It is interesting that the error increases as the number of beds decreases.)

As with any simulation, there is a certain variability in results due to pseudorandom generation of numbers. In order to estimate the variance in occupancy within the simulation, 20 100-week runs were performed with identical inputs but with different seeds for the random number generators. The sample standard deviation in mean percent occupancy between runs was found to be 0.033 percent.

It is possible to estimate the total error in predicted maximum average occupancy due to the simulation. Assuming that the error due to pseudorandom number generation is normally distributed and using a 99-percent confidence interval, the maximum error in occupancy (percent) is

$$2.58 \times (\sigma_{EM}^2 + \sigma_{EL}^2 + \sigma_{RN}^2)^{\frac{1}{2}} \times 100$$

where σ_{EM} = the standard deviation of the error due to variation in percent emergency. The maximum error is already computed as 0.43 percent. Assuming the error to be normally distributed and using 99-percent confidence intervals

$$\sigma_{EM} \cong 0.43/(2.58 \times 100) = 0.00167$$

σ_{EL} = the standard deviation due to variation in percent electives. The maximum error is computed as 0.40 electives. Using the same assumptions as for σ_{EM}

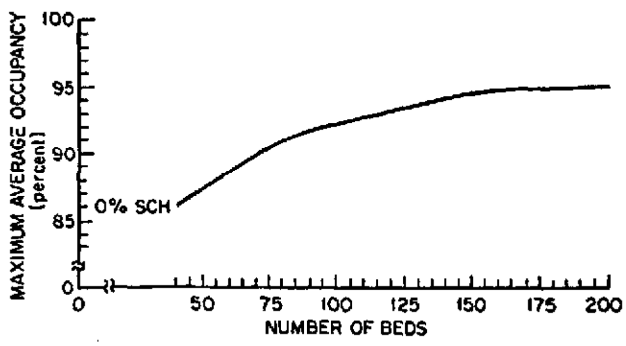
$$\sigma_{EL} \cong 0.40/(2.58 \times 100) = 0.00155$$

σ_{RN} = the standard deviation due to pseudorandom number generation. This has previously been found to be 0.00033. Therefore, maximum error in occupancy = $\pm 2.58 \times (0.00167^2 + 0.00155^2 + 0.00033^2)^{\frac{1}{2}} \times 100 = \pm 0.594$ percent occupancy. This figure, ± 0.594 percent occupancy, is an estimate of the maximum error due to the sources discussed above.

Results

The results of the procedures outlined are shown in Figs. 2–5, and the effects of the different parameters on occupancy are discussed

Fig. 4. Maximum average occupancy vs. number of beds with 95 percent emergency arrivals and all electives called in.



below. Perhaps it would be desirable to use Fig. 4 as a benchmark when examining the remaining figures since it is closest to the historical Poisson arrival process.

In all cases, maximum average occupancy increases with increasing facility capacity. The higher percent occupancy in larger facilities results from a decreasing coefficient of variation of census (standard deviation/mean) as the number of beds increases. When the coefficient of variation is small, the maximum average mean census may be closer to the facility capacity while still maintaining the turnaway and cancellation constraints.

For a given number of beds maximum average occupancy is observed to increase with decreasing percent scheduled. This is a result of the fact that as percent scheduled decreases, the number of call-ins is increased. Thus the ability of call-ins to reduce census variance is improved as percent scheduled decreases.

In this study when patients are scheduled five days a week, a lack of call-ins is felt most heavily on weekends. With no patients scheduled for Friday and Saturday, it is necessary to call patients in to maintain a high maximum average occupancy. If the call-ins are unavailable due to a high percent scheduled, a weekend drop-off in average occupancy occurs. (Remember that the ratio of call-ins to scheduled patients is constant.)

In most cases, maximum average occupancy is observed to increase as percent EMG decreases. This is because, in general, emergency arrivals introduce a greater variance in the census than elective arrivals. In certain cases, however, this is not true. In this study, when percent scheduled is 90 percent, the variance in census caused by elective admissions is larger due to the drastic weekend drop-off in average census. Thus, when percent scheduled is high and patients are scheduled five days a week, maximum average occupancy will *decrease* as percent EMG decreases and the percent of elective patients increases.

Fig. 5. Maximum average occupancy vs. number of beds with 50 percent emergency arrivals and 40, 70, and 90 percent of electives scheduled. Results of model are compared with Hill-Burton formula and Poisson model in (c).

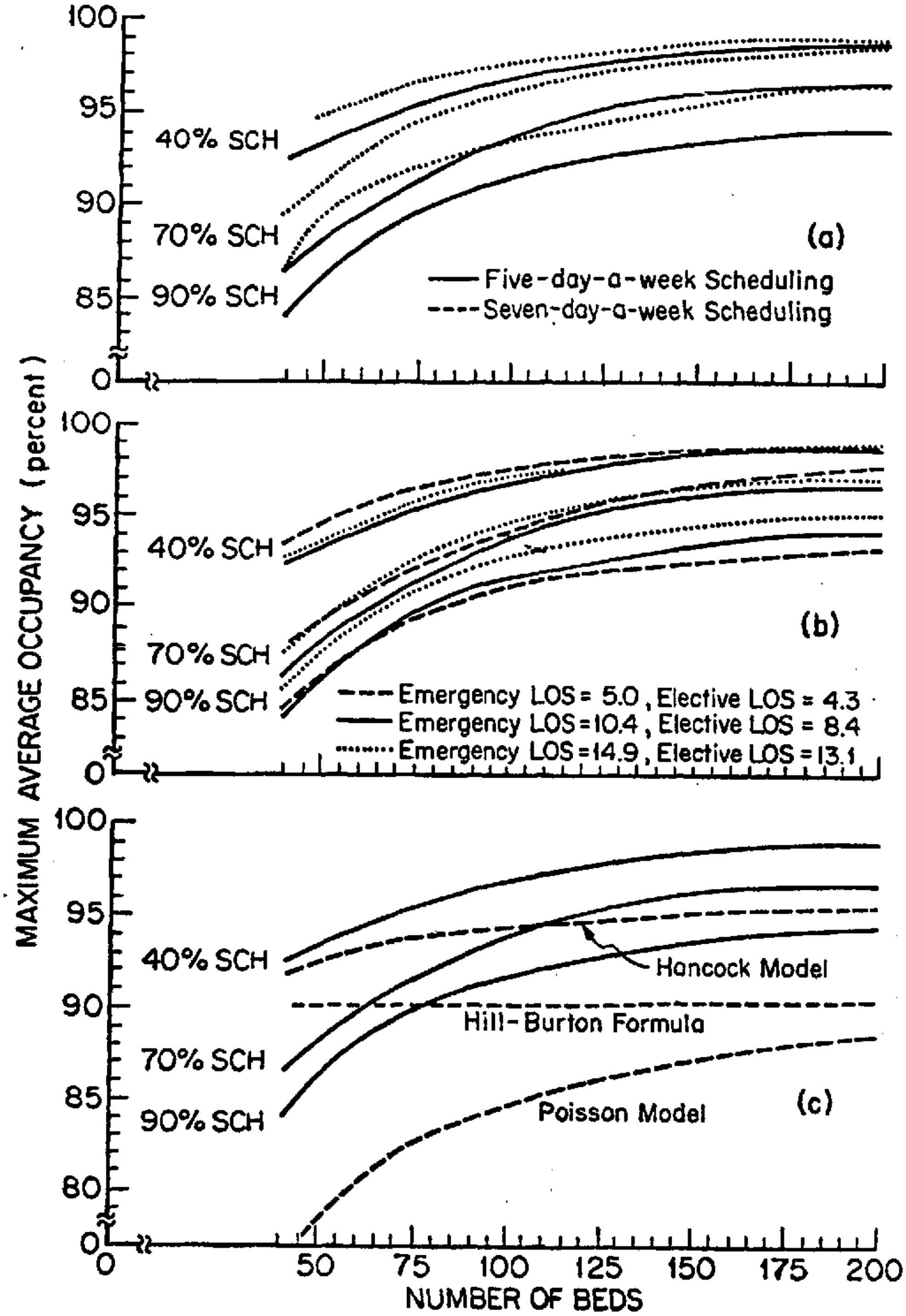


Table 2. Effect of Change in Scheduled Cancellation (SC) and Emergency Turnaway (ET) Rates on Maximum Average Occupancy

No. of beds	% EMG	% scheduled	Maximum average occupancy		Average no. SC + ET for 1-3% constraint
			SC and ET: 1-3%	SC and ET: 2/mo*	
320	30	30	99.3	98.7	8.3
320	30	90	94.6	90.6	25.6
320	90	90	96.1	90.6	21.9
80	30	30	97.6	96.5	3.2
80	30	90	90.7	87.3	5.4
80	90	30	90.8	85.0	4.7

* Fewer than two per month but as close to two as possible.

In general, if percent scheduled is below a certain level (approximately 80–90 percent) or if patients are scheduled seven days a week, maximum average occupancy will increase as percent EMG decreases.

It is advantageous (from a maximum average occupancy standpoint) to schedule seven days a week because no weekend drop-off then occurs. Figure 5a shows that when percent scheduled is high, the use of seven-day scheduling allows a higher maximum average occupancy than does the use of five-day scheduling. This is to be expected since the weekend drop-off for five-day scheduling has a greater effect when percent scheduled is high. Thus, when sizing a facility it is necessary to be aware of how patients are to be scheduled.

The results of varying mean LOS are shown in Fig. 5b, where it can be seen that mean LOS has a much smaller effect on maximum average occupancy than the other parameters do. In sizing larger facilities, the mean LOS will have little effect on the results, but the same cannot be said for small facilities (fewer than 40 beds).

As mentioned previously, the maximum average occupancy is affected by percent cancellations and turnaways. In Table 2 the maximum average occupancy with the 1–3 percent constraint on cancellations and turnaways is compared with the maximum average occupancy with the sum of the cancellations and turnaways being less than but as close as possible to 2 per month. For 320 and 80 beds, the constraint change results in a decrease in the cancellations and turnaways and the maximum average occupancy also decreases. For comparison, the sum of cancellations and turnaways for the 1–3 percent constraint is also given in average occurrences per month. For purposes of planning or determining the number of beds needed, the scheduled-cancellation and emergency-turnaway rates should be viewed as a matter of policy. Once agreement is reached on their acceptable level, then maximum average occupancies can be computed using the simulator. It should be noted here that the occupancy data for 320 beds are not strictly part of this research and are furnished only for their value in the sensitivity analysis of the cancellation and turnaway constraint.

Application of Results in Facility Sizing

The results of this study may be applied in determining the correct size of a hospital. The method is as follows.

A. Obtain an estimate of the average daily census (ADC) of the facility. For planning purposes, this is frequently obtained by taking the present ADC and adjusting for demographic factors over the planning horizon.

B. Estimate the parameters for the facility (percent EMG, percent scheduled, mean LOS, and scheduling pattern). The usual assumption is that the percentages will not change over the planning horizon.

C. Find the graph in Figs. 2–5 that is closest to the parameters estimated in B above.

D. Determine the necessary number of beds (NB) using successive approximations as follows.

1. Use ADC as an initial estimate of the number of beds, and determine a percent occupancy from the appropriate graph.
2. Determine the number of beds needed using NB = ADC/(% occupancy/100).
3. Use the number of beds found in step 2 to determine a revised percent occupancy.
4. Return to step 2 and compute a new NB using the revised occupancy of step 3.
5. Repeat steps 2, 3, and 4 until the bed-number estimates converge.

Use of the algorithm described above assumes, of course, that the ASCS system will be used to admit patients to the facility. The specific schedules, CRA and CA, which are specific for day of the week, can be quickly obtained using the admissions simulator. These values, of course, will vary for any point on the particular curve used.

As an example, consider a facility with average daily census = 180.0, percent EMG = 66, and percent scheduled = 70. To determine the optimal number of beds, find the occupancy estimate of 96.3 percent from Fig. 2b using number of beds = 180. Then,

$$NB = 180.0/(96.3/100) = 186.9$$

This rounds to 187 beds. Figure 2b gives the occupancy estimate of 96.5 percent for 187 beds. Thus the second estimate of beds is

$$NB = 180.0/(96.5/100) = 186.5$$

This again rounds upward to 187 beds, the sequence has converged, and the optimal number of beds of such a facility is 187.

The results of this study should not be extrapolated to small (fewer than 40 beds) facilities, which should be simulated individually since percent occupancy is extremely sensitive to the number of beds in the facility.

Discussion

These results may be used to determine occupancy factors for bed-planning methods. In comparison, other current planning methods, such as that of Shonick [4], the Hill-Burton formula, and the Poisson sizing assumption, overestimate the number of beds needed in most cases and are thus inappropriate. Shonick's methodology is the same as the one in this study, but his model is different, and his results cannot be put in the format used here. The lack of a call-in algorithm as a census restorer will cause Shonick's method to have substantially lower maximum average occupancies under identical cancellation and turnaway constraints.

In Fig. 5c the results of the present study are compared with the Hill-Burton formula, the Poisson assumptions, and the results obtained by Hancock, Martin, and Storer [12], who used a similar but less extensive approach. It is apparent that facilities can operate at occupancies much higher than those predicted by Shonick, the Poisson asssumptions, and the Hill-Burton formula. The exceptions occur in cases where percent scheduled is high and the number of beds is less than 75. In these cases the Hill-Burton formula predicts a somewhat higher occupancy than is possible. Both the Hill-Burton and Poisson models ignore important facility parameters that determine maximum average occupancy.

Figure 5c shows that the occupancy curve derived by Hancock falls in the same range as the results of this study, but its different shape is attributable to the fact that Hancock used different turnaway and cancellation constraints. His constraints were set at two cancellations and two turnaways per month. Thus, in small facilities, two per month represents a large percentage of arrivals, whereas in larger facilities the percentage becomes smaller. This explains the "flatness" of the curve derived by Hancock and also serves to point up the sensitivity of the turnaway and cancellation constraints mentioned earlier.

When sizing hospital facilities, all important parameters and characteristics of these facilities must be evaluated. All parameters must be considered collectively since their effects on occupancy are not independent. The results of this study may be applied to the sizing of individual facilities although it is important that factors not dealt with in this paper also be taken into account. Specifically, if the turnaway and cancellation constraints differ from those used here, one must expect the occupancy to differ as well. Other factors such as scheduling pattern, seasonal variations that cannot be smoothed by admission scheduling, and (to a lesser extent) mean lengths of stay should also be considered.

REFERENCES

1. Magerlein, D.B., W.M. Hancock, F.W. Butler, G.M. Mallett, and D.R. Young. New systems can mean real savings. *Hosp Financ Manage* 32:10 Apr 1978 and 32:18 May 1978.
2. Magerlein, D.B., R.J. Davis, and W.M. Hancock. *The Prediction of Departmental Activity and Its Use in the Budgeting Process.* Report No. 76-1, Bureau of Hospital Administration, University of Michigan, Dec. 1976.

3. Facilities Engineering and Construction Agency, DHEW. *Representative Construction Costs of Hospitals and Related Health Facilities*, p. iv. DHEW Pub. No. (OS) 73-6. Washington, DC: U.S. Government Printing Office, 1972.
4. Shonick, W. Understanding the nature of random fluctuations of the hospital daily census: An important health planning tool. *Med Care* 10:118 Mar.–Apr. 1972.
5. Hancock, W.M., D.M. Warner, S. Heda, and P. Fuhs. Admissions Scheduling and Control Systems. In J.R. Griffith, W.M. Hancock, and F.C. Munson (eds.), *Cost Control in Hospitals*, pp. 150–185. Ann Arbor, MI: Health Administration Press, 1976.
6. Johnston, C.M., W.M. Hancock, and D. Steiger. *The First Three Months of Implementation of the Admissions Scheduling and Control System at Delta Hospital*. Report No. 75-9. Bureau of Hospital Administration, University of Michigan, 1975.
7. Yannitelli, P.F. and W.M. Hancock. *Implementation of the Admissions Scheduling and Control System into Charles Hospital*. Report No. 75-3, Bureau of Hospital Administration, University of Michigan, 1975.
8. Handyside, A.J. and D. Morris. Simulation of emergency bed occupancy. *Health Serv Res* 2:287 Fall-Winter 1967.
9. Young, J.P. Stabilization of inpatient bed occupancy through control of admissions. *Hospitals* 39:41 Oct. 1, 1965.
10. Balintfy, J.L. *Mathematical Models and Analysis of Certain Stochastic Processes in General Hospitals*. Doctoral dissertation, Department of Engineering, Johns Hopkins University, 1962.
11. Hamilton, R.A., W.M. Hancock, and K.J. Hawley. *The Admissions Scheduling and Control System: The Admissions Simulator*. Report No. 75-1, Bureau of Hospital Administration, University of Michigan, 1974.
12. Hancock, W.M., J.B. Martin, and R.H. Storer. Simulation-based occupancy recommendations for adult medical/surgical units using admissions scheduling systems. *Inquiry* 15:25 Mar. 1978.

#12

Estimating the Need for Additional Primary Care Physicians

By Anthony Hindle, Nicholas Dierckman, Charles R. Standridge, Harry Delcher, Raymond Murray, and A. Alan B. Pritsker

A systems approach is used to assess the primary health care delivery system in Indiana. The output (office visits) of primary care physicians is estimated and compared with the demand for their services. Indexes of demand, supply, cost, and need are derived and used to determine the additional number of primary care physicians needed in each area. The results of this study are being used to encourage graduating medical students to practice in areas in need of additional primary medical care.

Primary medical care has become an area of concern because of the apparent decrease in the number of primary medical care providers, especially general and family practitioners, in the United States and specifically in the state of Indiana [1–3]. A technique was developed in this research to assess the primary health care delivery system in Indiana and to provide information regarding the need for additional primary care physicians in different areas of the state.

Discussions between systems analysts at Purdue University and physicians at the Regenstrief Institute for Health Care revealed that in order for the model to be useful, its parameters would have to be estimated from actual data. Thus the modeling process was constrained by the availability of data, precluding, for example, consideration of the effect of financing mechanisms such as Blue Cross/Blue Shield, Medicare, and Medicaid. A discussion of these modeling issues can be found in Standridge, Pritsker, and Delcher [4].

Methodology for Assessing Primary Medical Care Service

Primary Medical Care Service Areas

Researchers in Indiana [5,6] have described the distribution of primary care services by using ratios of population per primary care

This research was supported by the Regenstrief Institute, a division of the Regenstrief Foundation, Inc., Indianapolis.

Address communications and requests for reprints to Charles R. Standridge, Assistant Professor, College of Engineering, Systems Division, University of Iowa, Iowa City, IA 52242. Anthony Hindle is a lecturer in the department of operations research, University of Lancaster; Nicholas Dierckman is an implementation analyst in St. Louis employed by Technicon Medical Information Systems Corporation; Harry Delcher is assistant professor of medicine, Emory University; Raymond Murray is professor and chairman of the department of medicine, Michigan State University; and A. Alan B. Pritsker is professor in the School of Industrial Engineering, Purdue University.

0017-9124/78/03029015/$02.00/0

physician in a county. However, a political unit like a county may not accurately represent a medical service area since it is possible for cities of substantial population to be located relatively close to a county line, thus providing medical care to residents of contiguous counties, or for physicians to be located within easy access of people from another county.

In this study, primary medical care service areas are defined as population centers and their environs, specifically, towns or cities with populations of at least 2,000 (the approximate number of people per primary care physician in the United States [7]) plus the people in the immediate environs of these towns or cities. The environs are defined as the area within a five-mile radius of the town or city since that is a reasonable distance to travel to a physician. If the boundary drawn around the city overlaps that of another city, a combined service area results. According to this definition, Indiana contains 79 primary care areas. The state's counties are grouped into 12 districts, each of which is divided into primary care centers and the residual rural area surrounding the centers (see accompanying figure, p. 292).

Estimating the Capacity of the Full-Time Primary Care Physician

A full-time primary care (FTPC) physician is defined as a general or family practitioner in the age group 35–39, which has the highest output in terms of visits per year for this specialty [8]. (Tables 1 and 2, p. 293.) Primary care physicians are considered to comprise nonfederal, office-based physicians, both medical and osteopathic, who are general and family practitioners, pediatricians, internists, obstetrician-gynecologists, or general surgeons. The output of one FTPC physician is defined as the number of primary care visits per year for a general or family practitioner in the age group 35–39. The output of physicians with other age and specialty characteristics is expressed as a percentage of the output of the FTPC physician. Visits per week and weeks practiced per year by an FTPC physician were computed from *Reference Data on the Profile of Medical Practice* (1972–74) [9–11] and were 183.7 and 47.7, respectively. The product of these numbers, which is the annual output of an FTPC physician, is 8,762 visits per year.

Potential demand is defined as the need for primary health care as perceived by the population, disregarding factors that would interfere with satisfaction of that need [12,13]. This is the demand the population would place on the providers of primary care if the cost of obtaining such care, both monetary and nonmonetary, were zero. Expressed demand is defined as the amount of medical services actually sought by the population. This demand is backed up by an ability and willingness to pay not only the cost of the health care itself but also other costs incurred in procuring the care [12–14].

The ideal population is defined as having full geographic accessibility and no economic barriers to primary care, which implies that all of its potential demand is expressed and all of its expressed demand is satisfied. The yearly demand rate for the ideal population is estimated from the yearly number of visits made by those with family

Primary care service areas in Indiana.

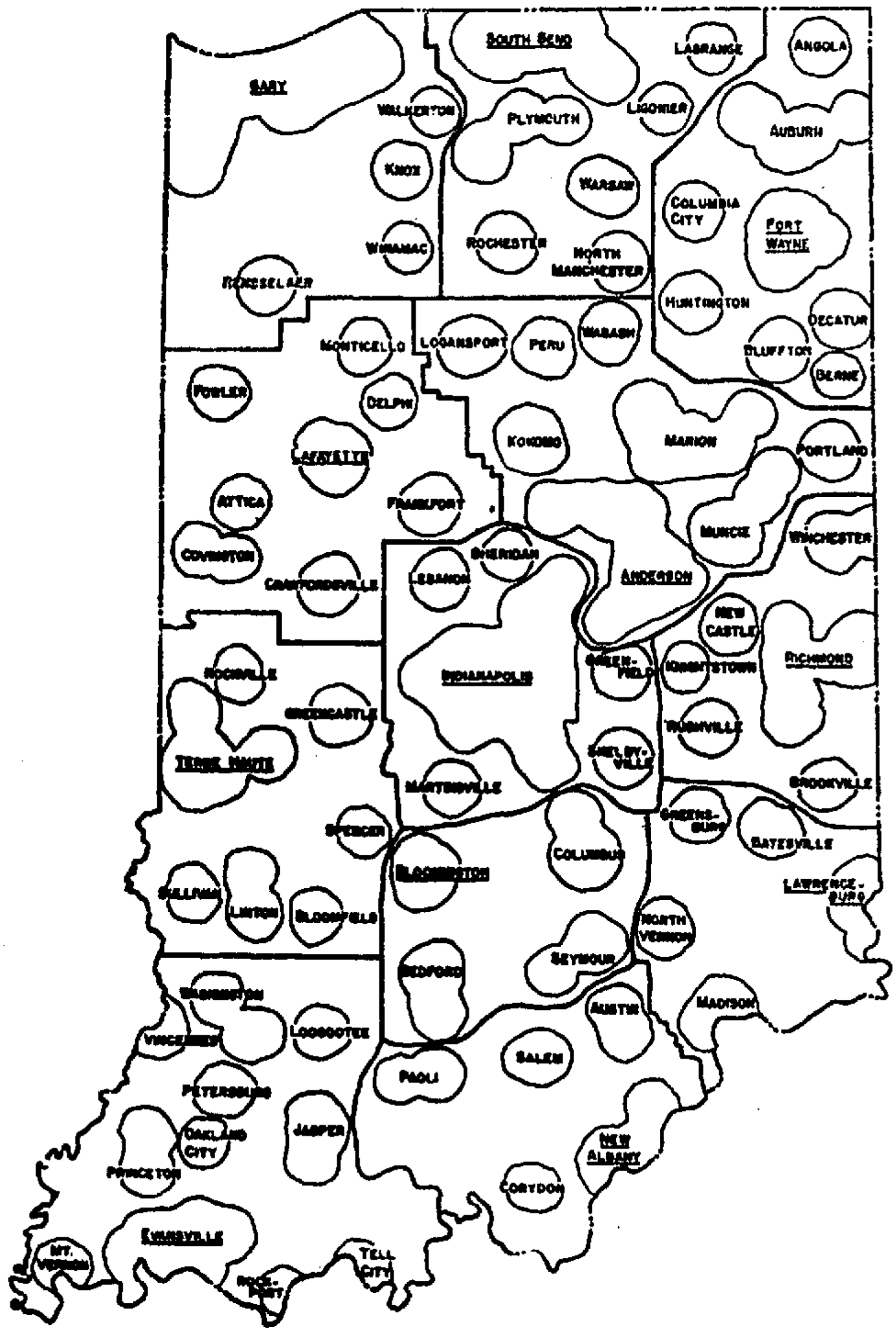


Table 1. Relative Number of Primary Care Visits for Different Specialties

Specialty	Visits/ day*	Weeks/ year†	Relative visits/ year
General and family	39.51	47.73	1.000
Internal medicine	20.96	47.20	0.524
Pediatrics	36.51	47.77	0.925
Obstetrics-gynecology	27.34	47.70	0.692
General surgery	18.73	47.03	0.467

* Estimated from Delcher, Raykovich, and Murray [19].
† Estimated from *Reference Data on the Profile of Medical Practice* [9–11].

incomes of $15,000 and over since it is assumed that this income group has all of its expressed demand satisfied and would be able to bear the full monetary cost of receiving care. In 1971 people in this income group visited all physicians at the rate of 5.1 visits per year, and, of these, 3.0 visits were to primary care physicians [15].

For the purposes of this study, the ideal population is also defined as having demographic characteristics similar to the population of Indiana and it is assumed that geographic accessibility to primary care (percentage of the potential demand that is expressed after allowing for the cost and inconvenience of traveling to a physician) for this group is the same as the average geographic accessibility for Indiana, which was established as 0.9375. The demand rate for primary care for the ideal population can then be calculated as 3.0/0.9375 or 3.20 visits. Therefore the yearly capacity of an FTPC physician is 8,762/3.20 or 2,700 individuals of the ideal population. With this information, an algorithm has been developed for obtaining indexes of primary care delivery for the previously defined service areas.

Table 2. Relative Number of Visits Attended by General and Family Practitioners of Different Age Groups
(Source: Standridge et al. [8])

Age group	Relative visits/ year
Under 35	0.812
35–39	1.024
40–44	0.958
45–49	0.892
50–54	0.825
55–59	0.759
60–64	0.693
65–69	0.626
Over ...	0.527

Algorithm for Determining Primary Care Service Index

1. Calculate the Equivalent Number of FTPC Physicians in Each Service Area. The visits attended by each physician can be expressed relative to those attended by the FTPC physician as a function of age and specialty [8–12,16–18]. For comparison among specialties (Table 1), the number of visits attended per day is estimated from the Delcher, Raykovich, and Murray study of Indiana physicians [19]. The weeks worked per year are estimated from *Reference Data on the Profile of Medical Practice* [9–11]. The days worked per week are assumed not to vary by specialty.

Similarly, the output of physicians can be compared among age groups. Standridge et al. [8] report relative values for general and family practitioners in Indiana (Table 2). Age and output are assumed to be independent of specialty because the available data do not allow estimation of the relationship between age and output as a function of specialty.

The output of each physician in a service area, expressed as a percentage of the output of the FTPC physician, is computed as the product of the relative number of visits of those in the specialty and the relative number of visits of those in the age group of that physician (see Tables 1 and 2). Then, the number of FTPC physicians in a service area is the sum of these products for all the physicians in that area.

Physician extenders, that is, physician assistants and nurse practitioners, are not included in the model. Standridge [1] estimated that these providers supplied less than 0.3 percent of the primary care visits in Indiana in 1975. However, it would be easy to extend the model to include these providers of primary care.

2. Compute a Geographic Accessibility Measure. Geographic accessibility is defined as the percentage of the potential demand that becomes expressed demand after allowing for the cost and inconvenience of traveling to a physician. The accessibility measure is derived by assuming that all the people in a service area are served by the physicians located in that area. The simplest case is the small, approximately circular, single-physician area. If the population is randomly located within the area, the minimum expected distance to the physician is given by locating the physician in the center of the area. The appropriate model for this situation was developed by Eilon, Watson-Grady, and Christofides [20].

The problem becomes more difficult if the area contains more than one physician location, but, for most areas, use of a nearest-location algorithm produces a set of polygons with the physician centrally located in each polygon. The area of the individual polygon is estimated as the area of the service area divided by the number of physician locations. Although a precise result for expected distance to the physician is not possible, a lower bound is given by assuming circular subareas. Experimental procedures and a sensitivity analysis led to the selection of this procedure [21]. Thus the expected Cartesian distance to the physician in miles is given by

$$0.667(A/(\pi L))^{\frac{1}{2}}$$

where A is the area of the service area in square miles and L is the number of physician locations in the service area. Expected distance was estimated for each service area allowing for the different types of practice—solo, partnership, and group [22]. A conversion factor, estimated to be 1.3 [22], was employed to convert expected distance into expected road distance.

Lacking information specific to Indiana, we assumed a simple linear relationship between distance and demand rate. Using data obtained from a study by Kane [23], we estimated the slope of this line and assumed that the relationship observed for rural Kentucky would also hold for rural Indiana because of similar topography. This set of assumptions led to the following equation for geographic accessibility in a service area:

$$ACCESS_j = 100 - 2.5D_j$$

where $ACCESS_j$ is the geographic accessibility in service area j, expressed as a percentage, and D_j is the expected road distance to physicians in service area j. The interpretation of the accessibility calculations is that $100 - ACCESS_j$ is the percentage of potential demand that does not become expressed demand because of distance to a physician.

3. Compute the Effect of Economic Status on Expressed Demand. A portion of the potential demand for primary medical care is deflected by the monetary costs of obtaining care. The economic feasibility of obtaining care is defined on the basis of restricted activity days per year, using the data and procedures of Weiss and Greenlick [24]. As shown in Table 3, persons in low-income families need more medical care than persons in families with higher incomes. It is assumed (and this assumption is borne out by the data in Table 3) that the higher a family's income, the larger the percentage of needs (potential demand) expressed as visits to physicians. It is further assumed

Table 3. Computation of Economic Feasibility Index

Income category (dollars)	Physician visits/person/year* (1)	Percent primary care visits* (2)	Primary care physician visits (3) = (1) × (2)	Restricted activity days/person/year† (4)	Primary care contacts/100 restricted activity days (5) = 100/(4) × (3)	Economic feasibility (6) = (5)/PC15 × 100
Under 3 000 ...	6.2	54.3	3.4	33.7	9.99	37.1
3 000–4 999	5.1	56.4	2.9	20.7	13.9	51.6
5 000–6 999	4.6	60.2	2.7	15.5	17.8	66.1
7 000–7 999	4.8	57.6	2.8	12.8	21.5	80.1
10 000–14 999 ..	4.7	59.8	2.8	11.9	23.8	88.5
15 000 and over .	5.1	59.2	3.0	11.3	26.9 = PC15	100.0

* Source: *Physician Visits, Volume and Interval Since Last Visit, United States—1971* [15].
† Source: *Disability Days* [31].

that all the primary medical care needs of those persons whose family incomes are $15,000 or greater are expressed as demands for physician services—that is, economic feasibility is indexed as a function of their satisfied demand for services (see Table 3 for calculations).

The following logarithmic function is used to calculate the economic feasibility index for each service area:

$$FEAS_j = a \times \ln(I_j) + b$$

where *FEAS* is the index of economic feasibility (expressed as a percentage) for service area *j*, I_j is average family income in service area *j*, and *a* and *b* are constants to be estimated by regression analysis. The values of *I* used were the midpoints of the income categories except for the highest category, for which $15,000 was used. Using the data in Table 3, regression analysis yielded values of 0.274 and −1.68 for *a* and *b*, respectively, with an R^2 of 0.955.

The interpretation of these calculations is that $100 - FEAS_j$ is the portion of the potential demand in a service area that does not become expressed demand because of the monetary cost of primary health care.

4. Compute the Potential Demand. Potential demand is the need for primary health care as experienced by the population, disregarding factors that could interfere with satisfaction of that need [12,13]. This is the demand the population would place on the providers of primary care if the cost of obtaining such care, both monetary and nonmonetary, were zero and, of course, assuming that the population is uniformly able to recognize conditions for which medical care is appropriate.

Potential demand is considered here as a function of two characteristics of the population: age and need for care [15, 25–29]. The ideal population is assumed to have the same age distribution as the state of Indiana. The average utilization rate for the ideal population of Indiana ($\bar{V}_o$) was computed from NCHS data [15] as 2.87 visits/person/year. This is the base of comparison for the age factor. A utilization rate for each service area ($\bar{V}_j$) was computed as the population-weighted average of the utilization rates of the age groups. Those utilization rates ($\bar{V}_a$) have been estimated from Indiana and U.S. data [15,30] and are presented in Table 4. Using utilization rates by age in the calculation of the potential demand requires the assumption that values for *ACCESS* and *FEAS* are invariant across age groups and that the percentage of expressed demand that is satisfied is the same for each age group.

A similar procedure is used for family income. By assuming that restricted activity or bed days, as discussed previously here and in ref. 24, are proportional to the need for primary care services in a population, it is possible to estimate the need for primary care in a service area in relation to the ideal population. The mean number of restricted activity or bed days for a person whose family income is $15,000 or over is 11.3 days/person/year [31]. This variable, labeled $\bar{B}_o$, is the base of comparison for the income factor.

Table 4. Primary Care Utilization Rates by Patient Age Group

Age group (a)	Visits/person/year ($\bar{V}_a$)
Under 18	2.29
18–64	2.95
Over 64	4.00

The mean number of restricted activity or bed days for the population of the service area $\bar{B}_j$ is determined as the population-weighted average of restricted activity days over the income groups. The mean number of restricted activity days by income group is estimated from data in refs. 31 and 32 and presented in Table 5.

These calculations for age and family income are used to determine the potential demand in numbers of the ideal population in a given service area by the following equation:

$$V_{0j} = N_j \times (\bar{V}_j/\bar{V}_0) \times (\bar{B}_j/\bar{B}_0)$$

where V_{0j} is the potential demand in units of the ideal population for service area j and N_j is the actual population in service area j.

5. Determine the Expressed Demand. All of the potential demand, however, will not be expressed as visits to physicians. Factors that may impede some people from making visits to physicians must be taken into account [13]. Two such factors, geographic accessibility and economic feasibility, are taken into account in the model and are assumed to act independently of one another. Thus the expressed demand in a service area, V_j, is obtained by multiplying the potential demand in that area, V_{0j}, by the accessibility and economic feasibility factors for that area as computed in steps 2 and 3:

$$V_j = V_{0j} \times FEAS_j \times ACCESS_j$$

Table 5. Estimated Restricted Activity Days of Patients in Different Income Groups

Income category	Estimated restricted activity days/person/year*
Under poverty level†	30.9
Over poverty level and less than $15 000	14.5
$15 000 and over	11.3

* Estimated from data in *Disability Days* [31] and *Census of Population: 1970* [32].
† Source: *Census of Population: 1970* [32].

6. Compute the Availability Index. Availability is defined as the percentage of the expressed demand that is satisfied by the suppliers of primary care, that is, that becomes satisfied demand. The latter is defined as the amount of medical services actually given by the providers of primary health care in response to population demand [15, 33]. An availability index is constructed based on a model of the services that flow across the service area boundaries. The basic assumptions of this model are:

1. If the physicians have spare capacity after all the expressed demands of the individuals in the same service area are satisfied, this spare capacity is made available to the adjacent rural area.
2. If people in a defined service area can receive better service in the rural area, they will make demands on the primary care physicians located there.
3. Persons do not cross district boundaries in order to obtain primary medical care services.

The availability index is defined as the ratio of the supply of visits available in the service area to the expressed demand. The supply available is the initial supply plus (minus) the amount of supply transferred into (out of) the service area. The availability index for a service area ($AVAIL_j$) is computed by the following formula:

$$AVAIL_j = (MD_j \times P)/V_j$$

where MD is the number of FTPC physicians in area j, P is the capacity of one FTPC physician in terms of numbers of the ideal population, and V_j is expressed demand in service area j (the total number of physician visits for the year). An iterative procedure was developed to find the number of FTPC physicians in service area j using the three assumptions described above and was applied to each of the districts independently.

7. Compute the Primary Care Service Index. The primary care service index for a service area ($PCSI_j$) is defined as the ratio of the satisfied demand to the potential demand—that is, it is the product of the expressed demand and the availability index, divided by the potential demand, expressed as a percentage:

$$PCSI_j = (V_j \times AVAIL_j)/V_{0j}$$

This index is thus the percentage of the potential demand that is satisfied.

8. Compute the Physician Requirement and Need Factors. The physician requirement factor, XMD, is the number of additional FTPC physicians that would be required in order to satisfy the expressed demand that is not currently being met:

$$XMD_j = (V_j/P) - MD_j$$

where V_j is the expressed demand in service area j (the total number of physician visits per year for the area, in numbers of the ideal population); MD_j is the total number of FTPC physicians in service area j; and P is the patient capacity of one FTPC physician in numbers of the ideal population. The physician requirement factor for each district is the sum of the requirement factors for the service areas in the district.

The need factor is the estimated average number of members of the ideal population who are beyond the capacity of each FTPC physician to serve but who would have to be served in order to satisfy that portion of the expressed demand that is not currently being satisfied. The need factor, XP_j, is expressed as follows:

$$XP_j = V_j/MD_j - P$$

The need factor for each district, XP_d, is

$$XP_d = (XMD_d \times P)/MD_d$$

where XMD_d is the number of additional FTPC physicians required to satisfy that portion of expressed demand that is not currently being met in district d and MD_d is the total number of FTPC physicians in district d.

Measures of Effectiveness of the Primary Medical Care System

Effectiveness indexes provide a measure of the capability of the primary care providers in each service area to meet the demand of the population of that area. To assess the performance of the primary medical care delivery system of a state, a policymaker needs statewide measures of effectiveness. Several measures of performance that can be derived from the indexes are discussed below.

Indexes Related to Minimum Levels of Service or Availability. Formulation of these measures of effectiveness is based on the assumption that the goal of the policymaker is that some minimum level of either the service index or availability index is obtained. The percentage of service areas that have index values below a defined minimum level is one such measure. Alternatively, a population-weighted average of index levels below the minimum level could be used.

Population-weighted Index for the State. This measure combines the indexes for the service areas by multiplying either the primary care service index or the availability index for a particular service area by the fraction of the state's population in that area and summing the products of all areas.

Physician Requirement Factor for the State. This measure is the number of additional FTPC physicians needed in the state to satisfy all of the expressed demand and is the sum of the requirement factors of the service areas.

The Analysis for Indiana

The procedure developed in this research is being used to evaluate the primary health care system of Indiana. It was first used in 1976 by the Board of Trustees of the Indiana Medical Distribution Loan Fund to identify areas in need of primary care physicians.

Table 6. Measures of Effectiveness of the Primary Medical Care System: Indiana, 1975

Measure	Value
Percent service areas with service index below 70%	43.9
Percent service areas with availability below 85%	30.8
Population-weighted service index (%)	69.6
Population-weighted availability index (%)	87.1
Statewide physician requirement factor	251

Data Sources for Physicians and Population

Data concerning primary care physicians in Indiana and their characteristics were obtained from the *Indiana Physician Profile* [7]. Data concerning the size and age distribution of Indiana's population were based on the *Indiana County Population Projections* [30]. The level of family income in each service area was derived from data in the 1970 United States Census [32].

Results for the State of Indiana

Table 6 summarizes the measures of effectiveness of the primary medical care system of Indiana. It shows that 43.9 percent of the service areas had a service index of less than 70 percent or that 43.9 percent of the service areas were meeting less than 70 percent of the potential demand. Furthermore, 30.8 percent of the service areas had physician availability indexes that were lower than 85 percent.

The population-weighted service index shows that less than 70

Table 7. Physician Requirement and Need Factors for Indiana Districts

District	Physician requirement factor	Need factor: ideal pop./physician
1. Gary	38.0	348
2. South Bend	4.5	59
3. Fort Wayne	35.4	622
4. Lafayette	8.7	261
5. Anderson	32.7	415
6. Terre Haute	21.4	748
7. Indianapolis	44.1	291
8. Richmond	6.6	226
9. Bloomington	3.3	97
10. Lawrenceburg	13.4	989
11. Evansville	28.9	610
12. New Albany	14.3	521

Table 8. Service Areas with Need Factors Greater than 900

Service area	Accessibility index	Economic feasibility index	Availability index	Primary care service index	Physician req. factor	Need factor
Lawrenceburg-rural	77	79	57	22	11.8	4 418
Fort Wayne—rural	83	64	44	31	17.6	3 502
Bluffton	91	84	44	33	2.9	3 302
Angola	92	79	46	34	2.3	3 060
New Albany—rural	79	78	49	30	10.1	2 753
Terre Haute—rural	80	80	52	33	11.3	2 428
Loogootee	69	79	54	38	1.0	2 291
Evansville-rural	77	80	54	33	12.7	2 291
Huntington	94	84	62	49	3.6	1 595
Mt. Vernon	93	79	63	46	1.3	1 527
Bloomfield	92	80	52	33	11.3	1 473
Gary-rural	84	80	66	45	9.2	1 379
Knightstown	89	81	67	48	0.8	1 320
Richmond-rural	81	81	67	44	5.8	1 320
Anderson-rural	83	83	68	47	9.8	1 215
Spencer	92	77	71	50	0.7	1 099
Linton	93	78	72	52	1.6	1 029

percent of the potential demand can be satisfied. The population-weighted availability index shows that 87 percent of expressed demand is satisfied. The value of the statewide physician requirement factor is 251, which shows that an additional 251 FTPC physicians are needed in order to meet the 13 percent of the expressed demand that is not satisfied. Since a full-time physician by definition provides the most visits per year of any physician, the requirement factor may be viewed as the minimum number of additional physicians required.

Results for Indiana Districts

The requirement factors and need factors for each of the 12 Indiana districts are presented in Table 7. The need factor reflects both primary care availability and the extra work each physician has to perform if he or she is to satisfy all of the expressed demand. Table 7 shows that 8 of the 12 districts have need factors of 10 percent or more of the capacity of an FTPC phyiscian (2,700) and four districts have need factors of 20 percent or more of this value. These results indicate that the demand for care placed on primary care physicians in dis... in various parts of Indiana significantly exceeds their capacity for providing such care.

Table 9. Cumulative Distribution of Service Area Need Factors by Percent of the Equivalent FTPC Capacity

Range (percent of 2,707)	Number of service areas	Percent of service areas
All	91	100.0
0	40	44.0
0–10	15	16.5
10–20	11	12.1
20–30	8	8.8
40–60	9	9.9
70–80	0	0.0
90–100	3	3.3
100–170	5	5.5

Results for Service Areas

Values for the geographic accessibility index, the economic feasibility index, the availability index, the primary care service index, the physician requirement factor, and the need factor were compiled for each service area. Table 8 (p. 301) shows those 17 areas with need factors greater than 900, which is one-third of the capacity of one FTPC physician.

Finally the distribution of the service area need factors in terms of percentages of the FTPC physician capacity (2,700) is shown in Table 9. The table shows that in order to meet all of the expressed demand in the state of Indiana, each physician in 36 of the service areas would have to carry a patient load at least 10 percent higher than his or her estimated maximum. In 17 of the service areas, each physician would have to care for at least 40 percent more patients than his or her maximum. Table 8 shows that an additional number of physicians equivalent to 114 FTPC physicians is required to meet the expressed demand in these 17 service areas. On the other hand, 40 service areas were found to have no need of additional primary care physicians.

Application of Results in Indiana

Beginning in April 1976, this research has been used to assist the members of the Board of Trustees of the Indiana Medical Distribution Loan Fund to determine the areas in Indiana in greatest need of additional primary care physicians. The fund is a loan-forgiveness program that supports medical students who agree, in exchange, to practice primary medical care in an area of need. Students funded by this program receive a list of the areas in Indiana in greatest need of additional primary care physicians, and medical graduates select practice locations jointly with the loan fund board of trustees. Indiana has about 300 medical graduates per year, and 95 students have participated in this program to date. Sixty-four of these are in postgraduate training and five are practicing physicians. The remaining 26 are still in medical school. It is hoped that this research will continue in helping to rectify the maldistribution of primary care physicians in Indiana.

REFERENCES

1. Standridge, C.R. *Primary Health Care Manpower Projections and Policy Assessment.* Master's thesis, School of Industrial Engineering, Purdue University, Aug. 1977.
2. *1970 American Medical Directory.* Chicago: American Medical Association, 1970.
3. *1973 American Medical Directory.* Chicago: American Medical Association, 1973.
4. Standridge, C.R., A.A.B. Pritsker, and H. Delcher. Issues in the development of a health manpower planning model. *Simulation* 31:9 July 1978.
5. Ackerson, S. and R. Murray. *Indiana Physicians and Rural Health.* Regenstrief Institute Report. Regenstrief Institute, Indianapolis, IN, Oct. 1974.
6. Boruff, J. Distribution of physicians. *ISBH Bull* 76:3 July 1972.
7. *Indiana Physician Profile.* Seven Area-Wide Comprehensive Health Planning Agencies of Indiana and Health Services Management, Inc., Indianapolis, IN, Dec. 1975.
8. Standridge, C.R., C. Macal, A.A.B. Pritsker, H. Delcher, and R. Murray. A Simulation Model of the Primary Health Care System of Indiana. In H.J. Highland, R.G. Sargent, and J.W. Schmidt (eds.), *Proceedings of the Winter Simulation Conference*, pp. 348–359, Gaithersburg, MD, Dec. 1977.
9. Center for Health Services Research and Development. *Reference Data on the Profile of Medical Practice, 1972.* Chicago: American Medical Association, 1973.
10. Center for Health Services Research and Development. *Reference Data on the Profile of Medical Practice, 1973.* Chicago: American Medical Association, 1974.
11. Center for Health Services Research and Development. *Reference Data on the Profile of Medical Practice, 1974.* Chicago: American Medical Association, 1975.
12. Donabedian, A. *Aspects of Medical Care Administration: Specifying Requirements for Health Care.* Cambridge, MA: Harvard University Press, 1973.
13. Wan, T. and S. Soifer. A multivariate analysis of the determinants of physician utilization. Paper presented at the 101st annual meeting of the American Public Health Association, Statistics Section, San Francisco, CA, Nov. 1973.
14. Aday, L. Economic and noneconomic barriers to the use of needed medical services. *Med Care* 13:447 June 1975.
15. National Center for Health Statistics. *Physician Visits, Volume, and Interval Since Last Visit, United States—1971.* Vital and Health Statistics Publications Series 10, No. 97. DHEW Pub. No. (HRA) 75-1524. Washington, DC: U.S. Government Printing Office, Mar. 1975.
16. A Profile of Primary Care Physicians. In *National Disease and Therapeutic Index*, Vol. 6, No. 1. Ambler, PA: IMS America, Ltd., June 1975.
17. Vahovich, S.G. Utilization of Physician Services. In S.G. Vahovich and P. Aherne (eds.), *Reference Data on the Profile of Medical Practice, 1973*, pp. 47–49. Chicago: American Medical Association, 1974.
18. Lorant, J. Physician Age in Relation to the Practice Characteristics and Performance. In R.J. Walsh, P. Aherne, and G.A. Ryan (eds.), *Reference Data on the Profile of Medical Practice, 1972*, pp. 89–94. Chicago: American Medical Association, 1973.
19. Delcher, H., T. Raykovich, and R. Murray. *A Survey of Indiana Physicians' Practice Characteristics and Task Delegation.* Regenstrief Institute Report. Regenstrief Institute, Indianapolis, IN, Mar. 1976.
20. Eilon, S., C. Watson-Gandy, and N. Christofides. *Distribution Management: Mathematical Modelling and Practical Analysis.* New York: Hafner, 1971.
21. Hindle, A. and N. Dierckman. *Manpower for Primary Medical Care in Indiana.* Regenstrief Institute Report. Regenstrief Institute, Indianapolis, IN, July 1975.
22. Dierckman, N.B. *Primary Medical Care in Indiana.* Master's thesis, School of Industrial Engineering, Purdue University, Aug. 1976.
23. Kane, R.L. Determination of health care priorities and expectations among rural consumers. *Health Serv Res* 4:142 Summer 1969.
24. Weiss, J. and M. Greenlick. Determinants of medical care utilization: The effect of social class and distance on contacts with the medical care system. *Med Care* 8:456 Nov.–Dec. 1970.
25. Aday, L. and R. Eichhorn. *The Utilization of Health Services: Indices and Correlates.* DHEW Pub. No. (HSM) 73-3003. Washington, DC: U.S. Government Printing Office, Dec. 1972.

26. Anderson, O. and R. Andersen. Pattern of Use of Health Services. In H.E. Freeman, S. Levine, and L.G. Reeder (eds.), *Handbook of Medical Sociology*, pp. 353–360. Englewood Cliffs, NJ: Prentice-Hall, 1972.
27. Anderson, O. and J. Newman. Societal and individual determinants of medical care utilization in the United States. *Milbank Mem Fund Q* 51:95 Winter 1973.
28. Andersen, R. and O. Anderson. *A Decade of Health Services*. Chicago: University of Chicago Press, 1967.
29. Bice, T. and K. White. Factors related to the use of health services: An international comparative study. *Med Care* 7:124 Mar.–Apr. 1969.
30. *Indiana County Population Projections, 1975–2000*. Prepared for the Indiana State Board of Health by the Division of Research, School of Business, Indiana University, Bloomington, IN, June 1976.
31. National Center for Health Statistics. *Disability Days*. Vital and Health Statistics Publications Series 10, No. 90. Washington, DC: U.S. Government Printing Office, 1974.
32. U.S. Bureau of the Census. *Census of Population: 1970*. Vol. 1, *Characteristics of the Population*, Part 16 (Indiana).
33. National Center for Health Statistics. *National Ambulatory Medical Care Survey: Background and Methodology*. Vital and Health Statistics Publications Series 2, No. 61. DHEW Pub. No. (HRA) 76-1335. Washington, DC: U.S. Government Printing Office, 1976.

#13

MANAGEMENT SCIENCE
Vol. 19, No. 12, August, 1973
Printed in U.S.A.

PHYSICIAN SUPPLY AND SURGICAL DEMAND FORECASTING: A REGIONAL MANPOWER STUDY*

A. REISMAN**, B. V. DEAN**, A. O. ESOGBUE†, V. AGGARWAL**, V. KAUJALGI**, P. LEWY‡ AND J. S. GRAVENSTEIN**

This paper discusses the methods and the results of a forecast for the demand for operative and obstetrical procedures and the supply of anesthesiologists in Cuyahoga County, Ohio. The techniques and results of a ten-year forecast for the demand for and supply of anesthesiologists and auxiliary personnel in greater Cleveland are discussed. Several regression models were used to forecast supply based on population, number of physicians, and the income per capita. The demand models were based on population, age and sex distribution projections, and historical data regarding operative and obstetrical procedures. The results of these "objective" models were then compared to forecasts under uncertainty generated by a panel of experts using the Delphi Method. Alternative states of health care delivery were investigated and implications for future anesthesiologist manpower requirements detailed.

1. Introduction

It is generally believed that medical manpower availability is a key element in the never ending struggle to solve the nation's health problems [13]. Many regions are facing severe shortages of health care professionals. There is general agreement that personnel qualified to administer anesthesia are badly maldistributed; some believe that even today many thousand anesthesia specialists are lacking. With the average age of anesthesiologists in the nation being forty-five years and an attrition rate of 30% over the next 10 years, a definition of what constitutes a properly balanced supply and demand becomes important. In Cleveland a group of concerned anesthesiologists[1] addressed itself to the problem of current shortages in anesthesia manpower. Under a grant from the U. S. Public Health Service,[2] the present authors in cooperation with the Cleveland Committee on Anesthesia Manpower critically investigated the current and future status of the anesthesia manpower situation in the county. Specifically, the problems of supply and demand of anesthesiologists in Cuyahoga County, Ohio up to 1980 were investigated and models were developed to analyze these problems.

2. Supply Model

The number of anesthesiologists in Cuyahoga County in the year 1980 can be forecast by considering different supply models. These models are essentially regression models. Parameters for these models were estimated using data from 1963–1969. In the study by Dougherty [4], a set of regression models for the supply of physicians in the state of Arkansas was developed. The models proposed here correlate the number of physicians to the population and per capita income, with the data obtained from the

* Received June 1972; revised October 1972.
** Case Western Reserve University, Cleveland, Ohio 44106.
† Georgia Institute of Technology, Atlanta, Georgia 30332.
‡ Balliol College, Oxford, England.
[1] Directors of Anesthesia Departments in nine major Cleveland Hospitals.
[2] Contract Nos. NIH 70-4033 and NIH 71-4022 (Anesthesiology Manpower Problems) for the period January 1, 1971 to March 31, 1972.

nation and the state. Since our work was confined to supply predictions for Cuyahoga County, two translateral prediction approaches were possible. One could either relate national figures first to the state and then to the county or relate the state figures directly to the county. A basic hypothesis in this study was that a relationship exists between the numbers of anesthesiologists, surgeons, and physicians in general.

A. *Model I*

This model considered population and per capita income as the most important factors which affect the number of physicians in a given region. Historical data on the number of anesthesiologists in Cuyahoga County were not available. However, the ratio of the number of anesthesiologists to the number of physicians was approximately constant for the state of Ohio from 1963 to 1969. The average ratio was 0.038, with a range from 0.036 to 0.040. We therefore accepted 0.038 as a reasonable ratio for our projections and treated Cuyahoga County as a microcosm of the state. In Model I the number of physicians in any year was assumed to be related to per capita income and population exponentially. Specifically, if $P(t)$ = number of physicians in year t, $S(t)$ = population of the county in year t and $I(t)$ = per capita income of the county in year t then $P(t) = A_1 S(t)^{B_1} I(t)^{C_1}$, where $t = X$-1900 (X = 1960, 1961, 1962, $\cdots$) and A_1, B_1, C_1 are equation constants. Note that both population and per capita income were assumed to be related to year t. Choice of base year as 1960 is irrelevant and does not change the results. $S(t) = a_1 t^{b_1}$ and $I(t) = a_2 t^{b_2}$ where a_1, b_1, a_2, b_2 are equation constants.

B. *Model II(A)*

This model predicts the supply of physicians, assuming an exponential relation between the ratio of per capita income of the county and the per capita income of the nation. If

$P(t)$ = number of physicians/1000 population,
$I(t) = I_c(t)/I_n(t)$,
$I_c(t)$ = per capita income of the county in year t,
$I_n(t)$ = per capita income of the nation in year t,

then $P(t) = A_2 I(t)^{B_2}$, A_2, B_2 are equation constants. It was assumed that $I(t)$ is exponentially related to year t, $I(t) = A_4 t^{b_4}$.

Now if $D(t)$ = number of physicians in the county in year t, and $A(t)$ = (population of the county) $\times 10^{3}$ then $D(t) = P(t) \cdot A(t)$.

The population of the county was assumed to be linearly related to year t, as $A(t) = a_3 t + b_3$, where a_3, b_3 are equation constants.

C. *Model II(B)*

A similar model can be built where the estimation of the supply of physicians was from the state to the county which is called Model II(B). If

$P(t)$ = number of physicians/1000 population,
$I(t) = I_c(t)/I_s(t)$,
$I_c(t)$ = per capita income of the county in year t,
$I_s(t)$ = per capita income of the state in year t,

then, $P(t) = A_3 I(t)^{B_3}$ and $D(t) = P(t) \cdot A(t)$.

The remaining symbols are similar to those used in Model II(A).

D. *Model III*

In this model the supply of physicians was assumed to grow exponentially with year t. If $P(t)$ = number of physicians in the county in year t and $t = X$-1900 (where

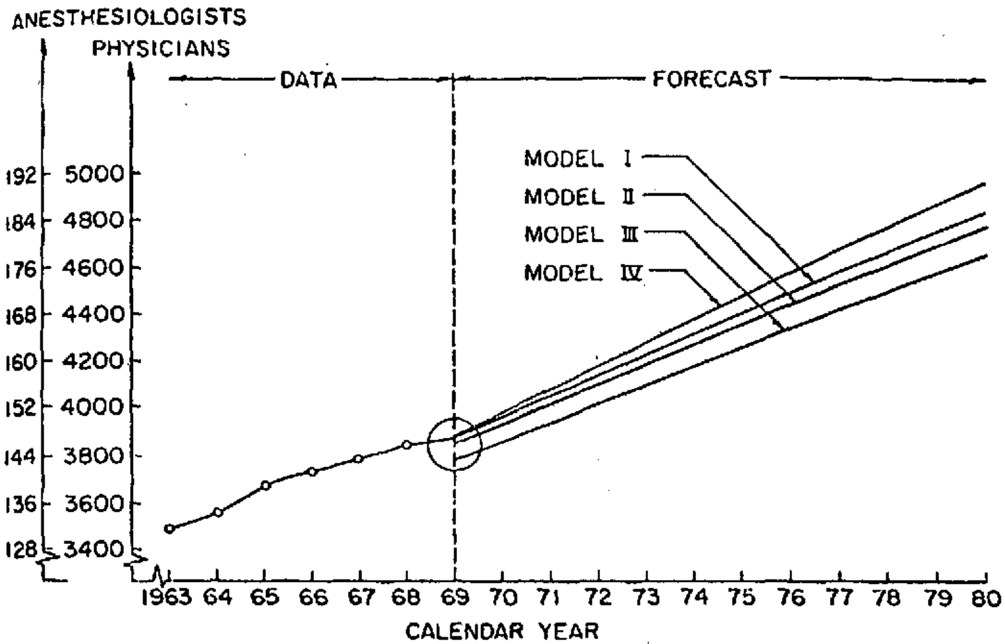


FIGURE 1. Supply of Anesthesiologists/Physicians in Cuyahoga County up to 1980.*

X = 1963, 1964, 1965, $\cdots$) then $P(t) = A_4 e^{B_4 t}$ where A_4 and B_4 are equation constants.

The detailed analysis of the models is reported by the present authors in [6]. The results of these models are shown in Figure 1.

It can be seen from the graph that Model III forecasts 182 anesthesiologists, the highest, and Model II (A) 164 anesthesiologists, the lowest estimate. The predictions of Models II (B) and I with 170 and 177 anesthesiologists fall within these values. The estimation based on time gives the highest, those based on per capita income, the lowest value.

According to these models, in 1980 there will be a 12% to 24% increase in the number of anesthesiologists since we counted 147 in 1969 in Cuyahoga County. The average annual rate of increase in the number of anesthesiologists is 2 to 3 anesthesiologists per year.

3. Age-Sex Specific Demand Model

In order to effectively plan for the supply of manpower and other resources in any professional discipline, it is necessary to have a means of assessing the future demand of the variables of concern. Manpower demand in any discipline over a period of time depends on demography, industrial mix, manpower utilization, and economic as well as social factors operative in a given community. Since these factors are subsumed in a model predicting incidence of operative procedures in any area, any meaningful projections of the demand for anesthesiology services in Cuyahoga County would normally include elaborate consideration of the demography of the area. This part discusses the development and application of a demand model for anesthesia personnel based on the demand for surgical procedures. In particular, an age-sex specific model is considered.

* The apparent discontinuity at 1969 occurs because different regression models were used on the historical data from 1963–1969.

Let the population be grouped by sex and age distribution, where the age groups 0–4, 5–9, $\cdots$, 80–84, 85 years and over are represented by $j = 1, 2, \cdots, J - 1, J$ respectively. A total of 18 age groups were considered. The surgical procedures were similarly classified using the standard 16 international codifications (i.e., Neurosurgery, Opthalmology, etc.). The following matrix of demand model was then developed.

Let A^m be the matrix of surgical procedures for males and A^f be the corresponding matrix of surgical procedures for females. This distinction is necessary since past studies, [1] for example, indicated differences in demand indices between the sexes. The cells of the matrix $[a_{ij}^m]$ then give the coefficient of demand and represent the number of procedures of type i performed on males in age group j in Cuyahoga County in 1970. Similarly, cells $[a_{ij}^f]$ represent the number of procedures of type i performed on females in age group j in Cuyahoga County in 1970. In performing the calculations we must consider all values of i and j where $i = 1, 2, \cdots, 16, j = 1, 2, \cdots, 18$.

Let b^m be a vector giving the distribution of males in Cuyahoga County in 1970 and b^f be a vector giving the distribution of females in Cuyahoga County in 1970.

Thus, b_j^m = number of males in age group j in Cuyahoga County in 1970, and b_j^f = number of females in age group j in Cuyahoga County in 1970.

Similarly let d^m, d^f denote the vectors of projected distributions for males and females respectively in Cuyahoga County in 1980.

Define a vector p^m as the ratio of the projected male population by age group in 1980 to that in 1970 in Cuyahoga County. Similarly define p^f for females. Thus,

$$p_j^m = d_j^m / b_j^m \tag{1}$$

and

$$p_j^f = d_j^f / b_j^f, \qquad j = 1, 2, \cdots 18. \tag{2}$$

We must then compute a vector $(X) = (X^m, X^f)$ defined as

$$(X^m) = [A^m](p_j^m), \tag{3}$$

$$(X^f) = [A^f](p_j^f), \tag{4}$$

which was obtained by direct matrix multiplication.

It is clear that (X^m) gives the projected demands for surgical procedures by males in Cuyahoga County in 1980, with (X^f) giving the corresponding projections for females. Thus, X_i^m gives the projected number of surgical procedures of type i to be performed on males in Cuyahoga County in 1980, and X_i^f the similar projections for females.

If we define N^m as the total number of surgical procedures required by males in 1980 and N^f the corresponding figure for females, then $N^m = \sum_{i=1}^{16} X_i^m$ and $N^f = \sum_{i=1}^{16} X_i^f$, hence $N = N^m + N^f$ is the projected total number of surgical procedures required in Cuyahoga County in 1980, if it is assumed that the current demand patterns continue till that time.

The data for matrices A^m and A^f (following pages) were collected from the QUEST division of Blue Cross of Northeast Ohio. The data cover 79% (estimated) of all surgical and obstetrical procedures performed in 1970.[3] This factor was used throughout the initial phases of the calculations and then adjusted accordingly to account for its partial coverage. The matrices A^m and A^f are shown in Tables 1 and 2 respectively.

The data for vectors b^m and b^f were obtained from [10]. The vectors d^m and d^f giving

[3] In Northeast Ohio.

the age-distribution of the population in Cuyahoga County in 1980 were not readily available from any known source. Consequently, estimates were obtained by comparing the population distributions of Cuyahoga County and the whole U.S.A. Data were collected from [8].

Two types of projections from this report, namely Series B and Series E, were employed. The basic difference between these two projections is due to the assumption under Series B projections of an average of 3.1 children per female in the population and for Series E, an average of 2.11 children per female. Employing these two projections leads to two types of results for the expected number of surgical procedures for both the male and female segments of the county.

Statistical tests were first applied to the null hypothesis: "The distribution of the population of Cuyahoga County, by age and sex in 1970 was the same as for the U.S. in 1970." The null hypothesis was accepted on the basis of Kolmogorov-Smirnov Nonparametric test and Chi-Square Goodness of Fit test.

The vectors p^m and p^f were calculated using relations (1) and (2), and they are shown in Table 3.

The total number of surgical procedures in 1970 in Cuyahoga County for males was 51,199 and for females (including obstetrics) 96,229. Hence the total number of procedures for both sexes was 147,428. The estimated demand for surgical procedures for males in 1980 is 57,666 by Series B and 55,941 by Series E projections. Similarly, the estimated demand for surgical procedures for females in 1980 is 116,041 by Series B and 114,606 by Series E. This estimation predicts an increase of from 9 to 13 percent in the surgical procedures for males and an increase of from 19 to 21 percent in surgical and obstetrical procedures required by females in the county. The results of this analysis also present data on the projected number of procedures for each surgical category up to the year 1980. These values are shown in Table 4.

The models thus provide us with the demand figures for Surgical Procedures for each category of surgery. While these figures are interesting in their own right, the greater aim of this study was the derivation of the number of anesthetics required by the population in the county in 1980. Consequently, to obtain this number, we proceed as follows:

Let α_i = the conversion factor giving the number of anesthetics required by the ith category of surgery.

Rational estimates for the values of α_i, $i = 1, \cdots, 16$, were obtained from [2] and are shown in Table 4. The total number of 132,575 anesthetics in 1970 was then computed. According to [7], there were 149 anesthesiologists in Cuyahoga County in 1970. This implies that 890 anesthetics were performed on the average by an anesthesiologist.

Using the above conversion factors and the precomputed projected number of surgical procedures, the number of anesthetics estimated by Series B forecasts in 1980 is 156,227. Under these assumptions and assuming identical workload in 1980 as in 1970, the estimated number of anesthesiologists N_A demanded in 1980 is given by

$$N_A = \text{Number of anesthetics in 1980} / \text{Number of anesthetics per anesthesiologist} = 176.$$

Similarly the number of anesthetics estimated by Series E in 1980 is 154,824. Hence the number of anesthesiologists demanded in 1980 is 174. Thus the assumption of 3.1 children per female in 1980 as opposed to a figure of 2.1 accounts for a difference of two anesthesiologists demanded in 1980 as provided by these models.

TABLE 1

$A^m = (a_{ij}^m)$: Matrix of Surgical Procedures and Age Groups for Males in 31 Hospitals of Cuyahoga County in 1970

Type of Surgical Procedures (i)	Age Group (j)																		$\sum_{j=} a_{ij}^m$
	(1) 0-4	(2) 5-9	(3) 10-14	(4) 15-19	(5) 20-24	(6) 25-29	(7) 30-34	(8) 35-39	(9) 40-44	(10) 45-49	(11) 50-54	(12) 55-59	(13) 60-64	(14) 65-69	(15) 70-74	(16) 75-79	(17) 80-84	(18) 85+	
1) Neurosurgical	61	45	30	29	26	37	40	51	63	73	88	100	64	62	35	29	14	6	853
2) Ophthalmalogical	221	179	102	66	64	56	51	43	74	103	147	174	191	208	172	167	98	34	2,150
3) Otorhino.*	854	1801	621	495	295	213	179	157	162	174	169	165	132	73	48	31	15	5	5,589
4) Thyroid etc.†	17	9	3	7	6	12	11	7	18	17	13	14	13	9	2	1	0	0	159
5) Vascular & Cardiac	214	103	61	85	81	71	126	272	541	801	818	762	493	299	149	118	52	27	5,073
6) Thoracic	27	5	8	40	41	39	36	31	41	55	82	85	98	90	47	45	26	13	809
7) Abdominal	759	385	382	352	375	296	296	353	468	633	772	826	701	522	380	291	152	62	8,005
8) Proctological	16	9	8	137	202	207	170	178	204	222	185	157	129	61	45	24	22	11	1,987
9) Urological	343	243	214	130	162	193	154	139	201	241	263	348	413	399	371	329	190	89	4,433
10) Breast	1	1	7	12	6	5	3	5	1	12	11	9	6	4	4	2	1	0	90
11) Gynecological	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12) Obstetrical	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13) Orthopedic	200	341	475	719	506	387	405	439	592	689	745	596	407	248	156	105	82	53	7,145
14) Plastic	196	168	212	297	231	177	166	165	173	217	205	212	147	84	87	46	25	12	2,820
15) Oral‡	62	26	31	39	53	40	40	46	44	40	32	23	26	20	10	4	6	3	545
16) Dental	14	25	24	40	64	59	43	49	58	63	51	49	34	23	14	14	8	2	635
Total = $\sum_{i=1}^{16} \sum_{j=1}^{18} a_{ij}^m$ =																			40,447

Total number of surgical procedures in Cuyahoga County for males = 40447/0.79 = 51,199 (Blue Cross Data included 79% of cases which had surgery).

* Otorhinolaryngology.

† Thyroid, Parathyroid, Thymus, Adrenals.

‡ Oral and Maxillofacial.

TABLE 2

$A^f = (a^f_{ij})$: Matrix of Surgical Procedures and Age Groups for Females in 31 Hospitals of Cuyahoga County in 1970

Type of Surgical Procedures (i)	Age Group (j) (1) 0-4	(2) 5-9	(3) 10-14	(4) 15-19	(5) 20-24	(6) 25-29	(7) 30-34	(8) 35-39	(9) 40-44	(10) 45-49	(11) 50-54	(12) 55-59	(13) 60-64	(14) 65-69	(15) 70-74	(16) 75-79	(17) 80-84	(18) 85 +	$\sum_{j=1}^{18} a^m_{ij}$
1) Neurosurgical	56	20	15	25	29	42	32	46	66	85	101	78	63	48	41	21	14	3	785
2) Ophthalmological	218	144	85	68	50	49	53	53	73	74	132	196	209	282	292	310	166	63	2,517
3) Otorhino.*	608	1637	765	706	416	249	162	137	159	162	132	106	67	67	44	19	8	3	5,447
4) Thyroid etc.**	15	8	10	10	36	44	34	35	50	78	75	46	31	29	13	4	2	1	521
5) Vascular & Cardiac	190	72	54	50	64	89	154	214	259	409	417	386	241	171	101	101	42	21	3,035
6) Thoracic	10	8	3	12	22	21	18	14	39	28	53	67	37	42	43	34	15	15	481
7) Abdominal	286	196	189	272	408	410	366	366	509	565	564	552	505	408	365	253	136	61	6,411
8) Proctological	13	4	10	101	115	86	111	113	135	139	139	94	86	67	48	36	13	9	1,319
9) Urological	161	76	29	43	71	77	98	86	96	101	134	112	88	98	65	36	21	14	1,406
10) Breast	4	0	6	117	193	134	144	196	282	333	196	137	115	77	78	53	31	14	2,110
11) Gynecological	16	23	47	428	1877	2158	1933	1833	2021	1871	1182	652	384	306	236	141	45	16	15,169
12) Obstetrical	0	0	132	4270	9964	7289	3291	1494	435	38	0	0	0	0	0	0	0	0	26,913
13) Orthopedic	213	221	325	289	257	200	254	260	424	516	604	598	432	358	293	277	215	180	5,916
14) Plastic	138	145	174	261	231	169	167	164	213	260	207	199	181	104	94	71	52	22	2,851
15) Oral***	43	14	21	30	33	33	16	27	24	41	32	19	12	17	20	9	2	1	394
16) Dental	10	19	29	75	102	67	55	62	46	52	55	46	32	28	21	15	4	2	720
Total = $\sum_{i=1}^{16} \sum_{j=1}^{18} a^f_{ij}$ =																			75,995

Total number of surgical procedures in Cuyahoga County for Females = 75,995/0.79 = 96,190 (Blue Cross data included 79% of cases which had surgery).

* Otorhinolaryngology.

** Thyroid, Parathyroid, Thymus, Adrenals.

*** Oral and Maxillofacial.

TABLE 3

Ratios of Projected Male and Female Population by Age Group in 1980 to that in 1970 in Cuyahoga County

Age Group	Males		Females	
	$(p^m)_B$*	$(p^m)_E$**	$(p^f)_B$*	$(p^f)_E$**
0–4	1.579	1.199	1.648	1.252
5–9	1.110	0.983	1.154	1.021
10–14	0.835	0.835	0.865	0.865
15–19	1.106	1.106	1.104	1.104
20–24	1.500	1.500	1.253	1.253
25–29	1.511	1.511	1.446	1.446
30–34	1.611	1.611	1.558	1.558
35–39	1.297	1.297	1.280	1.280
40–44	0.909	0.909	0.913	0.913
45–49	0.803	0.803	0.811	0.811
50–54	0.869	0.869	0.902	0.902
55–59	0.952	0.952	1.023	1.023
60–64	1.027	1.027	1.102	1.102
65–69	1.199	1.199	1.195	1.195
70–74	1.193	1.193	1.161	1.161
75–79	1.037	1.037	1.121	1.121
80–84	1.146	1.146	1.336	1.336
85+	1.243	1.243	1.358	1.358

* Subscript B refers to Series B of Projection of Population.
** Subscript E refers to Series E of Projection of Population.

TABLE 4

Demand for Surgical Procedures and Anesthetics in 1980 by Age Groups of Males and Females in Cuyahoga County

Type of Surgical Procedure	Males					Females				
	$(X^m)_B$	$(X^m)_E$	(a)	$(a)(X^m)_B$	$(a)(X^m)_E$	$(X^f)_B$	$(X^f)_E$	(a)	$(a)(X^f)_B$	$(a)(X^f)_E$
1) Neuro	952	923	0.9	857	830		858	0.9	795	772
2) Ophthalmalogy	2,446	2,339	0.75	1,835	1,754	2,965	2,860	0.75	2,224	2,145
3) Otorhino.	6,597	6,272	0.95	5,958	5,959	6,386	5,928	0.95	6,067	5,632
4) Thyroid, etc.	183	175	1.00	183	175	576	569	1.00	576	569
5) Vascular & Cardiac	5,211	5,117	1.00	5,211	5,117	3,233	3,148	1.00	3,233	3,148
6) Thoracic	912	901	0.97	884	874	540	535	0.97	524	519
7) Abdominal	9,022	8,685	1.00	9,022	8,685	7,316	7,177	1.0	7,316	7,177
8) Proctological	2,306	2,299	0.50	1,153	1,148	1,490	1,484	0.50	745	742
9) Urological	5,060	4,897	0.90	4,554	4,407	1,672	1,598	0.90	1,505	1,438
10) Breast	97	96	0.60	58	58	2,325	2,323	0.60	1,395	1,394
11) Gynecological	0	0	0.95	0	0	17,795	17,786	0.95	16,905	16,897
12) Obstetrical	0	0	0.97	0	0	35,356	35,351	0.97	34,295	34,290
13) Orthopedic	7,928	7,809	0.90	7,135	7,028	6,594	6,480	0.90	5,935	5,832
14) Plastic	3,255	3,159	0.50	1,628	1,579	3,245	3,171	0.50	1,622	1,586
15) Oral	659	632	0.95	626	600	461	442	0.95	438	420
16) Dental	716	708	1.00	716	708	835	829	1.00	835	829
Total*	45,556	44,193	—	39,820	38,921	91,672	90,539	—	83,600	83,390
Total in Cuyahoga County	57,666	55,941	—	50,405	49,267	116,041	114,606	—	105,822	105,557

* 31 Hospitals surveyed by Blue Cross.

4. Conclusions

Both the supply and demand models presented above are primarily projection models and the results can thus be compared. The demand for anesthesiologists under current practices has been predicted to be from 174 to 176 by 1980. The range for the supply of anesthesiologists in 1980 has been estimated to be from 164 to 182. The results for the supply level are somewhat less than the demand that is expected to exist for anesthesiologists in 1980.

The main limitation of the above models is that they neglect the effects of technological, legal, ethical and economic changes. It is, however, conceivable that these factors would influence the supply of anesthesiologists and the demand for surgical procedures and hence the demand for anesthesiologists in 1980.

To incorporate various states of nature that could conceivably occur ten years hence into the prediction for the desired level of anesthesia personnel, another model for forecasting demand under uncertainty was constructed, which is reported in detail in [6]. In that model, the subjective opinions of a panel of experts were used. The panel included individuals affiliated with a representative group of institutions or organizations, and having expertise in all areas that have a bearing on the problem under consideration. The panel was provided with factual data and information obtained from the supply models. Through several successive rounds of a DELPHI exercise, consensus of the experts in the health care area was reached on a set of questions concerning the future. The questions were concerned with the following: the attrition rate among anesthesiologists; the probabilities of occurrence of various futures over the next decade in regard to the quality of anesthesia care and the number of anesthetics per capita; the perceived requirements of anesthesia personnel with the change in the quantity of anesthesia care assuming no change in the population and the number of anesthetics per capita; and the possible uses of the new class of anesthesia personnel currently in training at Case Western Reserve University and elsewhere. Based on the results of the DELPHI certain predictions were made. It was observed that in 1980 there will be an estimated need for about 50 per cent more anesthesiologists, about 50 per cent more anesthesia residents in training, and a sufficient number of trained nonphysician personnel in anesthesia comprising a work force of about 70 in 1980. Through the DELPHI exercise, the constraints that contribute to the gap between current and "ideal" manpower levels in anesthesiology were also identified and ranked in order of their importance.

The DELPHI method predicted a demand for 1980 of 35 anesthesiologists more than the mathematical projections. A high demand would exist for anesthesiologists if an improved level of health care were provided by 1980. The DELPHI model considered this aspect explicitly and in fact reached a consensus on the figure of 0.57 as the probability for improved quality of health care with 10 per cent increase in the number of anesthetics per capita by 1980. The projection model, on the other hand, made the predictions under the current state of health care delivery, hence its predictions are lower than those obtained through the DELPHI.

ACKNOWLEDGMENT. The present work was supported under contract and grant, made available through NIH 70-4033 and NIH 70-4002. The authors are greatly indebted to Drs. J. Potter, N. DePiero, S. Katz, S. Kovacs, H. Kretchmer, E. Malia, B. Sankey, J. Smart, J. Viljoen; all members of the Cleveland Committee on Anesthesiology Manpower, and to Dr. M. Rhoten, for their guidance, cooperation, and assistance in conducting this study.

References

1. ADLER, G. R., "An Analysis of Patient Load in a Medium Size Community Hospital, M.S. Thesis, Technical Report E.1028.02, Department of Industrial Engineering, Syracuse University, June 1970.
2. Communications with Dr. J. Potter, Chairman, Department of Anesthesiology, Cleveland Clinic, Cleveland, Ohio.
3. DEPIERO, N. G. ET AL., "The Cleveland Pilot Program in Anesthesia Education," Marymount Hospital, Garfield Heights, Ohio 44125.
4. DOUGHARTY, L. A., "The Supply of Physicians in the State of Arkansas," Rand, RM-6385-APC, August 1970.
5. GRAVENSTEIN, J. S. et al., "Analysis of Manpower in Anesthesiology," *Anesthesiology*, Vol. 33, No. 3, (September 1970).
6. GRAVENSTEIN, J. S. et al. "Analysis and Forecasting of Anesthesia Manpower in Cuyahoga County, Ohio," Technical Memorandum No. 259, Department of Operations Research, Case Western Reserve University. March, 1972.
7. "Health Related Data," Northeast Ohio Regional Medical Program, Part II, Sections V, VI and VII.
8. "Population Estimates and Projections: Projections of the Population of the United States by Age and Sex, 1970 to 2020," U. S. Department of Commerce Service, P-25, No. 470 (November 1971).
9. REISMAN, A. et al. "Analysis and Codification of Pre- and Post-Operative Anesthesia Related Tasks," Technical Memorandum No. 229, Department of Operations Research, Case Western Reserve University (June, 1971).
10. "Report of Anesthesia Manpower Committee," New York State Society of Anesthesiologists' Committee on Total Anesthesia Care, 1970.
11. "Statistical Abstracts of Ohio—1969," Economic Research Division, Development Department, State of Ohio, 1970.
12. "Statistical Abstracts of the U.S.," U. S. Bureau of Census, 90th Edition, 1969.
13. STEWART, W. H., "Manpower for Better Health Services," *Public Health Report*, Vol. 81, No. 5 (May 1966), pp. 393–396.

A Spatial-Allocation Model for Regional Health-Services Planning

William J. Abernathy and John C. Hershey

Stanford University, Stanford, California

(Received August 12, 1971)

In planning for health services, the need arises to determine the location, capacity, and number of health centers for a geographically defined region. The present paper formulates this problem in a form convenient for solution and presents results from the model to clarify some important aspects of this allocation decision. The planning region is assumed to consist of geographically defined subareas or census blocks, of known location. The population is stratified in such a way that each stratum exhibits relatively homogeneous patterns of health-care utilization. The model characterizes the effects of center locations upon aggregate utilization and utilization of individual centers, and gives optimal locations of centers with respect to several alternative criteria. An example illustrating computational feasibility and the implications of various criteria for the location decision is presented.

THE PROBLEM OF determining spatial locations of primary health-care centers for a dispersed population is an important issue in primary health-care planning. This paper formulates the problem in terms of four different criteria. The importance of accepting census statistics as source data is recognized in the formulation and in the choice of optimization procedure.

The distance from individuals in a community to sources of care is a recognized barrier to utilization.[4,10,11] Several types of decay functions have been proposed to describe the inverse relation between distance to medical facilities and utilization of these facilities.[8,12] Recent research by the authors analyzes choice behavior among facilities when several facilities at varying distances are available.[2] It is demonstrated that the influence of distance upon choice behavior is much stronger than is the influence of distance upon overall utilization.

In addition, it is well known that demographic, socioeconomic, and health-status factors all influence individuals' medical-facility utilization. Strata that are differentiated by these factors show different utilization patterns as a function of distance.[2,4,11,13,16] That is, distance is a larger barrier to utilization for some strata than for others. Substantial progress has been made in understanding the complex interactions of distance and socioeconomic variables. However, a method for using these interactions to suggest alternative locations of medical facilities has not heretofore been developed.

A frequently stated goal of regional primary-health-care planners is to allocate facilities to locations in such a manner as to provide as much of the primary care demanded by the population as possible. Yet this may conflict with other goals. For example, if the health planner maximizes total utilization, centers may be placed closer to areas that contain large numbers of people for whom distance is a stronger barrier to utilization than to areas with populations that are more mobile. Such an allocation of centers may be far from optimal in terms of the goal of minimizing the average distance patients must travel for each visit.

The point is that alternative objectives in the spatial-allocation problem will often conflict. Any proposed locations of health centers must be studied in the light of each objective.

This paper presents a model that can assist a health-systems planner in determining the location and capacities of a specified number of health-care centers within a region in accordance with stated criteria. The formulation relies upon census data as input and recognizes different effects of distance upon strata identified within the population and the separate effects of distance upon the user's overall utilization and his choice of a particular center.

The process of making tradeoffs among multiple objectives is also explored. An example offering numerical experience with the model is given that demonstrates the manner in which alternative objectives dictate different spatial allocations of resources.

A MODEL OF UTILIZATION

SUPPOSE THAT a geographically defined medical-service area is delineated. This service area is assumed to consist of many geographically defined subareas—such as census blocks as defined in a federal census or special enumeration—about which L demographic data are available. Suppose that the medical-service area contains such census blocks.

The health planner has data about the demographic, economic, cultural, social, and/or health-status characteristics of individuals in the community. The community is stratified by these characteristics in such a way that each of M strata exhibits relatively homogeneous patterns of utilization.

We are particularly concerned about homogeneity with respect to two parameters of utilization behavior. The first, v, is the number of primary-care visits per time period that can be expected when a primary-care center is in immediate proximity to the individual. The second parameter, d, is a measure of the extent to which propensity to seek care decreases as distance to a primary-care facility increases. Both the strata and these two measures can be derived through analysis of regional utilization data or special survey information.[11,18]

Suppose the number of people in census block i and in stratum j is defined to be n_{ij}, where $i=1, \cdots L$, and $j=1, \cdots, M$. If the total population of the medical service area is N, then $\sum_{i=1}^{i=L} \sum_{j=1}^{j=M} n_{ij}=N$.

An example of a hypothetical medical service area with a total population of 50,000 is shown in Fig. 1. The horizontal and vertical axes are parallel to the general directions of the major highways in the area. This area is approximately 40 miles by 25 miles, and contains 32 census blocks. In Fig. 1, the number, border, and center of each census block are shown. Three small cities with populations of approximately 17,000 (City A), 9,000 (City B), and 13,000 (City C) are contained within the medical-service area. City A has its central location approximately at the point $(-10.0, 11.2)$; City B, located between the other two cities, is at $(2.7, -2.0)$; City C is at $(21.8, -3.3)$.

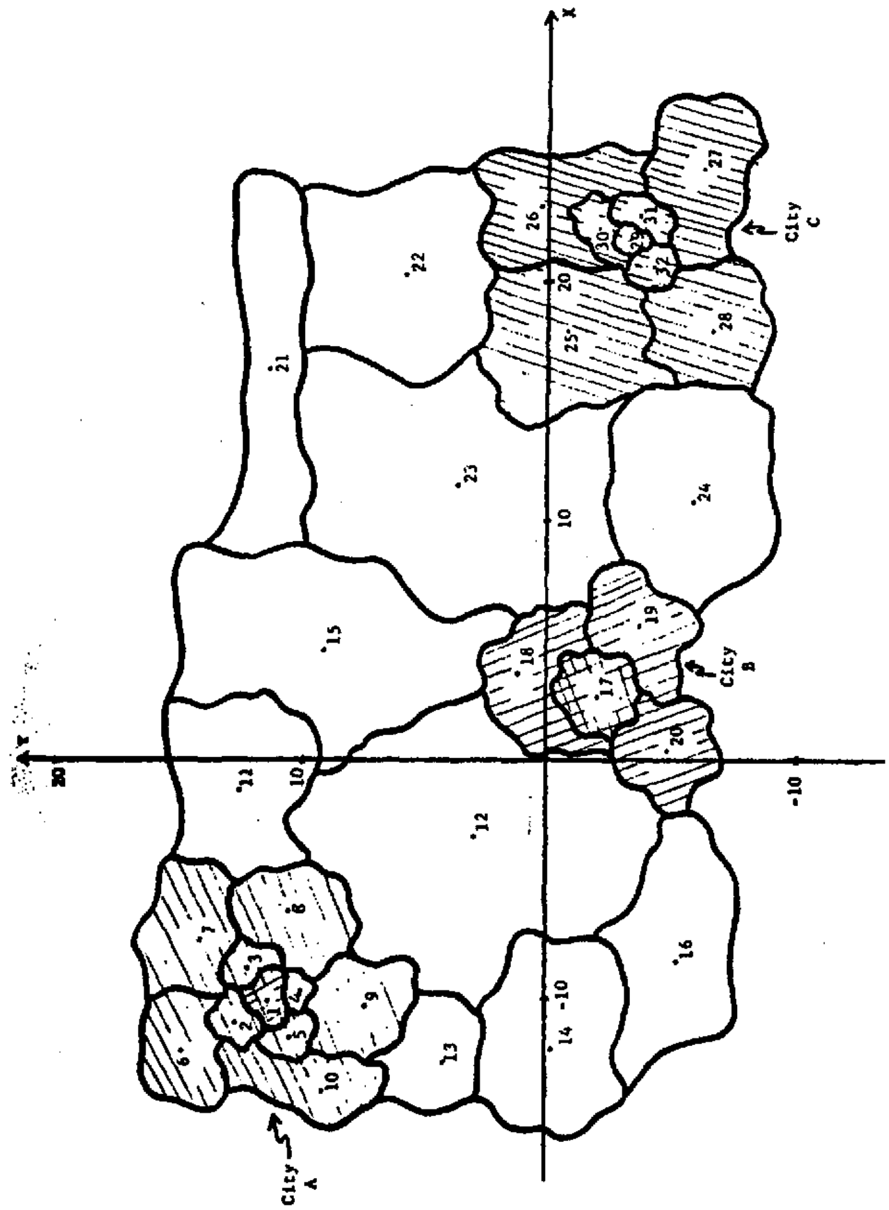


Fig. 1. A hypothetical medical-service area.

Suppose G primary care facilities (or centers) are to be located in the area. We define distance between two points in the area as the rectangular separation (the sum of the vertical and horizontal distances) between the two points. This is considered more valid than the direct distance between the two points, since individuals will travel on highways in the general directions of the two axes. Therefore, the distance between the center of census block i located at (z_{i1}, z_{i2}) and a particular health-care facility k located at (x_k, y_k) is

$$b_{ik} = |z_{i1} - x_k| + |z_{i2} - y_k|. \qquad (1)$$

For convenience, the location of each individual in a given census block is approximated by the center of the census block in which he resides. Thus, in Fig. 1, the distance from an individual located in census block 24 to a health-care facility at the point $(-6.3, 10.4)$ is 33.4 miles.

The only exception to this rule is for computing the distance individuals in a particular census block travel to a primary-care facility in the same census block. In this case, an adjustment is required to eliminate the possibility of estimating that all individuals in the census block travel zero, or a very small distance. This would be assumed when the facility is located at or near the center of the census block. A correction for this possibility is made by assuming that the average distance from the center of the census block to the border is the lower bound on distance traveled within a census block.

The rate of utilization by an individual in a particular stratum j residing in census block i depends upon (a) the expected number of visits per time period if the primary health facility is in immediate proximity to the individual, v_j; (b) the characteristic effect of distance on utilization, d_j; and (c) the actual distance between the census block and the health-care facility, b_{ik} (assuming one center, $k=1$). In the present formulation utilization is assumed to be inversely proportional to distance, as modified by a stratum's sensitivity to distance. Therefore, overall utilization for the n_{ij} individuals in stratum j and census block i is defined as

$$u_{ij} = n_{ij}v_j/(1+d_j b_{ik}), \qquad (d_j \geqq 0) \quad (2)$$

[The constant 1 in equation (2) provides that utilization approaches $n_{ij}v_j$ as b_{ik} approaches zero.] It should be noted that in some formulations of similar distance models, the denominators of expressions showing the effect of distance are raised to a power. The proper value of the power has not been empirically established for the general case, and various values are reported.[18] The present formulation assumes a power of 1 for simplicity. The numerical method used for assigning facilities to locations places no constraints on the use of different values, should they be appropriate for a particular application.

When more than one center is included within a planning region, the formulation must be extended to recognize the effect of distance on both the user's choice of a particular center among those available and the change in overall utilization that results when more centers are available. The effect of distance on utilization is best measured in terms of the proximity to the closest center. The minimum distance from a census block i to a center is defined as $b_i = \min_{k=1,\ldots,N} b_{ik}$.

A user's choice of a particular center from among those available is strongly influenced by the proximity of the center to the user in relation to the proximity of other centers. Recent research by the present authors shows that a regression on distance can explain over one-half of the variance in the percent of all families that choose a particular center.[2] In the present model the choice of a particular center k by individuals in census block i is represented by the probability P_{ik}. This probability is conditional upon the specific spatial relations among centers. Thus, if u_{ijk} is defined as the total number of visits to center k by individuals in stratum j in census block i, then

$$u_{ijk} = P_{ik} n_{ij} v_j / (1 + d_j b_i). \tag{3}$$

The particular probability distribution function that is used in the numerical examples discussed later is as follows:

$$P_{ik} = \{1/[\sum_{k=1}^{k=G} (b_i/b_{ik})]\} \cdot (b_i/b_{ik}). \tag{4}$$

This states that the probability of choosing a particular center k is a function of the distance to center k relative to the distance to the nearest center, or b_i/b_{ik}. The probability function is formed by normalizing this ratio by the term $\sum_{k=1}^{k=G} (b_i/b_{ik})$, to satisfy the constraint $\sum_{k=1}^{k=G} P_{ik} = 1$.

The specific functional relations that will best represent patient behavior in the choice of a facility [equation (4)] and in determining overall utilization [equation (2)] may vary from application to application. The two relations in the present example represent the effect of distance as a nonlinear function, since distance is included in the denominator of each. In any specific application, different relations would probably be appropriate, representing the effect of distance as a linear function, a different nonlinear function, or perhaps with discontinuities.

A simple example demonstrates the use of equations (1), (2), (3), and (4). Suppose a primary care center is located at the point (5, 5) in Fig. 1. An individual in census block 22 [with its center at (20.3, 5.7)] travels 16 miles, on the average, from his home to the primary-care center. Suppose that there are 100 people in stratum $j = 1$, with $v_1 = 4$ and $d_1 = 0.01$. Using equation (1), the total number of visits for stratum 1 in census block 22 would be 400 if the center were in immediate proximity. At 16 miles, however, the utilization is $100 \cdot 4/(1 + 0.01 \cdot 16) = 345$ visits.

If two more centers were available, one at $(-6, 5)$ and another at $(8, -7)$, then $b_{i2} = 27$, $b_{i3} = 25$, and $b_i = b_{i1} = 16$, for $i = 22$. By using equation (4), the proportion of visits to centers 1, 2, and 3 would be 0.44, 0.27, and 0.29, respectively. From equation (3), 152 visits would be to center 1, 93 visits would be to center 2, and 100 visits would be to center 3.

Regional Allocation of Primary Health Centers

The total demand U for primary health care in a region can be determined by summing over all centers, and then over all strata and all census blocks. Thus

$$U = \sum_{i=1}^{i=L} \sum_{j=1}^{j=M} \sum_{k=1}^{k=G} u_{ijk}.$$

The actual service to be provided at any given center u_k can be determined as follows:

$$u_k = \sum_{i=1}^{i=L} \sum_{j=1}^{j=M} u_{ijk}. \qquad (k = 1, \cdots, G)$$

The general statement of the regional allocation decision described here is to determine the x_k and y_k coordinate locations, $k = 1, \cdots, G$, of G centers so as to maximize or minimize a specified criterion. Several reasonable classes of objectives for a regional planner are apparent:

1. *Maximize utilization.* Maximize the total utilization of primary health services in the region, i.e., $\max U$.

2. *Minimize distance per capita.* This criterion might be represented in two different ways: (a) Minimize the average distance that each individual in the community must travel to reach the center of choice. (b) Minimize the average distance per capita to the closest center. The latter alternative is chosen for subsequent analysis. Mathematically, this takes the form $\min(\sum_{i=1}^{i=L} n_i b_i)/n$.

3. *Minimize distance per visit.* Minimize the average per-visit travel distance to the nearest center, where all visits as computed in equation (2) are included, i.e., $\min \sum_{i=1}^{i=L} \sum_{j=1}^{j=M} (u_{ij} b_i / U)$. An alternative statement of this criterion is to minimize the average distance traveled per visit to the user's center of choice. The first version is considered preferable for most applications, however, since the minimum required distance exlusive of choice better represents the effect of distance as a barrier.

4. *Minimize percent degradation in utilization.* This criterion minimizes the average reduction in the number of visits individuals in the community make as a percentage of the visits made if they were in immediate proximity to a primary care center, i.e., $\min(1/n) \sum_{i=1}^{i=L} \sum_{j=1}^{j=M} [n_{ij}(v_j - u_{ij})/v_j]$.

These criteria, or others to which a health planner might subscribe, are likely to dictate distinctly different spatial locations of a fixed number of facilities. The use of criterion 1 will tend to place centers near those high utilizers for whom distance is a strong barrier to utilization. On the other hand, criterion 3 is the most favorable to all high utilizers. It may minimize distance per visit at the expense of individuals for whom distance is a major barrier. By locating centers at a considerable distance from such individuals, distance per visit can be minimized through a reduction in the utilization rates. Criterion 2 will favor areas of large population density and will altogether ignore differential utilization patterns by different census blocks and strata.

Once a criterion is chosen, a method is needed to assign a given number of facilities to locations in such a manner that the criterion is maximized (or minimized). Difficulty is posed by the fact that functions in the model are multimodal. Recent work with problems of comparable difficulty has shown that a modified Hooke-Jeeves pattern-search algorithm[8] performs well (*see*, for instance, references 3, 14, and 15). However, it must be recognized that the gain in formulation flexibility is achieved at the expense of a guaranteed optimum. Obstacles to optimization are addressed by programming the search algorithm to function interactively with the problem solver. The problem of many local optima is effectively approached in this way. This problem can be overcome through the use of multiple starting points and effective tests for local optima.

The Hooke-Jeeves algorithm is used in this paper to solve the spatial allocation model with respect to the first three criteria stated above; that is, optimum values of x_k and y_k, where $k=1, \cdots, G$, are obtained with respect to criteria 1, 2, and 3, for $G=1$ to 4 centers.

The use of this model does not automatically solve the allocation problem. The burden of decision is still on the health planner. The main advantages of the model are that it provides the health planner with an explicit statement of information requirements, a mechanism for examining the incremental gains to be expected from adding additional centers, and a means to explore the differential effects of alternative criteria upon solutions.

Suppose that the health planner can select a criterion, and wishes to determine the optimal locations of G primary care centers, where G can range from one up to some maximum feasible number. (If some centers already exist in the region, he may want to hold these allocations constant and consider only the optimal locations of additional centers.) By using the model, he can compare the expected incremental gains in the criterion value for each additional center. If the choice among several objectives is unclear, the process of allocating centers can be repeated to maximize each objective, for each value of G. This facilitates a comparison of different locations in terms of their relative effects on multiple objectives.

Such a use of the model is illustrated by the example in the following section. The numerical results are obtained for a complete, on-line, interactive, FORTRAN IV version of the model, programmed to be suitable for several hundred census blocks, and up to 20 centers and 8 substrata. The program determines the values for all

TABLE I

UTILIZATION CHARACTERISTICS OF EACH STRATUM

		Utilization rate at immediate proximity	
		High	Low
Barrier effect of distance	Low	Stratum $j=1$ $v_1=5.2$ $d_1=0.0005$	Stratum $j=4$ $v_4=1.9$ $d_4=0.0008$
	High	Stratum $j=3$ $v_3=6.2$ $d_3=0.165$	Stratum $j=2$ $v_2=2.1$ $d_2=0.115$

four criteria described above although only one is optimized at a time. Use of the model provides understanding of the problem and insights into the impact of various criteria on the solution in a hypothetical example.

AN EXAMPLE

ASSUME THAT health care centers are to be located within the region described in Fig. 1. The region has 4 strata, with v_j and d_j as given in Table I. Strata 1 and 3 have markedly higher utilization rates at immediate proximity than strata 2 and 4. Distance is a much stronger barrier to utilization for strata 2 and 3 than for strata 1 and 4. The location of each census block, the average distance from the center of the census block to the boundary, and the population in each stratum are as given in Table II.

A comparison of general utilization characteristics in the three cities is useful. The people in and around City A have higher utilization at unit distance than residents of cities B or C. Of the three cities, the residents of City A are the most mobile in seeking health care, whereas distance is a very strong barrier to utilization for most people in City C. City B is smaller and falls between cities A and C in this respect.

It is useful to explore the implications of the various criteria for a given number of health-care centers and then to examine the effect of increasing the number of centers. The effect of alternative criteria is the most obvious when one center is positioned to satisfy the entire region's requirement. Figure 2 shows the three different locations for one center that are

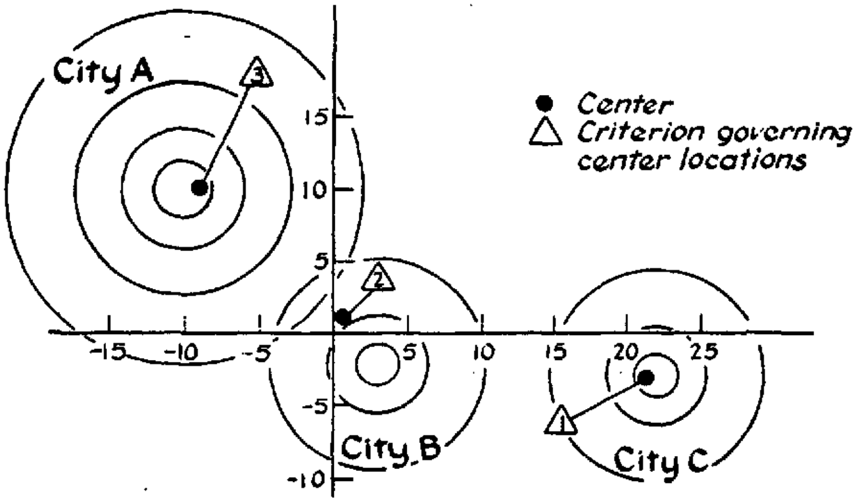
TABLE II

LOCATION, RADIUS, AND POPULATION OF CENSUS BLOCKS BY STRATUM

No.	Coordinates		Radius	Stratum 1	Stratum 2	Stratum 3	Stratum 4
	x	y					
1	−10.0	11.2	1.0	1000	500	450	600
2	−11.0	12.5	1.0	634	167	100	533
3	−8.7	12.0	1.0	633	167	100	533
4	−9.6	10.0	0.7	850	175	100	1000
5	−11.5	10.4	1.0	633	166	100	534
6	−12.0	14.8	2.0	600	200	130	450
7	−7.5	14.0	2.0	700	400	200	500
8	−6.3	10.4	2.0	900	200	120	150
9	−10.2	7.3	1.8	650	200	130	450
10	−13.7	9.1	2.0	650	200	120	450
11	−1.2	11.5	2.2	700	200	125	600
12	−3.2	3.0	5.0	563	100	37	450
13	−12.5	4.2	1.8	850	200	125	700
14	−12.0	−0.3	4.0	512	100	38	50
15	4.6	9.0	4.0	463	200	37	350
16	−6.3	−5.2	3.5	362	200	38	250
17	2.7	−2.0	1.7	500	1300	450	500
18	3.6	1.2	2.0	600	575	275	500
19	5.6	−3.7	2.0	600	375	275	400
20	0.5	−4.8	2.0	750	375	250	800
21	16.4	11.3	2.5	112	450	137	150
22	20.3	5.7	3.5	112	350	288	150
23	11.5	3.6	5.5	112	200	288	150
24	10.0	−5.8	4.0	600	700	600	400
25	18.0	−0.8	3.0	113	350	787	150
26	23.0	0.3	1.8	113	450	438	150
27	24.8	−6.3	1.5	113	450	537	150
28	18.0	−6.6	2.0	225	450	475	200
29	21.8	−3.3	0.8	250	1600	850	350
30	22.4	−1.9	1.0	107	534	600	217
31	22.8	−3.7	1.0	167	533	600	216
32	20.6	−4.3	1.0	166	533	600	216

dictated by criteria 1, 2, and 3 with respect to the three cities. The three dots indicate center locations; associated triangles identify respective criteria. The precise locations with respect to the coordinates in Fig. 1 are given in Table III. Data offering a comparison of each location with respect to all criteria, including criterion 4, are provided in Table IV.

The three criteria yield entirely different locations. When criterion 1 (maximize utiliza-



Location of 1 Center for 3 Different Criteria

Fig. 2. The effect of each criterion upon the location of one health center.

TABLE III

LOCATION COORDINATES IN MILES FOR SPECIFIED CRITERIA AND SPECIFIED NUMBERS OF CENTERS

(See Fig. 1 for locations of coordinates.)

Center number	Criterion					
	(1) Maximize utilization		(2) Maximize distance per capita		(3) Minimize distance per encounter	
	x	*y*	*x*	*y*	*x*	*y*
I With 1 center						
1	21.00	−3.00	0.64	1.20	−8.70	10.00
II With 2 centers						
1	21.4	−3.7	17.6	−3.30	18.50	−3.30
2	−9.89	10.4	9.89	10.4	−9.90	10.40
III With 3 centers						
1	22.40	−3.1	21.52	−2.78	22.30	−3.20
2	−10.16	10.40	−10.20	10.40	−10.20	10.40
3	3.63	−2.75	3.60	−2.80	3.60	−2.80
IV With 4 centers						
1	22.40	−3.14	22.00	−3.50	21.23	−3.08
2	−10.20	10.40	−10.10	10.30	−9.80	10.40
3	3.59	−2.78	2.69	−4.80	3.61	−2.70
4	11.32	−2.25	3.76	3.04	−11.35	3.00
V With 5 centers[a]						
1	22.40	−3.10				
2	−9.72	10.61				
3	3.24	−3.19				
4	−11.62	3.24				
5	11.04	−2.00				

[a] Determined only for the first criterion.

tion) is used, the center is located near the center of City *C*. This is due to the fact that this area contains a large number of individuals for whom distance is a strong barrier. At the same time this choice of a location tends to increase overall distance per capita vis-a-vis other criteria greatly, as can be seen from Table IV. One might argue that this criterion is most equitable, since it favors those whose utilization is most curtailed by large travel distances. However, a counter argument is that the resulting location shifts the cost of transportation to those who are less sensitive to distance barriers, irrespective of their ability to accept it.

The third criterion (minimize distance per encounter) leads to a location near the center of the largest city (*A*). This location is traditional but not particularly equitable. Overall

TABLE IV
CHARACTERISTICS OF OPTIMUM LOCATIONS

	Criterion		
	Maximize utilization	Minimize distance per capita	Minimize distance per encounter
With 1 Center			
Utilization, U	126,239	119,076	118,572
Distance/capita	25.82	18.00	21.18
Distance/encounter	24.86	16.56	14.60
Index—% decrease in utilization	33.2%	35.4%	36.20%
With 2 Centers			
Utilization, U	146,891	143,237	144,195
Distance/capita	8.53	8.14	8.17
Distance/encounter	7.39	7.41	7.39
Index—% decrease in utilization	21.7%	23.4%	22.96%
With 3 Centers			
Utilization, U	155,086	154,575	155,051
Distance/capita	5.32	5.31	5.32
Distance/encounter	4.90	4.94	4.90
Index—% decrease in utilization	16.9%	17.15%	16.91%
With 4 Centers			
Utilization, U	156,932	156,090	155,514
Distance/capita	4.87	4.77	4.79
Distance/encounter	4.58	4.43	4.39
Index—% decrease in utilization	15.84%	16.26%	16.50%

utilization is the lowest with this criterion and the percent reduction in average utilization is highest (criterion 4). This third criterion is appropriately described as a local one. That is, it does not consistently represent needs throughout the region. In the present example, the optimum results from a simultaneous decrease in utilization in some areas and an increase in others. As the center is moved away from distance-sensitive remote locations, distance per encounter decreases over the range considered. This is because utilization varies inversely with, and at a higher marginal rate than, distance. At the same time, the center is moved toward large population centers so as to increase utilization at close proximity. The criterion is analogous to the action that well-intentioned medical practitioners might take in independently locating their individual facilities—minimize the distance that their patients must travel. The essential point is that such a rule leads to much different results than are obtained with more global criteria.

The second criterion yields a location that might be anticipated. Per capita distance is minimized when the center is located in City *B*, between the two larger cities. It is interesting to note that this rule, like the other two, tends to locate the center in an area of high population rather than in a sparsely populated census block.

There is little difference in the effects of the three criteria upon locations when two or three centers are allocated. As shown in Table II for the two-center case, locations are near the centers of the two large cities (*A* and *C*) for all criteria. Similarly with three centers, they are consistently located near the centers of the three cities.

The four-center case, like the one-center case, again shows a distinct difference among all three criteria. This is depicted in Fig. 3. Three centers are consistently located in the

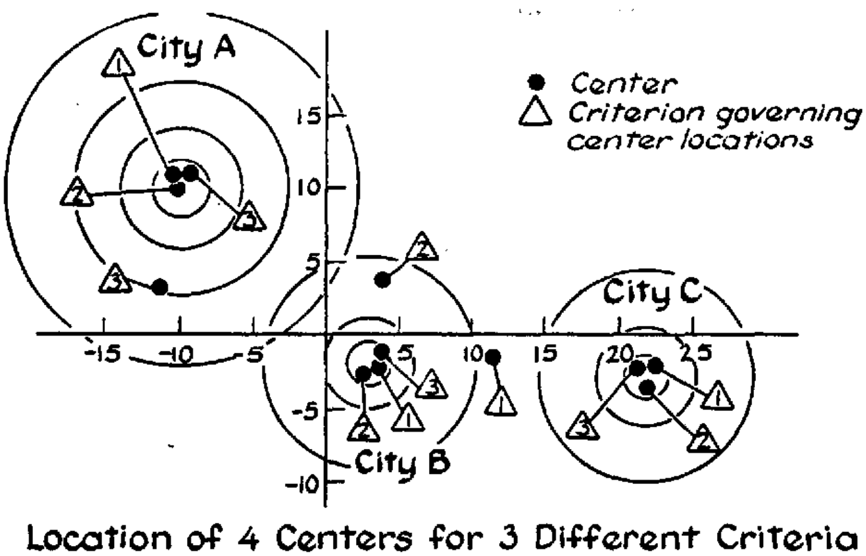


Fig. 3. The effect of each criterion upon the location of four health centers.

cities but the location of the fourth one depends critically on the criterion. The effect of each criterion upon the location of this fourth center is similar to the effect of each criterion upon the location of one center.

The differential effects of different criteria for health services tend to diminish as the number of centers increases. The impact of different criteria is much less when there are four centers than when there is only one. Figure 4 provides a direct comparison of the criteria as the number of centers is increased from one to four.

The three separate graphs shown in Fig. 4 present the value of each criterion when it serves as the basis for optimizing the location of from one to four centers. Each graph also provides a comparison of this with the degradation in the same criterion measure when each of the other two criteria is optimized. The common pattern with each criterion is a significant improvement as the number of centers is increased above 1 but a marked decrease in the rate of improvement as more centers are added—a strong tendency toward diminishing returns to

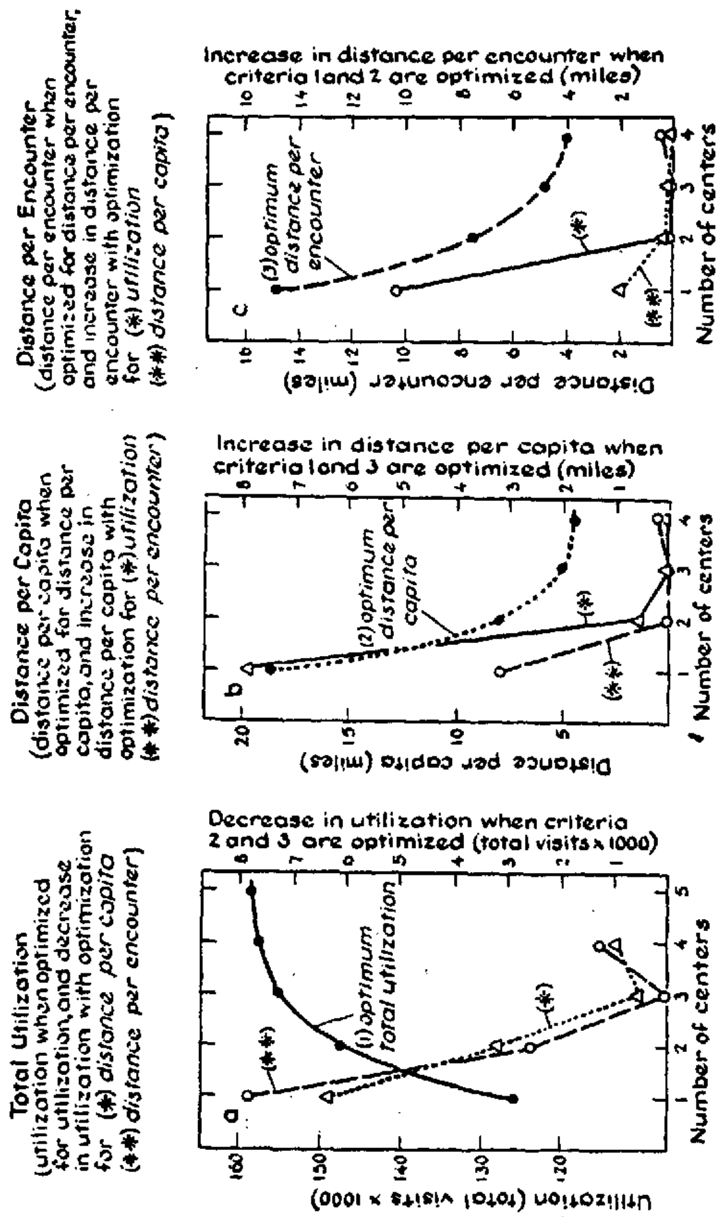


Fig. 4. Differential effects of three criteria—one to four centers.

scale. When more than four centers are allocated, there is little relative improvement in any of the measures, and differences among the region-wide measures are negligible.

The three graphs included in Fig. 4 demonstrate the use of the model in the important analytic role of determining the incremental benefits from increasing the number of centers in a region. The specifics of analysis can be varied to accommodate the planning problem at hand. One or more centers can be fixed in location or capacity to represent existing facilities, different criteria can be studied, and minimum or maximum center sizes can be included in the formulation. The model also gives volumes by strata for each center so that issues related to minimum size and economies of scale can be dealt with. While specific refinements can be introduced to accommodate a particular planning problem, Fig. 4 illustrates the type of analysis that can be performed to determine incremental benefits from adding more centers. A health planner can compare the cost implications of suggested numbers, sizes, and locations of centers with the marginal benefits of a given level of service, where benefits are measured in terms of such criteria as utilization, travel distance, or per capita percentage decrease in utilization on a region wide basis.

In the final analysis, the decision maker must decide how to weigh different objectives. No one criterion is best. The ultimate decision as to the number, capacity, and location of health centers must be based upon a broad range of considerations. However, formal analysis can be extremely useful in reducing the number of variables that must be considered subjectively. Furthermore, by forcing the model to evaluate decisions that are proposed on other bases, the model can serve to quantify the consequences of alternative choices for comparison with the optimum solutions derived from explicit criteria.

CONCLUSION

THE REGIONAL allocation problem is a central issue in research on health-care delivery systems. This paper has undertaken to develop a method that can be applied practically by a health planner in bridging the gap between research results and implementation. Emphasis is on providing a formulation that usefully characterizes the problem within a region rather than on the elegance of optimization procedures. (As a benchmark for computational feasibility, each optimization for the above example required about 20 seconds of CPU time on an IBM 360-67.) The formulation distinguishes the separate impact of distance on various socioeconomic strata with respect to their choice of a particular center as well as overall utilization. The model can be extended to accommodate considerations of importance in particular applications by changing the criteria used for optimization and the constraints imposed in the formulation.

ACKNOWLEDGMENT

RESEARCH REPORTED in this paper was supported in part by the National Center for Health Services Research and Development, Department of Health, Education, and Welfare.

REFERENCES

1. W. J. ABERNATHY AND J. R. MOORE, "Regional Planning of Primary Health Care Services," *Medical Care*, forthcoming (1972).
2. ——— AND E. L. SCHREMS, "Distance and Health Services: Issues of Utilization and Facility Choice for Demographic Strata," Research Paper #19, Stanford University, Graduate School of Business, July, 1971.
3. E. S. BUFFA AND W. H. TAUBERT, "Evaluation of Direct Computer Search Methods for the Aggregate Planning Problem," *Indust. Management Rev.*, 19-36 Fall, 1967.
4. A. CIOCCO AND I. ALTMAN, "Medical Service Areas and Distances Traveled for Physician Care in Western Pennsylvania," Public Health Monograph No. 19, U.S. Public Health Service, 1959.
5. J. D. FORBES AND W. T. ZIEMBA, "Estimation of Supermarket Drawing Power: An Extension of Location Theory and Practice," *Annals Regional Sci.* (forthcoming).
6. W. L. GARRISON, D. F. MARBLE, J. D. NYSTUEN, AND R. L. MORRILL, *Studies of Highway Development and Geographic Change*, University of Washington Press, Seattle, Washington, 1959.
7. P. F. GROSS, "Urban Health Disorders, Spatial Analysis and the Economics of Health Facility Location," Paper presented at the 40th National Conference of the OPERATIONS RESEARCH SOCIETY OF AMERICA, Anaheim, California (October 28, 1971).
8. P. Haggett, *Locational Analysis in Human Geography*, St. Martin's Press, New York, 1966.
9. R. HOOKE AND T. A. JEEVES, "Direct Search Solution of Numerical and Statistical Problems," *J. Assoc. Computing Machinery* 8, 212–229 (1961).
10. P. J. JEHLIK AND R. L. MCNAMARA, "The Relation of Distance to the Differential Use of Certain Health Personnel and Facilities and to the Extent of Bed Illness," *Rural Soc.* 17, 261–265 (1952).
11. R. L. KANE, "Determination of Health-Care Priorities and Expectations among Rural Consumers," *Health Services Res.* 4, 142–151 (1969).
12. G. W. SHANNON, R. L. BASHSHUR, AND C. A. METZNER, "The Concept of Distance as a Factor in Accessibility and Utilization of Health Care," *Medical Care Rev.* 26, 143–161 (1969).
13. J. A. SONQUIST AND J. N. MORGAN, *The Detection of Interaction Effects*, Survey Research Center, University of Michigan, Monograph 35, 1964.
14. W. H. TAUBERT, "A Search Decision Rule for the Aggregate Scheduling Problem," *Management Sci.* 16, 343–359 (1968).
15. ———, *The Search-Decision-Rule Approach to Operations Planning*, Unpublished Doctoral Dissertation, U.C.L.A., 1968.
16. J. E. WEISS AND M. R. GREENLICK, "Determinants of Medical Care Utilization: The Effects of Social Class and Distance on Contacts with the Medical Care System," *Medical Care* 8, 456–462 (1970).

How to determine the optimum number of operating rooms

Jay Goldman, D.Sc. and H. A. Knappenberger, Ph. D.

How many operating rooms do you really need? Is it necessary to acquiesce meekly when the surgical staff says more are needed? How can you determine objectively the point at which more rooms can be added at economic cost?

To keep within the total hospital budget, the decision to add an operating room must be based on several factors. First, there is the cost associated with building, equipping and staffing an additional room. Second, there is the cost associated with overtime which will occur due to lack of flexibility in handling cases which run longer than estimated in a tightly scheduled system. Finally, as the utilization of the existing facility increases, the number of patients placed on the waiting list will increase. While it may be difficult to assign a cost to patient waiting, it does exist and must be taken into consideration.

One could easily state the variables of a simplified break-even analysis to determine when an additional operating room is warranted. Such an analysis would generally consider opening an additional room when the demand on the operating suite reaches the point at which the following equality holds:

$$C_1 = 250C_2(OT_b - OT_a) + 365C_3(PW_b - PW_a),$$

where C_1 is the total annual cost for building, equipping and operating an additional room, C_2 is the cost per hour of overtime, C_3 is the cost for one day of patient waiting, OT_b is the average hours of overtime per day before a new room is opened, OT_a is the average hours of overtime per day after a new room is opened. PW_b is the average number of patients waiting before a new room is opened, PW_a is the average number of patients waiting after a new room is opened, 250 represents the number of operating days per year and 365 represents the total number of days per year.

Number of Operating Rooms	3	4	5	6
Demand in Minutes/Day	1977.4	1977.4	1977.4	1977
Capacity in Minutes/Day	1440	1920	2400	2880
Load Factor (Demand/Capacity)	1.373	1.030	0.824	0.687
Average Patient Days Waited	109	78	0.98	0.08
Average Overtime (Hours/Day)	0.91	1.08	0.43	0.25
Percent Utilization	81.4	80.9	69.0	57.6

TABLE 1. Given a demand per day of 1977.4 minutes, this simulation run displays the effects of using 3, 4, 5 and 6 operating rooms

In order to use the above break-even analysis at the current level of demand, the cost variables and C_1 and C_2, the present overtime requirement (OT_b), and the length of the present waiting list (PW_b) can be determined easily. While it is difficult to determine the exact cost of one day of patient waiting (C_3), the general order of magnitude of this cost can be estimated. The real difficulty in performing the break-even analysis lies in the prediction of the amount of overtime and the length of the waiting list (OT_a and PW_a) if an additional room were in operation.

Thus far the break-even analysis has been used to determine if the current demand warrants opening an additional room. The more general problem facing the hospital administrator is that of predicting the level of demand at which an additional room is justified. In this case, the variables OT_b and PW_b must be predicted for levels of demand other than the current one. The prediction of OT_a and PW_a at these additional levels of demand continues to be exceptionally difficult.

The problem of deciding when to open an additional operating room faced the administration of The Memorial Hospital of Wake County, N. C. Memorial Hospital is a 380 bed community general hospital which last year admitted 11,134 patients and operated at a 63.4 per cent occupancy. The operating suite performed over 3,500 surgical procedures with an average daily demand of 1,977.4 operating minutes.

In order to provide the required estimates of the variables in the break-even analysis, a simulation

---

Dr. Goldman is professor of industrial engineering at North Carolina State University at Raleigh. Dr. Knappenberger is assistant professor of industrial engineering there.

Demand in Minutes/Day	*2076.3*		*2175.1*		*2274*		*2372.9*	
Number of Operating Rooms	5	6	5	6	5	6	5	6
Capacity in Minutes/Day	2400	2880	2400	2880	2400	2880	2400	2880
Load Factor	0.865	—	0.906	—	0.918		0.959	—
Average Patient Days Waited	2.61	0.13	4.65	0.30	7.03	0.54	23.09	0.63
Average Overtime (Hours/Day)	0.50	0.26	0.78	0.39	0.87	0.44	0.91	0.48
Percent Utilization	72.1	61.2	75.3	63.6	77.4	65.5	80.1	67.4

TABLE 2. For four different levels of demand per day, this table displays dependent variables for use of 5 or 6 operating rooms.

model* was developed. The model consisted of a digital computer program written in Fortran IV for an IBM 360 Model 75 Computer. The logic of the program was a detailed accounting, in five-minute increments, of the flow of each patient from the time the decision was made to schedule surgery until the time that surgery was completed or the case canceled. Memorial Hospital provided a general framework for the basic structure of the model as well as data for estimating the form of the various distributions required in the simulation.

Several preliminary simulation runs were made using the model. These runs indicated that the average values of the required dependent variables stabilized within 180 simulation days. Thus, 180 days was chosen as the simulation period.

In order to test the validity of the model, a sequence of four simulation runs was carried out, each run using the average daily demand of 1,977.4 minutes and a different level of operating suite capacity. The four levels of operating suite capacity were 1,440, 1,920, 2,400 and 2,880 minutes per day. These four levels represent the use of three, four, five, and six operating rooms scheduled for a 480 minute operating day. The results of these runs are shown in Table 1.

From the data in Table 1, several factors are apparent. For extremely high load factors (around 1 and above) both the average daily overtime and the operating room per cent utilization tend to reach a plateau value. For this study, the overtime remained about one hour per day and the per cent utilization leveled off at about the 80 per cent level. However, when the level of utilization approaches its maximum, the length of the waiting line increases at a rapid rate. For lower load factors, the value of each variable increased as load factor increased.

The results in Table 1 agree with the results obtained from similar industrial systems and, more importantly, agree with Memorial Hospital's experience. There is little doubt that the use of five operating rooms is the correct choice given the average current daily demand of 1,977.4 operating minutes. Therefore, the decision facing the hospital administrator is when to open the sixth room. In making this decision, the simulation model can be used to provide estimates of overtime and patient waiting which would result from operating five and six rooms over a series of increased levels of demand.

As shown in Table 1, the load factor for five operating rooms is 0.824 and the load factor for four operating rooms is 1.03. Since the use of four operating rooms would result in an undesirable level of patient waiting a load factor between 0.824 and 1.03 would include the critical load factor at which point an additional room is desirable. Thus it was decided to evaluate the effect of using five rooms and six rooms at the intermediate levels of daily demand of 2,076.3, 2,175.1, 2,274 and 2,372.9 minutes. These levels of demand are respectively equivalent to load factors of 0.865, 0.906, 0.948 and 0.989 (five-room capacity).

For each of the above levels of demand 180 day simulation runs were performed using both five and six operating rooms. The results from these eight runs are shown in Table 2.

Using the results shown in Table 2, a break-even equation can be written for each of the four increased demand levels tested. They are:

*Moore, E. W.: An investigation of operating room utilization and performance through the use of simulation. M.S. Thesis, N. C. State University, 1967

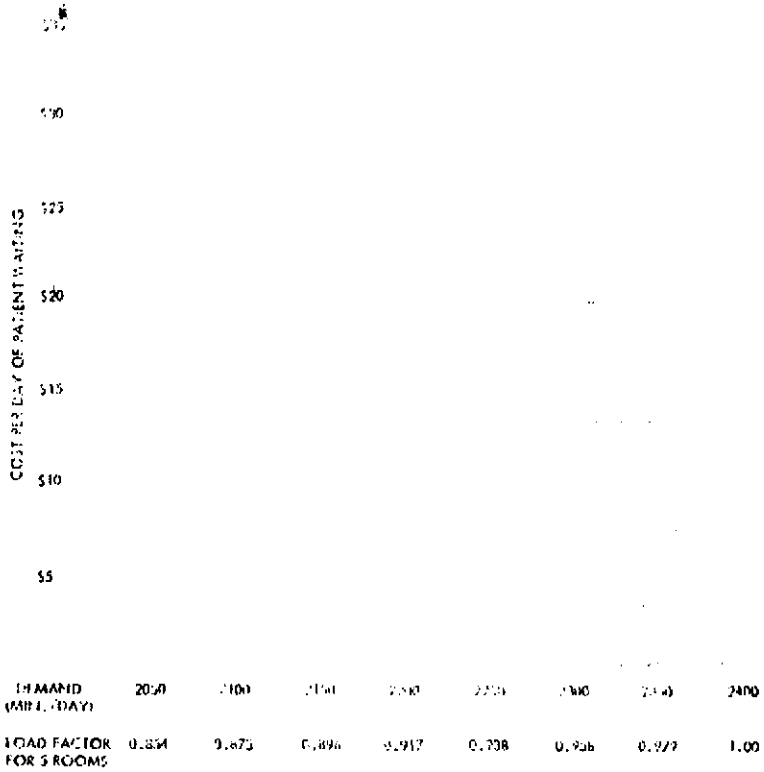


FIGURE 1 Level of demand at which an additional room is required as a function of cost per day of patient waiting.

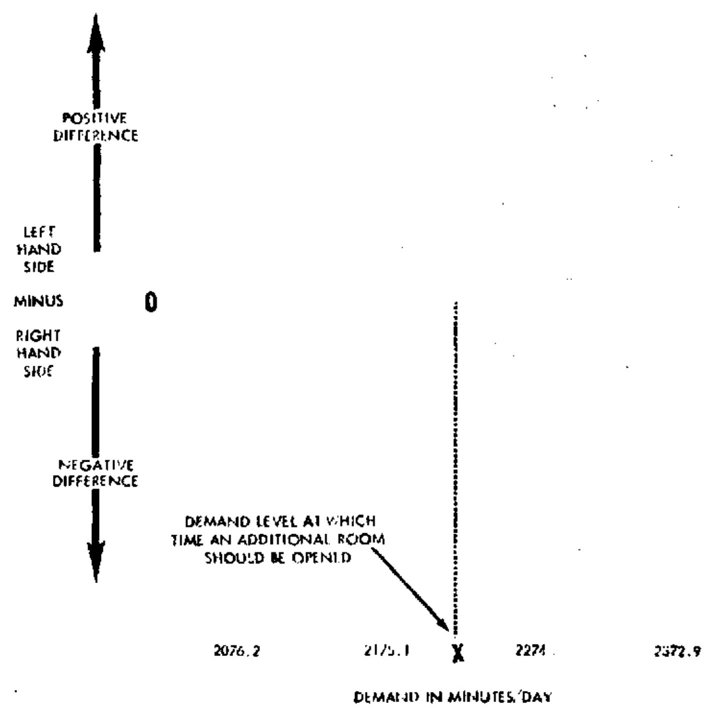


FIGURE 2 Break-even curve for determining the level of demand at which an additional room is required.

$$C_1 = 250C_2 (0.50 - 0.26) + 365C_3 (2.64 - 0.13)$$

for a demand of 2,076.3 minutes per day,

$$C_1 = 250C_2 (0.78 - 0.39) + 365C_3 (4.63 - 0.30)$$

for a demand of 2,175.1 minutes per day,

$$C_1 = 250C_2 (0.85 - 0.44) + 365C_3 (7.03 - 0.51)$$

for a demand of 2,274 minutes per day, and

$$C_1 = 250C_2 (0.91 - 0.48) + 365C_3 (23.09 - 0.62)$$

for a demand of 2,372.9 minutes per day.

If values can be assigned to C_1, C_2 and C_3, the left and right hand sides of each of the above four equations can be evaluated. If the right hand side of each equation is subtracted from the left hand side and the resultant value plotted against the appropriate level of demand, a break-even curve similar to that shown in Figure 1 can be drawn. The point at which the curve drawn through the four data points intersects the zero difference line indicates the demand (X) at which it is economically desirable to open the additional room.

As was previously discussed, it is often difficult to assign a specific cost value to patient waiting (C_3). However, it is possible to develop a picture which indicates the effect of alternative values of C_3 upon the decision to open an additional room. This can be accomplished by setting the left hand side minus the right side of the above four equations equal to zero and solving each for C_3. For the four equations this would result in:

$$C_3 = \frac{C_1 - 250C_2 (0.50 - 0.26)}{365 (2.64 - 0.13)}$$

for a demand of 2,076.3 minutes per day,

$$C_3 = \frac{C_1 - 250C_2 (0.78 - 0.39)}{365 (4.63 - 0.30)}$$

for a demand of 2,175.1 minutes per day,

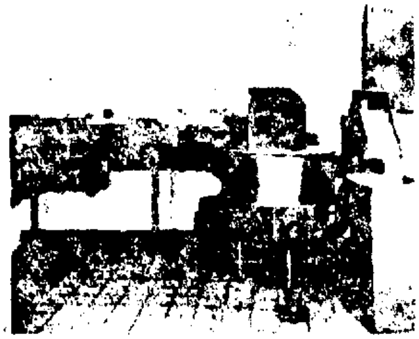
$$C_3 = \frac{C_1 - 250C_2 (0.85 - 0.44)}{365 (7.03 - 0.51)}$$

for a demand of 2,274 minutes per day, and

$$C_3 = \frac{C_1 - 250C_2 (0.91 - 0.48)}{365 (23.09 - 0.62)}$$

for a demand of 2,372.9 minutes per day.

(Continued on Page 198)

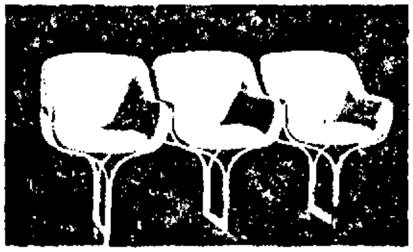


Disposal System 326

Engineered to dispose of paper and plastic food service items, the Somat Mesm Master combines a pulper and extractor capable of reducing up to 300 pounds of waste per hour to approximately 20 per cent of the original volume. A conveyor belt feeds items into the system, which processes the waste into semi-dry, odorless pulp and ejects it into haul-away receptacles. **American Can Co.**

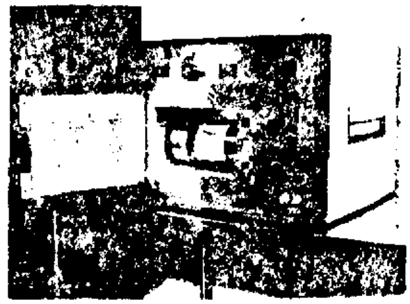
Respirator 327

Model LS-104-150 pediatric respirator, a volume limited, positive-pressure device designed for newborn, premature and very young children, provides continuous readouts indicating volume, breathing rate, and line pressure. Controlled and assist modes of operation are provided, and the unit features an apnea alarm system, adjustable maximum-pressure relief valve, and low-pressure alarm to indicate system leaks. **Bourns, Inc.**



Multiple-Seating 328

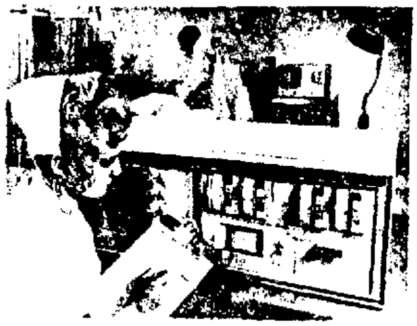
The new multiple-seating System 900 was specially developed for professional reception areas where quality, style and attractive appearance are essential. Combining a mirror-polished chrome frame with ruggedly-built, heavy-duty seating arrangement, System 900 is available in a variety of fabrics and colors with two seat styles. **Madison Furniture Industries**



Blood Warmer 329

Designed to operate on a counter or table top, the micro save blood warmer is capable of raising the temperature of refrigerated blood from the typical blood bank storage temperature of 4° C. to near body temperature. Blood is heated while still in its plastic container, then administered directly to the patient. **Holaday Industries, Inc.**

(Continued on Page 202)

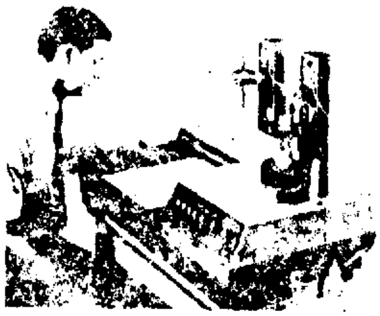


Monitoring System 330

A new monitoring system incorporates advanced alarm circuitry for detection and display of critical patient data in surgery recovery rooms and intensive and cardiac care units. Electronic circuitry is designed to recognize only those alarm conditions generated by unusual heart rate activity and reject the artifacts. **Honeywell**

Urinal & Bed Pan 331

Constructed of polyethylene plastic, two new patient-care items have been introduced by Vollrath. The low-cost disposable male urinal, featuring a wide flat base, prevents spillage and includes a handle which hooks on bed rails and a tight-fitting plastic cover, and a new fracture bed pan designed for use with immobile patients. **Vollrath Co.**

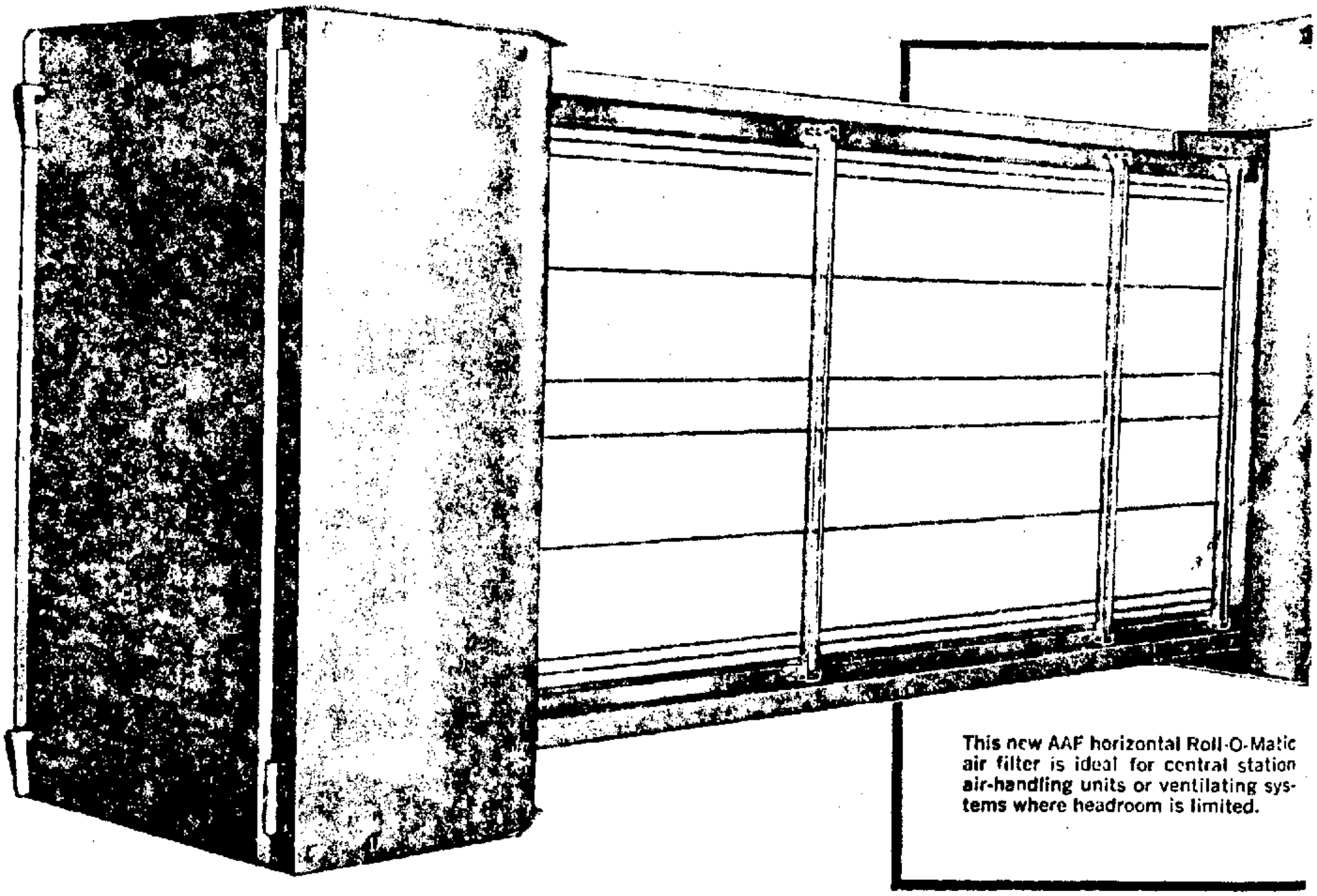


Recording Titrator 332

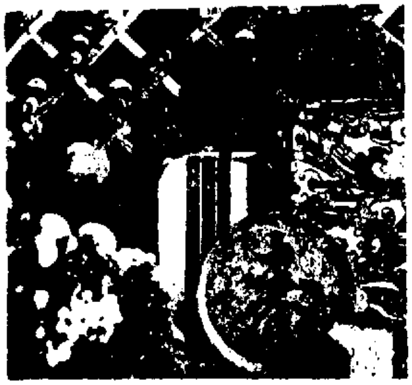
A new automatic recording titrator is operable in a variety of modes, such as derivative, manual, pH control and automatic. The titrator accepts any form of electrode and accommodates high or low resistance systems, aqueous or nonaqueous media, and samples ranging in size from 50 microliters up. All materials in contact with titrant and sample are pretested for durability and are resistant to corrosion and chemicals. **Beckman Instruments, Inc.**

Calculator 333

Calculations can be performed in as little as 10 milliseconds with the new Model 362 electronic calculator. Suitable for applications with potential growth to desk-top computing systems, the calculator features multi-register operation with 14 independent adders and 24 core storage registers. Square, root, logarithm and exponent operations facilitate scientific calculation. **Wang Laboratories, Inc.**



This new AAF horizontal Roll-O-Matic air filter is ideal for central station air-handling units or ventilating systems where headroom is limited.

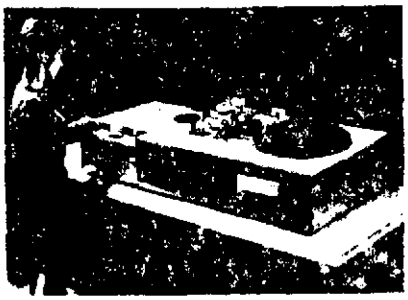


Trays 334

Colorful patterns and oblong, round and oval shapes designed to match or blend with any decor, mood or menu are offered with the new Motif collection of decorator trays. Suitable for hospitals and nursing home, the trays are made of a 17-layer laminate process which is safe in commercial dishwashers. **Bolta Products Group**

Enzyme Analysis Unit 335

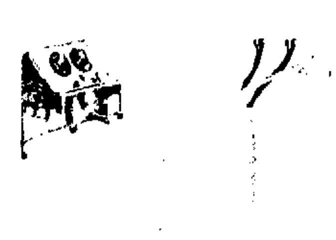
Zymat 340, a new laboratory instrument, utilizes kinetic technics to per-



form major enzyme analyses required in the clinical laboratory. The measurement of change is printed out on tape in International Enzyme Units, together with a sample identifying serial number. The device can operate as a programmed unit, without constant monitoring, and can analyze one sample every two minutes — up to 47 samples. **Bausch & Lomb**

Whirlpool Controls 336

Temp-O-Matic permits a technician to "program" whirlpool therapy in four ways — (1) Controlled filling to any selected level, with no overflow. (2) Controlled temperature by a thermostatic mixing valve preset



for 105° F. and a switch that permits manual adjustment. (3) Timed hydrotherapy, with shut-off at any time up to 30 minutes. (4) Timed emptying and automatic shut-off of water pump. **Ille Electric Corp.**

Ice Dispenser 337

Model MDG-290 pushbutton ice dispenser releases cubes into ice buckets, pitchers or other containers without contamination by human hands. The unit produces up to 340 pounds of ice cubes per day and has a maximum dispensing and storing capacity of 275 pounds. **Manitowoc Equipment Works**

For more facts circle 132 on reader card

LITERATURE & SERVICES

Partitions 338

Introducing the new Hospital Door, "Flexible Space for Nursing Care" brochure describes Modernfold folding partitions and doors to suit every hospital and nursing home requirement. **Modernfold, New Castle Products, Inc.**

Furniture 339

A wide range of planned patient room arrangements, featuring Borg-Warner beds and Drexel furniture, are illustrated in a new four color catalog. The catalog offers decorating suggestions and describes the home-like look in hospital furnishings. **Borg-Warner Corp.**

Parking Brochure 340

The utilization of TV monitoring systems, roving patrols with two-way radios, and pedestrian walkways is detailed in a new parking brochure. The brochure also describes architectural and engineering services available for parking structures of all types, including hospitals. **T. A. Constantine & Associates**

Fire Detection 341

A new catalog describes Pyr-A-Larm automatic fire and smoke detection systems. Detectors, control panels, annunciators and supplementary equipment are illustrated and their functions are detailed. **Pyrotronics**

Kosher Foods 342

Published in paperback size, the new Insitutional and Industrial Kosher Products Directory contains a list of more than 1,850 rabbinically endorsed and supervised kosher food service products. **Union of Orthodox Jewish Congregations of America**

Carpet Guide 343

"Modern Carpet Care," a 20 page illustrated booklet, outlines carpet care programs and suggested procedures for various cleaning methods. **World Floor Machine Co.**

Recipes 344

"Standardized Quantity Recipes" contains 230 recipes for various institutional requirements. Each recipe includes a yield in portions and suggests the size of portions and pans to use. The book is available for $10 from **College Printing & Typing Co., Inc.**

Chairs 345

A full line of adjustable medical-surgical chairs and stools is illustrated in a new 12 page catalog. Complete descriptions, dimensions and range adjustment data are included. **Ajusto Equipment Co.**

Snow & Ice Control 346

The "Wintermobility and You" booklet explains how institutions can plan for effective control of snow and ice. A simple, six-part program offers specific procedures and control tips, and the booklet lists the complete Meyer line of snow plows, spreaders and other equipment. **Meyer Products, Inc.**

(Continued on Page 298)

Emergency Lighting 347
A new manual covering the selection and installation of emergency lighting equipment explains how to determine the number of lampheads required from floor plans, illustrates recommended lamp placement; discusses battery selection and describes the method of determining equipment capacity. **Carpenter Mfg.**

SUPPLIERS NEWS

Addressograph Multigraph Corp., Cleveland, Ohio, announces opening of a **new** A-M branch at 413 N. Wabash Ave., Chicago, Ill. 60611

Borg Warner Corp., Chicago, Ill., announces formation of the **Health Products** division to manufacture and market products for hospitals, nursing homes and related health care facilities.

C. H. Dexter Div., The Dexter Corp., Windsor Locks, Conn., announces formation of the **Medical Products Dept.** to consolidate its development and marketing of medical disposables.

Hoffman LaRoche Inc., Nutley, N.J., announces acquisition of RCA's **medical electronics activity.** However, RCA will continue to sell medical support equipment, such as computer systems and CCTV.

Johnson & Johnson, New Brunswick, N.J., announces opening of its new **Midwestern Distribution Center,** located in Santa Fe Industrial Park, Will County, Ill.

Puritan Compressed Gas Corp., Kansas City, Mo., announces change of its corporate name to **Puritan-Bennett Corp.**

Ritter Pfaudler Corp. and Taylor Instrument Co., both of Rochester, N.Y., jointly announce an approved merger of their respective companies into a new corporation. **Sybron Corp.**

William H. Rorer, Inc., *Fort Washington, Pa.* and Amchem Products, Inc., jointly announce an approved merger of their respective companies into a new corporation, **Rorer-Amchem, Inc.**

Southern Equipment Co., St. Louis, Mo., announces acquisition of the **Stearnes portion of Stearnes-Imperial, Inc.,** a subsidiary of Bastian-Blessing Co.

Univac Div., Sperry Rand Corp., *Philadelphia, Pa.,* announces formation of the **Computer Graphics Group** to strengthen and unify the company's activities in the development and production of computer display terminals.

(*Continued on Page 212*)

#16

Facility Location: A Review of Context-free and EMS Models

By Charles ReVelle, David Bigman, David Schilling, Jared Cohon, and Richard Church

EMS location models are those formulated to address specific problems of emergency medical services systems; context-free location models are those developed without reference to particular applications. The literature on these two types of public facility location models is reviewed, and the development of the maximal covering model from several earlier context-free models is described, with emphasis on problem statements and articulation of service objectives. An application of the maximal covering model to fire truck location points up the ability of this model to handle multiple objectives; its ability to compare alternative solutions gives it great utility for planning and evaluating EMS systems of a wide range of complexity. Potential applications of the maximal covering model are discussed regarding EMS problems involving multiple time standards and service objectives, location of special equipment, and siting of fixed facilities.

The placement of facilities on a network to respond to demands or to attract consumers has become a subject for both study and teaching. The extensive bibliography by Lea [1] indicates how the literature in this area has burgeoned in recent years. The literature includes investigations of the location of both public and private facilities; much of the literature concerning both the public and private sectors through about 1970 was brought together in an article by ReVelle, Marks, and Liebman [2]. Private sector facilities are often located to fulfill precisely stated objectives, such as minimum cost or maximum profit. In contrast, the goals and objectives of public facility location are more difficult to capture and translate into quantifiable terms.

The gradually evolving interpretation of the goal of maximum public welfare has led to a number of possible problem statements in the sphere of public facility location. Each of these different statements has the potential to fit the perceptions of some decision maker. Recognizing the validity of many objectives has brought us to what

Address communications and requests for reprints to Charles ReVelle, Program in Systems Analysis and Economics for Public Decision Making, Johns Hopkins University, Baltimore, MD 21218.

David Bigman is an economist with the International Monetary Fund; David Schilling is assistant professor at the Center for Technology and Administration, American University; Jared Cohon is associate professor in the Johns Hopkins Program in Systems Analysis and Economics for Public Decision Making; and Richard Church is assistant professor of civil engineering at the University of Tennessee.

we believe to be the most useful of problem statements for the public facility problem. It is a statement that admits the multiplicity of locational objectives and seeks those spatial patterns that are superior under a number of these objectives.

The purpose of this article is to review the location analysis literature of the last few years that deals with public emergency facilities. Some of this literature represents studies directed toward specific problems of emergency medical services (EMS) systems, primarily location of ambulance dispatching points. Much of the literature, however, reports context-free models developed to address location problems in general; these models are not tied to specific applications. Nevertheless, each of the context-free models we shall discuss has been applied in planning some form of emergency facilities. We shall trace the development of the context-free models and discuss their potential for further application in emergency medical recovery systems. There is an additional body of location literature that treats locations on a plane, using the Euclidean metric; these works will not be reviewed here because we feel that the approach lacks realism for the types of problems we wish to address.

Location-Allocation Models for EMS Systems

The location-oriented literature that has been developed in the specific context of planning emergency systems may be divided into two categories: the first is concerned mainly with locational and spatial considerations and occasionally includes notions of random events that are explored in an iterative fashion through simulation or queueing. Several of these works draw on the context-free location literature. The second category addresses the event-service sequence, with emphasis on the random components of the events. Location analysis, when it appears, is generally an ad hoc rather than an integral feature of these investigations.

Spatial Analysis Applied to EMS

Focusing on the average response time, Volz [3] applied a "steepest descent" procedure to determine locations of ambulances. The procedure, unfortunately, is subject to entrapment at local optima. Although best solutions are possible, there is no way short of enumeration to verify whether a solution is inferior to the true optimum. The model's suitability for planning purposes is hindered by its requirements for data and computer time. Its complex structure makes it very difficult to introduce other considerations and additional objectives. The assumption that emergency calls arrive as a Bernoulli process (i.e., with stationary independent increments) may be unrealistic in certain contexts. Volz added other assumptions that reduce the mathematical complexity but impair the model's usefulness. Nevertheless, this model was the first to treat the interaction between location and response time in a random environment.

Fitzsimmons [4] studied the activity levels of ambulances distributed on a plane with rectangular distances separating the vehicle sites from the demand sites. He used both a simulation and a queueing model to estimate the activity levels; the mean response time was iteratively improved by relocation of the ambulances on the basis of a pattern search method, which is one kind of "descent" technique. As with Volz's model [3], globally optimal solutions could not be shown. Ambulance locations chosen by the search routine had to be modified to correspond to the actual locations of firehouses.

Berlin and Liebman [5] also emphasized a location/simulation approach, arguing that the determination of efficient location-allocation patterns for ambulances must take into account the spatial, temporal, and random distributions of demand. Their problem was subdivided into two separate parts for sequential analysis: a facility location problem and an ambulance allocation problem. For the first, they used the location set covering model developed by Toregas and ReVelle, which we shall discuss later, to determine the minimum number and the location of ambulance depots, so that all demand points were covered within stated distance levels; they then used a stochastic simulation to estimate the utilization level of ambulances allocated to each depot.

The merger of several location models with standard queueing models has been achieved by Stidham and his colleagues; their results are contained in two unpublished works [6,7]. In these works they viewed demands as arising in a random process and service as a similarly random-length event. They examined the effects of congestion on the efficiency of the system, using known results from queueing theory. They showed how constraints on the levels of the reliability of service availability and on maximum travel time give rise to familiar deterministic location models (the p-median problem and the location set covering model) for which solution methods are known.

Gladstone [8] designed a study to investigate the costs and benefits resulting from regional integration of service providers into a coordinated system of emergency medical care. This regionalization was expected to ease access to the system, increase the response capability, and reduce costs. The study examined costs and benefits using mainly process measures and some outcome measures. Although it is descriptive rather than prescriptive, the study may prove important to EMS planning, in which regionalization is often neglected.

Landau [9] explored some of the planning considerations associated with the wide variation among hospitals in ability to render high-quality emergency care. This variation implies the need for regional coordination among hospitals, so that patients can be taken within a reasonable time to hospitals appropriately equipped and staffed to provide the necessary care. This study provides an important reminder in the planning of regionalized emergency systems: comprehensive spatial-process analysis should take into account variation in hospital services.

Willemain [10] provided a formal treatment of the categorization of a region's emergency facilities by care capabilities within diagnostic classes. The need for appropriate categorization in EMS planning is apparent in light of accusations about wasteful duplications and underutilization of critical care units. Willemain's model illustrates the interrelationships among important planning variables but does not lend itself directly to empirical use.

Event-Service Analyses

A multiple-server queueing model was used by Bell and Allen [11] to determine the number of ambulances needed to achieve specified response times. They assumed that the temporal distribution of calls could be described by, and meet the assumptions of, a Poisson distribution; that is, that each event occurs independently of the preceding one. Unfortunately, the extension to multiple servers distributed across the network posed severe complications beyond the mathematical complexities because of the need to define rules of precedence and responsibilities for each area.

Savas [12] analyzed the cost-effectiveness of proposed changes in New York's ambulance service with a descriptive simulation model. The simulation was of a standard sort, developed with IBM; alternative locations were investigated by trial. The cost-effectiveness was displayed as a benefit/cost ratio of minutes saved to additional dollars expended per month. This was a straightforward but pioneering effort to use quantitative methods on an emergency service system.

Measuring benefit in lives saved, Smith [13] developed a benefit/cost analysis of a coronary emergency rescue service. He used both a queueing and a simulation model, with no appreciable difference in results. Only aggregate data were available, which reduced the reliability of the analysis.

Stevenson [14] used queueing theory to estimate dispatch delay in ambulance service operations. For a given number of ambulances, he provided order-of-magnitude estimates of the minimum expected response time, using aggregate data and a simple queueing model. The data reflected general regional characteristics such as average speed, area, and total number of calls. The results were expected to help the decision maker determine equipment allocation levels. Explicit locational considerations would probably have provided more useful results.

Chaiken and Larson [15] explored the characteristics of urban emergency services and reviewed the operational problems of these systems. Their concern was with vehicle deployment and application of queueing (chiefly) and simulation approaches to these problems. They considered allocation policies related to the number of units, location and relocation of units, response areas, and patrol patterns.

Acton [16] studied the cost-effectiveness of five alternative heart attack programs. Many of the uncertain parameters in the analysis were based on estimates by a panel of experts. He used a Monte Carlo simulation to calculate expected lives saved and associated costs of each program. The benefits due to saved lives were estimated by livelihood saved and by willingness-to-pay measures; the latter were obtained from questionnaires and require further study to gauge their validity. Cretin [17] analyzed the process of death following myocardial infarction and developed a model of risk of death from infarction. Her model combined features of both the disease process and the medical response and was used to evaluate policy alternatives. The disease was modeled as a discrete Markov process, with transition probabilities and other parameters obtained from statistical regression. This is an interesting approach, which could form an integral part of broader models.

Stevenson and Willemain [18,19] viewed the screening process as a triage system, in which calls were allocated to primary service if they were emergencies and to secondary service if they were not. Calls were also routed to the secondary service if all primary service vehicles were busy. The analysis consisted of combining assumed decision and event probabilities. Although the approach was straightforward, the assumed probabilities were not justified by the authors.

These models and conceptualizations represent the state of the art as reflected by the EMS-oriented literature up to about one year ago. It is clear that there is great potential for further application of location models to EMS problems. We believe such applications will draw heavily on the context-free location literature specifically on the maximal covering model and its refinements, which are discussed below.

Context-free Location Analysis

An interlocking set of three network location models will be described here. All three models have been applied, at least conceptually, to emergency services. The third model (the maximal covering model), whose further development and refinement we shall discuss, has recently been applied to the location of fire stations and equipment in Baltimore, MD. This is, to our knowledge, the first application of this model to such a setting, although it is characterized by the most versatile and adaptable objective of all three models.

The *p*-Median Problem

The location model that underlies, theoretically and historically, the others we shall discuss is the *p*-median problem, which was proposed by Hakimi [20,21] in the context of locating switching centers in a communications network. It was recognized, however, that the problem statement could be extended to many other frames of reference. We will state it in the setting in which people come to facilities for service, but it can easily be turned around to pertain to the situation of servers going from facilities to points of demand.

Locate *p* facilities on a network of demands so that the average travel time of all users is a minimum. Since the facilities are not distinguished by size or specialization, it is assumed that every user will travel to his nearest facility.

Numerous researchers have proposed solution methods for the *p*-median problem, and solution procedures are important: a well-posed problem statement without an adequate solution procedure is of little use. (Nevertheless, our primary concern has been with problem statements. Capturing the location objectives of decision makers has been a consistent theme of our research.)

Hakimi did not, in fact, contribute a solution procedure, but his posit of the problem did stimulate solution approaches. The first explicit attack on the problem was a heuristic procedure suggested by Maranzana [22]. Another heuristic procedure, which utilized node substitution, was described by Teitz and Bart [23]. This method approached the optimum mathematical solution more closely than did the method of Maranzana.

The first procedure that yielded mathematically optimal solutions to the p-median problem was provided by ReVelle and Swain [24]; they showed that linear programming could be used to derive solutions to even fairly large problems. The sequence of development from Hakimi's theorems through the heuristics of Maranzana and of Teitz and Bart to the linear programming formulation of ReVelle and Swain has been reviewed by ReVelle, Marks, and Liebman [2].

In 1972 Järvinen, Rajala, and Sinervo [25] proposed a branch-and-bound algorithm for the p-median problem, and El-Shaieb [26] followed with a related algorithm. Neither of these investigations, however, reported reaching the problem sizes that can be handled by linear programming. Two methods using the technique of decomposition were devised by Swain [27] and by Garfinkel, Neebe, and Rao [28] on the basis of the formulation of ReVelle and Swain; Khumawala suggested a heuristic procedure [29]. A probabilistic branch-and-bound solution procedure was devised by Swoveland et al. [30] when they chose the p-median framework for a study aimed to minimize ambulance response time in Vancouver, Canada. They used simulation output for successive approximations of mean response times.

Carbone [31] extended the p-median concept to the situation in which demand for service in a particular area arises randomly with some mean and variance. For this realistic setting, Carbone showed how to apply chance-constrained programming to determine facility location patterns. His consideration of the uncertainty in the demand structure can be extended to the other models we shall discuss; it provides a potentially useful tool to apply in the EMS setting.

Although the p-median problem is of interest, its emphasis on mean response time or mean travel time is not entirely sufficient in the context of emergency services because it makes no provision for the extremes of travel or response time. Hence it is entirely possible that a solution to this problem may leave some points of demand at excessive times from the nearest facility. This difficulty motivated Toregas et al. [32] to modify the p-median problem by the addition of maximum time constraints to ensure that no point of demand would be left with its closest facility further than the maximum time away. Rushton, Goodchild, and Ostresh [33] have since added the same feature to the Teitz and Bart heuristic.

The Location Set Covering Model

A possible outcome of gradually tightening the maximum time constraints is that, at some point, the p facilities will become insufficient in number to cover all points of demand within the time constraint. The logical question that follows is, given the time standard, what is the least number of facilities required? This question led to the formulation and solution of the location set covering problem by Toregas et al. [32]. A statement of the problem is:

> Find the minimum number of facilities and their locations such that each point of demand has a facility within S time units.

Interestingly, several specialists in emergency health services have also posed this problem, although they were unaware of the tractable mathematical structure underlying it. Huntley [34] gave the following verbal framework:

> How many ambulances? Are there enough? What is an acceptable response time? Usually in a metropolitan area, it is 15 minutes. . . . If a 15 minute response time is demanded, how many ambulances are required; where must they be positioned to provide reasonable assurance that this criterion is met?

The same fundamental questions were asked by Cassidy and Wilson in their study of the London Ambulance Service [35]:

> How many vehicles should the Service have and how should the total fleet be divided between stretcher ambulances and the various sizes of sitting-case vehicles? . . . Where should the vehicles be located . . . so as to bring a vehicle quickly to the scene of an emergency?

In addition to the linear programming solution method, Toregas and ReVelle suggested a simple, fast, and exact procedure [36,37] for solving location set covering problems. The procedure, known as "reductions," has made possible the solution of problems up to 50 times the size of the largest p-median problems that have been solved.

The point, however, is that the location set covering problem is an attractive statement of the needs of an EMS system. Recall that the time standard was applied to the time between the dispatch station and the point of demand. ReVelle, Toregas, and Falkson [38] extended this formulation to the location of ambulances; they showed how the problem could be modified to account for the additional service link from the demand point to the hospital where care is finally rendered.

This model is referred to in the literature as the location set covering model because it derives, for each point of demand, a *set* of facility sites that are within the distance standard of and are thus eligible to *cover* that demand point. It is a fundamental requirement of the model that at least one site must be chosen from the eligible locations in each set, thus ensuring the coverage of each demand point. The objective of the model is to determine the fewest number of facilities and their locations. A facility site that is eligible to cover a number of demand points is generally a more efficient choice than a site with only a few appearances in the many demand sets.

When the location set covering problem was first stated and solved, its precision of problem statement, its strong relation to stated needs, and its straightforward solution method quickly established it as a widely recognized model for the location of facilities. The successful efforts of Public Technology, Inc. in promoting this model for locating fire stations fostered this recognition. The model has been requested for use in nearly 100 cities.

The location set covering model was the first to use the maximum distance (or time) as a determinant of the spatial configuration of facilities. As such, it was the first to meet the need for distance or time standards that had been stated in the areas of ambulance deployment [34] and fire protection [39].

Jarvis, Stevenson, and Willemain [40] analyzed an EMS transportation plan in a four-county region encompassing the Wheeling, West Virginia, SMSA. The objective of their project was to determine the minimum number of ambulances required to meet coverage constraints stated by a planning authority. The constraint required response to 95 percent of requests within a specified time. They used the location set covering model and found the reduction algorithm [36,37] very useful in problem solving. They reported that the model was flexible enough to allow modifications due to nontechnical constraints and generated several location schemes that reflected operational criteria in conjunction with political considerations. The simple form of the model made it difficult, however, to meet other guidelines of the planning authority. These further needs could probably have been met by the maximal covering model discussed below.

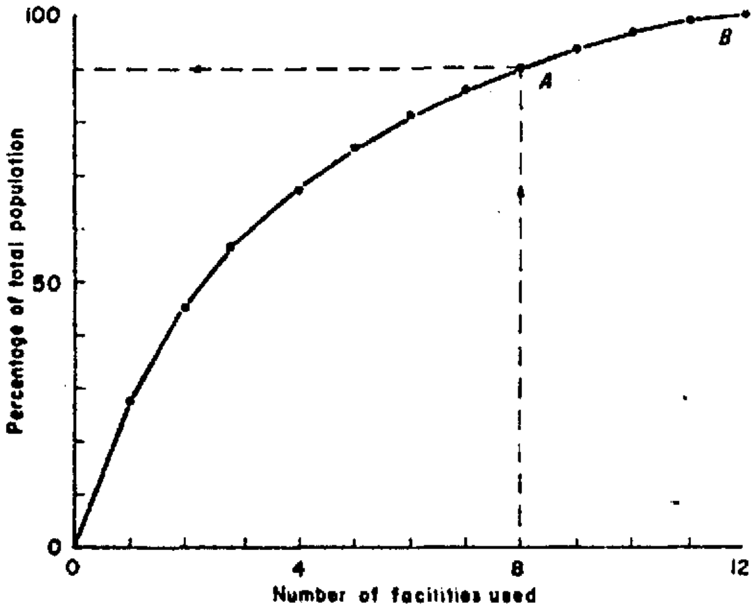
Maximal Covering Model

The p-median problem and the location set covering model were both benchmarks in the development of location models. Exploration continued, however, and a new model, promising significantly greater utility than either of these models, was devised. The new entry, the maximal covering model, was developed by Church and ReVelle [41]. The statement of the maximal covering location problem is:

> *Allocate p facilities to positions on the network so that the maximum population will find service within a stated time or distance standard.*

Solutions to the p-median problem minimize the product of population and travel time for a given number p of facilities. The location set covering model ignores population and finds the minimum number of facilities needed to serve a set of demand points within a response-time standard. The maximal covering model readmits the importance of population and retains a response-time standard the

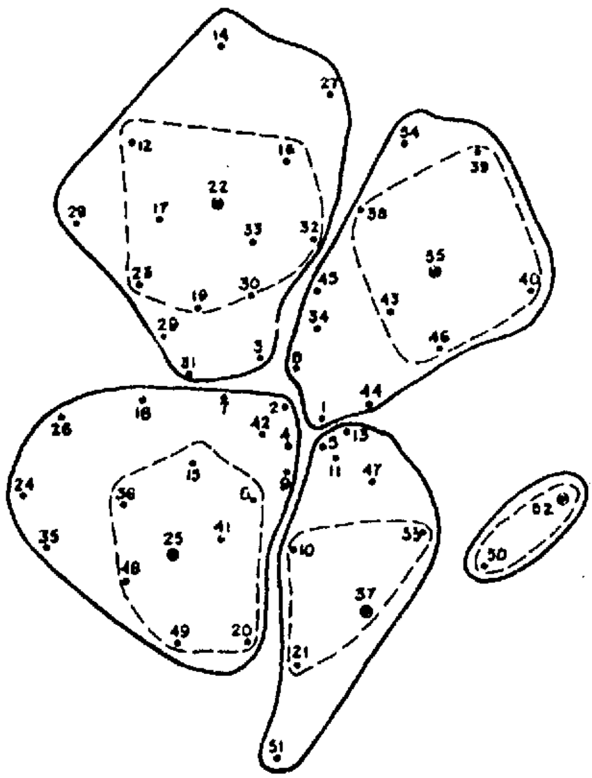
Fig. 1. Maximal covering model results: Percentage of population covered within 2.0 time units by various numbers of optimally placed facilities. Point B is a solution to the location set covering problem.



facilities are to be placed so as to cover as much of the population (or as many demand points) as possible. When p, the limit on facilities, is increased to the point that 100 percent of the demand is covered by the maximal covering model, its value is the same as the minimum number indicated by the location set covering model; the facility placements indicated by the two models, however, will not necessarily be the same.

The maximal covering model fills an important need in location analysis. It does not entail the excessive computational burden of p-median problems and can deal with problems as large as those handled by the location set covering model, using readily available linear programming codes. One great advantage of the maximal covering model is that it can indicate the maximum extent of demand coverage to be expected from a number of facilities less than the minimum needed to cover all demand. For example, Fig. 1 displays, for a particular set of demand points, the proportion of population covered within a distance standard by different numbers of optimally placed facilities. The graph was generated from 12 solutions of the maximal covering model. Point B on the graph is a solution to the location set covering problem: 12 facilities are needed to cover 100 percent of the demand. The remaining points, e.g., point A, are results the location set covering model could not have produced: eight optimally

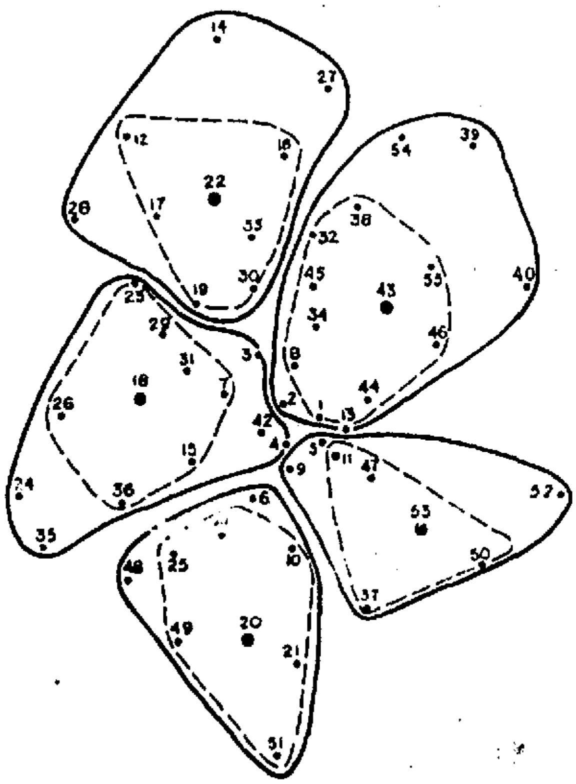
Fig. 2. A solution to the location set covering problem. Facilities (circled dots) are placed so that all demand points and 100 percent of the population are within 15-unit limits (solid lines) from some facility. Dashed lines are 10-unit limits from each facility; they enclose only 31 percent of the population.



placed facilities will cover 90 percent of the demand. The question is immediately apparent: are the four additional facilities worth their cost, when they cover only the last 10 percent of the demand? The location set covering model would have yielded only the information underlying point B of the graph.

Solutions available from the location set covering model are rigid, in that all points *must* be covered; in addition, the solutions are not "tight." Extensive computational experience indicates that the minimum number of facilities to provide cover may be arranged in many ways without significantly violating the fundamental coverage standard. Application of the maximal-covering model can often improve on solutions obtainable from the location set covering model. Consider the pattern of demand points shown in Fig. 2. This

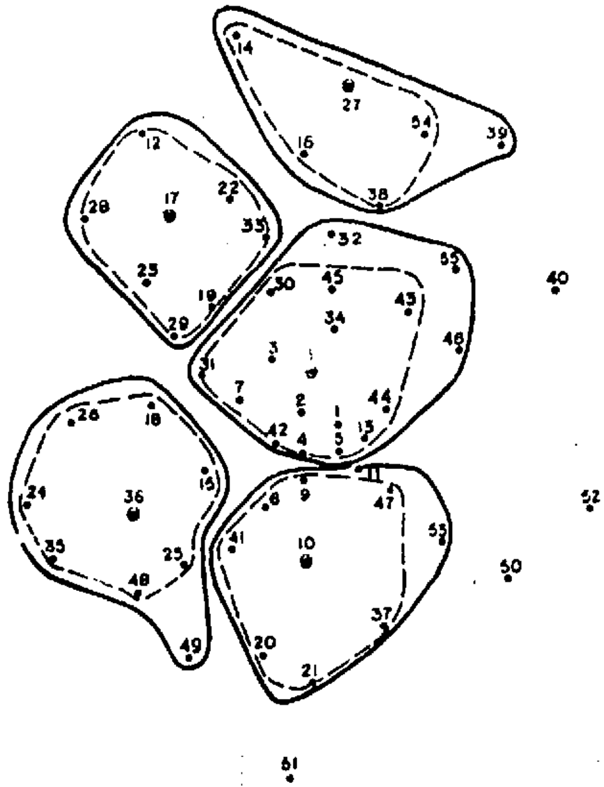
Fig. 3. A maximal covering solution to the location problem of Fig. 2. Facilities (circled dots) are placed so that all demand points and 100 percent of the population are within 15-unit limits (solid lines) from some facility. Dashed lines are 10-unit limits from each facility; they enclose 55 percent of the population.



is one of the many possible solutions to the problem of minimizing the number of facilities (five) needed to cover all points within 15 time (or distance) units. The clusters surrounded by solid lines include points within 15 units of the facility to which that cluster of points is assigned. The dashed lines surround points that are within 10 units of each cluster's facility; these points represent only 31 percent of the area's population. The remaining 69 percent of the population, although it falls within the 15-unit limit, is more than 10 units away from any facility.

Figure 3 represents another solution to the problem, obtained with the maximal covering model. This solution was computed with the same number of facilities, and all points still lie within 15 units of their closest facility. The computation, however, sought to maxi-

Fig. 4. Another maximal covering approach to the location problem of Fig. 2, with 15-minute constraint relaxed. Facilities (circled dots) are placed so that 95 percent of the population is within 10-unit limit (dashed lines) from some facility; 98 percent is still within 15-unit limits (solid lines).



mize the population within 10 units, given the constraint of keeping the entire population within 15 units. In the new configuration 55 percent of the population is within 10 units of a facility. Although the location set model might have found this solution by chance, the probability of its doing so is low; in general, only explicit optimization using the maximal covering model can be expected to yield such solutions.

If the rigid constraint of complete coverage within an outer limit is relaxed, alternative configurations with possibly more useful characteristics may be derived. Figure 4 depicts a solution obtained with the standard of 15 units discarded. Five facilities can cover *95 percent* of the population within 10 units while leaving only 2 percent of the population outside a 15-unit perimeter.

Multi-objective Maximal Covering and Fire Station Location

In the past year we have been assisting the fire department of the city of Baltimore in planning its future system configuration. The National Board of Fire Underwriters (now the Insurance Services Organization) has stated guidelines for fire protection [39] that are similar to those given by the EMS regulations except that they are given in terms of distance. The fire standards are absolute: they do not include the 95-percent specification that gives the EMS standards their realistic flexibility. In practice, however, it is accepted that absolute attainment of the standards is not within reach; thus the maximal covering model is appropriate and useful in this setting.

Although the fire standards are expressed for coverage of property, coverage of the population is another obvious objective. The fire hazard to population and the fire hazard to property, determined as the product of population (or property value) and fire frequency for subareas of the city, constitute still other measures against which to gauge coverage. Coverage of fires, as reflected in fire frequency, is also an important objective. These issues were worked out in consultation with the fire department, and coverage of property value, property value hazard, population, population hazard, and fires became five objectives to be maximized within the distance standards.

Recognition of multiple objectives in the system directed our work toward seeking those configurations that satisfied all the objectives to a high degree and examining, for each configuration, the trade-offs between the various objectives, rather than merely recommending a unique theoretical optimum. In ambulance deployment problems too, multiple goals are likely to be identified. We believe that more realistic planning can result from examination of the achievements of each alternative relative to a number of objectives.

Equipment Assignment. Another aspect of the Baltimore fire study was that different distance standards exist for fire trucks and for fire engines, and both standards had to be met simultaneously. The problem was further defined by the constraint that engines and trucks must be located at stations, which were the entities to be placed on the network. The problem, then, was one of locating facilities and then assigning specialized equipment to the facilities to achieve coverage for an area by the proximate presence of the equipment.

There are parallels to this situation in the EMS system. Coverage of an area might well be defined by the simultaneous presence of different services (e.g., helicopters and ambulances with telemetry) within time standards for that area. The ability of the maximal covering model to assign the locations of equipment is likely to be a useful tool in the EMS setting.

The Maximal Covering Model and EMS Needs

The maximal covering location problem is of special and important interest to EMS research because its solution explicitly uses the evaluative structure specified by regulations associated with the EMS

Act of 1973. The meaning of an "adequate number" of emergency vehicles, called for in the legislation, is interpreted in the regulations as a number sufficient that 95 percent of requests for assistance can be met within 30 minutes in rural areas and 10 minutes in urban areas. To the location analyst, the framers of the regulations for emergency services appear to be among the most articulate of decision makers, for their specifications of effectiveness measures are clear, concise mathematical entities. The analyst is seldom fortunate enough to encounter such precise problem statements.

The coincidence of viewpoints of the framers of the regulations and the authors of the maximal covering model suggests, correctly, that this model was structured to meet real needs. In fact, the origin of the model lay in a 1970 report of the Carnegie Commission on Higher Education [42]. That report called for establishment of 55 health care centers across the nation, located in such a way that 95 percent of the population was within an hour's drive of a facility. The numbers aside, this concise statement sparked the development of the maximal covering problem by Church and ReVelle [41] out of the location set covering problem.

The maximal covering model offers additional information beyond that requested by the EMS regulations. As discussed previously in connection with Fig. 1 (p. 137), the maximal covering model can indicate not only the number and locations of vehicles required for 95 percent of responses to occur within 30 minutes, but also the number and locations needed for any coverage level one wishes, allowing examination of the trade-off between the number of optimally located vehicles and demand coverage.

There are two other features of the maximal covering model that deserve attention. As Fig. 3 (p. 139) implicitly indicates, solutions to the location set covering problem are not unique; there are many ways in which all points of demand could be covered by a given number of facilities within a stated time. Surprisingly, the same is true of solutions to the maximal covering problem, as we have learned from our computational experience in incorporating secondary objectives into this model. If many solutions exist in which, for example, seven vehicles cover 95 percent of demands in 30 minutes, each of these solutions can be evaluated relative to secondary criteria, such as the percentage of demands that can be responded to within 15 minutes. In fact, for a given number of vehicles, one can explicitly maximize the fraction of demand covered within an inner time of 15 minutes while maintaining the constraint that 95 percent of demands are covered within 30 minutes. Extremely desirable solutions can be found in this way.

A parallel use of the maximal covering model is to append secondary outer standards to the basic response time standards. For example, the available vehicles may first be required to provide service to 95 percent of demands within 30 minutes, after which one can maximize the proportion of calls that can be answered within 45 minutes. The use of secondary inner standards may benefit most of a population center; the use of secondary outer standards could ameliorate the situation of those who represent the 5 percent of demand that is exempted from the regulation standard.

Still other uses of the maximal covering model are possible: after maximizing coverage within a stated time, one may seek the solution that minimizes the average service time for those outside the standard. The economic and geographic options available to planning regions differ enough that each region might benefit, both in service and in cost, from "fine-tuning" its EMS plan by examining many such secondary solutions within the regulation standards.

Some work has already been done on applying the other two context-free models, the p-median and the location set covering models, to emergency ambulance services. The work of ReVelle, Toregas, and Falkson [38] cited earlier extends the location set covering model to two more detailed aspects of the EMS location problem. One of these aspects concerns travel time on the link from demand point to care facility after the patient is picked up, which needs to be considered in determining ambulance locations. The other aspect is that demands in the real world tend to occur along continuous roads rather than at discrete points. Berlin, Re elle, and Elzinga [43] used the p-median framework to investigate the minimum average time when the travel from demand to point of care is included. We have not yet applied the maximal covering model to these more detailed expressions of the problem, but it will be useful to discuss here some possibilities inherent in such an application. These possibilities sh uld be of interest to EMS policy makers, since they allow the evaluative standards discussed in the regulations to be extended to include the link from demand point to point of care.

When an ambulance arrives at the scene of a call, certain limited care may be given the patient, depending on the training and equipment of the arriving personnel. The time standards for response to calls apparently apply only to the arrival of the ambulance at the scene. The purpose of the ambulance is twofold, however, and the second function is transporting the patient to a staffed and equipped emergency care center.

Without meaning to suggest different values for the time standards, it would seem useful to apply a time standard to the entire time between dispatch of the vehicle and the patient's arrival at the emergency care facility. If the 30-minute response standard were extended to cover both functions, more vehicles and locations would be required to achieve the stated coverage. Two standards might be formulated, one for the time from ambulance dispatch to the emergency scene, the other for the entire time from dispatch to the appropriate site of care. The addition of this new standard provides another yardstick with which to measure the effectiveness of an emergency recovery system. Modeling with this added standard does not seem to present any insurmountable challenge.

Still another standard could be applied for the time from demand scene to hospital. Application of such a standard, however, is not possible within the location concepts discussed thus far, which focus on the dispatch location of emergency vehicles. Achievement of a standard for the demand-point-to-hospital leg of the transportation triangle (depot to demand to hospital to depot) can only be influenced by the locations of emergency care hospitals.

The locations of such hospitals in turn affect the vehicle locations needed to achieve standards on the entire depot-to-point-of-care time. The twin problems are interlocking and complex, but potentially solvable. Together the two problems of ambulance location and location of emergency care facilities comprise almost all of the transport segment of emergency care. Seen as two interlocking problems, they give rise to a more basic question:

> With limited resources for purchase and operation of vehicles and for establishment and operation of emergency care facilities, determine the number and locations of both that will maximize the proportion of responses occurring within the time standard.

A complementarity of achievement exists between emergency vehicles and emergency care facilities. An additional emergency care facility might obviate the need for four (to choose a number) emergency vehicles by diminishing the total time from dispatch to final care site. The maximal covering model will permit these trade-offs to be examined. The relative costs of mobile and fixed facilities would indicate which choices were appropriate and open the possibility of cost-effectiveness comparisons that could lead to EMS systems that would be both functionally and economically optimal.

Summary and Conclusions

We have discussed the remarkable harmony between the structure of the maximal covering model and the goals of policy makers. The ability of the model to deal with multiple objectives and to optimally assign special equipment to optimally-located depots, as in our fire department study, indicates the model's general utility in a wide range of location problems. It should be particularly useful in planning and evaluating emergency medical systems.

In location research (perhaps in other types of research as well) a symbiosis exists between the setting of a particular application and the development of models. Location models aid in conceptualizing a problem in a particular setting, and the special characteristics of the setting feed back to enrich the location model. We expect such symbiosis from the EMS setting. We believe that location analysis can make a significant contribution to the evaluative structure of such services and that this setting will add new and important dimensions to location research.

REFERENCES

1. Lea, A. An annotated bibliography of location-allocation. Working paper, Department of Geography, University of Toronto, 1973.

ReVelle, C., D. Marks, and J.C. Liebman. An analysis of private and public sector location models. *Manage Sci* 16:692 July 1970.

3. Volz, R.A. Optimum ambulance location in semi-rural areas. *Transp Sci* 5:198 May 1971.
4. Fitzsimmons, J. A methodology for emergency ambulance deployment. *Manage Sci* 19:627 Feb. 1973.
5. Berlin, G. and J.C. Liebman. Mathematical analysis of emergency ambulance location. *Socio-Economic Plann Sci* 8:323 Dec. 1974.
6. Stidham, S., C. Toregas, and G. Fisher. Operational analysis of demand-responsive systems. Department of Environmental Systems Engineering, Cornell University. Report to the New York State Science and Technology Foundation, Jan. 1971.
7. Stidham, S. Stochastic design models for location and allocation of public service facilities. Department of Environmental Systems Engineering, Cornell University. Report to the New York State Science and Technology Foundation, May 1971.
8. Gladstone, R. *A Study to Determine the Impact of EMS Regionalization.* Santa Monica, CA: Rand, Jan. 1975.
9. Landau, T. *Considerations of Regionalization and Categorization in Hospital Emergency Planning.* Master's thesis, Massachusetts Institute of Technology, 1975.
10. Willemain, T. Analysis of a model of emergency facility categorization. Working paper. Department of Urban Studies and Planning, Massachusetts Institute of Technology, June 1975.
11. Bell, C. and D. Allen. Optimal planning of an emergency ambulance service. *Socio-Economic Plann Sci* 3:95 Aug. 1969.
12. Savas, E. Simulation and cost-effectiveness analysis of New York's emergency ambulance service. *Manage Sci* 15:B608 Aug. 19[illegible]
13. Smith, B.G. Coronary emergency rescue services treated as a queueing system—A cost-effectiveness analysis. Paper presented at the 27th annual meeting of the Operations Research Society of America, Washington, DC, Apr. 1970.
14. Stevenson, K. *Operational Aspects of Emergency Ambulance Services.* Technical Report No. 61. Cambridge, MA: M.I.T. Operations Research Center, May 1971.
15. Chaiken, J. and R. Larson. Methods for allocating urban emergency units: A survey. *Manage Sci* 19:P110 Dec. 1972.
16. Acton, J. *Evaluating Public Programs to Save Lives: The Case of Heart Attacks.* R-950-RC. Santa Monica, CA: Rand, Jan. 1973.
17. Cretin, S. *A Model of the Risk of Death from Myocardial Infarction.* Technical Report No. TR-09-74. Cambridge, MA: M.I.T. Operations Research Center, Nov. 1974.
18. Stevenson, K. and T. Willemain. *The Attitudes and Perceptions of Groups Engaged in the Provision of Emergency Medical Services.* Technical Report No. TR-11-74. Cambridge, MA: M.I.T. Operations Research Center, Dec. 1974.
19. Stevenson, K. and T. Willemain. *Analyzing the Process of Screening Calls for Emergency Ambulance Service.* Technical Report No. TR-08-74. Cambridge, MA: M.I.T. Operations Research Center, 1974.
20. Hakimi, S. Optimum locations of switching centers and the absolute centers and medians of a graph. *Oper Res* 12:450 May–June 1964.
21. Hakimi, S. Optimum distribution of switching centers in a communications network and some related graph theoretic problems. *Oper Res* 13:462 May–June 1965.
22. Maranzana, F. On the location of supply points to minimize transport costs. *Oper Res Q* 15:261 Sept. 1964.
23. Teitz, M. and P. Bart. Heuristic methods for estimating the generalized vertex median of a weighted graph. *Oper Res* 16:955 Sept.–Oct. 1968.
24. ReVelle, C. and R. Swain. Central facilities location. *Geogr Anal* 11:30 Jan. 1970.
25. Järvinen, P., J. Rajala, and H. Sinervo. A branch-and-bound algorithm for seeking the p-median. *Oper Res* 20:173 Jan.–Feb. 1972.
26. El-Shaieb, A. A new algorithm for locating sources among destinations. *Manage Sci* 20:221 Oct. 1973.
27. Swain, R. A parametric decomposition approach for the solution of uncapacitated location problems. *Manage Sci* 21:189 Oct 1974.
28. Garfinkel, R., A. Neebe, and M. Rao. An algorithm for the m-median plant location problem. *Transp Sci* 8:217 Aug. 1974.

29. Khumawala, B. An efficient algorithm for the p-median problem with maximum distance constraints. *Geogr Anal* 5:309 Oct. 1973.
30. Swoveland, C., D. Uyeno, I. Vertinsky, and R. Vickson. Ambulance locations: A probabilistic enumeration approach. *Manage Sci* 20:686 Dec. 1973.
31. Carbone, R. Public facilities location under stochastic demand. *INFOR* 12:201 Oct. 1974.
32. Toregas, C., R. Swain, C. ReVelle, and L. Bergman. The location of emergency service facilities. *Oper Res* 19:1363 Oct. 1971.
33. Rushton, G., M. Goodchild, and L. Ostresh. *Computer Programs for Location-Allocation Problems*. Monograph No. 6, Department of Geography, University of Iowa, July 1973.
34. Huntley, H. Emergency health services for the nation. *Public Health Rep* 85: 517 June 1970.
35. Cassidy, P. and P. Wilson. Operational research in the London ambulance service: Simulation of total service. *Greater London Counc Res Intell Unit Q Bull* No. 9, p. 10 Dec. 1969.
36. Toregas, C. and C. ReVelle. Location under time or distance constraints. *Pap Reg Sci Assoc* 28:133 Fall 1972.
37. Toregas, C. and C. ReVelle. Binary logic solutions to a class of location problems. *Geogr Anal* 5:145 April 1973.
38. ReVelle, C., C. Toregas, and L. Falkson. Applications of the location set covering problem. *Geogr Anal* 8:65 Jan. 1976.
39. National Board of Fire Underwriters. *Standard Schedule for the Grading of Cities and Towns of the U.S. with References to their Fire Defenses and Physical Conditions*. New York: National B[illegible]rd of Fire Underwriters, 1956.
40. Jarvis, J., K. Stevenson, and T. Willemain. *A [illegible]mple Procedure for the Allocation of Ambulances in Semi-Rural Areas*. Technical Report No. 13-75. Cambridge, MA: M.I.T. Operations Research Center, Mar. 1975.
41. Church, R. and C. ReVelle. The maximal covering location problem. *Pap Reg Sci Assoc* 32:101 Fall 1974.
42. Carnegie Commission on Higher Education. *Higher Education and the Nation's Health*. New York: McGraw-Hill, 1970.
43. Berlin, G., C. ReVelle, and J. Elzinga. Determining ambulance locations for on[illegible] hospital care. *Environ Plann A* 8:553 Aug. 1976.

#17

INTERFACES
Vol. 9, No. 5, November 1979

0092-2102/79/0905/0003$01.25

THE LONG ISLAND BLOOD DISTRIBUTION SYSTEM AS A PROTOTYPE FOR REGIONAL BLOOD MANAGEMENT

Eric Brodheim

Lindsley F. Kimball Research Institute, The New York Blood Center, New York, New York 10021

and

Gregory P. Prastacos

Department of Decision Science, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

ABSTRACT. Each year over two million hospitalized Americans depend upon the timely availability of the right type of blood products at 6,000 hospital blood banks (HBB's) in the United States. If the right blood products are not available at the HBB when required, then medical complications or postponements of elective surgery can result which translates to extra days of hospitalization and expense. On the other hand, since most blood products may only be administered to a patient of the same blood type within 21 days of collection, overstocking at HBB's leads to low utilization, which increases costs and is wasteful of the scarce blood resource.

The Long Island blood distribution system was set up as a prototype of a regional blood center and the hospital blood banks that it services collaborating to preplan regional blood flow. It maximizes blood availability and utilization according to a Programmed Blood Distribution System (PBDS) model and strategy that has been shown to be generally applicable. PBDS schedules blood deliveries according to statistical estimates of the needs of each HBB and monitors actual requirements to adjust deliveries when indicated by control chart techniques. In addition, it provides a daily forecasting of short-term shortages and surpluses for the next several days that results in controlled movement of blood to and from adjoining regions. Finally, the system is able to adjust the regional strategy so that availability is reduced uniformly at all HBB's during periods of seasonal, regional shortages.

PBDS has drastically improved utilization and availability of blood on Long Island: wastage has been reduced by 80%, and delivery costs by 64%. This prototype is acting as a model for other regional blood centers in the United States and for other national blood services as a basis for planning and controlling blood flow in a geographic area. It usually replaces preexisting procedures where a regional blood center collects blood based upon gross estimates and reacts to requests for blood by individual HBB's on the basis of experience and on the currently prevailing inventory situation.

Introduction

The Operations Research Laboratory of the New York Blood Center, in collaboration with the Long Island Blood Services (LIBS) division of the Greater New York Blood Program (GNYBP), has been studying the problem of providing a high availability of perishable blood products at each of the hospitals in a region, while

HEALTH CARE—BLOOD BANK

assuring that the maximum number of these products are utilized during their 21-day lifetime. This problem requires that the dual and seemingly conflicting concerns about availability and utilization be reconciled. It also requires a radical change in management decision concepts so that the regional blood center (RBC) that is responsible for collecting and distributing blood and the hospital blood banks (HBB's) that stock it for possible use recognize common objectives and collaborate in implementing a strategy that optimizes the regional utilization and availability of these scarce products.

The first section of this paper describes the national problem. The Management Science techniques that were utilized to create a transferable model for a distribution plan with near-optimal characteristics are described in the second section. This is followed by a description of the implementation of this model by the Programmed Blood Distribution System (PBDS) and of its testing at LIBS as a regional prototype for the evolution of a national blood distribution network and operation. The managerial and economic impact and the implications and extensions of this prototype model are then discussed in the concluding sections.

Background

The national problem

Most blood products in the United States are derived from whole blood that is collected in units of one pint from volunteer donors by approximately 200 RBC's. After laboratory processing and testing, blood products derived from the whole blood are distributed to the HBB's in the region where they are stored to be available for transfusion when requested. This paper is restricted to whole blood and red blood cells (i.e., whole blood from which plasma and other components have been separated), which together account for over 95% of all transfusions. Both have a lifetime of only 21 days during which they can be transfused to a patient of the same type and after which they have to be discarded.

Historically, HBB's have maintained high inventories of most of the eight different types of each of these products in order to provide high availability to satisfy patient needs and have accepted the low utilization resulting from spoilage. Consequently, the national utilization rate of whole blood and red blood cells prior to expiration was estimated to be only 80% in 1974. At that time, the federal government adopted a national blood policy that called for an all volunteer blood supply to be accessible to all segments of the public. The blood supply was to be efficiently administered through the formation of regional associations of blood service units in each of which a RBC and the HBB's that it serves would collaborate to achieve these objectives.

The role of Management Science

It had long been recognized that Management Science techniques were required to improve blood utilization and availability. Strategies with desirable characteristics were formulated for individual hospital blood banks and pragmatic regional approaches that improved performance in a certain region had been implemented and reported. The National Heart, Lung and Blood Institute Resources Panel in its 1973 recommendations, which were a major factor in formulating the National Blood Policy [4], stated that:

> One of the strongest arguments favoring the introduction of systems management to blood service is that it will result in improved service, in more economical utilization of the blood resources, and most important, in improved effectiveness and efficiency. For these reasons it is recommended that emphasis be placed on developing measures of activity and relating these to objectives. Only in this way will it be possible to obtain secure and credible evaluation of the changes to be made.

This study addressed these issues by:

(1) creating a model for relating measures of blood banking activity to availability and utilization;

(2) developing a management tool called the Programmed Blood Distribution System (PBDS) to implement a regional blood distribution strategy that is based upon this model;

(3) evaluating the effectiveness of PBDS in a prototype region (LIBS).

Complexity of the problem

The complexity of the blood distribution problem is primarily due to the perishable characteristics of blood, to the uncertainties involved in its availability to the RBC, and in the variable daily demand and usage for it at each of the HBB's. The situation is complicated by the large variation in the size of the HBB's supplied, the incidence of the different blood groups, and the requirements for whole blood and red blood cells at individual hospitals.

Since it is a national policy for blood to be derived from volunteer donors, its availability is uncertain and is a function of a number of factors that cannot be controlled by the RBC. The demand and usage of blood at HBB's are also uncertain and vary from day to day and between hospital facilities. The HBB's within a region may range from those transfusing a few hundred units per year to those transfusing tens of thousands of units per year. The most frequently occurring blood type (O positive) occurs in approximately 39% of the population, while the least frequently occurring blood type (AB negative) occurs in only about 0.5% of the population. While most medical authorities agree that at least 90% of all blood transfusions could be in the form of red blood cells, some hospitals transfuse almost entirely red blood cells while others transfuse entirely whole blood with the ratio of whole blood to red blood cells frequently changing with time as transfusion practices improve.

Approach

The transfusion services throughout the nation are characterized by diversity. Each RBC has independently evolved its own philosophies and techniques for blood distribution. Each region strives for "self-sufficiency" in supplying the blood needs of the hospitals in its region from donors who also reside in approximately the same area. Because of these factors, it is essential that any strategy devised be defensible from the point of view of both the RBC and each of the wide range of HBB's that it serves. Furthermore, any strategy that involves interactions between RBC's must provide for clearly defined benefits for all participants. In addition, it is desirable that the implemented strategy be characterized by two management concepts: rotation of blood products between HBB's, and prescheduled deliveries to the HBB's.

Any strategy that allocates blood products to be retained until transfused or outdated will result in low utilization, especially in the case of the small usage HBB's which, in aggregate, account for the largest part of overall blood usage. Consequently, some form of blood "rotation" is required whereby freshly processed blood is sent to a HBB, from which it may be returned, some time later, for redistribution according to the regional strategy. It is also desirable that a significant portion of the periodic deliveries to the HBB's be prescheduled. This way the uncertainty of supply faced by the HBB's is reduced, with a resulting improvement in the planning of operations, and the utilization of their resources.

The model

The blood needs of a HBB can be expressed in terms of the *demand* for blood (i.e., the number of units required to be on hand for possible transfusion to patients) and the *usage* of blood (i.e., the number of units transfused to patients). A model is required that translates demand and usage to availability and utilization as functions of the RBC blood distribution policy and the HBB's blood stocking policy. Such a model was established by a combination of statistical analysis and Markov chain modeling. This model was then used to derive regional allocation strategies with desirable properties regarding availability and utilization.

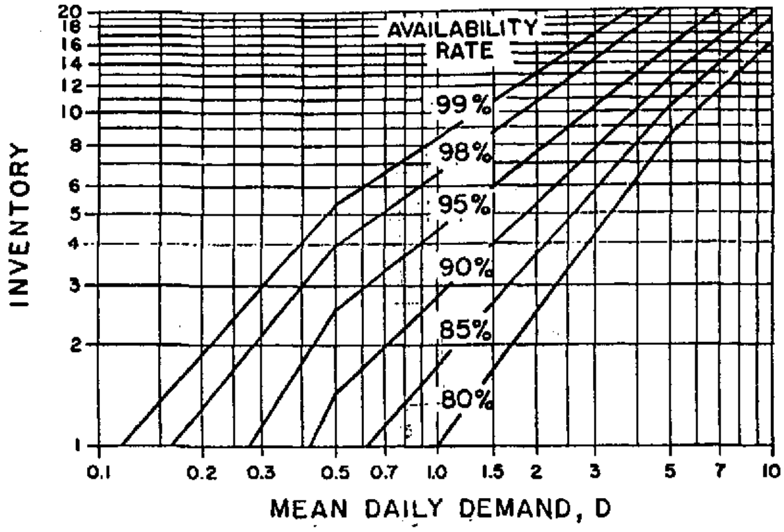
Availability

The availability rate (i.e., fraction of days when the inventory of a given blood type on hand is sufficient to meet the demand) at the HBB depends only upon the statistical pattern of demand and the total inventory level. To establish this relationship, data were collected on the daily demands for each blood type at a number of HBB's. These data, together with comparable information published by other researchers, provided a total of 49 data sets, each containing the daily demands for one blood type at one HBB for a period of at least six months.

Statistical analysis [2] established a "universal" piecewise linear relationship between inventory level (I) and mean daily demand (D)* with availability rate as a parameter as is illustrated in Figure 1. The ability of the model to predict the availability rate was verified by collecting additional daily demand data from 21 HBB's in collaboration with five RBC's throughout the country. The model predicted the availability rate to within approximately 10% of actual experience for availability rates in the range of 80% to 99%. This model was then utilized to establish the "acceptable" range of availability rates for HBB's. This was done by requesting a number of HBB's to concurrently provide estimates of their mean daily demand and of the inventory levels in each of the eight blood types which they considered adequate. In almost all cases, the levels that the HBB's considered adequate turned out to correspond to availability rates of between 90% and 95%.

*The mean daily demand is computed as 1/6th of the mean weekly demand based upon the observation that the mean demand for the three-day weekend (i.e., either for Friday, Saturday, and Sunday or for Saturday, Sunday and Monday) is close to twice the mean demand for each of the other four days in all instances.

FIGURE 1. Availability Rate Model.



It was similarly established that the daily usage could be modeled as an exponential-type distribution whose parameter is related to the mean daily usage (U). The demand-to-usage ratio (D/U) was found to vary between 1.5 and 4.0 for most HBB's with an average value of about 2.5 in most regions. These analyses showed that the parameters for the models of demand and usage could be readily estimated from records maintained by HBB's and further that availability rate could be reliably estimated by the model.

Utilization

The utilization rate (i.e., fraction of the periodic supply which is transfused) at an HBB depends on the size as well as the age mix of its blood supplies. In order to derive a model for utilization rate, the following basic distribution strategy was developed, based on the premise that periodic shipments are made to each HBB. At the beginning of every "period" the RBC ships a fixed number of fresh (1-2 days old) rotation units and a fixed number of older (6-9 days old) retention units to each HBB. The retention units are permanently retained by the HBB until transfused or discarded at the end of their useful life. The rotation units that are in excess of a fixed "desired inventory level" at the end of the period are returned to the RBC for redistribution as retention units. Since the rotation units are not subject to spoilage, the utilization rate is determined by the behavior of the retention inventory.**

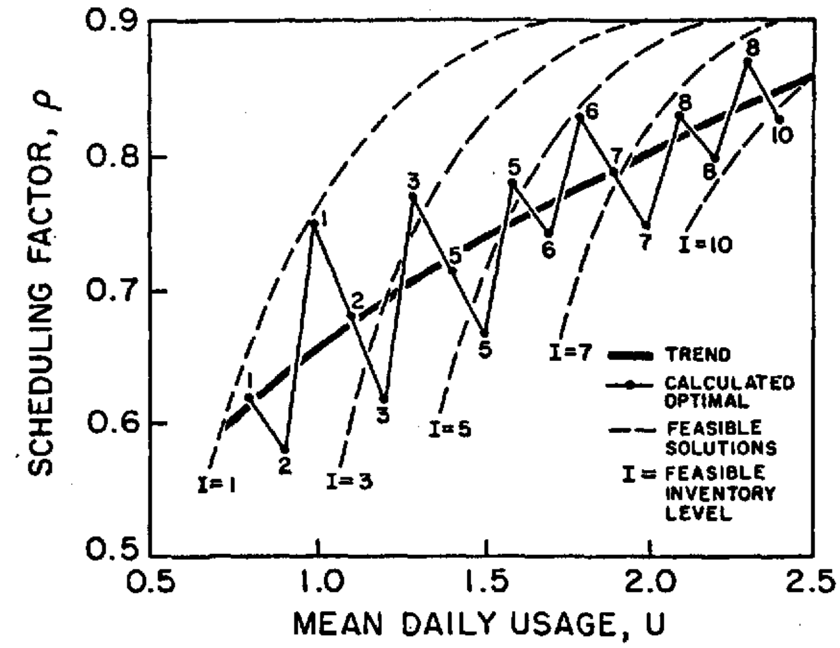
The number of retention units in inventory at a HBB immediately after each delivery can be represented by a finite-state Markov chain, whose transition probabilities are a function of the fixed periodic input (i.e., the fixed retention shipments), and of the variable demand and usage. Under the assumption that the oldest unit in inventory is transfused first, the steady state solution of the system can be computed and related to the utilization rate. This model was examined in [3], where

**Utilized Units = Total Supply − Units Spoiled.

analytical approximations were derived for Poisson usage, relating the number of retention units in each shipment to the resultant utilization rates with the desired inventory level as a parameter.

This relationship is illustrated for a fixed utilization rate of 98% by the family of broken lines in Figure 2 where the scheduling factor ρ is the fraction of mean daily usage that is replaced by retention shipments. As an example, if a HBB's mean daily usage for a given blood product is 1.5, then the HBB can achieve a utilization rate of 98% by any of the following combinations: desired inventory I=1 and ρ =0.89, or I=3 and ρ =0.82, or I=5 and ρ =0.70.

FIGURE 2. Utilization Rate Model.



It was shown that this stocking procedure maintains the mean inventory close to this desired inventory level most of the time. It was also shown that adding additional stages of returns and redistribution would make only slight improvements in the availability rate and utilization rate achieved. Since multiple redistributions introduce severe logistical problems and significant transportation costs, distribution strategies involving more than two stages of distribution were not investigated.

Properties of desirable regional allocation strategies

Having derived models enabling us to predict the availability and utilization rates of a HBB for any rotation/retention policy implemented by the RBC, the regional allocation problem was examined. It was assumed that some fixed penalty costs were associated with every nonavailable unit and every nonutilized unit, and the objective was to determine the distribution policy parameters so as to minimize the total expected regional cost.

First, the policy that minimizes the total expected one-period cost was derived [5]. It was shown that this policy involves the following operations:

(1) first allocate all available retention units so as to equalize the utilization rates at all HBB's;
(2) then allocate all available rotation units (which are not subject to spoilage) so as to equalize the availability rates at all HBB's.

It was shown that this policy is independent of the unit penalty costs, and that is maximizes both the availability and the utilization of blood in the region, simultaneously. That is, any deviation from the policy that would reduce utilization would also result in reduced availability for the next period, and vice versa. It was next shown [6] that this policy was not only myopically optimal but also approximately optimal in the long run. Further, in a large number of cases that were tested by computer, the utilization and availability rates computed from the myopic results also corresponded to the absolute optimal values computed.

This result established the principle that a distribution policy should seek to equalize utilization rates and availability rates. This is also a policy that has the essential elements of "fairness" in equally spreading the nonavailability and nonutilization risks among hospitals regardless of their relative size and is consequently a highly defensible policy.

Finally, it was shown that the highest possible regional availability and utilization rates are achieved when the desired inventory level for each blood type in each HBB is at the value that minimizes the total number of rotational units that are required to achieve these availability and utilization rates.

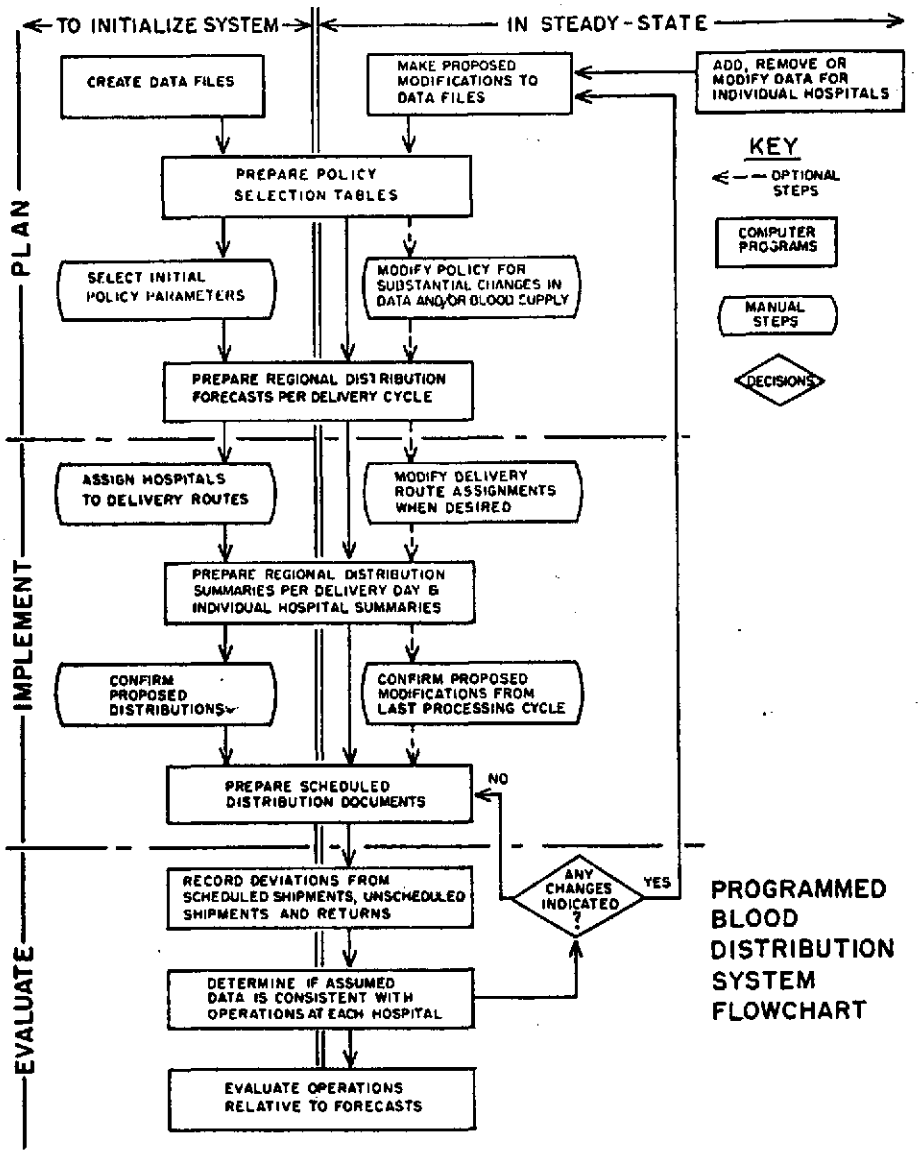
It is a straightforward effort by computer to calculate the combination of inventory level and scheduling factor that requires the minimum number of rotational units. The minimum number of rotational units required to achieve a fixed utilization rate of 98% and an availability rate of 95% are indicated by the points connected by the straight line segments in Figure 2. The irregular behavior of this solution is due to the fact that inventory levels must be integer values and rounding occurs on very small absolute values. As an example, the minimum rotational shipments required to an HBB of mean usage of 1.5 units daily to obtain the target goals above occur when the desired inventory is 5 units, and the scheduling factor is set to 0.67. The trend line which is drawn in the heavy line in Figure 2 is meant to indicate simultaneously the optimal values in inventory level and scheduling factor for given values of mean usage.

Adding operational constraints

The above distribution model of equalizing availability rates and utilization rates among the HBB's is illustrated by the two curved lines in Figure 3. The upper curved line shows the minimum total shipments required to achieve a fixed availability rate at a HBB of a given mean usage. The lower curved line shows the maximum retention shipments to achieve a fixed utilization rate. The area between the curves would have to be met by rotational shipments. As can be seen from the right end of the curves where the tails meet, this results in a situation where the larger usage HBB's receive almost all of their shipments in older retention units, while the smaller usage HBB's receive almost all of their shipments in fresh rotation units.

required for shipping and other purposes, is used to create data files. From these data files, and using the model described above, "policy selection tables" are generated. These tables indicate the minimum total fresh supply needed to be distributed on rotation in the region over a two-week period, in order to achieve certain "acceptable" values for availability and utilization. On the basis of these tables, of the amount and the stability of the collections, and of the reserve to be kept at the RBC, the attainable values are determined, and the "target" values are selected for these performance measures.

FIGURE 4. Programmed Blood Distribution Flowchart.



Once the planning phase is completed, hospitals are assigned to delivery routes. These delivery routes are constrained to provide for deliveries to all hospitals at a fixed time each delivery day. Delivery day intervals are either one, two, or four days, depending on the size of the hospital and special requirements or strong preferences expressed. From these assignments "regional distribution summaries" per delivery day are prepared for evaluation. These are revised as needed to equalize the amount of blood distributed each day as far as possible. Individual hospital "summaries of delivery schedules" and of "desired inventory levels" are then prepared and sent to all HBB's. After discussions during which the above distribution schedules are confirmed or modified as required and extensive educational sessions with the HBB's' management and operational personnel take place, the operation is ready to start. The final step is the preparation of packing documents which are prepared in the order in which deliveries are to be made.

Once operational, data files or distribution schedules are modified by one of two means. As new hospitals are added, as hospitals are removed, as changes in usage occur, such as those from increased bed capacity, or as changes in usage are detected by the control procedure, the data files are modified. The policy selection tables are revised each time there are revisions in usage estimates. The revised tables are then manually evaluated to determine if the changes are substantial enough to require a change in targets. A change in targets may also be required if a substantial increase or decrease in the blood supply is anticipated from other information. If such changes are required, then the regional distribution forecasting procedure is performed again as described above.

Scheduling deliveries

A major advantage of PBDS, both to the RBC and to the HBB's, is the ability to preschedule most deliveries. Prior to PBDS being implemented, a number of delivery vehicles were dispatched as orders came in. For urgent orders, vehicles were dispatched immediately, while for more routine orders an attempt was made to hold vehicles back until several deliveries in the same geographical area could be combined. This procedure was expensive and, perhaps more importantly, resulted in situations where even urgent orders were delayed, since delivery vehicles were not always available during peak delivery hours.

With the PBDS most deliveries are prescheduled, and take advantage of known traffic patterns in order to minimize delivery time. An interactive, computer-aided procedure was devised which assigns HBB's to delivery routes so as to meet their time and frequency of delivery requirements. The twelve delivery day planning cycle is split into three groups of four delivery days, after which the delivery cycle repeats. In each four-delivery-day cycle each HBB receives either one, two, or four deliveries. The procedure tries to satisfy the delivery requirements without leaving gaps in consecutive time slots, since an empty time slot indicates idle time.

An opportunity to test the flexibility of this delivery scheme occurred recently when the LIBS Blood Center was moved from one location to another several miles away. It was found that the delivery routes could be adjusted rapidly and the required reassignment of the HBB's was determined conveniently.

Controlling the system

The shortage deliveries and the rotational returns for each hospital are monitored in order to detect changes in hospital requirements. Every two weeks the "discrepancy" between the hospital's estimated† and expected usage during these two weeks is computed. This discrepancy is added to the cumulative discrepancy from prior weeks to form an updated cumulative discrepancy, which, together with the number of weeks included in it, and the value of the hospital's expected usage, are used to compute the "normalized cumulative discrepancy" of this week. This last value is compared with a statistically established "limit;" if it exceeds the limit, it is concluded that a shift in usage level has occurred. New usage estimates are computed, and new distribution schedules are prepared. Otherwise, no action is taken.

Since the mix of the eight blood types is a function of the ethnic mix of the population, which is in a state of transition for some of the hospitals tested, the above control procedure was established for each of the blood types as well as for all types of whole blood and all types of red blood cells. Adjustments were made either by blood type or for all blood types. It was found that this had the unfortunate characteristic that if overall usage was increasing or decreasing, individual blood types tended to go out of control in consecutive evaluation cycles before the total usage chart went out of control. This caused an excessive number of distribution changes. For this reason, the concept of "warning limits" and "action limits" was set up. A change is only made if one of the blood types exceeded the action limits. However, at that time if any other blood type also exceeded the warning limits, then that distribution would also be changed at the same time.

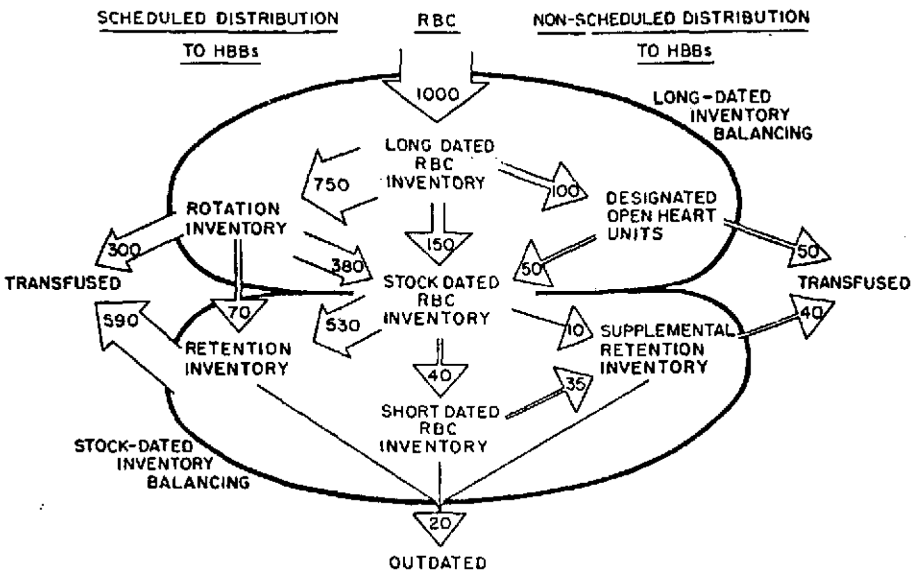
Daily inventory adjustment

The resulting regional blood flow is illustrated in Figure 5. In this figure the aging of the RBC inventory is indicated down the center of the figure with the scheduled movement of blood to HBB's indicated to the left of the figure and the nonscheduled movement to the right. The long-dated, stock-dated, and short-dated RBC inventories refer to blood units that are suitable respectively for rotation shipments, for retention shipments, and solely for supplemental shipments — which are filled by the oldest available units. The arrows indicate the blood flow that is normalized to 1,000 units collected.

On the basis of this anticipated regional blood flow, the RBC's inventory is evaluated and adjusted daily. Stock-dated inventory balancing is performed late each afternoon after all rotational returns have been received. It involves the part of the flow circled towards the bottom of Figure 5. The available stock-dated inventory is compared to the retention shipments that are scheduled, the anticipated supplemental shipments plus a small reserve for unusual circumstances which is shown as becoming short-dated inventory. When the inventory for any product exceeds these requirements, the excess units are designated as surplus, and transshipped to the New York Blood Services (NYBS) division of GNYBP. When stock-dated inventory is

†The usage is determined by combining the known weekly scheduled distributions with the recorded supplemental deliveries and rotational returns to form an estimate of actual usage.

FIGURE 5. Regional Blood Flow Balancing.



below requirements then either surplus long-dated units (if available) will be retained, or the shortage will be made up from the other divisions of GNYBP if possible.

Long-dated inventory balancing is performed each morning after the bulk of the bloods collected the previous day have been typed. It involves the part of the flow circled at the top of Figure 5. The long-dated regional inventory which is expected to become available during that day is compared to the commitment of units for scheduled rotation shipments plus units required to meet open heart surgery needs (which is a specialized procedure where only fresh blood units are suitable). Any units in excess of these requirements are either retained to make up for shortages in stock-dated inventory as discussed above, or are made available for transshipment to other divisions of GNYBP. Since LIBS collects in excess of its needs, there is generally a surplus of rotation units especially in the more common blood types.

Computer operation

The effectiveness of PBDS depends upon accurate and timely data on operations, which is achieved in part by running the system on a minicomputer, and by utilizing machine-readable bar codes on blood products and on test samples [1]. The machine-readable codes on test samples are used in conjunction with automated equipment which performs blood type determinations and links the data concerning the unit and its test results directly in the computer. This provides the earliest possible indication of what products will be reaching inventory during that day. The bar codes, which indicate product, blood type, identification numbers, etc., are also scanned as blood units are shipped out and returned, to maintain perpetual inventory. In this manner, the total inventory picture at the RBC is accurately maintained in real time.

The light pen techniques are also utilized to control computer operations and to identify the locations to which blood products are shipped or from where they were received. This is done by scanning bar codes on menu sheets that list all shipping locations, all types of transactions, and all available types of computer operations. This way, the computer is effectively used by nontechnical personnel, and any errors in entering data are minimized.

In order to maintain a modular approach that can be utilized by all RBC's regardless of size, a network of minicomputers is being used to handle all the RBC blood processing needs. In a larger operation, such as LIBS, the operations are split functionally (i.e., separate minicomputers in the laboratory and distribution areas), while in the smaller centers a single minicomputer would handle all functions.

Impact, implications, and extensions

Scientific impact

The successful operation of PBDS at LIBS has demonstrated the ability to set performance measures on the basis of regional planning. It establishes the first quantitative management guide for the selection of feasible targets and strategies, for the evaluation of options within a class of strategies, and for establishing the best possible performance within that class as a reference for the strategy selected. It further demonstrates the ability to identify deviations from anticipated performance and, consequently, the ability to manage by exception.

An example of this is the analysis of reported utilization by 28 HBB's that was performed recently. Statistical techniques indicated that the utilization rate pattern for 18 of the 28 HBB's were statistically not differentiable and showed an average utilization rate of 96%, which is close to the 98% value predicted; 6 HBB utilization rates were indicated as significantly below this norm and 4 utilization rates were indicated as significantly above the norm. Since the same utilization rate is achievable for all HBB's under PBDS, deviations from the statistical norm can be confidently attributed to assignable causes. The utilization rates above the norm are attributed to sophisticated techniques such as holding the same blood unit for several possible recipients concurrently. The poorer utilization performance of the 6 HBB's is attributed to poor blood banking practices such as failure to return untransfused units to inventory promptly, or to special hospital practices.

Management decision implications

PBDS implies major changes in management decision models both for the RBC and the HBB's. Most regions follow a procedure characterized by a decentralized and reactive distribution. In this mode of operation, the HBB checks its inventory status one or more times per day and, if it deems it to be low, places an order for additional blood. The RBC makes a decision whether its inventory is sufficient to fill the order. If it is, it delivers the requested quantity. If it is not, then it seeks to modify the order to a lesser amount, to substitute red blood cells for whole blood, etc. This "discussion" results in a modified order which is actually delivered. Both the HBB and the

RBC are making short-term decisions regarding each delivery, sometimes with conflicting objectives and are reacting to a situation more than anything else.

PBDS allows the system to operate in a predictive distribution mode. In this mode the RBC assumes responsibility for the long-term scheduling decisions, while each HBB assumes responsibility for the daily fine adjustments. The RBC starts with a distribution strategy as described above. A copy of this, in the form of a shipping schedule and recommended inventory levels to be maintained, is furnished to each HBB. The HBB inventory checking is then reduced to two steps. First, before each scheduled shipment is due, any inventory above the desired levels is returned and second, whenever the inventory is insufficient to meet demand, a supplemental order is placed which is automatically filled under normal situations since it is assumed that it is needed immediately.

The supplemental deliveries and the returns are monitored over a two-week period, at the end of which period a decision is made based on the control statistics on whether or not to revise the distribution. This is a further example of how PBDS has achieved a management by exception principle.

Economic impact

The economic impact of PBDS can be most directly measured in terms of improvement in blood utilization. Prior to the implementation of PBDS, the utilization rate in the LIBS region was 80%, which was also then the national average. Since the implementation of PBDS, the utilization rate for LIBS has improved by 16%, while the national average has improved little if at all during the same interval. The improvement in utilization at LIBS translates to 80% reduction in wastage, and therefore, to annual savings of $500,000 per year.

Of lesser economic impact is the reduction in the number of deliveries. Before PBDS was implemented, an average of 7.8 weekly deliveries were made to each hospital, all of which were unscheduled; after PBDS was implemented, the number of deliveries dropped to 4.2, but of which only 1.4 are unscheduled. By associating a $10 cost to an unscheduled delivery to an HBB and a $5 cost to a scheduled delivery (which is part of a route), PBDS has achieved a 64% reduction in delivery costs. This translates to annual cost savings of $100,000. Additional important, though less tangible, costs savings are achieved by the implementation of sounder blood banking practices to reduce discrepancies between actual and achievable performance for individual HBB's.

Probably the most important savings in the national health care bill brought about by PBDS is realized by improved blood availability to patients. Since deliveries to the hospitals are mostly prescheduled, elective surgeries can be prescheduled also, so as to minimize the number of surgeries postponed because of lack of the right blood products. However, savings from this improved availability are extremely difficult to estimate and quantify.

National implications

It can be seen from the model of Figure 3 that, in order to achieve high availability and utilization rates for the rarer blood types (small usage), most of the shipments have to be on rotation. In practice, however, RBC's can only schedule between 70% and 85% of the average blood collections for rotation shipments because of uncertainties in blood availability, and the need to keep an RBC reserve. Therefore, if the amount of rotational blood available for distribution is approximately equal to that which will be utilized within the region (i.e., if a strict self-sufficiency criterion is applied), then for these blood groups either very low availability or alternatively a very significant reduction in utilization must be accepted. Since neither of these alternatives is desirable, an alternative is to rotate more blood units in this blood group than will be eventually transfused in the region and redistribute some portion of the returned blood units to another region utilizing the inventory balancing techniques that were described earlier. This is a feasible approach in LIBS since, like most regions that are primarily suburban and rural in nature, it has the capacity for collecting blood significantly in excess of the overall regional needs.

LIBS collects approximately 20% more blood than will be ultimately transfused in the region. Virtually all of the rarer bloods are first rotated to HBB's in the region, while the excess collection in the more common bloods are mostly immediately sent to the New York Blood Services (NYBS) division of GNYBP. This division encompasses the metropolitan New York area and, like most large urban areas, has difficulty in collecting enough blood to meet the needs of the major medical centers and other hospitals in this area. The influx from LIBS is of significant help and the small fraction of older, rarer bloods that are received as stock-dated units can readily be absorbed. Thus the units collected in LIBS in excess of its own transfusion needs improves the quality of blood services in LIBS as well as in NYBS. On the other hand, the major medical centers in the New York Division of GNYBP also serve the needs of much of the population residing in the area covered by LIBS.

It should be noted that not all major regions such as LIBS require interaction with larger regions. A planning exercise performed for the regional blood center in Richmond, Virginia, which services only seven HBB's, indicated that this region was remarkably self-sufficient. This is primarily because it includes a major medical center which accounts for a significant part of the regional usage. However, for the most part there is a mutually beneficial interaction feasible between the smaller RBC's serving mainly suburban and rural areas and the larger ones which will mostly service larger urban areas.

These conclusions suggest that the concept of total regional self-sufficiency which has long prevailed in blood banking is in conflict with the goals of the National Blood Policy that calls for high availability with efficiency. Rather, a modest level of resource sharing between regions as outlined above will work to the benefit of all participants.

Extensions

The success of PBDS has fostered a favorable climate for the further application of Management Science techniques both on a national and international basis. The International Society of Blood Transfusion Expert Committee on Automation has recently endorsed the concepts inherent in PBDS and the Swiss Red Cross Transfusion Service is studying it as a basis for the creation of a national blood distribution system for that country.

On the national level, as the PBDS strategy is expanded to the other three divisions of the GNYBP, it will provide an opportunity to more thoroughly investigate a broader regional network. Concurrently, other regional blood programs are investigating the use of PBDS. The American Red Cross Blood Services — Northeast Region has already implemented a distribution plan for its six RBC's based upon the PBDS concept. These implementations represent broader geographic groupings of RBC's which, together with other such groupings, could eventually interact through a hypothetical national blood clearinghouse.

Such developments would contribute significantly towards meeting the objectives of the National Blood Policy. It would act to alleviate the "national blood shortage" which, in part, at least is thought to be not so much a national blood shortage but rather the lack of a national logistical system for moving blood from where it is available to where it is needed. They would thus contribute to a reduction in the consumer's hospitalization costs and the delivery of higher level health care.

Acknowledgements

We are indebted to Aaron Kellner, M.D., President of the New York Blood Center, Inc.; Johanna Pindyck, M.D., Director; Robert Hirsch, M.D., Medical Director of the Greater New York Blood Program; and Theodore Robertson, M.D., Director of Long Island Blood Services, for their wholehearted collaboration and encouragement. We are also grateful for contributions made by Professors Cyrus Derman, Ed Ignall, Peter Kolesar, and Donald Smith of Columbia University, at various phases of this work, and for helpful comments and encouragement provided by Professors Howard Kunreuther, William Pierskalla, and Morris Cohen of the University of Pennsylvania.

This research was supported in part by the National Institutes of Health, National Heart, Lung and Blood Institute Grant No. HL 09011-16 at The New York Blood Center.

References

[1] Brodheim, E., "Regional Blood Center Automation," *Transfusion*, Vol. 18, 1978, pp. 298—303.

[2] Brodheim, E., Hirsch, R., and Prastacos, G. P., "Setting Inventory Levels for Hospital Blood Banks," *Transfusion*, Vol. 16, 1976, pp. 63—70.

[3] Brodheim, E., Derman, C., and Prastacos, G. P., "On the Evaluation of a Class of Inventory Policies for Perishable Products such as Blood," *Management Science*, Vol. 21, 1975, pp. 1320—1326.

[4] National Blood Policy, Department of Health, Education and Welfare, Vol. 39, 1974, 176.

[5] Prastacos, G. P., "Optimal Myopic Allocation of a Product with Fixed Lifetime," *Journal of the Operational Research Society*, Vol. 29, 1978, pp. 905—913.

[6] Prastacos, G. P., "Allocation of Perishable Inventory," forthcoming in *Operations Research*. Also available as Technical Report No. 70, Operations Research Group, Columbia University, 1977.

[7] Prastacos, G. P. and Brodheim, E., "A Mathematical Model for Regional Blood Distribution," Working Paper 79-02-91, Department of Decision Sciences, The Wharton School, University of Pennsylvania, 1979.

THE GREATER NEW YORK BLOOD PROGRAM

The New York Blood Center • American Red Cross
310 East 67 Street, New York, N.Y. 10021, 794-3000.

August 3, 1979

Dr. Raj Nigam
RCA
David Sarnoff Research Center
Princeton, New Jersey 08540

Dear Dr. Nigam,

The Greater New York Blood Program, the largest blood program in the world, has assumed the responsibility for the blood needs of 18 million people residing in New York City and in much of its surrounding area. We recruit people to donate their blood, collect this living tissue from them, process it in our laboratories and distribute it to the 262 hospitals in our community where it must be available to care for the people in need. Our aim is to have enough blood of the right types on hand to meet the usual hospital demands and the emergencies which may arise, but not so much that it will be wasted since blood can only be stored for 21 days outside the body. We face the major problem of how to maximize the availability of blood to each of the 262 hospitals that we service while effectively discharging our implicit covenant to our donors to see that their gift is efficiently utilized.

We turned to management science to help us achieve this goal and provided the Long Island Blood Services Division of our program as a test site. The system that was developed has been in operation for two years. It integrates the blood supply of a region by developing a complex pre-planned rotational system so that the availability of blood of all types becomes the same in hospitals regardless of size and usage, and utilization of the blood also becomes the same.

Patients in the community have equivalent protection for blood services at all the hospitals without burdening some of them with excessive waste. This system has given us a mechanism to do exactly what we urgently needed - Take the Crisis Out of Blood Banking. We have offered to our hospitals a rational, effective, functioning system - which will soon be extended both throughout our service area and elsewhere in the United States and abroad. It is not unfair to state that this program can and will contribute to an improvement in health care throughout our nation, and therefore touch each and every one of us at some time in our life. We are proud of our achievement and appreciate this recognition.

Sincerely,

Johanna Pindyck, M.D.
Vice President and Director
The Greater New York Blood Program

JP:ck

#1○

Using Computer Simulation to Predict ICU Staffing Needs

by N. Duraiswamy, R. Welton, and A. Reisman

This article describes a computer simulation study to determine the number of full-time registered nurses required to deliver safe nursing care in a 20-bed medical intensive care unit. The authors' computer simulation model predicts different staffing levels for two six-month periods. The article illustrates how the computer simulation was accomplished and discusses the cost and feasibility of using computer simulation in other settings.

The first nursing service standard of the Joint Commission on Accreditation of Hospitals (JCAH) calls for "a sufficient number of duly licensed registered nurses on duty at all times . . . to give patients the nursing care that requires the judgment and specialized skills of a registered nurse"[1]. Many recent management studies[2–4] have attempted to define how many nurses constitute a sufficient number.

In many hospitals, nursing staffing is done almost by intuition; management adds, subtracts, and reassigns staff on the basis of a subjective perception of what seems necessary. Nurses are often reassigned from not-so-busy areas to cover busy areas[5]. However, in the intensive care unit (ICU), where nurses must be knowledgeable in extensive, observant, diagnostic, and therapeutic elements of critical care[6], reassigning nurses from other areas can jeopardize the unit's standard of care. This intuitive approach to staffing is also difficult to translate into quantifiable terms that are understandable to others, including hospital administrators, boards of trustees, and other nurses.

This article describes a system-analytic study of the nurse staffing needs of the Medical Intensive Care Unit (MICU) at the University Hospitals of Cleveland. In considering how to determine the efficacy of various staffing alternatives, we decided that a computer simulation would be the best method. Computer simulation creates a computer version of the real system and provides laboratory conditions in which decision makers can test alternative policies. Simulation allows the testing of staffing choices in the light of the conflicting objectives of cost containment and quality of care. Moreover, it allows consideration of variations in patient admission and discharge rates, length of stay, levels of acuity, and levels of nursing care requirements.

Traditional assessments for staffing

Systematic methods for staffing have taken several approaches. One approach takes such elements as diet, toileting, vital sign measurement, respiratory aids, hygiene, suctioning, and turning and/or positioning[7–9] and applies a weighted point system to these areas of care. The point total is then converted to some unit equivalent to a specific amount of nursing care time. Other approaches have lumped all care categories together and subjectively established required levels of care ranging from high to minimal[10] without defining the criteria to be used. An "hours per patient-day" formula multiplies the number of patients by standard hours of nursing care, such as 5.0 per patient, and then multiplies this total by 7 to obtain the required hours of nursing care per week. Another approach uses patient care categories such as minimal care, partial care, moderate care, complete care, and intensive care[11–13] but includes criteria for indirect and supportive elements of

N. G. Duraiswamy, Ph.D., is assistant professor of business at Indiana University, Kokomo. He received an M.S. and Ph.D. in operations research from Case Western Reserve University.

R. Welton, R.N., M.S.N., is a master degree candidate in the Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, Ohio.

A. Reisman, Ph.D., P.E., is professor of operations research, Case Western Reserve University.

TABLE 1. PATIENT CLASSIFICATION

Care Requirement Category	Weighted Points According to Level of Care			
	Intense	High	Moderate	Minimal
Physical Dependence	29.6	14.8	11.6	4
Observation	30.8	19.6	10.4	6.8
Social/Emotional Care	28.0	18.0	10.8	4.8
Rehabilitation/Teaching	22.0	15.2	9.24	4.4

the nursing process, thus allowing a comprehensive assessment of actual care required. Such a comprehensive approach forms the basis for the patient classification system developed at University Hospitals of Cleveland, and this approach was used in the study described here.

Background

The University Hospitals of Cleveland include a 1000-bed acute care medical center located on the east side of the city of Cleveland. Its six specialty hospitals, all under one central management, are closely affiliated with the professional and graduate schools of the health science disciplines of Case Western Reserve University.

The Medical Intensive Care Unit (MICU) of University Hospitals consists of two patient care areas: the intensive care unit (ICU) where critically ill patients are initially admitted and cared for during the acute stage of their illness, and the progressive care unit (PCU) where patients are cared for after the acute stage of their illness. Both units were designed to accommodate 10 patients each with 10 single rooms in the ICU and 5 double rooms in the PCU, yielding a total possible census of 20 patients for the MICU.

After the first year of operation at 100 per cent occupancy, the 44 budgeted full-time nursing positions proved to be inadequate to provide safe nursing care in both the ICU and PCU when both were at 100 per cent occupancy. For this reason, 4 beds in the MICU were closed to patients, reducing the total possible census from 20 to 16 patients for both units. We undertook our study to determine how the original 20-bed unit could be restored without resorting to overstaffing or understaffing.

TABLE 2. PATIENT CARE CATEGORIES ACCORDING TO TOTAL PCC SCORE

Total PCC Score	*Category*
80 or above	Intense
60–80	High
40–60	Moderate
less than 40	Minimal

Data collection

We used the department of nursing's Patient Care Categorization (PCC) assessment tool, with the charge nurse as the source of data. This classification system uses four categories of care requirements: physical dependency, observation required, social/emotional requirements, and rehabilitation and/or teaching requirement. Within each of these categories, there are four levels of required care. The charge nurse chooses which level of care each patient requires within each of the four categories.

A total score, called a PCC score, is then obtained by adding weighted points for the level of care chosen in each category. (See Table 1). Based on the *total* PCC score, a patient is assigned to one of four care categories. Table 2 lists total PCC scores and their corresponding care categories.

Using these four categories, the nursing department developed the policy for nurse-patient assignments (Table 3). This policy provides adequate patient-nurse rations to ensure safe care and is currently used for nursing assignments in the MICU.

The PCC tool had been in use for two years prior to this study. During that time, the PCC assessment was done one week of each month on the 7:00 A.M. to 3:00 P.M. shift by the MICU charge nurse; this sample was found to be representative of data obtained from initial daily assessments using the PCC tool. We computed the probability of a patient's being in a specific care category and the number of patients in a specific care category per day. Using these data, which spanned two calendar years, we identified seasonal variations, such as a greater number of admissions of patients with diagnosis of drug overdose or myocardial infarction during the spring and summer months.

The Daily Patient Census Forms for the ICU and PCU were used on a sampling basis to obtain admission and transfer or discharge data. The sampling included all patients admitted on 112 randomly selected days over a one-year period (1977–1978) to determine actual length of stay in both units. It was felt that this approach would provide a statistically significant sample yet minimize data collection time and effort. We then used this data base to determine the length of stay distribution in our simulation model.

There is a significant fluctuation in the daily patient load in the MICU. Consequently, one would expect a similar fluctuation in its daily nurse staffing needs. The

FIGURE 1. FLOW CHART OF THE SIMULATION MODEL

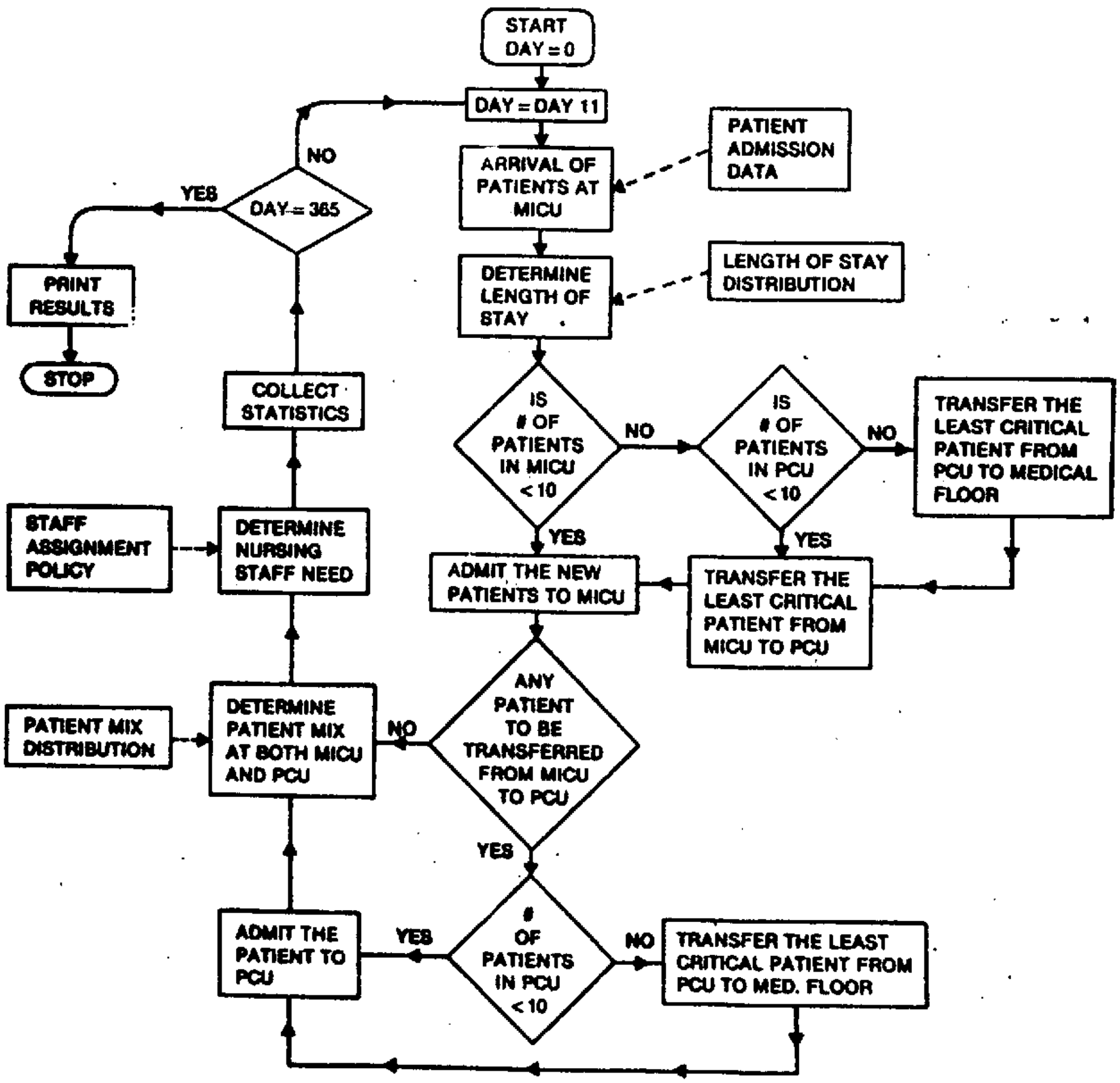


**TABLE 3.
MEDICAL INTENSIVE CARE
UNIT (MICU) POLICY
FOR NURSE-PATIENT ASSIGNMENT**

Nurse-patient assignment will be made in terms of patient PCC score and will comply with the score and nurse-patient ratios listed below.

Patient Care Category Score	*Nurse-Patient Ratio*
Intense	1:1
High	1:1
Moderate	1:2
Minimal	1:3

The above assignment policy is applicable for all shifts in both the ICU and PCU.

patient mix in addition to patient load determines the staffing requirement, but the random arrival of patients and the random length of stay suggest that the patient mix probabilities may not vary every day.

The simulation model

Figure 1 illustrates the simulation model. Patients are admitted to the ICU for a certain length of time, as determined from the length of stay distribution for ICU. At the end of this period, they are either transferred to PCU or discharged. The length of stay of the patient in PCU is again determined, using the length of stay distribution for PCU.

Since no addition of beds is planned for either the ICU or PCU, capacity constraints at the current ten-bed levels for each unit have been imposed in the model. According to current practice, when all beds are occupied, the least sick patient is transferred from the ICU to the PCU in order to admit a newly arrived patient. When a similar situation arises in the PCU, the least sick patient is transferred out of the PCU. The simulation model contains the logic to handle these situations.

The nursing care requirement for each patient is expressed in terms of hours and entered in the simulation model to determine the total nursing care hours needed in a shift. The model assumes that patients in an intensive care setting will require the same level of care during the other two shifts of the day.

The simulation covers a full year. We designed the simulation model to print out the number of nurses needed on each simulated day during the one-year period. In addition, the model generated descriptive statistical information about the average number of nurses required as well as the standard deviation for any chosen period of time (Figure 2a-d).

Validity of the model

The simulation model was validated using the historical data base from the MICU. The simulation results are therefore valid for prescriptive purposes if the utilization patterns do not change. However, if the utilization patterns are projected to increase or decrease, the simulation can be used as a vehicle to develop new plans for staffing.

Analysis

The total number of patient days for both the ICU and PCU is 308 as determined by the analysis of 112 randomly selected days during a one-year period. The patient mix probabilities used in this study are shown in Table 4. The values for patient mix probabilities were obtained from the data collected by the department of nursing over one full year during 1977. The patient mix probabilities for the winter, spring, summer, and fall seasons were obtained separately to capture seasonal variations in the staffing needs. During the course of the study, data on the patient mix probabilities were collected individually for both ICU and PCU through direct observation. However, these individual data were not used for the simulation because the nursing care required for a patient is the same in each unit for each patient category.

Staffing by seasonal demand

As evident in Figure 2, the variation in staffing requirements between fall and winter months is not significant. The average requirements for spring and summer were identical with only a small difference in the standard deviation. However, there is a notable variation in staffing requirements between the combined fall/winter and combined spring/summer periods. These results indicated a need for two staffing levels, each for a six-month period. Since there is no specific staffing level which meets the staffing requirement optimally on each day of the six-month period, we offered the data in Tables 5 and 6 to show the percentage of days of overstaff-

TABLE 4. PATIENT MIX PROBABILITIES

	PROBABILITY			
Patient Category	*Winter*	*Spring*	*Summer*	*Fall*
Intense	0.23	0.55	0.48	0.17
High	0.31	0.30	0.32	0.30
Moderate	0.33	0.12	0.18	0.40
Minimum	0.13	0.03	0.02	0.13

ing and understaffing associated with each staffing level. This information allows management to weigh the impact of alternative staffing levels against costs and at the same time to comply with HSA and JCAH standards.

Scheduling complications

The average staffing levels of 30 nurses per day for fall and winter and 36 nurses per day for spring and summer appear most desirable in terms of overstaffing and understaffing. The large turnover of nursing staff during the spring and summer months (approximately 15 terminations or transfers out of the MICU between March and September of 1977)[16] may have have implications for scheduling vacations for experien... nurses. New graduate orientation to the MICU and education in critical care, which also occur during spring and summer can increase the average number of nurses needed during these months.

While the simulation resulted in recommended staffing of 30 or 36 nurses per day, levels which are lower than the actual budgeted staffing (44 full-time nursing positions per day) for the MICU, several considerations must be noted. First, the staffing levels suggested by the model do not include provisions for vacations, days off, holidays, illness, absenteeism, and staff development commitments. Second, the staffing recommendations apply to situations in which every nurse is experien..d in critical care, has successfully completed the entire ICU orientation, and has demonstrated safe, independent

FIGURE 2. SEASONAL MICU NURSING STAFF REQUIREMENTS

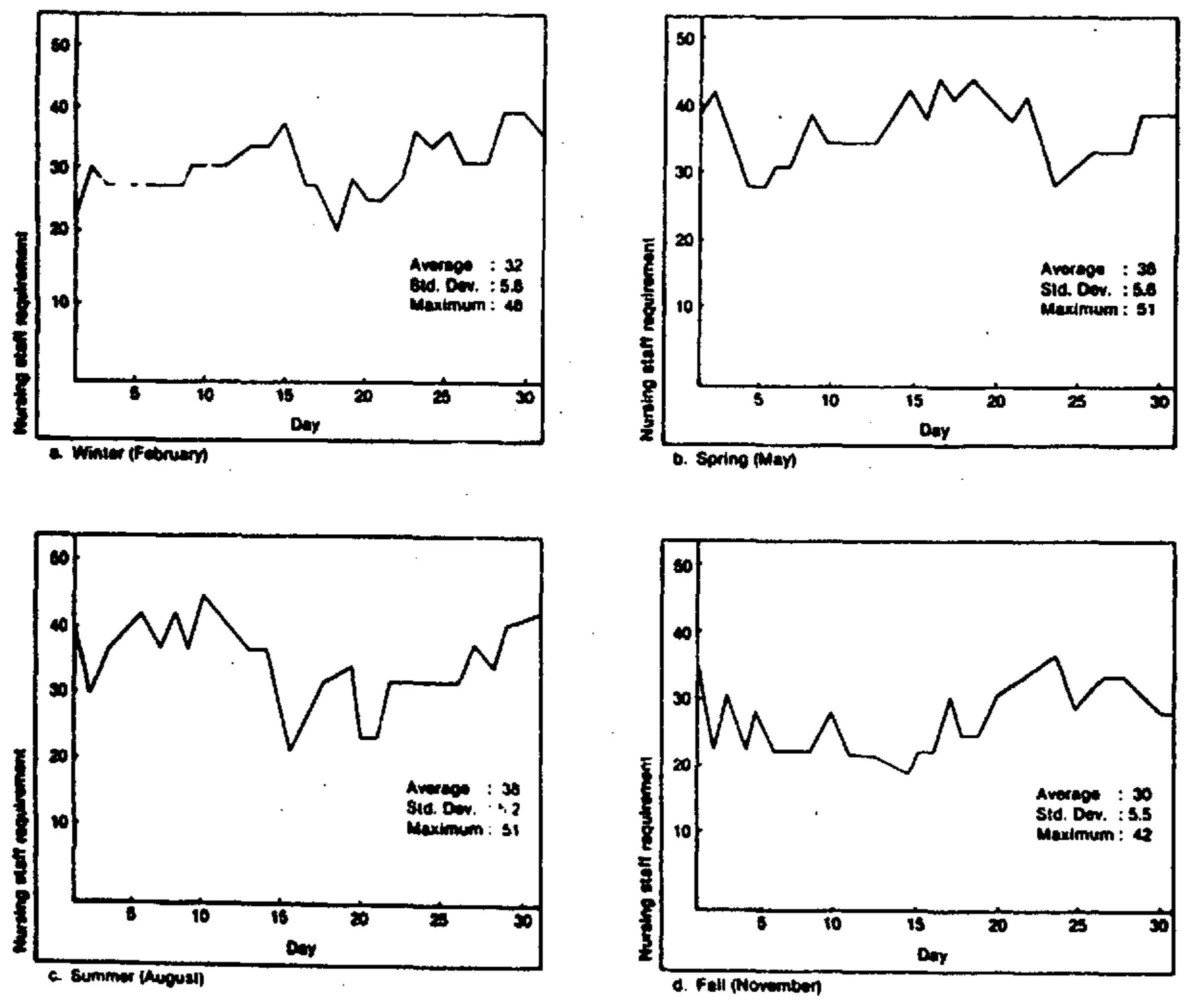


TABLE 5. ANALYSIS OF STAFFING LEVELS (FALL AND WINTER)

Nursing Staff Level (# of nurses/day)	% of Days of Overstaffing	% of Days of Understaffing	% of Days of Optimum Staff
26	18.8	81.2	0
28	37.2	62.8	0
30	39.4	41.2	19.4
32	58.8	41.2	0
34	72.2	27.8	0

TABLE 6. ANALYSIS OF STAFFING LEVELS (SPRING AND SUMMER)

Nursing Staff Level (# of nurses/day)	% of Days of Overstaffing	% of Days of Understaffing	% of Days of Optimum Staff
30	5.6	84.4	10.0
32	15.6	84.4	0
34	27.8	72.2	0
36	28.9	4[illegible].8	23.3
38	52.2	47.8	0
40	71.7	28.3	0

practice without close supervision. Finally, the recommended staffing levels are for fully qualified nurses for *direct* patient care; they do not include provisions for ICU nurses away from the bedside involved in administrative, educational, or other indirect patient care activities.

Costs of the study

Since this study was conducted to fulfill academic requirements to study real-life problems, there were no out-of-pocket costs for development of the model and its computer implementation. Data collection, analysis, computer programming, and computer time expenses would occur in other settings. Nursing time used for collection of data used in this study was approximately five minutes per day for seven consecutive days each month. However, this responsibility has been a regular part of the MICU charge nurse role for several years. Data analysis and programming of the simulation model required less than one person month.

Applying the computer simulation in other settings

Although this study was carried out in a large urban medical center, the same approach can be used in small community hospitals. A staff of inhouse engineers and programmers is not necessary. Small community hospitals with access to local universities can use university-based expertise to perform such studies. Moreover, many public or private universities are looking for real-life experiences for both faculty and students and will often perform studies on a gratis basis.

The seasonal variations in the data collected from a single institution should be considered illustrative only. The approach, however, is transferrable to other institutions, other services, and other settings.

References

1. *Accreditation Manual for Hospitals*, Joint Commission on Accreditation of Hospitals Chicago, 1973 (updated).
2. Ryan, T., Barker, B. L., and Marcianto, F. A. A system for determing appropriate nurse staffing. *J. Nurs. Admin.*, 5(5):30-38, 1975.
3. Norby, R. B., Freund, L. E., and Wagner, B. A nurse staffing system based upon assignment difficulty. *J. Nurs. Admin.*, 7(9):2-24, 1977.
4. Clark, E. L. A model of nurse staffing for effective patient care. *J. Nurs. Admin.*, 7(2):22-27, 1977.
5. Norby, R. B. et al. 1977, pp. 2-24.
6. Burrell, A. L., and Burrell, L. O. *Critical Care*. St. Louis: The C. V. Mosby Company, 1977.
7. Clark, E. L., and Diggs, W. W. Quantifying patient care needs. *Hospitals, J.A.H.A.*, 9:96-100, 1971.
8. Poland, M., et al. PETO: A system for assessing and meeting patient care needs. *Am. J. Nurs.*, 7:479-1482, 1970.
9. Williams, M. A. Quantification of direct nursing care activities. *J. Nurs. Admin.*, 7(8):15-18, 1977.
10. Hearn, C. How many high care patients: Deployment of nursing staff. *Nursing Times*, 17:504, 1972.
11. Norby, R. B. et al. 1977, pp. 2-24.
12. Phipps, W. J., and Philps, C. Definition of Terms Used to Classify Patients According to Their Nursing Care Requirements. Department of Nursing, University Hospitals of Cleveland, 1977.
13. Pardee, G. Classifying patients to predict staff requirements. *Am. J. Nurs.*, 3:517-520, 1968.
14. Zegeer, L. J. Calculating a nurse staffing budget for a 20-bed unit at 100% Occupancy. *J. Nurs. Admin.*, 7(2):11-14, 1977.
15. Newman, M. A., and O'Brien, R. A. Experiencing the research process via computer simulation. *Image*, 10:5-9, 1978.
16. Allen, M. Annual Report, MICU. Unpublished paper. Department of Nursing. University Hospitals of Cleveland: 1-6, 1977.

Michael B. Harrington

Forecasting Areawide Demand For Health Care Services: A Critical Review of Major Techniques And Their Application

Since well before Cassandra, people have sought to forecast the future. In modern times, forecasting methods have adopted fairly scientific guises. But it is not clear that their predictive accuracy has improved appreciably over that of their predecessors. This is especially true of methods applied to large-scale public problems, such as areawide health planning. Improved methods have been offset substantially by the greater complexity, more rapid change, and more pervasive interdependence characteristic of modern society. Forecasting skills have grown, but so has the difficulty of the task.

The impetus for developing improved methods for use by areawide health planners, begun under the Comprehensive Health Planning legislation of 1966,[1] no doubt will be reinforced by the Health Planning and Resources Development Act of 1974 (P.L. 93-641).[2]

Michael B. Harrington, Ph.D., is Senior Associate for the Systems Evaluation Department, CACI, Inc.-Federal (Arlington, VA 22209), and Professor of Health Economics, Southeastern University.

Basic research for this paper was completed under Contract HRA 230-76-0066 sponsored by the Bureau of Health Planning and Resources Development, Health Resources Administration, DHEW. In developing this paper, the author benefited greatly from participation in a series of three symposia sponsored by the Bureau of Health Planning and Resources Development, DHEW, during the fall of 1974 and spring of 1975. Of course, neither BHPRD nor the other symposia participants are responsible for shortcomings this paper may contain. In addition, valuable criticism was rendered by Sarah P. Frazier of the Washington Office of Arthur Young & Company; William Oreaf, Department of Medicine and Surgery, Veterans Administration, Washington, D.C.; and Dr. Chester McCall, Manager, Systems Evaluation Department, CACI, Inc.-Federal, Arlington, VA.

Somewhere between perfect forecasts of the health care future and none at all lies a "zone of feasible forecasting" (i.e., forecasts both useful and within the capabilities and budgets of health planners). This paper explores the zone of feasible forecasting, by considering the following points:

- ☐ The objectives of forecasting in areawide health planning;
- ☐ Specific factors contributing to the demand for health services that enter into any serious forecasting effort;
- ☐ Major sources of change and continuity in the areawide health care system that are identified in conjunction with demand forecasts;
- ☐ Six generic kinds of forecasting techniques in the health planning context, including an evaluation of the strengths and weaknesses of each;
- ☐ A simple framework through which each forecasting method might be brought to bear most effectively;
- ☐ Some apparent dimensions of the "zone of feasible forecasting," given the current state of the art.

Objectives of Effective Forecasting

At least three major objectives for forecasting areawide health service requirements recur in public discussion:[3]

1 Improving the foresight and skill with which public and private health care resources are allocated. Improved forecasts would help achieve this objective primarily by providing an areawide "picture" of future needs

Table 1. Factors in demand for medical care*

Factors affecting a patient's demand for health services	Factors affecting a physician's prescription of health services	Health services actually demanded as a result
1. Incidence of illness or injury 2. Severity of illness or injury 3. Cultural values and attitudes 4. Demographic factors 5. Cost factors 6. Perceived accessibility, availability and convenience of service provider 7. Family, peer group, or other authoritative pressure 8. Life style factors affecting health	1. Patient characteristics, including relative cost to the patient of using different health services 2. Physician's institutional affiliations 3. Physician's knowledge and attitudes 4. Relative costs to physician of prescribing alternative health services 5. Availability of prescribed or desired services	1. Physician care in an office or other outpatient setting 2. Emergency care 3. Inpatient hospital care 4. Referrals to specialized services, long-term care facilities, or other care-giving institutions 5. Home health services

* These factors are not necessarily ranked in order of importance.

and demands into which public and private objectives can be harmonized.

2 Augmenting both quality and cost control in medical care. Improved forecasts would do this by helping to avoid both allocation of "excessive" community resources to *some* health services in the future, and the allocation of "insufficient" resources to others.

3 Anticipating and planning for the impact of developing technology and changing methods of organization upon the health care system.[4]

These are ambitious objectives. Many argue that they are *too* ambitious for existing forecasting techniques, that our current skills are too limited to achieve them. This paper explores this question in greater depth.

Specific Factors Contributing to the Demand for Health Services

"Need" and "demand" have received considerable attention in the literature.[5] This paper concentrates upon the concept of demand, the more tractable of the two notions from a forecasting viewpoint.[6] Demand for health services —generally defined as the amount of services actually used by individuals in the community —is a function of a number of economic, social, physical and environmental factors (see Table 1). These factors are significant in any comprehensive forecast of requirements for particular health services. But taken together, they present formidable problems to the forecaster. First, the nature of each and the interaction among them are difficult to determine with precision, especially on an area-wide basis.[7] Moreover, the influence of less readily measured variables in Table 1 (e.g., peer group pressure, perceived accessibility) is very poorly understood from a predictive standpoint. Second, the importance of the two columns varies in terms of *relative* influence upon service demand. Factors in the first column tend to be more influential in the demand for emergency and ambulatory services, for example, while factors in the second column assume greater importance when forecasting demand for inpatient and long-term care. Third, the relative contribution made by factors *within* each column is different from service to service. Family or peer group pressure, for example, tends to be very significant in the demand for long-term care, but less important in emergency care situations.

To be fully useful, however, forecasting techniques should be able to deal with these demand-oriented complexities in a reasonably effective way.

Major Sources of Community Change and Continuity Affecting Demand for Service

Many factors affecting demand are fairly stable, particularly in the short run. Others are more

Inquiry/Volume XIV, September 1977

Table 2. Possible sources of change and continuity in the community health care system

Sources of change	Sources of continuity
1. Overriding problem(s); examples include:	1. Organizational and demographic superstructure; examples include:
• Improving quality of care for selected target populations	• Institutional and professional licensure requirements
• Gaining control over runaway costs	• The complex of other federal, state, and local regulations
• Redefining the balance of interests between providers and consumers of health care services	• Long-established federal, state, and local financing programs
• Broadening access to care for selected target populations	• The demographic mix of the community population
• Overcoming fragmentation of service delivery patterns	• The geographic distribution of the population
2. Prime Movers; examples include:	2. Kind and capacity of health care facilities; examples include:
• Medicare, Medicaid	• Short-term, general hospitals
• Blue Cross, Blue Shield	• Specialty hospitals and institutions
• NHI, if enacted	• Nursing, personal, or domiciliary care establishments
• Associations of health care professionals (e.g., AMA, AHA, ANHA, etc.)	• Custodial care homes
• PSROs	

volatile. When forecasting, it may be useful to estimate community forces promoting change and encouraging preservation of the *status quo* in these factors. Some commonly advocated strategies include:[8]

Forecasting Major Sources of Change:

At least two readily identified initiators of significant change in the health care system are susceptible to analysis using one or more of the forecasting techniques to be evaluated in the next section:

- ☐ *The Overriding Problem(s)*:. Identify and evaluate those problem(s) facing the community health care system that are widely thought to require solution and that, if "solved," will result in significant changes.[9] For a sample of common problems, see Table 2.
- ☐ *The Prime Movers*: Identify and evaluate these institutions and processes that strongly and *independently* affect the way health care is delivered (see Table 2).

Forecasting Major Sources of Continuity

While some aspects of the health care system will change within the planning time frame, perhaps dramatically so, most will not. It often is helpful to identify and evaluate these aspects explicitly.[10] Several generic sources of continuity include the so-called organizational and demographic superstructure and the kind and capacity of health service facilities in the area. Both sets of factors change relatively slowly and both strongly affect demand for services. Table 2 includes examples of each.

Of the problems in weighing these factors, perhaps the most difficult arise when appraising sources of change. There is a large amount of folk wisdom available on how to judge the strength of political and social forces. But successful measurement and precise forecasting of such forces is rarely (if ever) achieved. It is feasible to determine the general direction in which "trends" are going. It is quite difficult, however, to be correct about their magnitude and timing. This is complicated by the large number of decision-makers and their correspondingly large measure of autonomy and discretion.

Six Significant Forecasting Techniques

So far, factors that contribute to demand for health services, and the broader set of factors that contribute to both change and continuity in the community health care picture, have been described without reference to how each might be forecast. This section describes six generic techniques for dealing with each set more systematically, and suggests the major strengths and weaknesses of each in areawide health planning contexts.

It is possible to place all forecasting techniques into one of two categories: 1) those that are essentially intuitive and not derived from formal models, including historical analogies and Delphi; and 2) those that *are* derived from formal, frequently mathematical models, including trend extrapolations, morphological techniques, computer simulations, and operations research models (exemplified herein by decision theory).

The following section discusses six examples of these techniques. Each is promising in the areawide health planning context, particularly with respect to factors discussed in the prior two sections. Each technique is briefly described, and its strengths and potential weaknesses pointed out.

Historical Analogies

This method searches for, and examines the history of, some medical or organizational development "like" that being forecasted (e.g., the history of Medicare legislation as an analogue to the current debate over National Health Insurance). Once found, such an historical analogy can be examined in any desired detail. The objectives of this examination might be to discern evolutionary trends, key decisions, necessary resources, societal impact, the "lessons of history," and so forth. Whether implicit or explicit, this is the most commonly used method for forecasting in most fields of social affairs.[11]

Consider, for example, this hypothetical forecast that employs an historical analogue: In terms of many management and scientific services, the medical research field has been about 25 years ahead of the practicing physician. In 1960, approximately 65 percent of the nation's medical research facilities were using computers for data analysis and synthesis tasks similar to those in physician/computer diagnostic consultation. Based on an analysis of the history of the diffusion of techniques from medical research to practical medicine it might be predicted that, by 1985, approximately 65 percent of the nation's physicians will employ computer diagnostic consultation services.

Characteristically, the historical analogy mode of forecasting has a mix of strengths and weaknesses. Those most likely to be of interest in forecasting areawide demand for health care are listed below:

Strengths Claimed.

- ☐ If the historical situation is closely analogous to the one in which one is interested and is well documented, potentially all "relevant factors" may be considered and extrapolated
- ☐ One can get an appreciation for possible future consequences of proposed innovations by reviewing those of the existing analogue and the attitudes of relevant observers, constituents, and so forth before, during, and after its development.
- ☐ One can review the successes (and failures) of past efforts to control, channel, or plan for the analogous medical or organizational developments in order to improve present efforts.
- ☐ The results are comparatively easy for lay board members and the general public to understand and interpret.

Possible Weaknesses:

- ☐ Closely (or even distantly) analogous historical situations may be unavailable, unperceived as such, or insufficiently documented to support systematic analysis.
- ☐ If available, they may tempt the analyst to expect future development(s) in question to proceed more or less exactly as the historical analogy did.
- ☐ If available, they may reinforce possible tendencies to view medical or organizational evolution in terms of inexorable, tidy, cause-and-effect chains instead of being rather fluid, subject to chance or to the effects of unique circumstance.
- ☐ If available, they may present some or all of the problems of concern to historians (i.e., the general question of the accuracy, completeness, and resulting validity of the historical record upon which analysis is based).

- There is very little theory available about how one should forecast using historical analyses; little or no formal methodology exists and few criteria for judging "good" forecasts are available.

Delphic Techniques

Delphi is a method by which "expert opinion" is employed in systematic fashion. A well-structured forecasting problem is presented to a group of people who are well grounded in the field or who are otherwise "expert." Each one makes forecasts concerning specified problems by separately and anonymously employing specially prepared questionnaires. Results are summarized in a way that displays the range of responses (usually in some quantitative way); these then are circulated among forecasters. The process is repeated for as many cycles as desirable in order to focus the array of forecasts. All significant interactions among participants are in writing rather than face-to-face.[12]

Consider a hypothetical example: An HSA situated in a large, metropolitan area is concerned with planning to meet future demand for emergency medical services. It has reason to think that computer-assisted diagnosis techniques, currently being discussed for emergency room use, will strongly affect the way planning should be done. To explore these issues, EMS experts employed Delphi at a workshop session of a joint physician, computer-industry symposium. They predicted that by 1985 approximately 25 percent of local emergency room physicians will have access to computer assisted diagnostic services. They employed expert opinion and data on increasing local experimentation with the use of automated techniques in the medical profession, together with data on related EMS trends. They arrived at their forecast over a three-day period, employing Delphi techniques in three iterations.

Delphi has enjoyed considerable popularity in recent years, being employed in a wide variety of forecasting problems. Some of the major reasons are listed below:

Strengths Claimed:

- Use of more than one "expert" (and more than one kind of expert) is possible in a well-structured way.
- The use of experts is systematic; all confront essentially the same forecasting problem in the same context.
- There is specific, purposeful interaction among participants on the same problem. This tends to sharpen judgment and ideas and tends to reduce idle, unfocused speculation.
- Disturbances or biases, owing to group dynamics or dominant personalities, can be minimized.
- Individual responsibility for the accuracy of forecasts (and hence possible inhibitions) is avoided.
- The process itself can provide considerable educational benefit for those who participate through clarification of problems, issues, and goals. In addition, it can open communication lines between representatives of groups that ordinarily do not maintain contact with one another, or that have a history of inadequate communication.
- The process can be especially effective in dealing with prediction problems that require intuitive judgment, and the marshaling of subconscious processes; problems, that is, which do not centrally depend upon empirical data and inductive processes.
- The process can serve as both preparation and follow-up for face-to-face meetings devoted to forecasting matters.
- The technique can be applied in any forecasting situation wherein "expert" or otherwise desirable opinion exists, and can be reached by mailed or other questionnaires.

Possible Weaknesses:

- It is difficult to design questionnaires for Delphic application that meet acceptable scientific standards for validity, appropriate selection and use of "expert opinion," reliability, and the avoidance of error due to small or biased samples of respondents. In practice, little attention seems to have been paid to these standards in conducting Delphi.[13]

- ☐ The iterative convergence of expert opinion may reflect increasing conformity instead of increasing accuracy.
- ☐ "Expert" forecasts may be more susceptible to errors arising from "schools of thought," conflicts, or from unanticipated developments in fields outside the expertise in the group. The method's usual total reliance on expert opinion, unleavened by other methodological checks, requires special caution.[14]
- ☐ The strength and accuracy of predictions are heavily dependent upon the initial conditions specified and upon the quality and appropriateness of the information supplied to the participants during the iterative process. Poor information tends to produce poor results, in Delphi as elsewhere.

Trend Extrapolation

A wide range of methods fall under this heading. Each involves a review of the historical data (primarily quantitative) pertaining to some problem or issue. Using these data, trends are "determined," assumptions made, intervening variables are identified where possible, and the interaction of some combination of these factors is projected into the future as a forecast. Regression or time series analysis are among the most frequently employed statistical trend extrapolation techniques.[15]

Two major variations of this method are common. The first projects data that express directly the phenomena being forecasted. Most bed-need formulas used in health service forecasting are a variation on this theme. For example, the well-known Hill-Burton formula projects bed requirements on the basis of trends in population, use-rates, and average daily census.[16]

The second common variety is based upon data that are thought or known to be *correlated* with the phenomenon being projected. This method often is employed when little or no data exist for the phenomenon one wishes to forecast. Use of demographic data as a surrogate for the factors in column one of Table 1 is a common example. Or—citing the basic hypothetical example used throughout this paper—an in-depth study of EMS facilities that already have adopted computer diagnostic consultation services might show that their usage is related in a complex way to some ten different variables. These might include such factors as physician work load, degree of medical specialization, access to and use of other consultative services, cost of the computer service, and others. Previously completed studies make it possible to predict the growth rate through 1985 for these ten variables. Using these growth rates and the historical correlation between these factors and the use of computers by EMS personnel, it might be possible to predict that 28 percent of local EMS personnel will employ computer diagnostic consultation in 1985.

Trend extrapolation methods have a number of strengths to commend them to areawide planners. Here are some of the more important ones:

Strengths Claimed:

- ☐ Future developments often may be straightforward, predictable continuations of the present and immediate past, at least within "acceptable" error limits (e.g., population growth and mix, incidence of disease, level and distributions of income). Where this is true, trend extrapolations can be very effective at minimal cost, particularly over the short run.
- ☐ These methods encourage the search for trends and continuities and for factors that cause, augment, or inhibit such trends.
- ☐ Since these methods deal with "objective" data and with mathematical techniques, they tend to reduce (though, of course, not eliminate) the effects of analyst prejudice or bias.
- ☐ The average behavior of large numbers of people is much more stable and "predictable" than that of individuals or small groups. The capacity of quantitative (particularly statistical) methods for dealing with a large volume of data is very helpful on issues involving the "trend" behavior of large populations.
- ☐ Most statistical methods are based upon

well-established statistical theory that specifies the confidence one can have in their results (e.g., the various methods for appraising the dispersion of data about their central tendency). This theoretical superstructure is very useful for drawing attention to potential sources of imprecision or diffusion of focus in particular kinds of trend extrapolations.

Possible Weaknesses:

- ☐ Apparent trends in the data have no life of their own. Their continuity over time *may* be very tenuous. Though the temptation exists, such trends cannot be taken for granted.
- ☐ No *formal* means exists for taking "new" trends or factors emerging in the future into account in advance. Indeed, the search may be inhibited by the perceived "objectivity" of extrapolation techniques.
- ☐ Most quantitative techniques require data that are quite precise and reliable if their major strengths are to be capitalized upon. Data most frequently available to areawide planners tend to be of uncertain reliability and validity.
- ☐ If the essence of the forecasting problem cannot be expressed in quantitative terms *and* in terms for which the data are available on a year-to-year basis, most available techniques are only partially usable.

Morphological Analysis and Relevance Trees

The overall intent of methods falling under these two rubrics is to make a forecasting problem more manageable by breaking it into subproblems. One seeks to discern the relative importance of each of the subproblems and to improve prediction accuracy by working with more effectively defined subproblems. Specific attention is paid to enumerating the set of all significant (feasible) factors (alternatives) for explaining some outcome (achieving some objective). These factors or alternatives may be defined logically or through use of inductive mathematical techniques, such as multiple regression or the computerized Automatic Interaction Detector (AID) program. The overall process is called morphological analysis; the graphic display of subproblems showing their interrelationships is called a relevance tree.[17]

Consider the following hypothetical example: An HSA was concerned with the future effect on total demand for services that will result from a rapid expansion of ambulatory care facilities. A large amount of data on both the working behavior of physicians and on factors contributing to demand for local ambulatory facilities was analyzed using the AID program. This program worked through the data, establishing "factors" through a series of dichotomous splits. These factors were chosen by the program because the dichotomy splits that were selected maximized the between-group variance. The graphic display of the series of these dichotomized factors represented a "picture" of all factors that would contribute to changes in demand for ambulatory care, expressed in terms of relative importance. These factors then were appraised by "experts" who rendered judgments about the extent to which each factor would change over the next few years. Taken together, the graphic display and expert judgment resulted in the forecast that demand for outpatient services would increase directly with an expansion in such facilities, leading to a relative decline in use of key inpatient services.

Whether graphic or quantitative, morphological analysis has not yet seen many applications in the health care context. Its strengths largely remain to be tested, but might include the following:

Strengths Claimed:

- ☐ Broad, large-scale forecasting problems are much easier to handle in "pieces."
- ☐ Relevance trees readily show what is (thought to be) contingent upon what, thus chains of causality, potential arenas for conflicts of interest, and areas where organizational change are necessary (or are likely) can be shown more clearly.
- ☐ Recombination of subproblems may show a variety of ways in which changes might occur and might thus dispel possible tendencies to fall into "single possibility" ruts.

- Emphasis upon the explicit enumeration of many possible alternative factors or solutions to each subproblem has significant heuristic value, encouraging the development of fresh perspectives.
- Feelings for specific policies, organizational developments, or social factors that effectively enhance or slow progress toward areawide health goals, often emerge from relevance tree analysis.

Possible Weaknesses:

- Where quantitative techniques are *not* used, successful analysis is dependent upon asking the "right" questions (i.e., decomposing the problem properly). Few effective guidelines exist, particularly if the problem is ill-defined or poorly understood, as are most areawide health forecasting problems.
- Where quantitative techniques *are* used, most of the possible disadvantages described in connection with quantitative trend extrapolation methods apply.
- Few *a priori* quantitative or qualitative guides exist for chosing the factors that are relevant to a given forecasting problem.
- "All possibilities" for a given problem may be extremely large in number, particularly in areawide planning—so many that the capacity of HSA planning staffs to evaluate them is swamped.

Computer Simulations

A conceptual model is constructed of the major conditions and processes of the situation to be predicted. These elements are related to one another in logical and/or mathematical relationships that express, as accurately as possible, their "real world" interaction within specified constraints. Employing the model and the computer, one can generate time series (i.e., changes in the value of key variables from time period to time period) that express the interaction of the model's elements. Assumptions, variables, processes and the interaction among them, can be varied to simulate various hypothetical conditions or the adoption of proposed policies in the "real" world.[19]

Consider the following example: An HSA collaborated with a local university over a two-year period to develop a model that expresses the interaction of demand for various levels of health care, including ambulatory, inpatient, emergency and long-term. Confronting numerous pressures to restrict new construction of inpatient beds, the HSA sought to simulate future demand for all services, assuming a policy that held the number of inpatient beds constant. The results shed light on the year-by-year increased ambulatory and long-term care capacity that would be required by this policy:

Strengths Claimed:

- The continuing, simultaneous interaction of an indefinitely large number of quantitative elements can be considered. This interaction can be simulated throughout an indefinitely long planning period. The only limitations are data availability, cost, and analyst imagination.
- Complex interactions over substantial periods of time frequently produce results that are counter-intuitive. Given the limited ability of the human mind to trace out such interactions unaided, simulations may be the only fully effective means of projecting complex interactive situations.
- The model's conditions can be varied at will in order to test the effect that various hypothesized changes in the real world will have upon future developments.
- Unlike nearly all operations research and decision theory models, there is no requirement to limit analysis to the achievement of a single objective. Progress toward multiple objectives can be projected since the model need not be predicated upon notions of optimality.
- Both the construction of the model and its actual use are processes with impressive heuristic value. Issues and problems frequently are exposed and clarified that would not have emerged otherwise.
- The "tracing out" by the computer of implications from the initial conditions specified in the model is nearly immune from analyst prejudice and bias.

Possible Weaknesses:

- The situation to be simulated must be well understood (i.e., the model's elements, their relevance, and their interaction must correspond well with "reality") in order to develop a model whose results are not misleading or simply erroneous. This possible weakness is especially salient in areawide planning where the range of considerations is extremely wide and the amount of verified knowledge is typically rather narrow.
- The variables employed in the model must be measurable with considerable validity, reliability, and precision, since the mathematical manipulation involved in even fairly simple models can cause the error rate to "explode" in unpredictable ways.[19]
- Modeling a situation in a way that captures the essence of some dynamic situation, without misleading oversimplification or misrepresentation, is much more art than science. Few guides for constructing "good" models exist, particularly those of complex social situations.
- Time series data, upon which most simulation models are based, must be treated carefully. The kind of summary indices that usually must be employed (e.g., "visits" or beds per unit of population) can be misleading. Even where such indices do a creditable job of capturing the essence of the phenomenon they represent, such statistics may not readily reflect changes in *quality* as contrasted with changes in numerical value.[20] Improvements (or retrogressions) in efficiency or effectiveness are difficult to take into account, as well.
- The developmental and operating cost of a computer simulation of any real significance is quite high. For this reason alone, simulations may be beyond the reach of any but the largest HSAs for the immediate future.

Decision Theory

This approach is representative of many found in the field of operations research. It is useful in making what might be called conditional forecasts (i.e., predictions that if the strategies prescribed by the techniques are adopted, the decision-maker can expect the most favorable future outcome under the conditions assumed). The basic method is to define a series of alternatives among which a choice is to be made, and to appraise each alternative in terms of its performance during a specified, comparable future time period. The consequences of adopting a given alternative frequently are traced out explicitly in the form of a decision tree, whose end points are expressed in numerical value payoff(s). The future environment may be expressed either in terms of risk (i.e., the future assumes one of several possible known states, each with a *known* probability of occurrence), or it can be dealt with in terms of uncertainty (i.e., the future assumes one of several known states, each with an *unknown* probability of occurrence). Once the alternatives are defined, and the range of possible interactions with the future environment is expressed in terms of possible payoffs, decision theory offers a range of criteria for choosing the best alternative.[21]

Suppose, for example, the HSA is considering four proposals to expand long-term care (LTC) capacity in the area. Each proposal differs in bed size. HSA staff studies indicate that future demand for LTC is somewhat unpredictable. Three possibilities are projected: low, medium, and high demand—each with an estimated probability of occurrence. A matrix is constructed that relates the additional capacity of each proposal to the three possible future demand levels. The HSA then makes its decision using a variant of the maximum expected value criterion:

Strengths Claimed:

- Given its emphasis upon explicitly tracing out as many possibilities as feasible, the method encourages one to consider forecasting problems from as broad a perspective as possible, viewing the future as having multiple possibilities.
- It forces analyst and decision-maker alike to be as explicit as possible about the process by which alternatives are defined, examined, and evaluated.
- It facilitates use and communication of expert opinion by structuring forecasting

Table 3. A paradigm for individually employing forecasting techniques when projecting demands for health care services

		Informal theory			Formal theory		
		Historical analogies	Delphic techniques	Trend explorations	Morphological analysis	Computer simulations	Decision theory
Sources of change	"Overriding problem(s)"		High validity, moderate precision				
	"Prime movers"	Moderate validity, moderate precision					
Sources of continuity	Organizational and demographic "superstructure"	Moderate validity, moderate precision		High validity, high precision			
	Kinds and capacities of health service facilities				High validity, moderate precision	Moderate validity, high precision	
Factors affecting patient demand	Incidence of disease	Moderate validity, high precision		Moderate validity, high precision			
	Severity of illness or injury						
	Cultural and demographic factors	Moderate validity, low precision			Moderate validity, moderate precision		
	Cost factors						High validity, moderate precision
	Accessibility, availability, convenience		Moderate validity, moderate precision				
	Family and peer group pressure						
	Life style factors						
	Patient characteristics						
Factors affecting physician prescription	Physician's institutional affiliations						
	Physician's knowledge						
	Relative costs to physician of alternative prescriptions						
	Availability of prescribed services			Moderate validity, low precision			

problems so as to take advantage of specific expertise. In particular, it helps distinguish and sort out issues of fact, conjecture, preference, strategy, probability of outcome, and value of outcome. This is very helpful for gearing available expertise to those aspects for which it is uniquely suited (e.g., questions of fact vs. issues of strategy vs. matters of preference).

- ☐ It aids decision-makers in sorting out their own attitudes and preferences with respect to situations of risk, uncertainty, and indeterminancy (as these terms are defined by decision theorists).
- ☐ It helps keep debate within decision-making groups focused upon important, objective aspects of the problem at hand by providing a "picture" (e.g., the decision tree or table) from which to work.

Possible Weaknesses:

- ☐ The outcome or payoff corresponding to a given alternative often is a complex bundle of consequences not readily summarized by available quantitative measures, such as dollars or patients treated. Where such measures are used, the risk of oversimplification is substantial.
- ☐ Outcomes associated with possible alternatives often are highly uncertain entities for which only rough estimates can be made. The apparent precision of a single-valued payoff may be highly misleading.
- ☐ The discrete, noninterdependent alternatives required in decision theory often may not be appropriate, given the actual situation. This is especially true where the range of possible alternatives is actually continuous (as contrasted with discrete), where no means for measuring the results of a given alternative exists, or where the environment is highly uncertain or rapidly changing.
- ☐ Most decisions must be made in an organizational context. Such environments feature inter- and intra-organizational bargaining over objectives and the simultaneous pursuit of multiple, noncomplementary goals.[22] These factors tend to preclude exploitation of some of the major advantages that decision theory has to offer. Particularly vulnerable are such strengths as the concept of utility maximization, uniform notions of optimality, well-defined criteria for choice, and the like.
- ☐ The technique is not concerned with forecasting *per se* as much as it is with guiding decision making when the main dimensions of the future environment are known, or can be assumed with reasonable confidence.

A Paradigm for Applying Forecasting Techniques to Factors Influencing Future Demand for Health Services

At this point in the discussion, it is necessary to mention a framework within which existing forecasting techniques can be applied to advantage. Table 3 provides a first approximation of this framework.

Each of six major forecasting techniques discussed in the preceding section is represented in the columns of Table 3. Each of the major factors influencing demand is represented in the rows. The cell at the intersection of a given row and column represents the application of the column technique to the row factor. Table 3 represents a single technique applied to a single factor. In actual practice, of course, no such limitation would exist. In many cases, best results would be achieved employing two (or more) methods jointly.

A representative sample of cells have been selected because of the attention that the topics they represent receive in the literature. Blank cells indicate areas where little or no research activity has taken place. As Table 3 indicates, these unexplored areas predominate over areas where work has been done. Each cell is evaluated in summary fashion in terms of two criteria. These criteria are intended to express the overall usefulness of a given technique, as applied to a given factor. The criteria:

- ☐ Validity—to what extent can the technique be applied to the "essence" of the forecasting problem (as distinguished from its peripheral, if directly measurable or readily understood, aspects)?
- ☐ Precision—to what extent is the output of the technique expressed in precise terms?

Each of these is expressed along a single dimension—low, moderate, high.

There are three basic processes for gauging the validity and precision of a technique for forecasting where the technique is based upon formal theory.[23] First, one can verify the logical reasoning that led from the theory to the prediction in question. Second, one can evaluate the "reasonableness" of the assumptions that underlie the portrayal of the real world situation being forecasted. And third, one can verify some of the key predictions, comparing them against actual events as they unfold. These processes can be used either singly or in combination to appraise the quality of the output produced by the various forecasting methods.[24]

The results shown in Table 3 were *not* arrived at through rigorous application of these processes. Such an investigation is well beyond the scope of this paper, requiring examination of specific applications of each technique. Instead, Table 3 is intended to convey an overall impression based upon discussions noted earlier. With this caveat in mind, several interim conclusions emerge:

First, where possible, two or more forecasting methods should be employed in mutually reinforcing ways. In particular, historical analogies and Delphi are almost always more effective as exploratory techniques, and/or in conjunction with one or more of the remaining techniques. No single method should be relied upon exclusively.

Second, given the large amount of unavoidable uncertainty in forecasting, it is important to develop strategies and analytical techniques for estimating and hedging against possible error. Forecasts that provide an "optimistic-pessimistic" range rather than a point estimate, or that include a discussion of possible mitigating factors, serve this end.

Third, as Table 3 indicates, much of the field of demand forecasting remains largely unexplored at present. Especially deserving of attention appear to be those factors that affect the physician's impact on demand.

Fourth, the more sophisticated the forecasting technique, the better the situation to be forecasted must be understood to employ it, particularly on an areawide basis. Relatively simple methods, focused on the right issues, can be quite effective. Sophisticated methods without data of sufficient quality and quantity, or without the proper issue focus, may be of little use.

Some Apparent Dimensions of the "Zone of Feasible Forecasting"

An effective way to chart some of the major dimensions of this zone is to compare the state of the art against the objectives of areawide forecasting cited in the first section of this paper.

To Provide a Picture of the Future Context Helpful to All Planners, Public and Private

The new HSAs will be required to produce forecasts (perhaps updated annually) of the future areawide health care picture. These will be useful in carrying out their mandated functions and also can be used by all parties, public and private. The state of the art is perhaps strongest with respect to this objective. Fairly long-term forecasts, composed partly of quantitative and qualitative information, are well within the capability of existing methods. These probably can be made with "enough" assurance to provide a useful planning context for all interested parties.

To Improve Quality and Cost Control

The picture is less promising with respect to the capacity of current methods to meet this objective. Available methods can forecast demand for individual services with reasonable accuracy. But the state of the art reflects two major deficiencies: First, most methods to date have not been geared to forecasting demand for *two or more services* simultaneously or to forecasting demand for subspecialities within a given type of care (computer simulation being the major exception). This means the danger of ignoring the impact of changes in one service (or one subspecialty) upon the other is quite real. Second, most methods do not focus upon the *appropriateness* of utilization, hence the risk of forecasting patterns of demand that are inefficient, ineffective, or extremely costly (and making poor decisions based upon the results) is quite high. It should be noted that failure to gauge "appropriateness of utilization" is not an intrinsic failure of forecasting methods; skillful and imaginative use of existing methods could overcome this deficiency, especially if supported by appropriate utilization studies.

To Anticipate and Plan for New Technologies and Methods of Organization

Available forecasting methods, *based upon current demand data*, are poorly suited to achieving this objective. Demand is not a direct function of either technology or methods of organization. Hence, the use of current demand data (influenced to whatever degree by historical technology or methods of organization) is a poor way to achieve the objective.

To summarize, the state of the art in forecasting appears to support several limited but important forecasts on an areawide basis. These include:

- ☐ Short-term (two to three years) forecasts of demand based on current utilization, expressed in quantitative terms.
- ☐ Medium-range (four to six years) forecasts of possible (likely) changes in future demand, expressed in both quantitative and qualitative terms.
- ☐ Long-range forecasts (seven or more years) of the future context in which the health care system will find itself, expressed largely in qualitative terms, and in terms of broad quantitative ranges.

References and Notes

1 U.S. Congress, *The Comprehensive Health Planning and Public Health Services Amendments of 1966*, S-3008, 89th Congress, Nov. 3, 1966.

2 U.S. Congress, *The National Health Planning and Resources Development Act of 1974*, S-2994, 93rd Congress, Jan. 4, 1975.

3 See discussion of these objectives in the following: Arnold, M. F.: "Basic Concepts and Crucial Issues in Health Planning." *American Journal of Public Health*. 59: 1, pp. 1686–97, 1969. Colt, A. M.: "Elements of Comprehensive Health Planning." *American Journal of Public Health*. 60: 7, pp. 1194–1204, 1970. Dickey, W. J. et al.: "Comprehensive Health Planning—Federal, State, Local: Concepts and Realities." *Wisconsin Law Review*. pp. 839–78, 1970. Finally, note the legislative background to Section 1513 of the Health Planning and Development Act of 1974 (P.L. 93-641).

4 Horvath, W. J.: "Need for Estimating The Influence of Technological and Social Changes on Future Health Facility Requirements." *Health Services Research*. 3, Spring 1968, pp. 3–9. Fein, R.: "Medical Care Needs in the Coming Decade." *Bulletin of the New York Academy of Medicine*, 45, March 1969, pp. 255–70. Mechanic, D.: "Problems in the Future Organization of Medical Practice." *Law and Contemporary Problems*. 35: 233–51, 1970.

5 For three divergent perspectives on this issue see Kennedy, D. A.: "Perceptions of Illness and Health." *Social Science and Medicine*. 7: 787–805, 1973. Boulding, D. E.: "The Concept of Need For Health Services." *Milbank Memorial Fund Quarterly*, Part II. 44(4): 202–220, October 1966. Donabedian, A. V.: "Assessment of Need," Chapter 3 of *Aspects of Medical Care Administration: Specifying Requirements For Health Care*. Cambridge, MA: Harvard University Press, 1973.

6 There are a number of good reviews of this literature. See Berki, S. E.: "Demand and Utilization," Chapter 6 of *Hospital Economics*, (Lexington, MA: D. C. Health & Co.) 1972, pp. 121–165, and Feldstein, P. J.: "Research on The Demand for Health Services." *Milbank Memorial Fund Quarterly*, 44(3): 123–57, July 1966; and Hyman, J.: "Empirical Research on The Demand for Health Care." *Inquiry*, 3(1): 61–71, 1967; and Andersen, Ronald: *A Behavioral Model Of Families' Use Of Health Services*, (Chicago: Center For Health Administration Studies Research) Series #25, 1968. For a treatment of the distinction between the two concepts from the viewpoint of economic theory, see Jeffers, et al.: "On Demand vs. Need For Medical Services and the Concept of 'Shortage.'" *American Journal Of Public Health* 61(1): 46–63, January 1971.

7 Some representative examples: Anderson, J. G., and Bartkus, D. E.: "Choice of Medical Care—A Behavioral Model of Health and Illness Behavior." *Inquiry* 14: 348–62, December 1973. Wirick, G.: "A Multiple Equation Model of Demand for Health Care." *Health Services Research* 1(3): 301–46, Winter 1966. Hyman, J.: "Empirical Research on the Demand for Health Care." *Inquiry* 3(1): 61–71, 1966. Anderson, J. G.: "Demographic Factors Affecting Health Services Utilization: A Causal Model." *Medical Care* 11(2): 104–20, March/April 1973.

8 For a wide range of viewpoints on both the desirability and feasibility of making assessments of these factors, see the entire issue of *Daedalus*: "The Proceedings of The American Academy of Arts and

Sciences," 96: 639–998, Summer 1967. It reports the proceedings of a symposium devoted to issues in large-scale social forecasting.

9 For a discussion of how this concept has been developed and used in a variety of forecasting contexts, see: Bell, D.: "Twelve Modes of Prediction—A Preliminary Sorting." *Daedalus: The Proceedings of The American Academy of Arts and Sciences*, 93: 845–73, Summer 1964.

10 See the entire issue of: *The Annals of The American Academy of Political and Social Sciences.* 399: 1–160, January 1973, devoted to the general topic of: "The Nation's Health: Some Issues," edited by Sylvester E. Berki and Alan W. Weston.

11 This brief description, of course, severely understates the scope of a major methodological field—historiography. While not within the scope of this paper, a discussion of the methodological problems in reconstructing recent history is worth reviewing. A convenient survey of philosophies of history, which clearly indicates the importance of the "theory of history" employed with regard to selectivity of evidence and the variability with which it is weighted in evaluating a given historical situation, can be found in Patrick Gardiner (ed), *Theories of History*, (N.Y.: Free Press), 1959, and in Dray, William H., *Philosophy of History*, (Englewood Cliffs, N.J.), 1964, Chapter 3. See also Mazlish, Bruce (ed.): *The Railroad and The Space Program—An Exploration in Historical Analogy* (Cambridge, MA: MIT Press), 1965.

12 For one of the best of the early treatments of this topic, see: Helmer, Olaf, *Social Technology*, (N.Y.: Basic Books), 1966. Also see: Vickers, Sir Geoffrey, *The Art of Judgement: A Study of Policy Making*, (London: Chapman & Hall), 1965. For a more recent evaluation of Delphi in the context of the field of subjective scaling, see Pill, Juri: "The Delphi Method: Substance, Context, A Critique and an Annotated Bibliography." *Socio-Economic Planning Sciences.* Vol. 5, No. 1, pp. 57–71, 1971. Perhaps the most critical review of Delphi is contained in Harold Sackman's *Delphi Critique: Expert Opinion, Forecasting, and Group Process*, (Lexington, MA: Lexington Books) 1975. An application of Delphi in the long-term care context may be found in: Clark, L. H., and Cochran, S. W.: "Needs of Older Americans Assessed by Delphi Techniques." *Journal of Gerontology* 27(2): 275–78, April 1972.

13 This potential weakness of Delphi is perhaps most frequently overlooked in practice. A rigorous assessment of Delphi against standards for questionnaire design, as embodied in the *Standards For Educational and Psychological Tests and Manuals* published by the American Psychological Association (1966), is conducted by Sackman in Chapter 3 of his *Delphi Critique: Expert Opinion, Forecasting And Group Process*, (Lexington, MA: Lexington Books) 1975, pp. 11–27.

14 Exclusive use of expert opinion to make assessments and to validate tests has been evaluated in a number of contexts. The results have tended to show that, even in situations where the problems are fairly well structured and the information available is reasonably complete, expert opinion is not especially effective or reliable as compared with plausible alternative methods. For studies in several contexts, see Zarnowitz, Victor: "On The Accuracy And Properties Of Short-Term Economic Forecasts." *The Tasks Of Economics: The Forty-Fifth Annual Report Of The National Bureau Of Economic Research*, 1965; also Hersch, Charles. "The Discontent Explosion in Mental Health," *American Psychologist*, Vol. 23, 1968, pp. 497–506. While exclusive use of expert opinion applied to situations *in the present* has been largely discredited on scientific grounds, it could be argued that *future* situations (i.e., forecasting) are a different matter. This may be true, but the burden of proof seems to lie with those who advocate this view, at present.

15 For a good discussion of the major quantitative extrapolation methods in an econometric context, see Fox, Karl A., *Intermediate Economic Statistics*, (New York: Wiley & Sons) 1968, especially Chapters 7, 8, 9 and 11.

16 Report of The Committee of Consultants, *Evaluation Of The Hill-Burton Bed Need Formula: Short-Term General Hospital Beds*, Health Care Facilities Service Evaluation Project 71-4, August 1972.

17 For an example of morphological analysis applied in a nonhealth context, see Shillif, I. E., and Smith, R. D.: "A Forecasting Method for Setting Short-Range Research Objectives." *Research Management.* 12: 24–34, March 1972. For a description of the AID program, see Anderson, R., et al: "Automatic Interaction Detector Program For Analyzing Health Survey Data." *Health Services Research.* 6: 165–83, Summer 1971. For a comparison among multi-variate statistical techniques (including AID) from a methodological viewpoint, see Reinke, William A.: "Analysis of Multiple Sources of Variation: Comparison of Three Techniques." *Health Services Research.* 9: 311–21, Winter 1973. For a technical discussion of relevance trees see Fischer, Manfred, *Technological Forecasting and Social Change* I: 4 (1970), 381–89.

18 For a primer on the basic methods of computer simulation, see Forrester, J. W.: *Principles of Systems*, Cambridge, MA: MIT Press, 1967. For an application of simulation in a field related to area health planning, see Forrester, J. W.: *Urban Dynamics* (Cambridge, MA: MIT Press) 1969. For a specific areawide health planning application see Yett, D. E., Drabek, L. and Intrilligator, M. D.: "A Micro-Econometric Model for Regional Health Planning." *Business and Economic Bulletin*, 24(1): 1–21, Fall 1971.

19 This problem is discussed effectively by Alonso, W.: "Predicting Best With Imperfect Data." *Journal Of The American Institute Of Planners*, 34: 248–55, July 1968.

20 Bauer, R. A. (ed.), *Social Indicators*, (Cambridge, MA: MIT Press) 1966.

21 For a concise survey of the major conceptual issues in the field of formal decision theory, see: Ward Edwards and Amos Tversky (ed.), *Decision Making: Selecting Readings*, (Middlesex, England: Penguin Books), 1967, or Raiffa, H.: *Decision Analysis: Introductory Lectures On Choices Under Uncertainty* (Reading, MA: Addison Wesley) 1968. An interesting application of decision theory in regional health planning appears in Grimes, et al.: "Use of Decision Theory in Regional Planning." *Health Services Research*, 7: 73–78, 1974. Of particular inter-

est to health planners in the project review function is Reutlinger, Schlomo: *Techniques For Project Appraisal Under Uncertainty*, World Bank Staff Occasional Papers Number Ten, (London: Johns Hopkins Press), 1970. For a discussion of the context into which decision theory fits, see Chapter X of Kahn, H. and Wiener, A. J.: *The Year 2000: A Framework for Speculation*, (New York: Macmillan Co.), 1967.

22 For one of the most succinct discussions of this organizational tendency, see Cyert, R. N., and March, J. G.: *A Behavioral Theory Of The Firm*, (Englewood Cliffs: Prentice-Hall), 1963, Chapter 3. Also valuable is: Thompson, J. D., and McEwen, W. J.: "Organizational Goals and Environment: Goal Setting As An Interactive Process." *American Sociological Review*. 23: 23-31, 1958.

23 The epistemology of this matter is by no means resolved. For example, see Helmer, O. and Rescher, N.: "On the Epistemology of the Inexact Sciences" *Management Sciences*, 6: 1, 1959 pp. 25-52. A more general view of the major issues is contained in Frank, P.: *Philosophy of Science* (Englewood Cliffs: Prentice-Hall), 1957, especially Chapters 11 and 12, and in Ikle, F. C.: "Can Social Predictions Be Evaluated?" *Daedalus: Journal of the American Academy of Arts and Sciences*. 96(3): 733-58, Summer 1967.

24 For a rendering of "the basic philosophical problem facing the forecaster" from the perspective of each of five philosophically plausible points of view, see Mitroff, W. I., and Turoff, M.: "Technological Forecasting and Assessment—Science and/or Mythology" *Technological Forecasting and Social Change*, Vol. 5, 1973.

James B. Henry
Rodney L. Roenfeldt

Cost Analysis of Leasing Hospital Equipment

For many years, hospitals throughout the United States have used funds from grants, donations and other free money sources to underwrite capital expenditures for equipment. Unfortunately, these free funds have not kept pace with the rapid increase in demand for capital expenditures. As an alternative to the outright purchasing of equipment, some hospitals have turned to leasing. Although no exact figures are available, it is estimated that over [illegible] million worth of assets are leased to hospitals in the United States each year.[1]

In an attempt to determine why hospitals choose to lease, administrators and controllers of 25 non-profit hospitals in South Carolina were interviewed. Several reasons were offered for their choosing leasing as a financing vehicle. Six of the reasons most frequently suggested are listed below:

1. The asset is viewed to have a high degree of technological obsolescence, therefore the ability to trade in equipment is desired.
2. Service is better on leased equipment than on purchased equipment. The ability to stop making lease payments seems to influence performance on maintenance contracts.
3. The asset is only available under a lease agreement.
4. The hospital can be reimbursed faster under a lease than under a purchase agreement.
5. Leasing provides capital funds when other sources are scarce.
6. Leasing rates are lower than borrowing rates.

Each of the hospital officials interviewed indicated a strong interest in the cost of leasing, but all had little insight as to how one might determine financial charges. Most of those sampled who felt that leasing rates were less than borrowing rates did not compare the financial cost of leasing with debt financing using present value analysis. The objective of this paper is to provide some information concerning the financial cost of leasing for hospitals by examining 58 noncancellable contracts obtained from 17 of the 25 hospitals interviewed. The hospitals ranged in size from 52 to 603 beds, and the contracts varied from $3,990 to $424,500 in asset cost, totaling $3,067,498 in assets. The contracts were initiated in the years 1971–1976 and varied in length from 36 to 84 months.

The following sections describe the guidelines used to determine whether a financing arrangement is a true lease or lease-purchase, outline the cost measurement procedure used to evaluate the sample of leases and lease purchase financing alternatives for non-taxpaying facilities, and present the financial costs which are currently being incurred by the hospitals with these financing arrangements. It should be pointed out that this analysis is concerned only with determining the effective interest rates for the financing alternatives; it does not consider the impact of third party reimbursement. Many hospital officials are not aware of the effective rate of interest on their lease contracts, and the fol

James B. Henry, Ph.D., and Rodney L. Roenfeldt, D.B.A., are Associate Professors of Finance, College of Business Administration, University of South Carolina, Columbia, SC 29208.

[illegible] performed pursuant to contract [illegible] Security Administration [illegible]

lowing analysis is intended to explicitly indicate the range of these rates.[2]

True Lease Versus Conditional Sale

Although several types of arrangements are frequently referred to as lease financing by lessors and lessees, hospitals should be aware of the differences in financing arrangements, since some lease contracts are actually conditional sales. With a true lease, rental payments are expensed for reimbursement during the year in which they occur; on the other hand, with a lease-purchase (conditional sale financed by a leasing company), the asset cost is depreciated and reimbursed over its useful life. The guidelines for reimbursement of a true lease are basically the rules outlined by the Internal Revenue Service for corporate taxpaying entities.[3] Non-profit hospitals and taxpaying facilities, therefore, face similar rules.

In summary, these rules state that a lease will be construed as a sale if any equity interest in the property is built up by the lessee during the life of the contract; or secondly, if rental payments are not considered to be fair rent; or thirdly, if an asset is purchased at termination of the contract for a price less than fair market value. Furthermore, interest payments should not be stated as a separate cost in addition to principal.

The feature in a contract that most often causes a lease to be construed as a conditional sale is when the predetermined salvage value indicated in the contract is not equal to fair market value. Contracts allowing for transfer of title for an amount less than fair market value—such as a nominal fee, dollar option, 10% of original cost, or call for abandonment—would not likely qualify for true lease status, but instead be construed as a lease-purchase. In addition, side letters that are separate from the contract which state a specific purchase price less than fair market value do not change the nature of the contract; if honored, they will likely result in the financial arrangement being considered a conditional sale.

Cost of Leasing

Cost Measurement for Non-profit Facilities

Many lessees in the sample were of the opinion that the only relevant data to include when measuring an interest rate on a lease contract is the cost of the equipment and the rental payments during the period of the lease. Leasing companies typically compare this so-called "running rate," represented by r_L^* in equation (1), to the hospital's estimated cost of borrowing money in the debt market.

$$C_o = \sum_{t=1}^{N} \frac{L_t}{(1 + r_L^*)^t} \tag{1}$$

Where:

- C_o = Capitalized value or cost of the asset.
- L_t = Lease payment in period t
- r_L^* = Effective before-tax interest rate on lease contract with a zero salvage value.
- N = Length of lease contract.

An obvious flaw in this type of analysis is ignoring the value of the asset at the end of the lease. A true lease requires the lessee to pay fair market value for the asset at contract termination if the lessee desires to own the asset. If a debt financing arrangement is used and the asset is purchased under a conditional sale, the hospital has title to the asset at termination of contract with no additional payment necessary. To be consistent in measurement when comparing these alternatives, one must either assume the hospital sells the asset under the purchase agreement and will not exercise a purchase option under the lease, or assume that one keeps the asset if purchased and will exercise a purchase option under the lease. A running rate that ignores a fair market value purchase option cannot be compared with a debt financing rate unless the salvage value is zero. Inclusion of salvage value when calculating a measurement rate for a lease contract can be accomplished by adding a term to equation (1) as follows:

$$C_o = \sum_{t=1}^{N} \frac{L_t}{(1 + r_L)^t} + \frac{S}{(1 + r_L)^N} \tag{2}$$

Where:

- S = Estimated fair market value of asset at termination of lease contract
- r_L = Effective before-tax rate on lease contract assuming a salvage value of S.

In the event an asset has a salvage value greater than zero, it is evident that the true effective before-tax rate on the lease contract is greater than the running rate.[4] Since it is

Table 1. Before-tax rates for hospital equipment*, March 1975–June 1976, Prime Rate: 6½–8½%

Equipment cost	Estimated salvage value (Percentage of cost) 0%	10%	25%	40%	50%	Lease-purchase
$0–$49,999	8.8%	11.9%	15.8%	19.1%	21.0%	
	11.4	14.3	17.9	21.1	23.0	
	11.5	14.4	18.1	21.2	23.1	
	12.6	15.3	18.6	21.6	23.3	
	13.8	16.3	19.6	22.5	24.3	
						18.5%
						19.6
						26.1
$50,000–$99,999	12.6	15.3	18.7	21.6	23.3	
	83.7	84.6	85.6	86.7	87.3	
						10.7
$100,000 & over	10.2	[illegible]	16.8	19.9	21.8	
	9	14.6	18.1	21.0	22.8	

* All 60 month contracts

difficult to estimate the salvage value with great precision, a useful approach is to use more than one estimate.[3] This will result in a range of rates which may be compared to the debt rate. With this approach, the hospital analyst can be made aware of the sensitivity of effective lease [illegible] salvage value, as well as being provided with an indication of his margin for error in estimation before making a decision.

Effective Before-Tax Lease Rates for Sample Data

Effective before-tax rates for salvage values of 0%, 10%, 25%, 40%, and 50% of original cost are calculated for the true lease contracts in the sample. Before-tax rates are also calculated for lease-purchase contracts. It is assumed that the assets are purchased at termination of the lease-purchase contracts for the stated option price.

To identify changes in rates caused by changes in general interest rates, three periods—May 1971 to July 1973, August 1973 to February 1975, and March 1975 to June 1976—are established based on prime rates existing during the period. Prime rates ranged from 4¾ to 8¼%, 8½ to 12%, and 6½ to 8½% for each period, respectively. A contract is assigned to each period based on its origination date and further categorized by its initial cost. Tables 1 through 3 present before-tax measurement rates computed for 53 contracts [illegible] maturities of 60 and 84 months.

Table 1 contains data for the third and [illegible] ent period, with contracts originating between March 1975 and June 1976 when the prime rate ranged from 6½–8½% (three contracts were proposals submitted by leasing companies but not yet accepted by the hospital). Running rates ranged from 8.8 to 13.8% with the exception of one extremely high rate of 83.7% for a proposed contract. Although some hospital equipment may not be considered as a prime investment for commercial banks, not all assets should be considered as high risk, yet all but a few of the contracts had running rates which far exceeded the prime rates. If one assumes a 25% purchase option, the majority of the sample lease cases appear to be very expensive.

Importance of salvage value in determining effective lease rates is evident by the sensitivity of rates to varied salvage estimates. For example, the 13.8% running rate increases to 19.6% with an estimated salvage value of 25% of equipment cost. It is impossible to generalize about the size of salvage value which is appropriate for every piece of equipment, but estimates obtained from hospital administrators and leasing companies tended to range between 10% and 40% of original cost. Effective rates ranged from 11.9 to 16.3% and 19.1 to 22.5% for salvage values of 10 and 40% respectively (excluding the extreme case with a rate exceeding 80%).

Although no discernible relationship exists between equipment cost and [illegible] rates, the rate (10.7%) for the one lease-purchase contract for equipment [illegible]

Table 2. Before-tax lease rates for hospital equipment*, August 1973–February 1975, Prime Rate: 8½–12%

Equipment cost	Estimated salvage value (Percentage of cost) 0%	10%	25%	40%	50%	Lease-purchase
$0–$49,999	11.3%	14.1%	17.7%	20.8%	22.6%	
	11.9	14.7	18.2	21.3	23.1	
	16.8	19.2	22.4	25.2	26.9	
	67.1	68.0	69.3	70.5	71.3	
						9.1%
						14.5
						21.5
						26.1
$50,000–$99,999	8.8	11.6	15.2	18.3	20.1	
	10.0	12.9	16.6	19.7	21.6	
	16.9	19.3	22.5	25.3	27.0	
	17.2	19.6	22.7	25.4	27.0	
	23.2	25.4	28.2	30.7	32.2	
						12.0
$100,000 & over	20.5	22.7	25.7	28.3	29.9	

* All 60-month contracts.

$50,000 is considerably less than rates for three lease-purchase contracts with costs under $50,000 (18.5 to 26.1%). Including estimated salvage value in the lease analysis is a crucial factor when determining whether the lease is more or less expensive than a lease-purchase or alternative debt financing arrangement. It was clear in the sample interviews that restrictions on some hospitals, preventing them from borrowing directly from banks, undoubtedly increased their financing cost by forcing them to use true lease and

Table 3. Before-tax lease rates for hospital equipment, May 1971–July 1973, Prime Rate: 4⅜–8¾%

Equipment cost	Estimated salvage value (Percentage of cost) 0%	10%	25%	40%	50%	Lease-purchase*
$0–$49,999	6.2%	9.7%	14.1%	17.6%	19.7%	
	11.0	13.9	17.6	20.7	22.6	
	11.2	14.1	[illegible]	20.9	22.8	
	11.2	14.4	18.4	21.8	23.8	
	11.5	14.2	17.7	20.6	22.4	
	11.5	14.2	17.6	20.6	22.4	
	11.5	14.2	17.7	20.6	22.4	
	11.9	14.6	18.0	21.0	22.8	
	11.9	14.6	18.0	21.0	22.8	
	15.2	17.6	20.9	23.7	25.4	
	9.2*	11.0	13.3	15.2	16.4	
	9.2*	11.0	13.3	15.2	16.4	
	9.2*	11.0	13.3	15.2	16.4	
	9.2	11.0	13.3	15.2	16.4	
	9.2*	11.0	13.3	15.2	16.4	
	9.2	11.0	13.3	15.2	16.4	
	9.2*	11.0	13.3	15.2	16.4	
	9.2	11.0	13.3	15.2	16.4	
$50,000–$99,999	20.3	22.8	25.8	28.2	29.8	
	9.2*	11.0	13.3	15.2	16.4	
	13.0*	14.5	16.8	18.3	19.3	
	13.0	14.5	16.8	18.3	19.3	
$100,000 & over	9.2*	11.0	13.3	15.2	16.4	
	9.2*	11.0	13.3	15.2	16.4	
	9.2*	11.0	13.3	15.2	16.4	

* All rates with asterisk indicate 84 month contracts and are from the same hospital. All other contracts are [illegible] in duration. There were no lease-purchase contracts available during this time period.

lease-purchase financing agreements. The restrictions apparently put hospitals at a decided cost disadvantage for raising capital—without accomplishing the attempted objective of limiting debt financing—since both the true lease and lease-purchase are in essence noncancellable debt-type agreements.

Tables 2 and 3 contain effective before-tax lease rates for the second and first periods respectively. These data also indicate a wide variation in rates among contracts. These two periods have extremely different prime rates—8½ to 12% for August 1973 to February 1975, and 4⅞ to 8¾% for May 1971 to July 1973—with lease rates on the first (earliest) period generally lower than for the second period.

The misleading nature of considering only running rates is illustrated by additional information obtained for two true lease contracts initiated during the first period. Each contract was about to expire and the hospital was currently negotiating with the leasing company to determine the fair market value of the equipment. The leasing company had established a fair market value for one piece of equipment equal to almost 25% of original cost. The running rate for the contract was 11%, but the effective rate including the required salvage value was 17.5%. Similarly, the effective rate on a second contract which had a running rate of 11.2% was increased to over 20% due to a salvage value equal to 33% of original cost being established at the end of the contract.

Three of five shorter term contracts not included in the tables illustrate even more dramatically the importance of including salvage value estimates in cost evaluation. The contracts were terminated and fair market value purchase-options were exercised. The two 48-month contracts with running rates of 6.9% and 4.3% terminated with a hospital paying salvage values of 36 and 37½% of cost, respectively. These actual salvage values increased the effective lease rates to 19.3 and 18.5%. Also, a purchase option on a 36-month contract, with a running rate of 1.8%, was exercised at 52% of cost, resulting in an effective rate of 26.7%.

Summary

This paper provides information concerning interest rates on true leases and lease-purchase agreements for hospital equipment. Since the salvage value estimate is extremely important for determining the effective rate of a true lease contract, various estimates are used. This procedure provides the analyst with a range of rates comparable with the debt rate and also indicates the sensitivity of the lease rate to changes in salvage values.

The effective rates on the lease and lease-purchase contracts obtained from South Carolina hospitals were quite high. Most rates ranged from 12% to 25%, depending upon the salvage value estimate. But some rates were considerably higher, with one contract having a rate of 87%.

References and Notes

1. Henry, J. F., "Finding the Cost of a True Lease," *Hospital Financial Management*, December 1976, pp. 19-28.
2. Caplan, [illegible] A., "Leasing: Is It Really the Answer To Your Needs?" *Medical Instrumentation*, Vol. [illegible], No. [illegible], January-February 1977, pp. 48-50.
3. See Internal Revenue Service, Rev. Rul. [illegible].
4. For additional information see Henry, J. B. and Reynolds, R. J., "Hospital Equipment Financing," Contract [illegible] [illegible] with the Social Security Administration, Department of Health, Education and Welfare.

Although risk associated with the salvage value is not explicitly considered in equation (2), it is indirectly considered by examining the effect of different salvage values upon the effective interest rate. For a detailed comparison of alternative lease cost measurement techniques and alternative adjustments for risk, see Bower, Richard S., [illegible] *Financial Management*, [illegible], pp. [illegible].

21

Methodologic Articles

APPLICATION OF COST-BENEFIT ANALYSIS TO THE HEALTH SERVICES AND THE SPECIAL CASE OF TECHNOLOGIC INNOVATION

Herbert E. Klarman

As an economic technique for evaluating specific projects or programs in the public sector, cost-benefit analysis is relatively new. In this paper, the theory and practice of cost-benefit analysis in general are discussed as a basis for considering its role in assessing technology in the health services. A review of the literature on applications of cost-benefit or cost-effectiveness analysis to the health field reveals that few complete studies have been conducted to date. It is suggested that an adequate analysis requires an empirical approach in which costs and benefits are juxtaposed, and in which presumed benefits reflect an ascertained relationship between inputs and outputs. A threefold classification of benefits is commonly employed: direct, indirect, and intangible. Since the latter pose difficulty, cost-effectiveness analysis is often the more practicable procedure. After summarizing some problems in predicting how technologic developments are likely to affect costs and benefits, the method of cost-benefit analysis is applied to developments of health systems technology in two settings—the hospital and automated multiphasic screening. These examples underscore the importance of solving problems of measurement and valuation of a project or program in its concrete setting. Finally, barriers to the performance of sound and systematic analysis are listed, and the political context of decision making in the public sector is emphasized.

The purpose of this paper is to discuss the application of cost-benefit analysis to the assessment of technology in the health services. With the few exceptions that are noted, the focus of this paper is on services, not research.

In carrying out this task, there is no substantial body of empirical research literature to draw upon, analyze, and synthesize. Accordingly, the task will be approached in three distinct steps. First, the theory and practice of cost-benefit analysis in general will be reviewed. Second, applications to the health field will be discussed. Third, the potentialities and limitations of cost-benefit analysis for the assessment of health systems technology will be suggested, using concrete illustrations.

THEORY AND PRACTICE OF COST-BENEFIT ANALYSIS

As a formal and systematic approach to choosing among investments in public projects, cost-benefit analysis is only a generation old. It derives from the marriage of theoretical advances in the new welfare economics and the previously undernourished public expenditures branch of public finance (1). In reviews of the cost-benefit literature, few references are encountered that antedate 1958 (2-4). Most of the theoretical as well

This article is a revision of a paper presented at the Conference on Technology and Health Care Systems in the 1980s in San Francisco in January 1972, sponsored by the Health Care Systems Study Section, National Center for Health Services Research and Development.

as empirical research has been carried out in connection with water resources projects (5-10).

Aims and Criteria of Choice

Cost-benefit analysis aims to do in the public sector what the better known supply-demand analysis does in the competitive, private sector of the economy. When market failure occurs—whether through the absence of a market or through the existing market's behaving in undesirable ways—public intervention comes under consideration (11). Cost-benefit analysis is helpful in determining the nature and scope of such intervention.

The most egregious example of lack of a market is given by the case of the pure public good. Such a good is collective, usually entails governmental action, and is characterized by a particular feature: when more of it is consumed by A, B need not consume less (12). National defense is one example frequently encountered in the literature, and the lighthouse on the shore is another. Certain aspects of basic research and the dissemination of research findings share this feature, since the acquisition of new knowledge by D does not diminish its value for C, the original investigator who developed it.

In the context of cost-benefit analysis the most important cause of market failure is the presence of substantial external effects. Such effects are called economies if positive, and diseconomies if negative. Vaccination against a communicable disease is perhaps the most commonly cited example of benefits accruing to a third party or to the community, in addition to the benefits received by the patient and health worker, who are 'directly involved in the transaction' (13, p. 18). Still another example from the health field is the protection accorded to the community by hospitalizing persons with severe mental illness (14, p. 12).

The goal of public policy is to adopt those projects or programs of service that yield the greatest surplus of benefits over costs. Evaluation of projects is prospective, oriented toward the future. The criterion of choice, analogous to that of maximizing profits in the market economy, is to maximize present value. Stated differently, but meaning the same, the criterion is to equalize marginal benefit and marginal cost. Strictly speaking, as Stigler (15) notes, maximizing present value is also the criterion for optimum behavior in the private sector. As Fuchs (16) pointed out, this criterion is quite different from that of attaining the maximum amount of a particular indicator of benefit.

Of course, the notion of balancing benefits and costs is by no means alien to medicine. Lasagna (17) states, "Since no drug—free of toxicity—has ever been introduced that is effective for anything, those of us who are pharmacologists have learned to live reasonably comfortably with the notion of paying some sort of toxicological price for welfare."

A possible source of misunderstanding about cost-benefit analysis is that benefits are usually costs presently borne that would be averted if the program in question proved to be effective. It is essential to distinguish these present and potentially avertible costs from the resource costs required to conduct that program. Since the two types of cost are not always juxtaposed, this distinction is not obvious, and the failure to draw it is not evident.

There are two essential characteristics of cost-benefit analysis: breadth of scope, and length of time horizon. The objective is to include all costs and all benefits of a program, no matter "to whomsoever they accrue," over as long a period as is pertinent and practicable (3).

Counting, Measuring, and Valuing Benefits and Costs

When an agency wishes to undertake a project or program, it may be tempted to go far afield in counting benefits and to neglect some costs. For example, in water resources projects, certain secondary benefits may be included improperly. In vocational rehabilitation programs both savings in public assistance grants and income taxes on subsequent earnings are sometimes counted as benefits, even though neither item entails a saving in the use of resources. Both grants and taxes are transfer payments, that is, they represent a transfer in command over resources (18, p. 47).

The distinction between costs and transfers is not meant to suggest that the sole justification of public projects or programs is an increase in output or gross national product (GNP). On the contrary, there is increasing recognition that public projects or programs may carry multiple objectives, including income distribution, more jobs, and regional growth (19-22). There is no reason why, for purposes of cost-benefit analysis, earnings on a job cannot be assigned greater weight than the same amount of money received in public assistance grants (18, p. 167). What must be recognized is that such weights are judgmental, are likely to be arbitrary (at least initially), should be derived in the public arena, and, above all, must be clearly stated.

Similarly, as shown by the progressive individual income tax, we seem to act on the belief in the United States that an extra dollar accruing to a low income person is worth more than an extra dollar accruing to a high income person (19). Again, assigning relative weights may help to improve analysis for public policy. There is no reason to believe, however, and no intention to claim, that agreement on such weights is imminent.

On the cost side a good example of the tendency toward understatement is the neglect of compliance costs imposed on individuals and firms in calculating such costs as those of administering the individual income tax, and of Medicare for the aged. Once the installation of seat belts in automobiles is made mandatory, the temptation arises to disregard the cost of seat belts to car owners (23, 24).

Counting benefits and costs involves deciding what to include and what to exclude. When an item that may be properly included can be measured, the next problem is that of valuation. The ease of valuation, indeed its possibility, depends largely on whether the item in question is traded in the market and bears a price. In that case there are many good reasons for simply adopting that price (25). When market price is deemed to be a defective measure of value, however, an attempt is made to estimate an imputed or shadow price (26). One modification of market price that is widely accepted is to set a lower value on unemployed resources; the size of this adjustment may vary not only with the state of the economy at large, but also by geographic region and by occupation (20).

When an item, even an important one, lacks a market price, the tendency is to omit it from calculations. If total benefits are thereby understated, a program may be erroneously deleted. More important, perhaps, programs with a sizeable proportion of unvalued to total benefits stand to lose in competition for funds with programs that have few, if any, unvalued benefits (27). Among the items most likely to be omitted are so-called intangible benefits; such benefits are especially prominent in the health field. It is not that they can never be valued (28). Rather, one may distinguish between intangible benefits that for the time being remain difficult to value and pure public goods, which are not traded on the market and therefore cannot be valued.

The dilemmas of valuation can be escaped by retreating from cost-benefit analysis to cost-effectiveness analysis. The latter is the less demanding approach, since it does not require the valuation of all benefits in terms of a common numéraire. Cost-effectiveness analysis requires only that benefits be measured in physical terms. Once an objective or output is specified, the aim is to minimize the cost of attaining it. The cost data required for cost-effectiveness analysis are, however, the same as for cost-benefit analysis (29).

Retreating from the valuation of benefits to measurement alone entails a substantial loss: analysis can no longer assist in setting priorities among several fields of public activity. The reason is simple. While cost-benefit analysis cuts across diverse objects of public expenditure, cost-effectiveness analysis can only help in choosing among alternate means of achieving a given, presumably desired, outcome (30). It is cost-effectiveness analysis that was incorporated as a major element in the planning, program, and budgeting (PPB) systems of the federal government. After its initial development by the Rand Corporation, PPB was introduced by the Department of Defense in 1961 and extended to other departments and agencies by Executive Order in 1965 (31, 32).

Both cost-benefit analysis and cost-effectiveness analysis imply the measurement of outcomes that are associated with particular projects or programs of service. Presumably there is a link between inputs and outputs that is measurable and known. It does not matter whether behavior follows a deterministic or probabilistic pattern. In the development of water resources the design of a particular project almost guarantees the emergence of certain physical outcomes—so much land will be lost to flooding, so much more land will be protected from flooding, so much land will be irrigated, etc. In national defense the outcome of a proposed course of action is much more uncertain, since other countries can take evasive and retaliatory action. In the health field, as will be shown, the presumed link between inputs and outputs is sometimes tenuous (33). Too often, the task of measurement, which necessarily precedes valuation, has been neglected.

The Rate of Discount

There is wide consensus in economics that a dollar is worth more today than it will be worth a year or two from now; this holds true when overall price level remains constant. As long as assets are safe, consumers are believed to have a positive time preference, that is, prefer to consume now rather than later. For producers investment may be productive through either the lapse of time, as in wine making, or the adoption of more round-about methods of production. Borrowers are therefore willing to pay interest for the use of capital, and lenders, in a capitalist economy, expect to receive interest. In a socialist economy an accounting or imputed rate of interest is employed to help allocate resources over time. The interest rate that calculates the present values of future streams of benefits and future streams of costs for public projects or programs is the well-known discount rate of cost-benefit analysis (18, p. 165).

Although economists agree that a discount rate is necessary for rendering commensurate benefits accruing and costs incurred at different times, they do *not* agree on the size of the discount rate. Marked differences of opinion prevail for a number of reasons. One is that a diversity of interest rate structures exists in the real world, owing to capital market imperfections, differences in risk, and governmental monetary policies (2, 3, 34). There is controversy regarding which imperfections to allow for and how to allow for them. Another reason for differences of opinion, as Musgrave (35) makes clear, is the source of financing–private consumption or investment. Still another reason for differences of opinion is a value judgment: whether the proper measure of the discount rate for public projects is the opportunity cost of capital in the private sector, or, to the contrary, whether it is the social rate of time preference. The private rate may be high, well above 10 per cent, particularly when allowance is made for the corporation income tax of 50 per cent (36). The social rate of time preference is usually much lower, based on a longer time horizon or greater readiness by the community acting together than by individuals to postpone gratification in favor of future generations. The social rate, which has been justified in terms of future risk and uncertainty, probability of personal survival, and the diminishing marginal utility of additional income or consumption as per capita income grows over time, is not a number that we know how to ascertain empirically (34). Accordingly, still another procedure, which combines private opportunity cost and social time preference, is also not measurable.

In practice, the agencies of the federal government have employed a wide range of discount rates, usually without giving a reason (37). Nevertheless, the consequences of choosing a high or a low discount rate are clear. A low discount rate favors projects or programs with benefits accruing in the distant future. In effect, as Boulding (38) has suggested, a high interest rate favors the aged and a low one favors the middle aged. When a project or program is short-lived, with both benefits and costs concentrated in the near future, the choice of discount rate is of minor or no consequence; indeed, for a short-lived program discounting may be dispensed with. Some economists are averse to selecting a particular discount rate, on the ground that they are in no position to choose between generations (13, p. 57). The tendency is to display calculations of the present values of benefits and costs under two or more discount rates. It seems to me that such alternative calculations do not afford helpful guidance to the policy maker, unless he is advised when to employ one or the other.

Even in the present state of the controversy there may be some merit to employing a single number for all public projects or for all public human investment projects. The combined method, recommended by a panel of consultants to the Bureau of the Budget in 1961 (39), and subsequently developed by Feldstein (40), can furnish an adequate rationale even though it does not yet yield a specific number. Such a number admittedly would be arbitrary, a reflection of value judgment (2). Henderson (34) reports that the French have adopted a centrally determined rate of discount of 7 per cent, to be applied to all public enterprises. This rate is higher than that encountered in many American cost-benefit studies.

APPLICATIONS OF COST-BENEFIT ANALYSIS TO THE HEALTH FIELD

The health services literature contains many affirmations of the importance of cost-benefit analysis for improving the allocation of resources to and within the health field. It may prove to be a source of astonishment that relatively few complete cost-benefit studies of health programs have been carried out. Perhaps it is appropriate that fewer cost-benefit studies have been performed than advocated. Where the aim is to minimize the cost of producing a given good or service, or even of constructing a hospital of specified size and with suitable appurtenances, the apparatus of cost-benefit analysis is superfluous (33). It then suffices to compare unit costs.

Criteria for Inclusion

The major reason for the shortness of the list of complete cost-benefit studies is that most studies conducted to date are limited in one or more respects. First, in the 1970s, there seems to be little point to considering nonempirical analyses. Thus, today Mushkin's seminal work (41) in conceptualizing the application of cost-benefit analysis to the health field must be excluded from consideration.

A second, perhaps more critical, requirement for including a study is that both the benefits and costs of specified programs be measured and valued simultaneously, with their respective present values juxtaposed and compared. By this criterion, the majority of empirical studies so far performed in the health field are excluded, including that by Fein (42) on mental illness, by Rice (43, 44) on a number of diagnostic categories, and my own on syphilis (45) and on heart disease (46). While all of these studies attempt to measure and value the cost of a disease, thereby, in effect, measuring and valuing the total benefits of eradicating that disease, none attempts to estimate the costs of conducting programs with specified contents and aims. Although each study has made a contribution to the counting, measurement, and valuation of direct and indirect tangible benefits, and two have explored the valuation of intangible benefits, none has presented a comparison of costs and benefits under a specified set of conditions.

The above two requirements–quantification and juxtaposition of costs and benefits–impress me as being incontestable. A third requirement can be defended as equally necessary: that benefits and costs reflect a known link, alluded to above, between program and outcome, i.e. between inputs and outputs. Such a link should be empirically based. Today, speculative or hypothetical relationships do not suffice (47). To apply economic valuation to hypothetical relationships between programs and outcomes is to indulge in an academic exercise, since the results of such valuation cannot transcend the quality of the underlying measurements. Such an exercise is not only idle, in that it can make no contribution to policy formulation, but it may be counterproductive if it obscures the fact that the relationships between inputs and outputs are not yet known and remain to be ascertained (14, p. 29).

In an article discussing the contribution of health services to the U.S. economy, Fuchs (48) has demonstrated the importance of information concerning the efficacy of health services. The economist can indicate the types of data he requires, but he is seldom in a position to procure them by himself; he must rely on other investigators in health services research to help him to obtain them.

This third requirement implies an important corollary. The size of a problem, as measured by the total costs of a disease, is not a reliable guide for policy (49, 50). Even in communicable diseases, less than eradication may be an acceptable goal. For most diagnostic conditions it is essential to know the extent to which a given program is likely to reduce the size of the problem. This point is often overlooked. It lends itself to oversight particularly when benefits and costs are not juxtaposed. In the early cost-benefit studies in the health field there may have been a further tendency for economists to attribute greater efficacy to medical care than was perhaps warranted (51).

Weisbrod (13) performed the earliest of a small number of such studies and his remains one of the most systematic. He compared the benefits and costs of intervening in three diseases–cancer, polio, and tuberculosis. Drawing in a creative way on Bowen's work in deriving the demand curve for a public good (52), Weisbrod was frequently reduced to obtaining cost data and some notion of the link between inputs and outcomes from personal communications with clinicians and administrators. His threefold classification of benefits—direct, indirect, and intangible—followed Mushkin and has become the convention.

How such benefits are measured and valued, as well as an assessment of the current state of the arts, will be given below. Both accomplishments to date, and possible shortcomings in the accepted procedures will be presented.

Direct Benefits

Direct benefits are that portion of averted costs currently borne which are associated with spending for health services. They represent potential tangible savings in the use of health resources. Certainly in the long run manpower not required to diagnose and treat disease and injury does become available for other uses. It is reasonable to suppose that our economy, like others, has a vast variety of wants in the face of a totality of relatively scarce resources, so that freeing resources for other, desired, objectives represents a contribution to economic welfare.

In the absence of a specific program of services to be evaluated, the measure of direct benefits is usually taken to be total resource costs currently incurred. The appropriateness of this measure as a basis for policy is questionable, as indicated above. Nor is it helpful to take some fraction of the total. In terms of resource use, diminishing marginal productivity is likely to set in as a program expands beyond a certain point. In terms of valuation of benefits, diminishing marginal utility is often a plausible assumption.

While it is usually taken for granted that direct benefits, or the current costs of care that will be averted, can be measured with precision, this is true only when a firm produces a single good or service, such as maternity care in a special hospital. In most instances several goods or services are produced jointly. Under conditions of joint production it is possible to calculate the extra or marginal cost for each product, but not its average unit cost (10, pp. 44-45). When average unit cost figures are presented, they reflect an allocation of overhead and joint costs; and such allocation is necessarily an arbitrary accounting procedure, even where it is systematic and replicable. An alternative procedure, which is no less arbitrary, is to assign to a diagnostic category its proportion of total costs, with the proportion taken from the percentage distribution of patients or services. In the absence of facilities that produce only a single product, it might be helpful to analyze cost data for facilities with varying diagnostic compositions of patient load. However, other factors are also at play, and there is no logical solution to the problem of determining average cost under conditions of joint production of multiple outputs (18, p. 166).

Another complication, which affects the calculation of direct benefits and also of indirect benefits, is the simultaneous presence of two or more diseases in a patient. The presence of disease B when intervention is attempted in disease A serves to raise or lower the costs of intervention and therefore the corresponding benefits (45). The reason that indirect benefits, which represent gains in future earnings, are also affected is that the presence of diseases A and B in a patient may reduce the probability of successful outcome from the treatment of either. The effect is to overstate the benefits expected from reducing the incidence of one or the other disease (51). The magnitude of this effect is not known.

The prevailing tendency is to take direct benefits from a single-year estimate of costs (44). Since survivors will also experience morbidity in the future, some medical care costs are being neglected. Initially this procedure may have been associated with an emphasis on single-year estimates, to the exclusion of present value estimates (50). Once the necessity of present value estimates is recognized, other explanations must be sought for this shortcut. A possible explanation is that survivors will experience only average morbidity in the future; when extra morbidity is absent, there is perhaps no need to deal with morbidity. A more plausible explanation lies in the lack of longitudinal data on the morbidity experience of defined population cohorts.

The fact is that a single-year estimate reflects the prevalence of a disease, not its incidence. It may be that the prevalence figure is sufficiently greater than the incidence figure for chronic conditions, so that it makes ample allowance for future events. Indeed, the prevalence figure in the base year is the same as the sum of the incidence figures for all survivors to this year, if certain factors remain constant, such as the size of population, death rates for the particular diagnostic group, and the incidence rate. When any of these factors follows a rising trend, however, the prevalence figure exceeds the cumulative sum of the past and present incidence figures and falls short of the sum of incidence figures expected in the future.

To the extent that unit costs or prices tend to increase faster in the health services sector than in the economy at large, the value of direct benefits will also increase. In my own work I have incorporated an adjustment for this factor into the discount rate, deriving thereby a *net* discount rate (45, 46). If economic growth were to slow down in this country, the lag in productivity gains of the health services sector behind the economy at large would be reduced, and so would the size of this adjustment.

Transportation expenses for medical care are a resource cost which is disregarded in cost-benefit analysis, although they are allowed as deductions under the individual income tax. When the physician made home calls, his travel expenses were automatically included in health service expenditures. The foremost reason for neglecting them today is, most likely, lack of reliable data. There may be the further, implicit assumption that patients' transportation costs are of a small order of magnitude.

Indirect Benefits

Earnings lost due to premature death or disability, which will be averted, are indirect benefits. Debility as an impairing factor in production has not attained the prominence in empirical studies that Mushkin (41, 51) attached to it from a conceptual standpoint.

Since the publication of Rice's studies (53) it is no longer necessary to estimate loss of earnings on the back of an envelope. Drawing fully on the data resources of the federal government and using unpublished tabulations almost as much as published ones, Rice (43-44) prepared her estimates in systematic fashion. She applied labor force participation rates, employment rates, and mean earnings, inclusive of fringe benefits, to the population cohort in question. For men and women separately, she derived estimates of the present values of lost earnings due to mortality under alternative discount rates and a one-year estimate of lost earnings due to disability or morbidity.

Several elements of the benefit calculation that were still at issue a decade or so ago appear to be more or less settled now, some perhaps prematurely. These can be summarized as follows:

1. Our ordinary concern is with loss in earnings, not income. The latter includes income from property.

2. Consumption by survivors is no longer subtracted from gross earnings in order to arrive at net earnings. Viewed prospectively, everybody is a member of society, including the patient (54).

3. The value of housewives' services is recognized, despite the fact that such services are not traded in the market and are omitted from the GNP. Weisbrod (13, pp. 114-119) developed and applied a complex method for measuring the cost of a substitute housekeeper, but subsequent writers have followed Kuznets (55) in employing a simpler approach, putting the value of the services of a housewife at the level of earnings of a full-time domestic servant. To employ a single number is the more practical procedure by far. The magnitude of that number is a separate question, however. It is increasingly evident that the value given by the earnings of a domestic servant is not adequate (56). Thus, the accepted value of the housewife's contribution would increase substantially if day care centers for working women were expanded at public cost.

An alternative approach is to value the housewife's contribution at the opportunity cost of her staying out of the labor force (45). Implementation of this approach is impeded by two considerations (57). First, the method is complicated, since values would vary with the individual housewife's educational attainments, type of occupation, amount of job experience, full- or part-time employment status, etc. Second, nonpecuniary factors, which certainly influence the labor force participation rates of women, are difficult to measure and may behave erratically. When total family income permits, the pecuniary opportunity cost of the wife's staying home has been known to be as low as zero or even negative.

4. The employment rate has been typically taken at 96 per cent, or an overall level of 4 per cent unemployment at the level of "full" employment (44). In the 1970s the magnitude of this rate is at issue. Whatever the magnitude, Mushkin's argument is accepted that the health services system should not be charged with failures by the economy to provide jobs to all who seek them (41, 58).

What is often not taken into account is the tendency for persons rehabilitated after serious illness or injury to find fewer job opportunities than persons who have remained healthy and on the job. In my study of syphilis (45), I recognized the loss of earnings due to the "stigma" attached to this and similar diseases. When prevention is feasible, it seems appropriate to assign to it an extra weight or bonus for this reason.

5. Calculations of indirect benefits rest on the implicit assumption that the life expectancies of cohorts of potential survivors are known. Usually standard life tables are employed, separately for men and women. For diseases of low frequency it seems reasonable to disregard any effect on the total death rate occasioned by the deletion of a particular cause of death. For major diseases the problem is important, although simple deletion may be incorrect. As Weisbrod (13, pp. 34-35) recognized more than a decade ago, survivors who have avoided a particular cause of death may have a higher or lower susceptibility to other, competing causes of death. I compared the effects of simply deleting heart disease as a cause of death on life expectancy and on work-life expectancy. The former was large—11 to 12 years—and the latter was small—less than a year (46). For a disease with heavier impact at the younger ages, the effect on work-life expectancy would be relatively larger, and correspondingly greater attention would have to be paid to the effect of competing causes of death.

Intangible Benefits

Pain, discomfort, and grief are among the costs of illness currently borne, which constitute the intangible benefits of a program of health services that averts them. The benefits accrue partly to the patients and partly to their friends, relatives, and society at large, to the extent that we take pleasure in the happiness of others. That positive external effects in consumption exist is indicated by personal and philanthropic gifts, to the extent that they are not subsidized by the deductibility provisions of the income tax (59). Looming even larger is the averted premature loss of human life. Since none of these effects is traded on the market, none carries a price tag. In attempting to put a value on averting them the question arises: what would one be willing to pay to avoid them?

In my paper on syphilis (45), I estimated willingness to pay for escaping the early and late manifestations of the disease by examining expenditures incurred in connection with other diseases that met certain conditions. After consultation with clinicians I adopted psoriasis as the analogue for early syphilis, and terminal cancer as the analogue for its late stage. The conditions specified were that the expenditures for medical care represented principally a willingness to pay for freedom from the particular disease, since in neither case could direct or indirect tangible benefits, as defined above, be realized. To the extent that payments were made only by the patient (directly or through health insurance), willingness to pay by others was neglected and total willingness to pay was understated.

Neenan (60) has estimated the consumer benefit of a community chest x-ray program for tuberculosis. With the help of some fee data indicating willingness to pay, he obtained very high estimates of value.

Several years have elapsed since intangible benefits were valued. The analogous diseases approach has not been repeated; this suggests that neither the estimates themselves nor the procedures for obtaining them have been found useful. One reason is obvious: the approach is specific, calling for the development of estimates, disease by disease.

A larger body of literature is devoted to the value of human life than to the other types of intangible health benefit. Life insurance holdings are clearly not applicable to bachelors and jury verdicts are inconsistent (13, p. 37). The implications of public policy decisions or governmental spending are difficult to elicit in the absence of information on the alternatives that faced the decision makers (19). Moreover, such valuation may lack stability and consistency (24, pp. 133-134).

Schelling (61) has proposed a different approach. He would measure the value of human life, as distinguished from livelihood, by the amount people are willing to spend to buy a specified reduction in the statistical probability of death. Acton (24, p. 258) applied this approach, and derived an estimate of the value of human life at $28,000. This amount serves as a substitute for the net value of lost earnings and is not an additional sum.

I am not sanguine about the applicability of Acton's numerical estimate to the evaluation of program alternatives. Acton was the first to criticize the defects in his estimate, including the small size of his sample, and its apparent biases. While these defects can be remedied in the future, what troubles me is the likelihood that respondents to this type of question may not grasp its meaning. Do respondents know the actual probabilities of their dying in the coming year? How is a small—e.g. 1 per cent—reduction in statistical probability perceived? How much more is a 10 per cent reduction worth than a 1 per cent reduction? Is it plausible to postulate a strictly linear relationship between increase in risk and willingness to pay to cover it? (54). Moreover, does not the value of a gain depend somewhat on the starting point? (62). If all payments come from the consumer, the distribution of income must exert a sizeable influence; by how much would willingness to pay change if the task of reducing the death rate were viewed as a collective responsibility that is fully financed from public funds?

Titmuss (63) regards the value of human life as priceless and beyond valuation. Yet implicit values are being placed on human life whenever public policy decisions are made on highway design, auto safety, airport landing devices and traffic control measures, mining hazards, factory safeguards, etc. In emphasizing voluntary giving, the sense of community that the gift relationship in blood both reflects and promotes, Titmuss seems to be pointing to a large external benefits component that is neglected when life-time earnings are taken as the proxy for the value of human life. Although the concern for the altruistic motive is salubrious and appropriate, the conclusion does not follow that human life is priceless.

As Mishan (54) observes, a rough measure of a precise concept is superior to a precise measure of an erroneous concept. It is agreed that the notion of the value of human life, apart from livelihood, is sound. A numerical estimate of this value would be useful in comparing the worthwhileness of alternative programs. Comparisons of programs would gain in relevance and aptness if all benefits were counted, including the saving of human life or improvements in life expectancy. This potential gain is much more likely to be realized if all benefits are entered into the model, rather than if some appear only in footnotes.

I am unable to say at this time how such a number or set of numbers for the several age groups can best be derived. Certainly Schelling's questionnaire method (61) can be improved. Perhaps the implications of past or existing public policies will yield a narrower range than one expects. It is conceivable that a committee can do a better job in the realm of values than in the realm of fact. In any event, the value of human life is probably higher for identified and known individuals than for members of statistical populations. If so, incurring extraordinarily large expenditures in behalf of the former is far from conclusive evidence of irrational behavior.

Weisbrod (13, p. 96) avoided dealing with the problem of valuing intangible benefits by assuming proportionality to tangible benefits. This is an unsatisfactory solution, given the differential impacts of various diseases on life expectancy, disability, and morbidity. However, a solution to this problem was not needed when the emphasis of public expenditures analysis shifted from cost-benefit to cost-effectiveness. To repeat, in cost-effectiveness analysis outcome is expressed in physical terms, e.g. life years gained, and the task of analysis is to discover the program that will yield the desired outcome at the lowest unit cost. In the health services it goes without saying that desired outcome incorporates a constant level of quality of care, or at least an acceptable level.

Cost of Program

The estimate of the cost of a proposed program, with which benefits are compared, poses no special difficulties. A budget is prepared in terms of the market prices of inputs, which may be adjusted by shadow prices when warranted.

If programs vary in size, it is appropriate to examine the possibility that economies of scale exist (14, pp. 82-83). However, since health services are rendered in the local area, the prospects of realizing such economies are much more limited than in the manufacture of goods. Moreover, when the size of a program increases, factor costs may rise. Finally, as the scope of a program approaches the size of the total population at risk, the extra cost of additional units of output increases when increasingly resistant groups are encountered. Conversely, it has been suggested that in the early phases of a program unit cost is likely to be higher than later on, since administrators learn by doing (14, p. 24).

Cost-Benefit Versus Cost-Effectiveness Analysis

Although it is not so difficult to estimate the costs of programs, it is quite difficult to formulate the contents and expected outcomes of programs. In my judgment this has been the chief obstacle to the useful application of cost-benefit or cost-effectiveness analysis in the health field.

Elsewhere I have listed the data required by the economist for valuing outcomes (46). A clear statement of each type of outcome is necessary. Certain events, such as death, disability, extra unemployment, and the use of health services must be entered on a calendar, beginning with the base year, and assigned a duration. The data should extend for a period as close to a person's lifetime as possible, with particular attention to the possible recurrence of illness and its exacerbation.

This list of data requirements implies a degree of knowledge about the effects of health services on the health of a population that is often lacking. The obstacles to the attainment of such knowledge are many. Medicine is not an exact science, and physicians may disagree among themselves and the same physician may disagree with his own past findings. Field studies are complicated by what Morris (64) calls the iceberg phenomenon: members of the designated control group, who are presumably normal, may in fact have the disease under investigation in asymptomatic form. The possibility of inducing iatrogenic disease means that only studies performed on normal populations in the community, which are far more costly than studies of captive clinical populations, can yield valid results (65).

A serious gap in existing data arises from the lack of longitudinal studies of populations. Few investigators possess the requisite patience and dedication, or experience the necessary career stability. The funding agencies, under conditions of budgetary stringency, have even shorter time horizons. Although statistical manipulation of existing cross-section and time-series data is a much cheaper and almost always available approach, it may not afford an adequate substitute in many instances, especially when a high degree of correlation exists among the independent variables under scrutiny. In 1965 I reported that only one study met the longitudinal data requirements listed above—Saslaw's study (66) on rheumatic fever. Unfortunately, the report on this study was truncated in publication. Neenan's study of chest x-rays for tuberculosis (60), conducted in 1964, concentrated on the short term, on the ground that a recovered patient suffers no impairment of earnings while early detection alone does not alter the long-term outlook. No evidence is adduced for these assumptions.

Acton (24) has recently completed a cost-benefit analysis of alternative programs for reducing deaths from heart attacks. He considered five programs: an ambulance with specially trained nonphysician personnel; a mobile coronary care unit with a physician; a community triage center; a triage center combined with the ambulance; and a program to screen, monitor, and pretreat the population. The largest net benefit, whether measured by the number of lives saved, or valued by the criterion of earnings or that of willingness to pay, is given by the screening, monitoring, and pretreatment program. However, the value of personal time lost in screening is neglected, and the screening program seems to display great variability in outcome. Acton's cost-benefit analysis follows his cost-effectiveness analysis, in which lives saved are not assigned a value. The screening, monitoring, and pretreatment program yields the largest number of lives saved, but the average cost per life saved is second from the highest and the marginal cost of saving two additional lives is $24,000 each, compared with the estimated average and marginal cost of $3,200 for saving the first 11 lives under the ambulance program (24, p. 117). Acton's work is noteworthy for the wealth of detail on the epidemiology of heart attacks, physiology, treatment, and delivery systems for treatment or prevention. This economist drew extensively on the expertise of health services specialists and investigators.

The report of the Gottschalk Committee to the Bureau of the Budget (BOB) (67) contains a cost-effectiveness analysis of alternative modalities for treating chronic end-stage kidney disease. The problem facing the BOB, and posed to the Committee, was to define the appropriate role of the federal government in this field. The conclusion that a substantially expanded federal role was warranted was reached on other grounds, which did not entail an economic analysis. These included: some veterans were already receiving free care in veterans administration hospitals; several foreign countries, each poorer than the United States, were committed to delivering this service to all patients; voluntary insurance plans in the U.S. were not paying for the cost of prolonged hemodialysis and their leaders saw no prospect of doing so; the patients requiring treatment largely comprised middle-aged adults; and what was still a unique life-saving measure was available for application to known individuals, persons who would otherwise die in short order. Once the recommendation was made in favor of an expanded role for the federal government and a feasible mechanism was designed to finance the care rendered to individuals, the problem that remained for economic analysis was how best to discharge this responsibility–through hemodialysis in an institution, hemodialysis in the patient's home, kidney transplantation, or some mixture of these modalities.

The cost-effectiveness analysis clearly pointed to the superiority of the transplantation route, which incorporates hemodialysis both for initial and back-up support. When hemodialysis is necessary, doing it at home is much cheaper (29). These findings influenced the Gottschalk's recommendations to the BOB that kidney transplantation be expanded as far as possible. (The best mix of modalities was not solved for because the implied assumptions of constant unit cost and constant utility of a life-year made it unnecessary.) Again, as noted with respect to Acton's study, the economic analysis drew heavily on the underlying epidemiologic, physiologic, and clinical data developed by or in behalf of the Committee.

Had the Gottschalk Committee performed a cost-benefit analysis, it seems plausible to postulate that a shortage of kidneys for transplantation and the relatively greater ease with which hemodialysis facilities can expand might have yielded a higher net benefit value for dialysis, at least in the near future. However, allowance for the superior quality of life under transplantation would constitute a partial offset (29).

Contrary to some impressions (68), the Gottschalk Committee did inquire into available prevention programs and determined that, for the foreseeable future, the number of eligible patients with end-stage kidney disease would not change. The Committee did not inquire into the dispersion of the distribution of life years gained.

Thus, it did not consider whether an average gain of 10 years is worth the same when it is the product of 10 years each gained by 100 per cent of the population at risk, or of 20 years each gained by 50 per cent of the population, or of 40 years each gained by 25 per cent of the population. Can it be said that the marginal utility of an additional year is constant or does the principle of diminishing marginal utility govern?

The Committee did not have to deal with two problems that might arise under different circumstances. One is that even cost-effectiveness analysis is not so simple as it appears to be when two or more types of outcome are sought as goals. If only one outcome, such as life years gained, is preeminent, other outcomes may be neglected. Where all outcomes are important—reduced mortality, lower morbidity, and less disability—it becomes necessary once again, as under cost-benefit analysis, to arrive at common or weighted measures of outcome for alternative programs (69). Only the problem of valuing intangible benefits is escaped. However, in cost-effectiveness analysis the focus is confined to outcomes common to health services programs, and the weighting problem is serious only when the several types of outcome do not occur in the same proportions for every program. The second problem not faced by the Gottschalk Committee is the appropriate role for government to assume if expensive life-saving measures became practicable for other organs of the body. Nor did the Gottschalk Committee attempt to deal with the question of increases in patient load if the very success of the program it sponsored led to the relaxation of criteria governing patient eligibility for treatment. The quantitative effect of such relaxation may be appreciable.

In the years 1966-1967, during the early spread of PPB in the federal government, a number of cost-effectiveness studies were carried out in the Department of Health, Education, and Welfare (23, 49, 68). Although costs and benefits were calculated simultaneously, the link between the inputs and outputs of programs was measured too often by means of hypothetical numbers. Once the relationships were postulated, no effort was made to pursue the measurement problem through empirical inquiry in subsequent budgetary periods. In certain instances only expenditures chargeable to the federal budget were counted as costs, neglecting expenditures incurred by individuals and by other levels of government (23, 24).

PROBLEMS IN ASSESSING HEALTH SYSTEMS TECHNOLOGY

This discussion of the potentialities and hitherto modest achievements of cost-benefit analysis and cost-effectiveness analysis in the health field bears directly on the analysis of the development and spread of health systems technology. However, changes in technology bring to the fore an additional factor: a heightened degree of uncertainty concerning future benefits and costs. According to systems analysis, one appropriate response to the prospect of uncertainty is to perform a sensitivity analysis, concentrating on a few key factors or assumptions to which the measure of costs or benefits appears to be especially sensitive (34, 70). This proposition strikes me to be a formal one, awaiting empirical content.

Nevertheless, within the time frame of a decade, any allowance for uncertainty due to developments in technology may be excessive. It has been suggested that the technology that will be applied in the next ten years is already known, and that the pattern of technologic diffusion is discernible (71). This view may be too sanguine, but it is not contradicted by the record of the Gottschalk Committee. By wisdom or good luck, the

Committee's projections of survivorship of patients with transplanted kidneys and the cost of hemodialysis at home, both of which were originally supported by scanty data, have been borne out (72).

If technologic developments over the next decade are, in effect, already known to those gifted with early recognition, what can be said about prospective benefits and costs? In a plea at a health services research seminar in New York City for more research and development funds, Bennett (73) argued that the half-finished invention is the most costly product, so that technologic progress is bound to bring a lower unit cost of service, as well as improved performance.

In those cases where straightforward development takes place and serious adverse side-effects are not encountered, Bennett's view of the cost-reducing and benefit-enhancing effects of technologic progress is undoubtedly correct. However, in many respects the future is shrouded in uncertainties. Such factors as the size and geographic distribution of population, value structures, and political decisions are uncertain for the future, even if technologic developments are not. Public policies are also known to create unintended and unanticipated consequences. An accepted way to deal with uncertainty is to provide for flexible operation, that is, to avoid a finely tuned operation which yields a minimum cost only for a particular scale of output. Similarly, if manpower is to be used flexibly in the future, it must be endowed with a more general education than otherwise. Thus, flexibility, whatever its cause or source, imposes a modest extra cost over a moderate range of outputs (18, pp. 105, 123-124).

The Historical Record

Rather than pursue this argument of pros and cons, I propose to examine the historical record. What have been the effects of past changes in health systems technology on costs and on benefits? A review of the literature on this subject reveals sharp differences of opinion.

In a monograph on hospital expenditures sponsored by the National Center for Health Services Research and Development, Feldstein (74) attributes most of the postwar increase in hospital cost to an increase in demand, or, more precisely, to an upward shift in the demand curve. To paraphrase his argument, technical change in the absence of scientific progress may occur for two different reasons. Economic analysis has emphasized technical change in response to a shift in the relative prices of inputs (75). If wages rise faster than the prices of other inputs, for example, hospitals will economize on labor by using more disposable items, by automating laboratory procedures, etc. The effect of such substitution is to prevent costs from rising as fast as they otherwise would have.

The second reason for technical change without scientific progress, which Feldstein emphasizes, is a shift in the demand for hospital care. This type of change generally yields a new product. The spreading of high-cost techniques is primarily due to rising income and increased health insurance coverage. As income increases, patients tend to raise the valuation of more costly care by relatively more than the valuation of less costly care. An increase in the proportion of the hospital bill paid by insurance will shift hospitals to more expensive technology, as the out-of-pocket price per unit of benefit is lowered.

Gains in scientific knowledge, including managerial innovations, that have the potential of lowering the cost of care may actually have the opposite effect. This happens again if the new scientific knowledge raises the benefits of expensive care by relatively more than the benefits of inexpensive care. In addition, if patients' real preferences do not prevail but hospitals persist in producing services with the most expensive techniques for which benefits are not less than cost, scientific progress cannot lower cost per patient day.

In a monograph on physician expenditures, Fuchs and Kramer (76) draw a sharp distinction between the effects of demand factors and those of technology. Their arguments concerning technology reflect an historical perspective, and may be paraphrased as follows. The late 1940s and early 1950s were marked by the introduction and widespread diffusion of many new drugs, particularly the antibiotics, which had a pronounced effect on the length and severity of infectious diseases. Since the mid-1950s, advances in medical technology have not brought about a similar improvement in the ability of physicians to improve health. Renal dialysis, cancer chemotherapy, and open heart surgery may achieve dramatic effects in particular cases, but bring about only marginal improvement in general indexes of health. Moreover, the early advances tended to be physician-saving, while the later ones were characteristically physician-using. The improvement in health resulting from the early advances was so great, that it turned the anticipated slight rise in demand for physician services into a slight decline. The reason is, according to Grossman (77), that healthier people have less objective need for physicians' services. By contrast, Fuchs and Kramer conclude that changes in demand factors had little effect on expenditures for physician services before the advent of Medicare and Medicaid in the mid-1960s.

In effect, whereas Fuchs and Kramer view technology and the conventional demand forces as being independent of one another, Feldstein holds that the effects of technology may also be exerted through a shift in demand. Both positions are stated ably and forcefully. As often happens, each raises more questions than it can answer. It would be premature, therefore, to attempt to pass judgment on the validity of the respective findings concerning the effects of technology in the postwar era.

In a study focusing on the marked acceleration in the upward trends of costs and expenditures for hospital and physician services in 1966 (78), I have argued, though by no means conclusively, that the large expansion in cost reimbursement to hospitals and the adoption of a new, previously untried method of paying physicians at reasonable and customary fees, subject to the prevailing distribution of fees in a local area, must have exerted strong effects of their own. In the case of hospitals, cost reimbursement for most patients leads to an impairment of financial self-discipline, since a dollar need only be spent in order to be gotten back. In my judgment, this proposition holds true for any institution, whether it be under voluntary nonprofit, governmental, or proprietary auspices. So far I am not persuaded by the empirical studies that have reached conclusions to the contrary (79, 80).

A number of works have appeared that attempt to explain the behavior of the nonprofit hospital (81-85). They are, for the most part, far-ranging and enlightening. One is also entertaining, positing a theory of conspicuous production, with the hospital's objective taken to be the closing of a status gap (85). None really attempts to deal with the sharp discontinuity in hospital cost and price behavior beginning in 1966.

A rise in personal income may lead to greater reliance on technology for still another reason. For example, many persons are unable to stop smoking. A higher income enables them to pay more for cigarettes with a filter and reduced tar and nicotine contents.

Similarly, a higher income permits people to spend more on automobiles with safety gadgets, reducing the need to exert influence on the behavior of drivers. It may be more effective to operate on impersonal environmental forces than to try to change the behavior of individuals (86).

At this time no general answer is discernible to the question of how changes in health systems technology affect costs and benefits. It happens only once in a generation, perhaps even less frequently, that an idea such as early ambulation after surgery is born of necessity in wartime, effects huge savings in the use of health resources, and also exerts a positive effect on health. In most cases, the effects of technology will be mixed. Often the product is new, in the sense that a treatment is created that was not available previously and therefore could not have been demanded. The decision of whether or not to adopt a piece of technology, and the extent of its spread once adopted, depend on a number of factors, including the values of consumers, the motivations of providers, the availability of funds, methods of provider remuneration, as well as the cost and efficacy of the service in question.

Such a general formulation of the problem of assessing health systems technology, as provided above, affords practically no guidance to decision making. Only the concrete circumstances surrounding a project or program can indicate the special problems of measurement and valuation and the unique opportunities for solving them, what is to be emphasized in the analysis, and what may be neglected with only a moderate degree of trepidation. Accordingly, I will examine two examples in detail: hospitals and automated multiphasic screening (58).

The Hospital

Economists have offered essentially three views concerning capital investment in the hospital. First, hospitals invest too little capital, hence their productivity gains lag behind those of the economy at large (87). Second, hospitals invest too much, because grants and bequests accrue to them at zero price (88). Third, there is no optimum amount of investment in hospital beds, since there is no standard of appropriate hospital use (89). Conceivably, each position may have some merit to the extent that it reflects the situation in different sectors of the hospital.

For simplicity I shall employ a threefold classification of hospital capital investment—patient beds, supporting housekeeping services, and ancillary medical services (82). The unique problems of measurement and valuation facing the application of cost-benefit or cost-effectiveness analysis will be explored for each sector.

Patient Beds. The heart of the exercise in evaluating a project to expand hospital bed capacity, in my judgment, lies in one's explanation of the phenomenon of hospital use. At one pole, if the primary determinants of use are biologic in nature, an increase in bed supply beyond a certain point must result in additional empty beds. If hospitals are paid at stated charges, empty beds inflict a heavy financial burden on each institution (79). The reason is that fixed costs constitute two-thirds to three-quarters of total operating costs (90). Each institution would therefore be subject to financial self-discipline in building beds, and there would be little occasion for outside intervention beyond the provision of information on the plans of other hospitals. The effect of introducing more technology might well be to increase the proportion of fixed costs to total operating costs, thereby reinforcing the efficacy of financial self-discipline.

At the other pole, if all beds built tend to be used under conditions of prepayment, as Roemer (91, 92) first suggested, there is no automatic criterion for an optimum bed supply. In the absence of evidence that low hospital use has an unfavorable effect on health status, the appropriate public policy is to clamp a tight lid on bed supply (79). The application of more or less technology in the hospital is beside the point, although it does seem preferable to operate any extra beds as cheaply as possible.

Patient census is a function of bed supply in the long run; combined with patient mix, it sets the requirement for nursing personnel, which may be viewed largely as a requirement for personal services, with little or no substitution of equipment permitted. However, substitution is possible among levels of nursing personnel. The extent of actual substitution of low-paid for high-paid staff is perhaps overstated by the failure of hospital budgets to incorporate expenditures for special duty nurses.

Housekeeping Services. I do not see any problems of sophisticated analysis in the area of supporting housekeeping services. Here the appropriate criterion for decision making is that of cost minimization. Bed sheets and towels are to be washed as cheaply as possible, for a given specification of whiteness. Patients' rooms and corridors are to be kept clean as cheaply as possible. Meals of a given quality–nutrition, calories, hot or cold–are to cost as little as possible.

Once it is recognized that certain products or services need not be produced by the hospital but can be purchased from the outside, the problem is that of developing valid comparisons of unit cost. In addition, some administrators may wish to allow for certain risk factors. In the absence of competition among suppliers, the sales price may be quoted artificially low at the outset, only to be raised later. Also, in the absence of competition, purchases from the outside may increase the risk of running out of inventory.

Apart from an allowance for lower risk associated with production within the hospital, estimates of internal cost of production should include only differential cost. No portion of overhead cost should be attributed, because this would continue in entirety after internal production ceased. Moreover, top management will perform the same role as coordinator, whether some goods and services are produced inside the hospital or acquired by purchase.

In fact, the rise in hospital wages and gains in productivity attainable in large-scale manufacturing have led hospitals to increase the purchase and use of disposable items and ready packaged supplies. As Flagle (93) reports, gains in productivity from investment in large-scale plant have been achieved outside the health care system, which shares in them through purchase.

If the objective of cost minimization is for a given level of cleanliness or nutrition, how this level is to be determined must be established. I doubt whether much would be accomplished by searching for effects on the health of patients. Rather, the criteria must be either patients' satisfaction or acceptability to management. Expressions of satisfaction are somewhat suspect, since patients are likely to be impressed by any display of interest in their opinions. A more practicable approach would be to compare alternative standards of service, none of them falling below adequacy, with the additional cost of attaining successively higher levels.

In some respects the computer partakes of a supporting housekeeping service and in other respects, when participating in diagnosis, it is akin to an ancillary medical service (94). The computer is a housekeeping service when it processes the payroll and issues bills to patients and insurance plans. As a substitute for older ways of bookkeeping and billing, the evaluation of computer performance is straightforward. Does it reduce costs? If so, by how much?

Medical Services. Even when the computer helps in diagnosis the test is still cost reduction, if an older way of performing the same task is being replaced. There may be a complication, however. The cost of operating the computer falls on the hospital, while savings in physician time accrue to the attending physician. The presence of distributional considerations suggests that the decision reached is not independent of who the decision maker is, or who exerts predominant influence on him.

Apart from the distributional considerations of who pays and who saves, evaluation of the worthwhileness of the computer in assisting in diagnosis is no different from the way another ancillary medical service, the laboratory, is evaluated. With respect to services that were rendered in the past, the test is simple. Does the new equipment save money or does it expand services for the same amount of money? In the laboratory additional and more costly equipment does replace technical personnel. A possible offset is the tendency to prescribe more services (95), although within the limits of existing capacity of equipment and staff the marginal cost of additional units of service is low. What is not known is how much good is accomplished, particularly in the absence of information on the timeliness of delivery of the reports on these services.

Flagle (93) has reported economies achieved in patient surveillance due to continuity of use of the monitoring system in infusing blood. This finding strikes me as analogous to the finding in his early work (96) that a single channel is more efficient than two channels when the demand for services varies stochastically.

The intensive care unit is a more complex operation to evaluate. To the extent that it substitutes equipment for nurses it should cost less. However, the unit is also intended to save lives. The yield in life-years gained is properly subject to more sophisticated analysis.

From this discussion it appears that cost-benefit or cost-effectiveness analysis is a plausible approach only if the service rendered is a new one or if the old product has changed appreciably, gaining new dimensions. When all benefits take the form of savings in health resources, that is, are direct and tangible benefits, the appropriate form of analysis is cost-benefit. When the preponderant benefits are intangible or life-saving, the dilemma is to choose between cost-benefit and cost-effectiveness analysis. On the one hand, cost-effectiveness analysis is easier to perform, since intangible benefits need only be measured but not valued. Indeed, according to Feldstein (97), even the problem of choice of discount rate is simpler in the case of cost-effectiveness analysis, with only the social time preference rate being relevant. On the other hand, to resort to cost-effectiveness analysis is to give up in advance whatever help analysis can offer in choosing among several objectives or program areas. It then becomes necessary to make the choice among programs on other grounds, as the Gottschalk Committee did.

I am unable to discern a general resolution to this dilemma. It is certainly not evident how to establish priorities in a systematic way when cost-benefit analysis is abandoned. Perhaps the choice can still be made in a practicable way, with reasons explicitly stated, when remarkable benefits are under consideration, as in the treatment of end-stage kidney disease. When the benefits in question are modest but difficult to value, how is one to decide whether or not to adopt a particular piece of technology? To follow the lead of pace-setting organizations is almost always to say yes. Perhaps we should put trust in our ability to continue to improve the valuation of intangible benefits in the future (28). Setting standard values on gains in life expectancy at various ages would seem worth exploring. However, I can also see increasing difficulty in the future in valuing direct tangible benefits, if fewer market prices become available for health services in the event that provider reimbursement shifts away from fee-for-service toward capitation and salary methods.

Automated Multiphasic Screening

Often cited and discussed as an example of technologic development in the health field is automated multiphasic health screening (98). The reports issued from the Kaiser-Permanente laboratories in Oakland and San Francisco reveal a good deal about the organization and staffing of such a service and present data on unit costs (99-103). No evaluation akin to cost-benefit or cost-effectiveness analysis was attempted prior to 1973, when a preliminary cost-benefit analysis for middle-aged men was issued (104).

Collen and associates (101) report that total costs for screening an individual are $21.32; which, they note, is only one-fourth or one-fifth of the cost of a periodic health examination employing more conventional modalities. The position of the authors is that this comparison will serve for the time being, pending determination of the efficacy of multiphasic health screening. The fact is that some people do undergo a periodic health examination, whatever its efficacy may be.

Garfield's position (105, 106) differs from that of Collen, in that the effectiveness of screening in arresting or curing previously unknown disease is beside the point. For Garfield, automated multiphasic screening has assumed a useful social function, serving as a sorting mechanism for patients with prepayment who would otherwise flood the health services system.

I have difficulty with both positions. Collen's comparison of cost with that of the periodic health examination reminds one that the latter procedure is notoriously controversial, with the central issue revolving precisely about its effectiveness. Among physicians there appear to be true believers, persistent skeptics, and ambivalent prescribers (107-109). Furthermore, as emphasized in the Nuffield report (110), screening implies an invitation to the patient to come and see the doctor who promises him a favorable outcome. This is in contrast to the more usual visit initiated by the patient who has symptoms and seeks relief.

My criticisms of Garfield's position are more serious, for his view that automated multiphasic screening should be regarded as a sorting mechanism, a substitute for the rationing of services by price, raises a host of questions. Apparently, judging from a more recent presentation of his position (111), much of Garfield's argument is based on an interpretation of what happened under Medicare and Medicaid. To my knowledge, the Medicare program experienced only a modest increase in the use of services and a huge, unexpected, increase in unit cost. There is no way to interpret the unanticipated rise in expenditures under Medicaid in the absence of data on trends in size of the eligible population, per capita use, and unit cost. My own view is that the increase in eligible population may have been the major factor.[1]

Garfield (111) has hypothesized a difference in price elasticity of demand between the sick and his other three categories of patient—the well, the worried well, and the early sick. However, there have been no empirical studies of the demand for physician services

[1] See Klarman, H. E. Major public initiative in health care. *Public Interest* 34: 106-123, 1974.

in which people are so classified. From other studies it would appear that a host of factors, such as health insurance, earnings as an expression of the value of time, age, and the supply of providers, are important determinants of the demand for physician services (112).

The assertion that the supply of services for sick care is inelastic is not unique to Garfield. In the area of trends in the education of physicians, which takes longer and therefore responds more slowly than any other health occupation, my own reading indicates that even this system has been somewhat responsive, even while insisting that class size in medical school must be kept small (18, p. 101), and still more responsive after the policy decision to expand enrollment was made and implemented by funding. Whether the supply response has been sufficient to meet rising demand is, of course, a different issue.

The most serious reservation I have about Garfield's position touches closely on the nature and function of cost-benefit analysis. If complete prepayment serves to create a condition of perpetual excessive demand, then some rationing or control measures are clearly indicated. Why assume, without comparing alternatives, that automated multiphasic screening is the most appropriate instrumentality? It seems to me that when the stated purposes of a program change, so should the menu of alternatives to be considered.

Two reports by Collen (101, 102) on the cost of screening fill a real need. Two measures are presented—cost per test and cost per screening. Cost per test reflects only direct departmental costs, while cost per screening incorporates an allocation of overhead expense. The article published in 1970 (102) offers a costing rule: in order to allow for all costs incurred, double the reported cost per test. The earlier article (101), which appears to present essentially the same data, suggests a blow-up of 50 per cent; I am unable to account for the difference.

Since the screening process is automatic, the capital equipment is indivisible, and all procedures are schedulable, economies of scale are to be expected. The larger the scale of operation, the lower is the average unit cost. However, to achieve the lower cost, full utilization of existing facilities is essential. Accordingly, it is said to be advantageous to have available a source of stand-by patients, such as those awaiting admission to the hospital (103).

Collen's second article (102) goes beyond cost per test or per screening, and reports cost per positive case. For mammography a prevalence rate of 1.2 per cent converts the unit cost of $4.90 into a cost per positive case of $408. Since one-fifth of the women with positive mammograms have cancer of the breast, the screening cost per true positive case is $2,000. His doubling rule would raise the cost to $4,000. The cost of diagnosis for all five women and of treatment for one is still excluded.

The proportion of false positives is a function not only of the accuracy of the screening test but also of the prevalence rate (113, 114). There are two reasons for aiming to keep down the number of false positives: to avoid needless anxiety, and to prevent iatrogenic disease associated with the diagnostic process itself.

The data reported to date from the Kaiser-Permanente laboratories indicate that automated multiphasic screening is both feasible and affordable. The question is whether it is worthwhile. One answer is in terms of its effects on health. The Advisory Committee on Automated Multiphasic Health Testing and Services (AMHTS) (115) states that much of disease uncovered by testing will be chronic or not reversible; it will not yield a saving in the use of services or an improvement in health. There seems to be little point to using multiphasic screening if this is the case.

A second answer is that of Garfield (111), which I have criticized at length. He provides no persuasive reason for choosing this instrumentality to control the use of physician services.

A third answer is possible: that automated multiphasic screening is an integral part of a package of comprehensive health services to which everybody has a right. Usually a service is aspired to by the poor because the middle and upper classes are already getting it. This is not yet the case regarding automated multiphasic screening.

Clearly, a reasonable answer can only be provided through an evaluation of automated health screening for its worthwhileness. The report by the Advisory Committee (115) states, "There are elements of AMHTS that defy cost-effectiveness analysis, but which depend primarily on medical, social, and scientific objectives." If I understand the statement, I disagree with it. It may be, however, that I do not understand it. What are the medical, or social, or scientific objectives that defy measurement?

Following the formulation of data requirements given in the preceding section, I propose that data be compiled to evaluate automated multiphasic screening as follows: the volume of disease detected that was not previously known; what could be and in fact was done about all this disease; what the outcomes in terms of health status and subsequent utilization of services were; and at how much cost, inclusive of diagnosis and treatment, the outcomes were attained (116, 117). It must be added that, as indicated by a recent paper (118) which compares study and control groups for such measures of outcome as work and health services utilization, Collen's group is steadily compiling more and more of the requisite data. Still lacking is information on costs that correspond to the specified benefits.

Barriers to Systematic Analysis

To bring some focus to a discussion of the necessary steps ahead, I have prepared a list of barriers to the systematic and rational analysis of expenditures for health systems technology. At the same time I shall assess the prospects for lowering or overcoming each barrier.

1. When the costs of operation mount beyond all projections, the tendency is to argue that the computer or automated laboratory, as the case may be, is not merely providing services but is performing a research function. Yet doing things we know little about does not define research. Certain features of research, such as formulation of hypotheses, design of study, and capability for statistical analysis of data, are not necessarily available wherever services are rendered. Although some replication of research is desirable, it should be intentional and need not be universal (119). It follows that sources of research funds should exercise discrimination in allocating them. If the absorption of so-called research costs by patients is precluded, this tendency to encourage pseudoresearch will be minimized.

2. A tendency exists to expand the range of functions said to be performed by new equipment. Surely, data on payroll could assist management in controlling cost by department; data on billings could provide a proxy for cost data by diagnosis. The first of these applications can be evaluated according to a strict criterion: is potential cost control

achieved, so that savings are realized? The second application can be judged on its own merits as an intermediate good: of what value is such information and to whom?

3. In the health field there is a tendency to adopt the best available and latest technology in every institution. This drive is promoted by the medical ethic of doing the utmost for the individual patient and reinforced by current methods of paying providers by third parties. The voluntary nonprofit form of organizing hospitals is frequently mentioned as a factor. Still another factor is usually neglected, namely, the nature of the physician-hospital relationship in this country. Physicians who specialize in treating patients with a given disease will not accede to its exclusion from hospital A, where they hold a staff appointment, unless they are granted staff privileges in hospital B, where the planning agency would like to concentrate all facilities for diagnosis and treatment. Only in part are financial interests involved; equally, or even more important, is the preservation and application of professional skills.

4. Economic valuation has no meaning without a firm basis in the underlying data on the link between the inputs and outputs of specific programs. It is not often that economists can develop such data. Other investigators must be persuaded and enabled to do this by investing their time and energies in longitudinal studies.

5. It is discouraging to perform technical analysis, to persuade the decision makers of its usefulness, to have it adopted, and then to discover that funds for health services are cut off because total government spending is being curtailed. Adjusting aggregate demand in the economy through changes in total expenditures is bound to result in the stop-and-go operation of individual programs. This is both wasteful and frustrating, and poses a substantial threat to continuity in the provision of health services through public financing.

6. Since cost-benefit or cost-effectiveness analysis is economic evaluation of public projects or programs, it must inevitably take place in a political climate. While the economic tool of cost-benefit analysis implies a delineation of goals and an articulation of values, the imperatives of the political process may call for a blurring of differences and potential conflicts, in order to facilitate the building of coalitions aimed at the accomplishment of particular ends. Schultze (31) has observed this paradox: PPB has been applied most in an area, national defense, where future uncertainty is greatest but value differences among citizens have been traditionally least; PPB is not applied much in the human resources area, where the problem of uncertainty is not so serious, but differences in values among citizens prevail, as well as a great many vested interests.

Some political scientists, such as Wildavsky (120, 121), would agree with the above description and conclude that such are the facts of life. Most changes in governmental budgets are incremental anyway and do not—indeed cannot—derive from base zero (122). Within the boundaries set by defined political understandings, there are ample opportunities to improve decision making through systematic analysis. There is no reason to believe that politicians prefer to make poor decisions over good ones. In cases that are of vital importance to the body politic, many politicians, when persuaded of the right thing to do, would be willing to use up some of the credit they have accumulated and make the tough, though unpopular, choice. They cannot take such a stand on every issue, however. Therefore, the exceptionally capable practitioner of economic cost-benefit analysis must know how and when to make an allowance for the existence of a political cost-benefit calculus (120).

/ Klarman

REFERENCES

1. Haveman, R. H. Public expenditures and policy analysis: An overview. In *Public Expenditures and Policy Analysis*, edited by R. H. Haveman and J. Margolis, pp. 1-18. Markham Publishing Company, Chicago, 1970.
2. Eckstein, O. A survey of the theory of public expenditure criteria. In *Public Finances: Needs, Sources, and Utilization*, edited by J. M. Buchanan, pp. 439-494. Princeton University Press, Princeton, N.J., 1961.
3. Prest, A. R., and Turvey, R. Cost-benefit analysis: A survey. *Economic Journal* 75(300): 683-735, 1965.
4. Steiner, P. O. *Public Expenditure Budgeting*. The Brookings Institution, Washington, D.C., 1969.
5. Eckstein, O. *Water Resources Development*. Harvard University Press, Cambridge, 1958.
6. Hirshleifer, J., DeHaven, J. C., and Milliman, J. W. *Water Supply: Economics, Technology and Policy*. University of Chicago Press, Chicago, 1960.
7. Kneese, A. V., and Smith, S. C. *Water Research*. Johns Hopkins Press, Baltimore, 1966.
8. Krutilla, J. V., and Eckstein, O. *Multiple Purpose River Development*. Johns Hopkins Press, Baltimore, 1958.
9. Maass, A., Hufschmidt, M. M., Dorfman, R., Thomas, H. A., Marglin, S. A., and Fair, G. M. *Design of Water-Resource Systems*. Harvard University Press, Cambridge, 1962.
10. McKean, R. N. *Efficiency in Government Through Systems Analysis*. John Wiley & Sons, Inc., New York, 1958.
11. Arrow, K. J. The organization of economic activity: Issues pertinent to the choice of market versus nonmarket allocation. In *Public Expenditures and Policy Analysis*, edited by R. H. Haveman and J. Margolis, pp. 59-73. Markham Publishing Company, Chicago, 1970.
12. Samuelson, P. A. The pure theory of public expenditure. *Review of Economics and Statistics* 36(4): 387-389, 1954.
13. Weisbrod, B. A. *Economics of Public Health*. University of Pennsylvania Press, Philadelphia, 1961.
14. Sloan, F. A. *Planning Public Expenditure on Mental Health Service Delivery*. The New York City Rand Institute, New York, 1971.
15. Stigler, G. J. *The Theory of Price*, rev. ed., p. 150. Macmillan and Company, New York, 1952.
16. Fuchs, V. R. Health care and the U.S. economic system: An essay in abnormal physiology. In *Technology and Health Care Systems in the 1980's*, edited by M. F. Collen, pp. 169-175. National Center for Health Services Research and Development, Washington, D.C., 1973.
17. Lasagna, L. Quoted in *Decision-Making on the Efficacy and Safety of Drugs*, edited by J. D. Cooper, p. 2. Interdisciplinary Communication Associates, Washington, D.C., 1971.
18. Klarman, H. E. *The Economics of Health*. Columbia University Press, New York, 1965.
19. Freeman, A. M. Project design and evaluation with multiple objectives. In *Public Expenditures and Policy Analysis*, edited by R. H. Haveman and J. Margolis, pp. 347-363. Markham Publishing Company, Chicago, 1970.
20. Haveman, R. H., and Krutilla, J. V. *Unemployment, Idle Capacity and the Evaluation of Public Expenditures*. Johns Hopkins Press, Baltimore, 1968.
21. McGuire, M. C., and Garn, H. A. The integration of equity and efficiency criteria in public project selection. *Economic Journal* 79(316): 882-893, 1969.
22. Weisbrod, B. A. Income redistribution effects and benefit-cost analysis. In *Problems in Public Expenditure Analysis*, edited by S. B. Chase, pp. 177-209. The Brookings Institution, Washington, D.C., 1968.
23. United States Department of Health, Education, and Welfare. *Motor Vehicle Injury Prevention Program; Cancer Control; Arthritis Control; Selected Disease Control Programs*. Office of Assistant Secretary for Program Coordination, Washington, D.C., 1961.
24. Acton, J. P. Evaluation of a Life-Saving Program: The Case of Heart Attacks. Doctoral Dissertation, Harvard University, Cambridge, 1970.
25. McKean, R. N. The use of shadow prices. In *Problems in Public Expenditure Analysis*, edited by S. B. Chase, pp. 33-65. The Brookings Institution, Washington, D.C., 1968.
26. Margolis, J. Shadow-prices for incorrect or nonexistent market values. In *Public Expenditures and Policy Analysis*, edited by R. H. Haveman and J. Margolis, pp. 314-329. Markham Publishing Company, Chicago, 1970.
27. Fein, R. Definition and scope of the problem: Economic aspects. In *Assessing the Effectiveness of Child Health Services*, edited by A. B. Bergman, pp. 44-50. Ross Laboratories, Columbus, Ohio, 1967.
28. Weisbrod, B. A. Concepts of costs and benefits. In *Problems in Public Expenditure Analysis*, edited by S. B. Chase, pp. 257-262. The Brookings Institution, Washington, D.C., 1968.

29. Klarman, H. E., Francis, J. O'S., and Rosenthal, G. D. Cost effectiveness analysis applied to the treatment of chronic renal disease. *Med. Care* 6(1): 48-54, 1968.
30. Gorham, W. Allocating federal resources among competing social needs. *HEW Indicators* pp. 1-13, August 1966.
31. Schultze, C. L. *The Politics and Economics of Public Spending*, pp. 1-34, 77-102. The Brookings Institution, Washington, D.C., 1968.
32. United States Bureau of the Budget. Bulletin 66-3, Supplement, and Bulletin 68-2 on Planning-Programming-Budgeting, October 12, 1965, February 21, 1966, and July 18, 1967. Reprinted in *Planning, Programming, Budgeting: A Systems Approach to Management*, edited by F. J. Lyden and E. G. Miller, pp. 405-443. Markham Publishing Company, Chicago, 1968.
33. Klarman, H. E. Present status of cost-benefit analysis in the health field. *Am. J. Public Health* 57(11): 1948-1953, 1967.
34. Henderson, P. D. Investment criteria for public enterprises. In *Public Enterprise*, edited by R. Turvey, pp. 86-169. Penguin Modern Economics Readings, Penguin Books, Baltimore, 1968.
35. Musgrave, R. A. Cost-benefit analysis and the theory of public finance. *Journal of Economic Literature* 7(3): 797-806, 1969.
36. Baumol, W. J. On the discount rate for public projects. In *Public Expenditures and Policy Analysis*, edited by R. H. Haveman and J. Margolis, pp. 273-290. Markham Publishing Company, Chicago, 1970.
37. *Interest Rate Guidelines for Federal Decision-Making*. Joint Economic Committee, Congress of the United States. U.S. Government Printing Office, Washington, D.C., 1968.
38. Boulding, K. E. Notes on a theory of philanthropy. In *Philanthropy and Public Policy*, edited by F. G. Dickinson, pp. 57-71. National Bureau of Economic Research, New York, 1962.
39. Hufschmidt, M. M., Krutilla, J., and Margolis, J. Standards and criteria for formulating and evaluating federal water resources developments. In *Hearings on Guidelines for Estimating the Benefits of Public Expenditures*, pp. 135-212. Joint Economic Committee, Congress of the United States. U.S. Government Printing Office, Washington, D.C., 1969.
40. Feldstein, M. S. The social time preference discount rate in cost benefit analysis. *Economic Journal* 74(294): 360-379, 1964.
41. Mushkin, S. J., and Collings, F. d'A. Economic costs of disease and injury. *Public Health Rep.* 74(9): 795-809, 1959.
42. Fein, R. *Economics of Mental Illness*. Basic Books, New York, 1958.
43. Rice, D. P. *Economic Costs of Cardiovascular Diseases and Cancer*. Health Economics Series No. 5. U.S. Government Printing Office, Washington, D.C., 1965.
44. Rice, D. P. *Estimating the Cost of Illness*. Health Economics Series No. 6. U.S. Government Printing Office, Washington, D.C., 1966.
45. Klarman, H. E. Syphilis control programs. In *Measuring Benefits of Government Investments*, edited by R. Dorfman, pp. 367-410. The Brookings Institution, Washington, D.C., 1965.
46. Klarman, H. E. Socioeconomic impact of heart disease. In *The Heart and Circulation*, Vol. 2, pp. 693-707. Federation of American Societies for Experimental Biology, Washington, D.C., 1965.
47. Menz, F. C. Economics of disease prevention: Infectious kidney disease. *Inquiry* 8(4): 3-18, 1971.
48. Fuchs, V. R. The contribution of health services to the American economy. *Milbank Mem. Fund Q.* 44(4, part 2): 65-101, 1966.
49. Grosse, R. N. Cost-benefit analysis of health service. *Annals* 399: 89-99, January 1972.
50. Klarman, H. E. Conference on the economics of medical research. In Report of the President's Commission on Heart Disease, Cancer, and Stroke, Vol. 2, pp. 631-644. U.S. Government Printing Office, Washington, D.C., 1965.
51. Mushkin, S. J. Health as an investment. *Journal of Political Economy* 70(5): 129-157, 1962.
52. Bowen, H. R. *Toward Social Economy*. Rinehart and Company, Inc., New York, 1948.
53. Rice, D. P., and Cooper, B. S. The economic value of human life. *Am. J. Pub. Health* 57(11): 1954-1966, 1967.
54. Mishan, E. J. Evaluation of life and limb. *Journal of Political Economy* 79(4): 687-705, 1971.
55. Kuznets, S. *National Income and Its Composition, 1919-1938*, pp. 22-23. National Bureau of Economic Research, New York, 1947.
56. Walker, K. E., and Gauger, W. H. The Dollar Value of Household Work. New York State College of Human Ecology, Cornell University, Ithaca, New York, 1973 (processed).
57. Sirageldin, I. A-H. *Non-Market Components of National Income*. Institute for Social Research, University of Michigan, Ann Arbor, 1969.
58. Gelman, A. C. *Multiphasic Health Testing Systems: Reviews and Annotations*. U.S. Government Printing Office, Washington, D.C., 1971.

59. Vickrey, W. S. One economist's view of philanthropy. In *Philanthropy and Public Policy*, edited by F. G. Dickinson, pp. 31-56. National Bureau of Economic Research, New York, 1962.
60. Neenan, W. B. *Normative Evaluation of a Public Health Program*. Institute of Public Administration, University of Michigan, Ann Arbor, 1967.
61. Schelling, T. C. The life you save may be your own. In *Problems in Public Expenditure Analysis*, edited by S. B. Chase, pp. 127-162. The Brookings Institution, Washington, D.C., 1968.
62. Thurow, L. *Investment in Human Capital*, p. 134. Wadsworth Publishing Company, Inc., Belmont, California, 1970.
63. Titmuss, R. M. *The Gift Relationship*, p. 198. Pantheon Books, New York, 1971.
64. Morris, J. N. *The Uses of Epidemiology*, p. 45. E. & S. Livingston, Edinburgh, 1957.
65. Scheff, T. J. Preferred errors in diagnosis. *Med. Care* 2(3): 166-172, 1964.
66. Saslaw, M., Vieta, A., and Myerburg, R. Cost of rheumatic fever and its prevention. *Am. J. Public Health* 55(3): 429-434, 1965.
67. Report of the Committee on Chronic Kidney Disease (C. W. Gottschalk, chairman). Bureau of the Budget, Washington, D.C., 1967.
68. Grosse, R. N. Problems of resource allocation in health. In *Public Expenditures and Policy Analysis*, edited by R. H. Haveman and J. Margolis, pp. 518-548. Markham Publishing Company, Chicago, 1970.
69. Feldstein, M. S. Health sector planning in developing countries. *Economica*, New Series 37(146): 139-163, 1970.
70. Quade, E. S. Systems analysis techniques for planning-programming-budgeting. In *Planning, Programming, Budgeting: A Systems Approach to Management*, edited by F. J. Lyden and E. G. Miller, pp. 292-312. Markham Publishing Company, Chicago, 1968.
71. Sturm, H. M. *Technology and Manpower in the Health Service Industry, 1965-1975*. U.S. Government Printing Office, Washington, D.C., 1968.
72. Klarman, H. E. Economic Aspects of Chronic Kidney Disease Revisited. Paper presented at annual meeting of American Public Health Association, November 11, 1969 (processed).
73. Bennett, I. L. The Scientific and Educational Basis for Medical Care. Paper presented at New York Health Services Research and Policy Seminar, December 1, 1970 (processed).
74. Feldstein, M. S. *The Rising Cost of Hospital Care*. Information Resources Press, Washington, D.C., 1971.
75. Blaug, M. A survey of the theory of process innovation. In *The Economics of Technological Change*, edited by N. Rosenberg, pp. 86-113. Penguin Modern Economics Readings, Penguin Books, Baltimore, 1971.
76. Fuchs, V. R., and Kramer, M. J. *Expenditures for Physicians' Services in the United States, 1948-1968*, pp. 35-42. National Center for Health Services Research and Development, Washington, D.C., 1973.
77. Grossman, M. *The Demand for Health: A Theoretical and Empirical Investigation*. National Bureau of Economic Research, New York, 1972.
78. Klarman, H. E. Increase in the cost of physician and hospital services. *Inquiry* 7(1): 22-36, 1970.
79. Klarman, H. E. Approaches to moderating the increases in medical care costs. *Med. Care* 7(3): 175-190, 1969.
80. Pauly, M. V., and Drake, D. F. Effect of third-party methods of reimbursement on hospital performance. In *Empirical Studies in Health Economics*, edited by H. E. Klarman, pp. 297-314. Johns Hopkins Press, Baltimore, 1970.
81. Davis, K. A. Theory of Economic Behavior in Non-Profit Hospitals. Doctoral Dissertation, Rice University, Houston, 1969.
82. Ginsburg, P. B. Capital in Non-Profit Hospitals. Doctoral Dissertation, Harvard University, Cambridge, 1970.
83. Newhouse, J. P. Toward a theory of nonprofit institutions: An economic model of a hospital. *American Economic Review* 60(1): 64-74, 1970.
84. Reder, M. W. Some problems in the economics of hospitals. *American Economic Review* (Papers and Proceedings) 55(2): 472-480, 1965.
85. Lee, M. L. A conspicuous production theory of hospital behavior. *Southern Economic Journal* 38(1): 48-58, 1971.
86. Rosenstock, I. M. Why people use health services. *Milbank Mem. Fund Q.* 44(3, part 2): 94-124, 1966.
87. Ginzberg, E. *Men, Money, and Medicine*, p. 55. Columbia University Press, New York, 1969.
88. Long, M. F. Efficient use of hospitals. In *The Economics of Health and Medical Care*, edited by S. J. Mushkin, pp. 211-226. Bureau of Public Health Economics, University of Michigan, Ann Arbor, 1964.
89. Anderson, O. W. Trends in hospital use and their policy implications. In Fifth Annual

Symposium on Hospital Affairs, *Where is Hospital Use Headed?* pp. 2-5. Graduate School of Business, University of Chicago, Chicago, 1964.
90. Feldstein, P. J. *An Empirical Investigation of the Marginal Cost of Hospital Services.* Graduate Program in Hospital Administration, University of Chicago, Chicago, 1961.
91. Roemer, M. I. Bed supply and hospital utilization: A national experiment. *Hospitals* 35(21): 36-42, 1961.
92. Shain, M., and Roemer, M. I. Hospital costs relate to the supply of beds. *Mod. Hosp.* 92(4): 71-73, 168, 1959.
93. Flagle, C. D. Evaluation and control of technology in health services. In *Technology and Health Care Systems in the 1980's*, edited by M. F. Collen, pp. 213-224. National Center for Health Services Research and Development, Washington, D.C., 1973.
94. Shegog, R. F. A. Reviewing some applications of computers to medicine. In *Problems and Progress in Medical Care*, Third Series, edited by G. McLachlan, pp. 145-170. Oxford University Press, London, 1968.
95. Ahlvin, R. C. Biochemical screening–A critique. *New Engl. J. Med.* 283(20): 1084-1086, 1970.
96. Flagle, C. D., Gabrielson, I. W., Soriano, A., and Taylor, M. M. *Analysis of Congestion in an Out-patient Clinic.* Operations Research Division, The Johns Hopkins Hospital, Baltimore, 1959.
97. Feldstein, M. S. Choice of technique in the public sector: A simplification. *Economic Journal* 80(320): 985-990, 1970.
98. National Academy of Engineering. *A Study of Technology Assessment*, pp. 96-104. Committee on Science and Astronautics, U.S. House of Representatives. U.S. Government Printing Office, Washington, D.C., 1969.
99. Collen, M. F. Statement in Subcommittee on Health of the Elderly, Special Committee on Aging, U. S. Senate. *Hearings on Detection and Prevention of Chronic Disease Utilizing Multiphasic Health Screening Techniques*, pp. 214-222. U.S. Government Printing Office, Washington, D.C., 1966.
100. Collen, M. F. Periodic health examinations using an automated multitest laboratory. *J.A.M.A.* 195(10): 830-833, 1966.
101. Collen, M. F., Kidd, P. H., Feldman, R., and Cutler, J. L. Cost analysis of a multiphasic screening program. *New Engl. J. Med.* 280: 1043-1045, 1969.
102. Collen, M. F., Feldman, R., Siegelaub, A. B., and Crawford, D. Dollar cost per positive test for automated multiphasic screening. *New Engl. J. Med.* 283: 459-463, 1970.
103. Collen, M. F. Automated multiphasic health testing: Implementation of a system. *Hospitals* 45(5): 49-50, 56-58, 1971.
104. Collen, M. F., Dales, L. G., Friedman, G. D., Flagle, C. D., Feldman, R., and Siegelaub, A.B. *Multiphasic Checkup Evaluation Study: 4. Preliminary Cost-Benefit Analysis for Middle Aged Men.* Medical Methods Research, Oakland, California, 1973 (processed).
105. Garfield, S. R. The delivery of medical care. *Sci. Am.* 222(4): 15-23, 1970.
106. Garfield, S. R. Multiphasic health testing and medical care as a right. *New Engl. J. Med.* 283(20): 1087-1089, 1970.
107. Ingalls, T. H., and Gordon, J. E. Periodic health examination, 1900-1965. *Am. J. Med. Sci.* 251(3): 123-140, 1966.
108. Siegel, G. S. An American dilemma–The periodic health examination. *Arch. Environ. Health* 13(3): 292-295, 1966.
109. Wade, L., Thorpe, J., Elias, T., and Bock, G. Are periodic health examinations worthwhile? *Ann. Intern. Med.* 56(1): 81-93, 1962.
110. McKeown, T. Validation of screening procedures. In *Screening in Medical Care: Reviewing the Evidence*, pp. 1-13. Oxford University Press, London, 1968.
111. Garfield, S. R. A look at the economics of medical care. In *Technology and Health Care Systems in the 1980's*, edited by M. F. Collen, pp. 169-175. National Center for Health Services Research and Development, Washington, D.C., 1973.
112. Feldstein, P. J. Research on the demand for health services. *Milbank Mem. Fund Q.* 44 (3, part 2): 128-162, 1966.
113. Blumberg, M. S. Evaluating health screening procedures. *Operations Research* 5(3): 351-360, 1957.
114. Thorner, R. R., and Remein, Q. M. *Principles and Procedures in the Evaluation of Screening for Disease.* U.S. Government Printing Office, Washington, D.C., 1971.
115. Report of the AMHTS Advisory Committee to the National Center for Health Services Research and Development, Vol. 1, pp. 29, 31. U.S. Government Printing Office, Washington, D.C., 1970.
116. Pole, J. D. Economic aspects of screening for disease. In *Screening in Medical Care: Reviewing the Evidence*, pp. 141-157. Oxford University Press, London, 1968.
117. Suchman, E. A. *Evaluative Research*, p. 65. Russell Sage Foundation, New York, 1967.

118. Ramcharan, S., Cutler, J. L., Feldman, R., Siegelaub, A. B., Campbell, B., Friedman, G. D., Dales, L., and Collen, M. F. Kaiser-Permanente Multiphasic Health Checkup Evaluation Project: 2. Disability and Chronic Disease after Seven Years of Multiphasic Health Checkups. Paper presented at the Epidemiology Section of the American Public Health Association, October 11, 1971 (processed).
119. Etzioni, A., and Remp, R. Technological 'shortcuts' to social change. *Science* 175: 31-38, January 7, 1972.
120. Wildavsky, A. The political economy of efficiency: Cost-benefit analysis, and program budgeting. *Public Administration Review* 26(4): 292-310, 1966.
121. Wildavsky, A., and Hammann, A. Comprehensive versus incremental budgeting in the department of agriculture. *Administrative Science Quarterly* 10(3): 321-346, 1965.
122. Lindblom, C. E. Decision-making in taxation and expenditures. In *Public Finances: Needs, Sources, and Utilization*, edited by J. M. Buchanan, pp. 295-329. Princeton University Press, Princeton, N.J., 1961.

Manuscript submitted for publication, April 12, 1973

Direct reprint requests to:

Dr. Herbert E. Klarman
Professor of Economics
New York University
Graduate School of Public Administration
4 Washington Square North
New York, New York 10003

#22

MEDICAL CARE
November 1980, Vol. XVIII, No. 11

Original Articles

Cost-Benefit and Cost-Effectiveness Analysis in Health Care

Growth and Composition of the Literature

KENNETH E. WARNER, PH.D.,* AND REBECCA C. HUTTON, M.H.S.A., M.A.A.E.†

Concern about the escalating costs of health services is reflected in the rapid growth of the literature on cost-benefit and cost-effectiveness analysis (CBA and CEA, respectively) in health care. A search of that literature for 1966–78 produced a bibliography of more than 50[illegible] relevant references, growing from half a dozen per year at the beginning of the period to close to 100 each of the most recent 2 years. The literature growth has been more rapid in medical than nonmedical journals and a preference for CEA over CBA appears to be emerging. Studies related to diagnosis and treatment have gained in popularity, while the early prominence of studies with a substantive prevention theme has diminished. Consistent with the increasing medical focus of the literature, num[illegible] articles oriented toward individual practitioner decision making have grow[illegible] more rapidly than those oriented toward organizational or societal decision making. In addition to documenting these trends, this article identifies published reviews of health care CBA/CEA and books and articles attempting to convey the principles of CBA/CEA to the health care community. The article concludes with speculation on likely near-future trends in the literature and consideration of the quality implications of the rapid growth.

CONCERN about the high and rising costs of health care has prompted a wide variety of cost-containment efforts in both the public and private sectors.[1] Underlying all of these efforts is the search for the appropriate balance between the costs of care on the one hand and the amount and quality of care on the other. The challenge is to determine which medical interventions are effective and under what circumstances, and then to compare alternative interventions in terms of their costliness, to assess their relative cost-effectiveness. The task is a monumental one; determining effectiveness is itself immensely difficult.[2]

To identify and convey the meaning of cost-effective medical interventions, numerous scholars and policy makers are advocating the performance and use of two closely related evaluative techniques: cost-benefit and cost-effectiveness analysis (CBA and CEA, respectively). As measured by contributions to the literature, health professionals' interest in these techniques and in their findings increased exponentially through the past decade.

* Associate Professor, Department of Health Planning and Administration, School of Public Health, University of Michigan.

† Health Economist, Foundation for Medical Care Evaluation of Southeastern Wisconsin, Inc., Milwaukee.

Research for this article was supported by contracts OTA-C-78-128 and 933-3450.0 with the Office of Technology Assessment, U.S. Congress.

0025-7079/80/1100-1069 $01.90 © J. B. Lippincott Co.

The purposes of this article are to describe the growth and composition of the health care CBA/CEA literature, to speculate on possible trends in the near future, and to reflect briefly on the quality implications of the growth pattern we have documented.

The next section of the paper reviews published attempts to convey the principles of CBA/CEA to the health care community, and to acquaint that community with the substantive content of the literature. The following section describes the development of our bibliography of more than 500 items on health care CBA/CEA, and the method employed to classify literature contributions for purposes of the analysis presented in the following two sections. This discussion emphasizes the intent, meaning and limitations of our selection of items for inclusion in the bibliography. As we explain there, our objective was to capture the literature which health professionals perceive as CBA/CEA; this may diverge in significant ways from the literature which an expert analyst would classify as CBA/CEA on grounds of purely technical characteristics. The following section of the paper offers an empirical characterization of the magnitude and nature of the literature, examining the diffusion over time of health care CBA/CEA interest in several dimensions (e.g., number of publications, the mix of medical and nonmedical publication vehicles, and medical functions emphasized). The next section describes the specific substantive topics and areas of interest that have dominated the attention of health care CBA/CEA authors. The concluding section speculates on likely near-future trends in the health care CBA/CEA literature and offers some thoughts on the quality implications of the growth pattern that has emerged in recent years.

We wish to emphasize that by "health care" we mean personal (individual) health care services. CBAs and CEAs of such health-relevant subjects as highway safety, environmental protection and certain aspects of individual health behavior have been excluded from consideration. This restriction reflects the desire to examine carefully subjects of most direct and immediate concern to the medical and health-policy communities and to keep the study within manageable bounds. The restriction is not intended to suggest any biases as to the relative interest or importance of medical and nonmedical health subject areas.

Conveying Principles and Practice To the Health Care Community

Cost-benefit and cost-effectiveness analysis have come to refer to formal analytical techniques for comparing the negative and positive consequences of alternative uses of resources. The objective of CBA and CEA is to structure and analyze information in a manner which will inform and thereby assist decision making. Often enshrouded in technical jargon and mathematics, CBA and CEA are really nothing more than attempts to weigh logically the pros and cons of a decision. Each of us engages in CBA/CEA-like thinking every day, frequently subconsciously. "And ultimately, something like [CBA/CEA] must necessarily be employed in any rational decision."[3]

Both of these techniques require analysts to identify, measure and compare all of the significant positive and negative consequences of alternative programmatic means of addressing a given problem. The principal technical distinction between the two techniques lies in the valuation of the desirable consequences of programs. In CBA, all such consequences —benefits—are valued, like costs, in monetary terms. Conceptually, this permits an assessment of the inherent worth of a program—do the benefits exceed the costs?—as well as comparison of competing program alternatives, i.e., which of several programs generates the largest excess of benefits over costs? Because all costs and benefits are measured in the same (monetary) unit, CBA can be employed to compare similar or widely divergent types of programs. Thus, in theory at least, CBA might be used to decide whether public resources should be allocated to construction of a dam or construction of a hospital.

In CEA, desirable program consequences are not valued in monetary terms, but rather are measured in some other unit; in health care CEAs, common measures include years of life saved and days of morbidity or disability avoided. The reason for a nonmonetary measure of program effectiveness is either the impossibility or undesirability of valuing important outcomes in dollars. Thus, the "bottom line" of a CEA is not, like a CBA, a net monetary value; rather, it is expressed in units like "dollars per year of life saved." CEA permits comparison of cost per unit of effectiveness among competing program alternatives designed to serve the same basic purpose, but unlike CBA it does not allow comparison of programs having different objectives, because the effectiveness or outcome measures differ. Nor does it permit assessment of the inherent worth of a program: is a cost of $50,000 per year of life saved acceptable? Obviously, this question requires a social and political judgment; it is not simply a technical matter.

CBA and CEA have been studied for decades by economists and systems analysts, but these analytic techniques are novelties for much of the health care community. As we discuss below, significant numbers of health care applications date only from the 1970s and efforts to convey the principles of analysis to the health care community are few and recent. Methodology and review publications addressed specifically to an audience of medical professionals are still fewer in number and of very recent vintage.

Of publications written solely to present or evaluate the state of the art in health care CBA/CEA, the first we found appeared in the mid-1960s. In 1966, Crystal and Brewster[4] wrote an introduction to CBA/CEA in the health field. The following year, Klarman[5] published the first of two prominent reviews he has written, this one appearing in the *American Journal of Public Health*.

From then until 1972, no significant health care review or methodology contributions appeared in print, with the exception of a chapter by Grosse[6] in a 1970 book oriented toward students of economics and policy analysis. This chapter is particularly noteworthy for its review of CBA/CEA applications in the Department of Health, Education, and Welfare (HEW) during the author's tenure as an HEW official. Grosse conveyed much of the same material in an article published 2 years later,[7] though again the audience was not health care professionals. That same year, however, witnessed publication of a book which became one of the health care community's most widely read and frequently cited contributions, Cochrane's *Effectiveness and Efficiency: Random Reflections on Health Services*.[8] In our estimation, this short book had a profound impact in turning the thoughts of health care professionals toward issues of resource scarcity and the link between efficiency and equity. It is at least possible that Cochrane's book played a significant role in the rapid growth in health care CBA/CEA which began the following year.

In 1974, Klarman published two articles on CBA methodology and applications. One is the most often cited review and discussion of health care CBA/CEA[9]; the other was the first such article to appear in a medical journal.[10] The following year, Dunlop published a review in *Social Science and Medicine*.[11] 1975 also witnessed publication of a controversial issue of the *New England Journal of Medicine* which was devoted to a discussion of CBA/CEA and related methodology[12] and several illustrations of its application.[13-16] To many observers, this *Journal* issue stands as a landmark in the evaluation of medical practice.

Two years later, another issue of the *New England Journal of Medicine* offered readers a discussion of CEA methodology,[17] a sophisticated application of it,[18] and an important, thoughtful treatment of the limitations of formal analysis.[19] Many health services researchers adjudge this package, combined with Weinstein and Stason's book on hypertension policy,[20] to be a milestone in health care CBA/CEA. At the end of that same year, in December 1977, the Arthur D. Little Company completed an "Introduction to Cost-Benefit Analysis Applied to New Health Technologies."[21] This document, intended to assist health planners and others, was prepared under a contract with the Bureau of Health Planning and Resources Development in the Health Resources Administration, HEW.

In December 1978 at the Urban Institute *Conference on Medical Technology*, Weinstein[22] [illegible] the state of the art of the literature and reviewed "[illegible] nonrandom sample" of health care CBA/CEAs. The most noteworthy feature of the paper is the author's discussion of remaining methodologic issues. While several of these have been of concern since the inception of formal CBA/CEA, others represent subtle, sophisticated problems, the existence of which is testimony to progress on more basic issues. Indeed, the paper serves as a vivid reminder simultaneously of the frustrating, seemingly intractable problems of CBA/CEA and of their gradual yielding to sustained conceptual and empirical struggle.

Recent months have witnessed several attempts to convey the principles of CBA/CEA to various elements of the health care community. In the fall of 1979, Dittman and Smith published an article in *Health Care Management Review* which presented a conceptual framework for consideration of costs and benefits.[23] Like the Arthur D. Little "Introduction,"[21] this work was supported by the Bureau of Health Planning and Resources Development and was addressed to health planners. Around the same time, Shepard and Thompson contributed a discussion of CEA principles in *Public Health Reports*.[24] An emerging theme of national health policy—health promotion and disease prevention—constituted the substantive focus of two recent reviews of CBA/CEA understanding. One, by Scheffler and Paringer, first appeared in the Surgeon General's report on prevention and health promotion and subsequently published in *Medical Care*.[25] The other, by Warner, was published in *Social Science and Medicine*.[26]

Within the federal government, the Congressional Office of Technology Assessment has recently completed an 18-month study of the Implications of Cost-Effectiveness Analysis of Medical Technology,[27] one component of which discusses the methodology of CBA/CEA and reviews the health care CBA/CEA literature. In contrast with previous reviews, the OTA report devotes considerable attention to the numerous CBA/CEA contributions which have escaped attention in the review literature, but which, because of their numbers and publication vehicles (eg, medical specialty journals), "constitute the principal exposure of many practicing health professionals to the language, concepts, and application of CBA/CEA." The OTA study includes a volume of case studies reviewing understanding of the costs and effectiveness of several important medical procedures.

Finally, in a new book, Warner and Luce[1] present the methodology of CBA/CEA, review and assess its application in the health care literature, and examine its policy potential and limitations in the context of the current policy debate on how to moderate health cost inflation.

Development of Health Care CBA/CEA Bibliography and Classification of References

The succeeding two sections present an empirical analysis of the growth and content of the health care CBA/CEA literature. The analysis derives from counts and classification of more than 500 references in a bibliography† covering the period 1966 through 1973 and including CBAs and CEAs on personal health services topics, reviews and comments on such literature, and discussions of CBA/CEA methodology directed specifically to health care professionals. Excluded from the bibliography are scores of CBA/CEAs on health-relevant but nonmedical subjects (e.g., traffic safety and control of environmental pollution), as well as dozens of books and articles on CBA/CEA methodology which are addressed to a general nonhealth audience. Thus, certain prominent articles which have had profound impacts on health [illegible]/CEA were not included in the bibliography because they were not directed exclusively at an audience concerned solely with personal health services. The seminal work of Rice on measuring the cost of illness is a good case in point,[28,29] as is the related work of Acton and others on measuring the value of life.[30] These studies reside at the heart of a long-lived CBA intellectual debate, and each has formed the basis of attempts to value the health benefits of programs, but the issues and techniques transcend categorization as personal health care methodologies; they are equally relevant to numerous human welfare programs outside of the personal health services arena. Studies which examine the costs of health programs, but not their benefits, are also excluded from the bibliography. Similarly, several excellent studies of the social costs of specific illnesses are excluded because they are not CBA/CEAs of efforts to combat these costs.[31,32]

While exclusion was the rule for tangentially related methodology references, we adopted a policy of inclusiveness as regards classification of a reference as a CBA or CEA. We have not read many of the included references, but a reading of a few randomly selected articles suggests that some of the studies which purport to be CBAs or CEAs are not true CBAs/CEAs. At the extreme, at [illegible] one articles whose title advertises it as presenting a cost-benefit ratio does not include any comparison of economic costs and benefits.[33] More commonly, numerous articles have a clear CBA/CEA intent but approach analysis in a manner which would cause the CBA/CEA expert to exclude them from a listing of technically sound CBA/CEAs.

Rather than weed out such references, we have chosen to include all studies known to be CBA/CEAs or which are presented as such in their titles. Our reasons were two: first, in order to achieve consistency in the weeding process, we would have had to read each of the 500-plus articles with great care, which time did not permit; even if we had had the time, a consistent inclusion/exclusion rule would have been difficult to design. Second, and of greater importance, our objective was to identify and characterize the literature which is introducing health professionals to the ideas and analysis of cost-effectiveness in health care delivery. Whether an article is of high or low quality, it serves this function, however well or poorly. In several fields of practice (e.g., nursing administration[34] and certain medical specialties), conceptually inaccurate articles appear to constitute the principal exposure of professionals to the language and ideas of CBA/CEA. Thus, in order to capture both the growth and nature of health

† [illegible]

care professionals' introduction to CBA/CEA, we felt it appropriate to include all articles which purport to relate to this theme, and not only those which CBA/CEA experts would characterize as technically sound.§

Following compilation of the bibliography, we classified each reference according to the following dimensions:

(1) Year (1966–1978)

(2) Type (CBA, CEA, general or unknown)

(3) Publication vehicle (medical journal, journal intended primarily for nonphysician health professionals, administrators, or health services researchers, nonhealth journal, other)

(4) Medical function of program or technology (prevention, diagnosis, treatment)

(5) Physical nature of program or technology (technique; drug, procedure, equipment; personnel, system)

(6) Decision orientation, i.e., whose decision-making the paper is intended to assist (e.g., individual practitioner, organization, society)

(7) Subject matter (a specific program or technology; review article; methodology)

Classification involved numerous arbitary judgments. Many of the assignments depended on the content of abstracts or even the wording of titles. Where available information suggested that each of two (very occasionally three) categories were appropriate, half (or a third) credit was assigned to each. For example, in the "medical function" dimension, certain screening programs were recorded as half prevention and half diagnosis. (A comprehensive blood pressure control program was counted as one third for each of prevention, diagnosis, and treatment.) "Unknown" or "other" categories were used liberally when we lacked confidence in our ability to categorize references accurately.

While the possibility remains that numerous assignments were nonoptimal, we are unaware of any significant sources of bias. Thus, at minimum the quantitative analysis should provide an accurate qualitative characterization of the size, nature and contents of the literature.

Growth and Character of the Literature

Diffusion

The magnitude and rate of growth of the health care CBA/CEA literature are indicated in Figure 1. The lower line plots the annual sum of references clearly identifiable as CBAs and CEAs, while the higher line traces the total of all CBA/CEA-relevant references (i.e., adding to the former general methodology papers, references not clearly identifiable as CBAs or CEAs, etc.). As the data vividly demonstrate, significant interest in health care CBA/CEA is a phenomenon of the 1970s. Prior to 1970, the annual number of health care CBA/CEAs and related publications never exceeded [illegible]; after 1970 the number was never less than 25, and since 1976, the

§ References included in the bibliography were obtained from four sources: 1) computer-assisted literature searches, 2) published professional literature indexes, 3) reference lists of individual articles, papers and books, and 4) communication with leading health services researches. In the first category, two computer-assisted literature searches provided numerous references. MEDLARS covered relevant citations from *Index Medicus*. For the years 1966 through 1975, this search covered the subject heading "Cost and Cost Analysis" (which until 1976 included CBA and CEA). From 1976 to the present, the search was limited to the *Index Medicus* subject heading "*Cost Benefit Analysis*," a heading introduced in 1976 which includes both CBA and CEA. The second computer assisted search was conducted by the National Health Planning Information Center, using the key words "cost benefit analysis" and "cost effectiveness analysis." For the months post-dating the MEDLARS search *Index Medicus* was consulted directly. Published indexes also supplied many relevant economic studies not included in *Index Medicus*. Beginning with the 1966 editions, two indexes of economic literature were searched: the *Index of Economic Articles* and the *Journal of Economic Literature* [illegible] prior to 1969 entitled the *Journal of Economic Abstracts*.

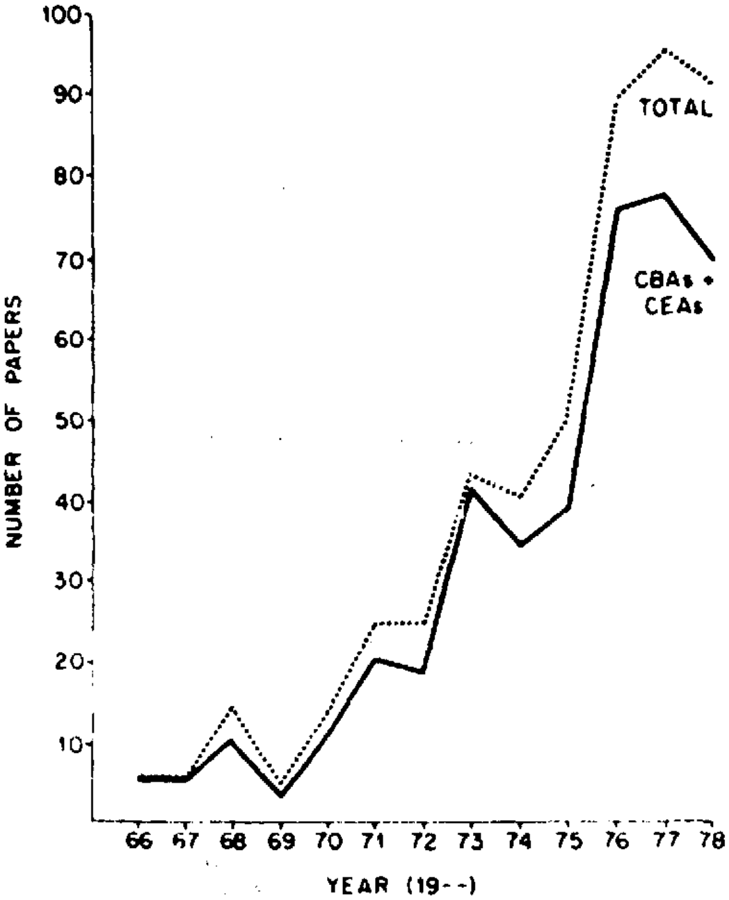


FIG. 1. Growth of health care CBA/CEA literature.

total has always exceeded 88. (See also Table 1.)•

Will the health-care CBA/CEA literature continue to grow? The number of contributions dropped in 1978 and preliminary analysis of 1979 data suggests that the number did not rebound significantly the next year. However several health care "environmental" influences in the early

The general proliferation of professional journals might be expected to result in increased numbers of publications on many subjects, without representing a genuine increase in relative interest in the subject. To provide perspective, one can compare the growth in the health care CBA/CEA literature with that of the overall number of citations in *Index Medicus*. Over the entire period studied (through 1977, since 1978 was not yet completed when we acquired our data), *Index Medicus* citations increased from 157 000 to 260 000 articles, an increase of two thirds. By comparison, the CBA/CEA literature grew by a factor of from 11 to 18. Even in very recent years, growth in the latter considerably outpaces that of the overall medical literature. For example, from 1975 to 1977 the number of contributions to the CBA/CEA literature nearly doubled, while *Index Medicus* citations rose less than 10 per cent.

• Table 1 highlights the findings discussed in this section. Breaking the period into the "early" years as those prior to 1974 and "recent" years as 1974-78 represents an arbitrary decision based on our observation of trends. Nevertheless, it is interesting to note that this dividing line or one year earlier seemed appropriate for all of the phenomena of interest.

TABLE 1. Trends in Health Care CBA/CEA, 1966–1973 and 1974–1978*

	1966–1973	1974–1978
1. Average annual number of publications	17.0	73.2
2. Publications in medical journals as % of total journal publications	40.2	62.7
3. CEAs as % of CEAs + CBAs	42.1	53.2
4. % of articles on:		
Prevention	44.7	22.0
Diagnosis	18.8	30.9
Treatment	36.5	47.2
5. % of articles with orientation of:		
Individual	8.3	15.[illegible]
Organization	21.3	10.8
Society	70.4	73.4

* All differences significant at p = 0.05

1980s seem likely to promote renewed growth. Establishment of the National Center for Health Care Technology, with its mandate to assess the safety, efficacy and cost implications of medical technologies, should foster analytical activity. Publicity associated with other governmental efforts should increase awareness and interest.[27,35-37] Of great potential importance, a variety of efforts within the medical community should spark interest and promote analysis.[38-40¶]

And most generally, continued public concern about the high and growing costs of care should itself generate numerous attempts to assess the cost-effectiveness of medical procedures and technologies.

¶ Interest within the medical community in issues pertaining to the costs of health care appears to be great. A recent American College of Physicians survey on the subject drew a 45 per cent response rate from the entire membership. The typical ACP questionnaire produces only a 15 per cent response rate. (Personal communication with Robert Moser, M.D.)

Publication Vehicles

Figure 2 plots the annual numbers of articles in medical and nonmedical journals. The purpose is to examine the proportion of the literature which has been intended primarily for a physician audience and how this proportion has changed over time. While the time paths follow each other closely, the graph shows a shift from a rough parity prior to 1973 to a clear majority of medical journal articles after 1973. In other words, the rate of growth of the medical literature has exceeded that of the nonmedical journal literature, particularly in recent years. This shift is statistically significant (p = 0.05). (See Table 1.)

In categorizing references by publication vehicle, we kept track of a subset of medical articles, namely those published in the *New England Journal of Medicine*. We isolated these articles because of the *Journal's* general position of leadership in the medical literature and because several of the best and most influential health care CBA/CEAs have been published in the *Journal*. It is interesting to observe that before 1975, the number of CBA/CEA-relevant contributions in the *Journal* exceeded one only once (in 1968, when two contributions were identified). The *Journal* published seven relevant articles in 1975 and four or more each year since.

Mix of CBAs and CEAs

Prior to the most recent years, the annual number of CBAs generally exceeded the number of CEAs. Since the mid-1970s the reverse has been true. The mix in the two periods is statistically significantly different. (See Table 1.) This provides support for Weinstein's conclusion that CEA "has been gaining in acceptance relative to benefit-cost." The reason however, is not obvious. Weinstein attributes the shift to "the conceptual limitations of the human capital approach and the empirical barriers to the willingness-to-pay approach" to

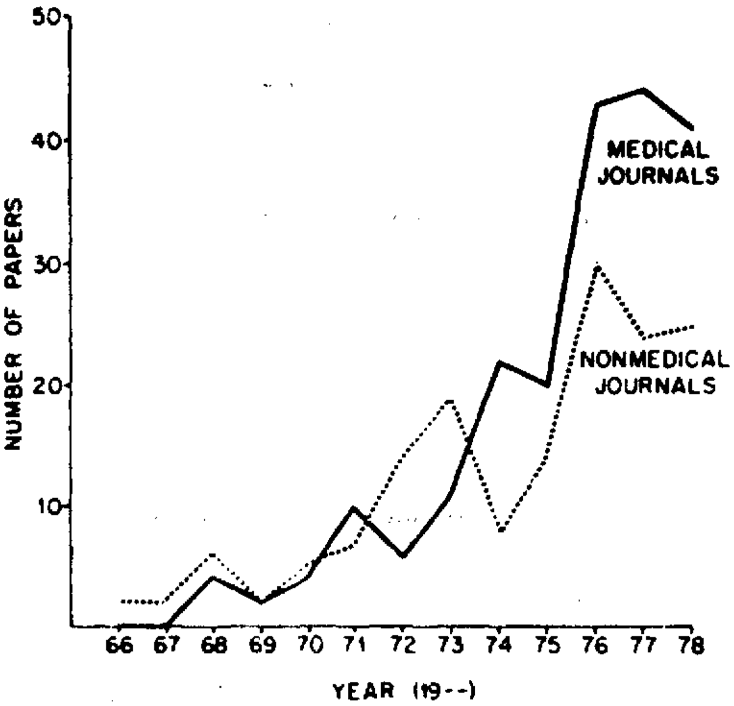


FIG. 2. Diffusion of CBA/CEAs by type of journal

valuing benefits in CBAs.[illegible] Complementary or alternative explanations relate to the apparent relative conceptual simplicity of CEA: analysts use CEA because it is easier for the economic layperson—e.g., the physician—to understand; also, the recent relative growth in the literature in medical journals appears to include relatively more contributions by physicians who, as economic laypersons, may find CEA easier to perform than CBA.**

Medical Function

Of three broad categories of medical functions, prevention and diagnosis each account for slightly more than a quarter of classifiable studies in the bibliography, while treatment accounts for just under half.†† However, if one divides the period into the years preceding 1974 and 1974 through 1978, there is a significant shift in the relative mix away from prevention and toward diagnosis and treatment. During the more recent 5 years, the numbers of both diagnosis- and treatment-oriented papers have exceeded the pre-1974 totals by a factor of 4 or 5. By contrast, the number of prevention-oriented contributions is only 50 per cent greater than that of the earlier period. (See Table 1.)

This shift seems consistent with the relative growth in the medical journal share of the literature, assuming that physicians are relatively more interested in diagnosis and

** The observation that physicians are making relatively more contributions to the literature in recent years is based on our impression. We made no attempt to categorize authors formally by degree or profession.

†† In his recent review, Weinstein[illegible] observed, "Diagnostic procedures, apart from screening tests, have received little attention." Our attribution of more than a quarter of the codable literature to diagnosis is not necessarily at variance with this observation, since we include many screening programs in the diagnosis category.

treatment, compared with prevention, than are nonphysician health professionals (including both providers and health services researchers). Also, consistent with the principal early non-health care applications of CBA/CEA, early health care CBA/CEAs concentrated relatively more on health care "public goods," including especially communicable disease control, rather than individual patient care, a growing concern today. Several excellent communicable disease prevention studies are found in the recent medical literature,[41,42] but this is one of the few substantive areas in which the number of pre-1974 papers actually exceeded the number from 1974 through 1978. (See the next section.)

The permanence of the shift away from prevention is problematic. The widespread perception that "technology" is a major villain in medical cost inflation, combined with the general medical orientation toward diagnosis and treatment, has contributed to growing interest in the CBA/CEA literature on diagnostic and treatment technology and in individual physician decision making concerning the use of such technology. These interests should be sustained in the near future. However, the new federal government emphasis on prevention,[43] increasing public acceptance of the idea of health promotion, and the conscious linking of prevention to cost containment[28] should promote renewed interest in prevention-oriented CBA/CEA.

In addition to examining the mix of contributions oriented toward prevention, diagnosis and treatment, we investigated the mix of types of treatment. Half of all treatment papers were concerned with curative treatments, with the remaining 50 per cent divided roughly equally between rehabilitation and maintenance. Perhaps reflecting the inherent subjectivity and difficulty of quantifying "pain relief," "comfort," etc., the literature included not a single contribution which we could identify as dealing with palliation. The relative mix of treatment functions does not appear to have changed significantly in recent years.

Physical Nature of Subjects of Study

In attempting to categorize subjects by their physical nature, we found ourselves incapable of definitively assessing the vast majority as either technique, drug, procedure, equipment, personnel or system. Most seemed to represent a mix of two or more categories; consequently we included them in a "miscellaneous" category.‡‡ Even some which we did manage to categorize left us with a feeling of discomfort. A study of the cost-effectiveness of computerized tomographic (CT) scanning appears on the surface to belong under "equipment" (where we did categorize it); yet that same study emphasizes the important role—and cost—of the new technicians needed to operate the scanner.

Our principal impression is that the literature covers a broad spectrum of types of programs and technologies, with procedures being the best represented category. In recent years, there appears to have been distinct growth in the attention devoted to equipment-embodied technologies, though that growth has been heavily concentrated on one such technology: CT scanning. (See the next section.)

Decision Orientation

The original intent of CBA/CEA was to assist in social decision making, to identify and value program costs and benefits from a societal perspective. Businesses and individuals have long employed the logic of CBA/CEA to grapple with decision problems, but the CBA/CEA label seems to be applied with increasing frequency to

‡‡ [illegible] year did we manage to categorize more than 10 per cent of the references in other than "miscellaneous," with 20 percent being more typical.

analyses whose decision-assisting perspective is narrower than that of "society."

Our data suggest that the social perspective has dominated the literature over the entire period studied, accounting for roughly 70 per cent of all publications in both the early and more recent years. However, articles oriented toward individual (e.g., practitioner) decision making have increased most rapidly in recent years. Comparing the pre-1974 period with the years 1974 through 1978, one observes a near doubling of the share of papers oriented toward the individual perspective. This growth has come at the expense of papers with an organizational orientation. While the two categories together account for fewer than 30 per cent of the literature contributions, the shift is statistically significant. (See Table 1.)

Substantive Topics and Areas of Interest in Health Care CBA/CEA§§

The health care CBA/CEA literature covers a vast array of disease problems and, as indicated in the previous section, all of the major medical functions. Yet despite the diversity, a few subjects and concerns account for a large share of the literature.

The single disease class which has captured the most attention in the literature is also the nation's number one killer: more than two dozen papers in the bibliography concern cardiovascular disease, while an additional 16 citations relate to hypertension screening and treatment. Other major disease problems have also received considerable attention. Cancer screening programs have been the subject of more than 20 papers, including 9 on breast cancer screening; cancer treatment by itself has not received attention. Eighteen papers have addressed mental illness problems and programs, with the same number of citations relating to dental care. Together, drug abuse and alcoholism account for a similar number of references. Renal disease has received an amount of attention (almost 20 papers) disproportionate to its health importance but reflective of the political and economic importance associated with public funding of dialysis. The federal government's mid-1960s interest in disease control programs, and in kidney disease in particular, made this the only disease problem to have more than one citation before 1969.

Two general classes of health problems have captured considerable attention. A variety of communicable diseases have been the subject of more than two dozen papers. Since communicable disease and its treatment have distinct "public goods" characteristics,‖ this is a logical subject of CBA/CEA, and it is not too surprising to find that half of all the communicable disease papers date from before 1974. By contrast, the second class of problems—the prevention of birth defects—has been studied much more in recent years, with only 2 of 15 papers predating 1975. Several birth defect disease problems have received isolated attention, but at least one—phenylketonuria—has been the subject of three studies.

With one exception (rheumatic fever), we found no other instances of multiple studies focused on a specific disease problem. In some cases, this seemed odd. For example, one might expect a fair amount of CBA/CEA interest in diabetes, yet we discovered only one directly relevant reference.

§§ The literature discussed in this section includes roughly 400 references. Rather than cite them here the authors have prepared lists for each topic mentioned in the section. The lists are available from the authors on request.

‖ The technical economic problem is one of significant externalities: these diseases are public health problems because of their communicability, and their prevention (for example through immunization) confers benefits not only on the immediate recipient of the prevention measure. Thus society has an interest in immunizing individuals which goes beyond the individuals' own private interest.

Several disease problems emerge in the guise of surgeries intended to treat them. Each of the following surgeries is the focal point of at least one reference in the bibliography: radical cystectomy, tonsillectomy, cholecystectomy, herniorrhaphy, appendectomy, synovectomy, joint replacement and hysterectomy. In addition, there is an equal number of papers relevant to surgery and CBA/CEA, but not identifiable with a specific surgery. Many of the surgery papers were contributions to a recent book on the subject.[44]

We classified some two dozen papers as nonspecific screening and prevention. Some of these related to particular activities (e.g., multiphasic screening), while others discussed CBA/CEA issues more generally. A few represented attempts to cover several separate activities.

In recent years, a great deal of policy discussion and regulatory activity has concentrated on the adoption and diffusion of expensive sophisticated capital equipment. Thus, it was with considerable interest that we explored whether such equipment has been the focal point of numerous CBA/CEAs. With one exception, the answer is a striking no. The exception, the CT scanner, was the most talked about medical technology of the 1970s, and both the quantity and nature of the general interest are reflected in the CBA/CEA literature on CT. Some 17 citations are on this technology, all but two of them published in 1977 and 1978.

Will other equipment-embodied technologies emerge as the subject of much attention in the literature? As controversy on specific technologies grows, particularly related to their cost implications, additional CBA/CEA papers can be anticipated. Electronic fetal monitoring is an example of one such technology which has already been the subject of a few papers. The work of the National Center for Health Care Technology, combined with general interest and concern, should increase the proportion of CBA/CEA literature focusing on equipment-embodied technologies.

A variety of services accounted for a significant proportion of articles. Some of these services have relatively tangible outcomes and hence are good candidates for CBA/CEA; for example, we found half a dozen studies of pharmaceutical services. When one abstracts from basic moral positions, some studies related to reproductive health lend themselves to reasonably objective analysis. Moral overtones, however, may dominate the potential for objective analysis in decisions on appropriate topics for CBA/CEA. An obvious example, abortion, was the subject of only one paper during the period covered. The literature review also yielded several articles on each of family planning and maternal and child health programs.

Other services address social needs which are extremely difficult to quantify in a meaningful fashion. In general, one would not expect such services to receive a great deal of attention in a literature which places a premium on quantification and measurement. Exceptions most likely would reflect an unusual policy or social importance. We noted previously that there are 18 CBA/CEA-relevant articles in the mental health area. Similarly, we found a dozen papers on geriatric services and an additional four papers on institutional versus home care, with the patient type not indicated. Given current problems and anticipated growth in the elderly population, continued interest in this subject matter would not be surprising. Another area of considerable current interest, occupational health and rehabilitation, includes both tractable and difficult-to-measure outcomes. We identified nine literature contributions on relevant topics.

Program services do not constitute the only area in which social importance recommends analysis while quantification problems limit it. Manpower programs of lustrate another instance in which technical innovations—often, in this case, substitution of one type of personnel for another—produce outcomes which are difficult to quantify usefully. Nevertheless, analysts have made a dozen contributions on this subject.

We noted previously a couple of conspicuous absences from the literature: certain major disease problems have received very limited CBA/CEA attention, diabetes and cancer therapies serving as good examples; and CBA/CEAs specific to sophisticated capital equipment are less common than policy interest might lead one to suspect, though, again, their frequency may increase due to both growth in interest and support of analysis.

Related to dearth of equipment-specific studies, apart from screening procedures relatively few diagnostic procedures have been the subject of CBA/CEA attention. A few procedures have received isolated discussion but only radiology has received [illegible] attention. Weinstein[22] has identified the evaluation of diagnostic procedures as deserving of CBA/CEA efforts. His plea is supported by the growing body of literature that indicts the increasing use of diagnostic tests as a major source of medical cost inflation.[45] The evidence suggests that everyday, mundane tests are at least as significant contributors to that inflation as the more sophisticated and expensive technologies,[46,47] yet the former have received very little CBA/CEA attention. Again, problems of measuring and valuing the outcomes of diagnostic procedures stand in the way of ready application of CBA/CEA.[48,49]

In closing this section, we note two other areas which seem underrepresented in the literature. For the last several decades, drugs have epitomized the scientific growth of medicine and dramatically altered the practice and outcomes of health care. Drugs have been the subject of hundreds of biochemical and medical studies and, within the social sciences, numerous analyses of medical technical change. Yet aside from implicit and tangential interest in them (e.g., as a component of hypertension management), drugs have not often concerned CBA/CEA analysts.

Finally, the literature reveals very little evidence of attempts to compare the costs and benefits (effectiveness) of specific medical and nonmedical interventions to deal with health problems. While our bibliography search focused on medical approaches, one might have anticipated identification of a few studies which cross the medical–nonmedical border. Yet with the exception of the early HEW efforts,[6,7] the principal mechanism for crossing that border is the reader's location and comparison of separate, independent studies. For both conceptual and empirical reasons, this is not a highly rewarding analytical strategy. Conceivably, heightened awareness of prevention alternatives will motivate formal efforts to grapple with medical–nonmedical comparisons in the future.

Conclusion

Reflecting concern with the high and rising costs of personal health services, the health care CBA/CEA literature has grown rapidly in recent years. The social milieu seems likely to encourage sustained interest in controlling the costs of care; with that interest should come continuing growth in analysts' efforts to grapple with issues of cost effectiveness.

The relative growth in medical journals' share of the literature reflects increasing professional concern about the spiraling costs of health services: medical professionals are expressing their interest in the issue and nonmedical analysts are exhibiting a desire to work with physicians to address it. Concentration of health care CBA/CEAs in the medical literature has had a number of ramifications: studies with a diagnosis or treatment focus—the principal orientations of physicians—have increased much more rapidly than those with a prevention theme; consistent with the increasing emphasis on personal health services, analyses oriented toward individual practitioner decision making appear to be growing more rapidly than those with an orientation toward social or organizational decision making; CEA, a more readily comprehensible technique, has come to dominate CBA.

Some of the recent trends seem likely to persist in the near future, while others are more problematic. We expect continued growth in the medical journals' share of the literature, through increasing interest in CBA/CEA among nonphysician health professionals might restrict such relative growth in the physician-oriented literature. In any case, we would anticipate intensification of the trend toward concern with individual practitioner decision making and increasing use of CEA relative to CBA. The growth in diagnosis and treatment orientation may or may not be sustained. Interest in the cost of specific capital-intensive medical technologies should promote a diagnosis and treatment focus, but the emerging concern for health promotion and disease prevention could reverse the relative decline in CBA/CEA interest in prevention.

Growth in the size of the literature should not be interpreted as implying commensurate maturation in the technical quality of published analyses. Indeed, two opposite quality-related trends are detectable in the literature of the 1970s. On the positive side, a number of well-trained analysts, including several physicians, have been grappling seriously with difficult CBA/CEA problems in the context of specific health care applications. As a consequence, methodological progress is detectable on several fronts; examples include the definition and quantification of multi-attribute effectiveness measures, incorporation of indirect economic costs and benefits into increasingly comprehensive CEAs, and, substantively, dealing with the unique problems of evaluating diagnostic technologies.[48,49] The result has been a conceptual and substantive enrichment of the literature, and evolution in the nature and sophistication of remaining methodological issues.[22]

On the negative side, the rapid growth of the literature has come about, in part, because new practitioners of CBA/CEA include many individuals with no training in analysis and only a rudimentary appreciation of its subtleties. Their enthusiasm is admirable, but it has not been matched by adequate skill in application. As a result, our impression is that the literature has included a higher proportion of technically low-quality analyses in recent years than in the early years of the period we studied. As CBA/CEA diffuses to nonphysician health care professionals, this tendency can be expected to persist.

In closing, we caution that neither the size of the literature nor its quality need correlate highly with its impact on medical practice or health policy. Technically poor studies may stimulate useful cost consciousness in those responsible for delivering health services. By contrast, superior analyses may achieve little impact as a constellation of factors—economic incentives, political concerns, etc.—may dominate considerations of social cost-effectiveness.[1]

Acknowledgment

Helpful comments on earlier drafts of this article were provided by Clyde Behney and Bryan Luce of the Office of Technology Assessment and by three anonymous referees. We are also grateful to the members of the Office of Technology Assessment's Advisory Panel on Cost Effectiveness for guidance and suggestions.

References

1. Warner K, Luce B. Cost-benefit and cost-effectiveness analysis in health care: principles, prac-

tice, and potential. Ann Arbor: Health Administration Press (forthcoming).

2. Congress of the United States, Office of Technology Assessment. Assessing the efficacy and safety of medical technologies. Washington: U.S. Government Printing Office, 1978.

3. Gramlich E. Unpublished manuscript on cost-benefit analysis. Ann Arbor: University of Michigan, 1979.

4. Crystal R, Brewster A. Cost/benefit and cost/effectiveness analyses in the health field: an introduction. Inquiry 1966;3:3.

5. Klarman H. Present status of cost-benefit analysis in the health field. Am J Public Health 1967;57:1948.

6. Grosse R. Problems of resource allocation in health. In: Haveman R, Margolis J, eds. Public expenditures and policy analysis. Chicago: Markham, 1970:518–48.

7. Grosse R. Cost-benefit analysis of health services. Annals of the American Academy of Political and Social Science 1972;399:89.

8. Cochrane A. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust, 1972.

9. Klarman H. Application of cost-benefit analysis to the health services and the special case of technologic innovation. Int J Health Serv 1974;4:325.

10. Klarman H. Application of cost-benefit analysis to health [illegible] technology. J Occup Med 1974;16:172.

11. Dunlop D. Benefit-cost analysis: a review of its applicability in policy analysis for delivering health services. Soc Sci Med 1975;9:133.

12. McNeil B, Keeler E, Adelstein S. Primer on certain elements of medical decision making. N Engl J Med 1975;293:211.

13. McNeil B, et al. Cost effectiveness in hypertensive renovascular disease. N Engl J Med 1975; 293:216.

14. McNeil B, Adelstein S. Value of case finding hypertensive renovascular disease. N Engl J Med 1975;293:221.

15. Neuhauser D, Lewicki A. What do we gain from the sixth stool guaiac? N Engl J Med 1975;293:226.

16. Pauker S, Kassirer J. Therapeutic decision making: a cost-benefit analysis. N Engl J Med 1975; 293:229.

17. Weinstein M, Stason W. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med 1977;296:716.

18. Stason W, Weinstein M. Allocation of resources to manage hypertension. N Engl J Med 1977;296:732.

19. Fein R. But on the other hand: high blood pressure, economics, and equity. N Engl J Med 1977;296:751.

20. Weinstein M, Stason W. Hypertension: a policy perspective. Cambridge: Harvard University Press, 1976.

21. Harris C. Introduction to cost-benefit analysis applied to new health technologies. Prepared for Bureau of Health Planning and Resources Development, HRA. Cambridge, Mass.: Arthur D Little, Inc., 1977.

22. Weinstein M. Economic evaluation of medical procedures and technologies: progress, problems and prospects. In: Wagner J, ed. Medical technology. Washington: U.S. Govt. Printing Office, 1979:52–68. (DHEW publication no. (PHS) 79-3254.)

23. Dittman D, Smith K. Consideration of benefits and costs: conceptual framework for the health planner. Health Care Management Review 1979;4:45.

24. Shepard D, Thompson M. First principles of cost-effectiveness analysis in health. Public Health Rep 1979;94:535.

25. Scheffler RM, Paringer L. A review of the economic evidence on prevention. Med Care 1980;18:473.

26. Warner K. The economic implications of preventive health care. Soc Sci Med 1979;13C:227

27. Congress of the United States, Office of Technology Assessment. The implications of cost-effectiveness analysis of medical technology. Washington: U.S. Government Printing Office, 1980.

28. Rice D. Estimating the cost of illness. U.S. DHEW, PHS. Health Economics Series No. 6, 1966.

29. Cooper B, Rice D. The economic cost of illness revisited. Soc Secur Bull 1976;39:21.

30. Acton J. Measuring the monetary value of lifesaving programs. Law and Contemporary Problems 1976;40:46.

31. Berry R, Boland J. The economic cost of alcohol abuse. New York: Free Press, 1977.

32. Klarman H. Syphilis control programs. In: Dorfman R, ed. Measuring benefits of government investments. Washington: Brookings Institution, 1965:367-410.

33. Bennett W. Cost-benefit ratio of pretransplant bilateral nephrectomy. JAMA 1976;235:1703.

34. Crabtree M. Application of cost-benefit analysis to clinical nursing practice: a comparison of individual and group preoperative teaching. J Nurs Ad 1978;8:11.

35. Altman S, Blendon R, eds. Medical technology: the culprit behind health care costs? Proceedings of the 1977 Sun Valley Forum on National Health. Washington, D.C.: U.S. Government Printing Office, 1979. (DHEW publication no. (PHS) 79-3216.)

36. Committee on Technology and Health Care. Medical technology and the health care system. Washington: National Academy of Sciences, 1979.

37. Wagner J, ed. Medical technology. Proceedings of the Urban Institute Conference, West Palm Beach, Florida, December 10-12, 1978. Washington, D.C.: U.S. Govt. Printing Office, 1979. (NCHSR Research Proceedings Series; DHEW publication no. (PHS) 79-3254.)

38. American Medical Association. Report of the National Commission on the Cost of Medical Care. Chicago: American Medical Association, 1978.

39. American Association of Medical Colleges. Cost containment education efforts—United States medical schools. Washington, D.C.: American Association of Medical Colleges, 1978.

40. Kridel R, Winston D, eds. Cost-effective medical care. Chicago: Resident Physicians Section of the American Medical Association, 1978.

41. Schoenbaum S, et al. Benefit-cost analysis of rubella vaccination policy. N Engl J Med 1976, 294:306.

42. Schoenbaum S, McNeil B, Kavet J. The swine-influenza decision. N Engl J Med 1976,295:759.

43. Healthy people: the Surgeon General's report on health promotion and disease prevention. Washington, D.C.: U.S. Government Printing Office, 1979. (DHEW publication no. (PHS) 79-55071.)

44. Bunker J, Barnes B, Mosteller F, eds. Costs, risks and benefits of surgery. New York: Oxford University Press, 1977.

45. Scitovsky A, McCall N. Changes in the costs of treatment of selected illnesses, 1951-1964-1971. Washington, D.C.: National Center for Health Services Research, 1976. (U.S. DHEW publication no. (HRA) 77-3161.)

46. Fineberg H. Clinical chemistries, the high cost of low-cost diagnostic tests. In: Altman S, Blendon R, eds. Medical technology: the culprit behind health care costs? Washington, D.C.: U.S. Government Printing Office, 1979:144-165. (DHEW publication no. (PHS) 79-3216.)

47. Redisch M. Hospital inflationary mechanisms. Presented at the meeting of the Western Economic Association, Las Vegas, June 10-12, 1974.

48. McNeil B. Pitfalls in and requirements for evaluations of diagnostic technologies. In: Wagner J, ed. Medical technology. Washington, D.C.: U.S. Government Printing Office, 1979:33-39. (DHEW publication no. (PHS) 79-3254.)

49. Wagner J. The feasibility of economic evaluation of diagnostic procedures: the case of CT scanning. Prepared for the Office of Technology Assessment, Congress of the United States, under Contract No. 933-2040, 1979.

23

Morgan N. Jackson
James P. LoGerfo
Paula Diehr
Carolyn A. Watts
William Richardson

Elective Hysterectomy: A Cost-Benefit Analysis

Hysterectomy is one of the most frequently performed major surgical procedures in the United States today.[1] The fact that incidence rates vary considerably between England and the United States,[2] and vary even within the same geographic area of the U.S.,[3] has caused it to be cited as an example of a procedure which is often performed unnecessarily. Although the surgery obviates a woman's concern about the risks of an unplanned pregnancy or cancer of the cervix or uterus and frees her from menstruation, it may expose her unnecessarily to the risks of mortality and morbidity from the surgery or its complications. This study was undertaken to analyze the direct costs and direct benefits of elective simple hysterectomy. Costs of alternative methods of contraception and their complications were not included in the analysis.

Rice[4] has investigated the determination of indirect costs of illness. However, an analysis of elective hysterectomy using direct cost, i.e., charges paid for inpatient and ambulatory services, has not been reported. It is assumed that such a report would provide information useful in making the decision as to whether or not the procedure should be discouraged. Analyzing only direct costs avoids the controversy surrounding an appropriate valuation of indirect costs. Such a strategy was chosen on the assumption that it would maximize the direct benefits of having an elective hysterectomy. We assume that if no benefits were shown using only direct costs, further analysis would not change the conclusions. In addition, the direct costs (amounts paid for surgical procedures and hospital care) or direct benefits (measured as foregone future amounts paid for treatment of uterine diseases) would have the greatest impact on public policy. Since much of the inpatient and ambulatory costs of medical care are paid through third parties,[5] they and, more generally, their enrolled groups would be most likely to reap potential benefit from the surgery were a net benefit to obtain. Conversely, if the analysis determined that the procedure incurred a net cost, they would be paying that cost.

Several recent studies have examined the question of whether hysterectomies should be performed when the usual medical indications are not present. Usual medical indications include cancer of the cervix or uterus, symptomatic non-malignant uterine conditions, uterine involvement with malignant or non-malignant disease of contiguous organs. For a complete list, refer to the discussion by Donohue.[6] Bunker et al.[7] analyzed the situation of a 40-year-old woman who had mild symptoms of uterine disease in the absence of life-threatening uterine pathology. They considered such a woman to have marginal indications for surgery and calculated that her life expectancy would be increased by 14 days if a hysterectomy were performed. In the absence of data on the number of women who think they have benefited from the surgery, the authors tentatively concluded that "elective hysterectomy for the relief of symptoms in a healthy 40-

Morgan Jackson, M.D., M.P.H., a former Robert Wood Johnson Clinical Scholar at the University of Washington, currently practices medicine in Gillette, WY 82716. **James P. LoGerfo, M.D., M.P.H.**, is Assistant Professor of Health Services and Assistant Professor of Medicine. **Paula Diehr, Ph.D.**, is Assistant Professor of Biostatistics. **Carolyn A. Watts, Ph.D.**, is Assistant Professor of Health Services and Assistant Professor of Economics. **William Richardson, Ph.D.**, is Professor of Health Services. All are with the School of Public Health and Community Medicine, University of Washington, Seattle 98195.

Inquiry Volume XV, September 1978

year-old woman is justified on the basis of known risks and benefits." Deane and Ulene[8] compared hysterectomy and tubal ligation as alternative sterilization strategies in young women who desired permanent contraception. Using decision analysis and cost-benefit analysis, they found that at interest rates above 5%, tubal ligation would be the preferred strategy for all women over 20 years old, since it would be less costly in both mortality and dollar dimensions. Cole and Berlin[9] examined the benefits and costs of elective hysterectomy for women at age 35 years. Assuming a 0.06% operative mortality rate and using *theoretical* cost projections, they concluded that the operation would increase life expectancy by 0.2 years and that the cost of performing the operation on one million women would exceed the benefit by $570 million when discounted to present value. They concluded that cancer prophylaxis was not adequate justification for elective hysterectomy.

Other studies have documented the adverse psychological sequelae[10-12] that may befall women undergoing hysterectomy; Barker[13] suggested that these were more likely to be found in a woman in whom the procedure had been elective. The consensus of current literature concerning intangible considerations favors the notion that the costs to a woman of having an elective hysterectomy outweigh the benefits.

Materials and Methods

Data for the study were obtained from the Seattle Prepaid Health Care Project,[14] a four-year social experiment designed to evaluate the effect of provider financing and organization on patient health status and utilization of medical services. Near-poor families residing in the Model Cities target area of Seattle who met specific income requirements were eligible for the Project. The charge data (amounts paid to providers) were collected on patients enrolled in a very comprehensive prepayment plan that fully covered all medical services provided by members of a Physicians' Service Bureau (King County Medical/Blue Shield) and community hospitals. The providers included all community hospitals, pharmacies, and virtually all non-federal, non-group health cooperative physicians in King County, Washington. There were no deductibles or other copayments. For the services considered in this paper, and within the time period examined, information on amounts paid to providers for inpatient and ambulatory medical services is essentially complete.

The data base consisted of all paid charges that occurred due to the specified uterine-related gynecologic diseases listed in Table 1. These data were collected by selecting information using three-digit diagnosis codes from the *Eighth Revision, International Classification of Diseases*.[15] Some overlap occurred in the diagnoses between conditions which could be attributed to an intact uterus and those which would not occur after hysterectomy. The categories included in the list are those where virtually all of the cases could be assumed to be attributable to the presence of an intact uterus. If most of the diagnoses in a category could occur in a woman posthysterectomy, this category was excluded. In this way, the charge data served as an estimate of the total cost of gynecologic services consumed because of uterine-related disease. In women who have had a hysterectomy, charges for none of these diagnoses should be incurred, and those foregone charges therefore would be considered tangible benefits of surgery.

Table 1. Uterine-related diagnoses used for data selection[15]

	H-ICDA code
Gonococcal infection*	098
Malignant neoplasms of the cervix	180
Malignant neoplasms of the uterus	182
Uterine fibroma	218
Other benign neoplasms of the uterus	219
Neoplasms unspecified of the uterus	234
Iron deficiency anemia**	280
Pelvic inflammatory disease	616
Infective disease of the cervix	620
Other diseases of the cervix	621
Uterine prolapse	623
Malposition of the uterus	624
Other diseases of the uterus	625
Disorders of menstruation	626

Notes

* Gonococcal PID or gonococcal cervicitis is exceedingly rare in women posthysterectomy (personal communication with Dr. King Holmes).

** Assumed to be due to menometrorrhagia for premenopausal women.

Table 2. Outpatient cost data

Age	Period 1 costs (6/71–6/73)		Period 2 costs (7/73–2/75)	Total outpatient costs	Patient years of exposure	Average annual cost per woman (in rounded dollars)
	1972 dollars	Adjusted to 1974 dollars				
30–34	$7,363	$8,320	$4,496	$12,816	349	$37
35–39	4,098	4,631	5,226	9,857	252	39
40–44	2,276	2,572	4,807	7,379	200	37
45–49	6,194	6,999	3,187	10,186	230	44
50–54	4,345	4,910	2,631	7,541	191	39
55–59	1,629	1,841	1,222	3,063	139	22
60–65	707	799	696	1,495	160	9

Within the age range of 30–65 years (the upper age limit of the Prepaid Health Care Project), 1,129 women were enrolled with King County Medical/Blue Shield under the Prepaid Health Care Project plan. Their enrollment ranged from 1 to 48 months, with a mean enrollment of 23.8 months (s.d. = 12.6). The charge data were used to compute the annualized age-specific average expenditure per woman for the selected diagnoses. The effect of not having charges for women over 65 is quite small because future foreborne costs are discounted and the present value of costs incurred after age 65 is expected to be quite small. Since there is a significant prevalence of hysterectomy in the general population, data from the professional activity survey[16] were used to adjust the annualized figure for women who had intact wombs.

The charge data were collected over a period from July 1971 through January 1975. Because of the substantial change in the medical component of the Consumer Price Index, all charges were adjusted to 1974 dollars.[17]

Data on charges for hysterectomies were collected by assessing patient files to obtain hospital charges and by selecting outpatient hysterectomy-specific charges from surgeons, anesthesiologists, and consultants. Since these charges were the only data entered in the cost side of the analysis, they represent a minimum cost to the individual or to third parties for the procedure because they do not include costs or increased utilization due to indirect effects of the surgery.

Several assumptions were made to facilitate the analysis: 1) The women undergoing hysterectomy are 30 years old and have completed childbearing. This age was chosen as the earliest reasonable age at which to perform the operation frequently. Although the age seems young, as of 1970 only 17% of women in the U.S. had planned pregnancies after that age.[18] Eighty-three percent had completed their planned families, and therefore could be assumed to be receptive to contraception. 2) The women desire, and obtain, effective contraception. It was assumed that the women

Table 3. Inpatient cost data

Age	Period 1 costs		Period 2 costs	Total inpatient costs	Patient years of exposure	Average annual inpatient costs per woman
	1972 dollars	Adjusted to 1974 dollars				
30–34	$6,157	$7,142	$6,790	$13,932	349	$40
35–39	8,673	10,061	4,787	14,848	252	59
40–44	3,050	3,538	6,019	9,557	200	48
45–49	13,903	16,127	4,111	20,238	230	88
50–54	2,992	3,471	1,308	4,779	191	25
55–59	3,924	4,552	283	4,835	139	35
60–65	1,887	2,189	1,778	3,967	160	25

Inquiry/Volume XV, September 1978

Table 4. Exposure and expenditure data from the Prepaid Health Care Project

Age	Calculated total woman years of exposure	Annualized average expected expenditure per woman (in rounded dollars) for uterine-related diagnoses: Outpatient costs	Inpatient costs	Total costs	Total amount of projected expenditures not discounted	PV of total cost for each age interval at 6.5% interest	PV of total cost for each age interval at 3% interest
30–34	349	$37	$40	$ 77	$ 385	$ 341	$ 363
35–39	252	39	59	98	490	317	399
40–44	200	37	48	85	425	200	299
45–49	230	44	88	132	660	227	400
50–54	191	39	25	64	320	80	167
55–59	139	22	35	57	285	52	128
60–64	160	9	25	34	170	23	66
	Total projected expenditure (direct costs)				$2,735 = Undiscounted	$1,240 Discounted at 6.5%	$1,822 Discounted at 3%

not choosing hysterectomy obtained contraception adequate to prevent childbirth. Thus, costs incurred due to pregnancy, abortion, or morbidity due to contraception were not included in the calculations. 3) The women undergoing hysterectomy have no uterine pathology and experience no morbidity referable to the uterus. This is the extreme case where there are no medical indications for the operation. If a benefit were found in analyzing an operation performed under these conditions, then less rigid criteria would show greater benefit. 4) The patients are healthy and experience minimum operative mortality. 5) There are no significant complications of the surgery, exclusive of fatality. This assumption would lead to a conservative estimate of the costs that might be realized by the operation.

Once the annualized cost projections were obtained, the stream of expected benefits were discounted to present value using a rate of return of 3%, as well as a 6.5% discount rate to test the sensitivity of the findings to the present value calculations. The rationale for using these two figures is discussed below.

Results

The calculations of age-specific costs for outpatient and inpatient treatment of uterine related disease are set forth in Tables 2 and 3. Table 2 shows that the average annual per capita costs for uterine-related ambulatory care ranged from a low of $9 after age 60, to a high of $44 for women age 45–49. For women between the ages of 30 and 55, the costs were stable at approximately $40 per year. The range of inpatient average annual per capita costs was larger. The peak of $88 occurred for women in the 45–49 age group. Women younger than 45 years old incurred costs that ranged from $40–$59, whereas women who were older than 49 years old incurred costs of $25–$35.

Table 4 sets forth the annualized expected age-specific costs per woman for the specified uterine-related diagnoses. This table also shows the net direct benefit as a result of a hysterectomy to be $2,735, undiscounted. When discounted at a 3% interest rate, the value of the benefit is reduced to $1,822. By discounting at a rate of 6.5%, the present value becomes $1,240.

The mean cost of a hysterectomy in the Prepaid Health Care Project was $1,637 (s.d. = $336, N = 22). Although both present values of the benefits are within one standard deviation of the cost, the importance of the discount rate can be seen by the fact that, at a rate of 3%, there is a small net benefit of $185—whereas with a discount rate of 6.5%, there is a net cost of $397. Recalling that the strategy of the analysis was chosen to maximize the expected direct benefit, relaxation of any of the restrictive assumptions of the analysis would decrease the net benefit (or increase the net cost).

Discussion

The findings of this study suggest that very little, if any, financial benefit would be expected to accrue from the purchase of an elective hysterectomy by or for an asymptomatic woman at 30 years of age. It is necessary, however, to realize the limitations the data may impose on the generalizability of the results. First, calculations were performed using direct costs exclusively. The authors acknowledge that many indirect costs as well as intangible factors would influence a woman's decision regarding hysterectomy. In the absence of uterine-related morbidity, and excluding considerations of contraception, the major intangible benefits would be freedom from fear of cancer of the cervix or uterus. However, the many possible psychological sequelae, as well as the small but existent operative mortality, suggest that from the point of view of intangible considerations, a net cost would be realized. Indirect costs which might be incurred would primarily be in time lost from work as a result of the surgery or its complications. There are also indirect benefits in the form of foregone time lost from work in the future as a result of prevented uterine-related morbidity. If these costs can be assumed to be larger than benefits, then calculation of the direct costs indicates the minimum estimated cost of the procedures.

Clearly, determination of the appropriate discount rate to calculate present value is crucial to the conclusions of the study. Although a net benefit of $1,098 might be expected at a 0% interest rate, use of this undiscounted figure would be misleading. Because the benefits accrue over time, future benefits and costs should be discounted to a standard value for comparison. Since all the costs and benefits in this paper were measured in 1974 dollars, the calculated benefits have been adjusted for inflation. Consequently, the real rate of interest would be the most appropriate figure to discount future benefits. Where the benefits and costs have not been adjusted for inflation (i.e., where the streams are in nominal terms) the nominal (observed) rate of interest is the appropriate discount rate to use if it can be assumed that the expected inflation rate in the sector of the economy in question is the same rate as that which influences the market rate of interest.[11]

A real rate of interest of 3% is the most appropriate rate of interest to use in this paper. The calculation of this real rate follows that discussed in Mandell and is equal to the observed market rate of interest minus the "expected" rate of inflation (which is assumed to be equal to the historical rate). For the time period under consideration, these calculations lead to a real rate of 3% (8% observed interest rate minus 5% historical inflation). The rate of 6.5% was also used to calculate present value. This figure was chosen as a combination of the prevailing loan and/or borrow interest rates for individuals.

Whether a patient had to spend savings or borrow money to obtain a hysterectomy, the fact that this operation would yield an internal rate of return of approximately 4% would make it an unattractive investment on strictly financial grounds. When the indirect and intangible costs of the procedure are added to the deliberation, the total of all three factors would be very likely to result in a net cost. However, since the valuation of the indirect and intangible costs is individual, the decision to have a hysterectomy might be facilitated by this analysis which specifies the direct costs, but certainly would involve many other considerations.

The major difference between this study and that of Cole and Berlin is that their analysis used theoretical estimates of cost as opposed to the observed costs that were used in the present study. Although the methods were different, the fact that the conclusions are similar lends strength to the argument against elective hysterectomy on financial grounds. The conclusions of these papers are not at odds with the analysis of Bunker et al., since they were analyzing the situation of a woman who was symptomatic. Even though final considerations of the direct and indirect costs of the procedure might not be adequate reasons for permitting the procedure, the presence of symptoms might make the intangible benefits of the surgery great enough to outweigh the net tangible cost. The findings of Deane and Ulene (that tubal ligations would be preferable to hysterectomies when sterilization is the goal) also mitigate against recommending a hysterectomy for other than medical indications.

In addition to these financial considerations, mortality calculations—performed by the authors—were found to be in agreement with those of Cole and Berlin. Our analysis showed that 30-year-old women undergoing elective hysterectomy could expect an average increase in life expectancy of 0.32 years, assuming a 0.06% operative mortality. However, half of those additional years of life would be lived after the age of 75, and 78% would be lived after the age of 65. Thus, the procedure would largely benefit elderly women at the expense of the young woman who would not survive the surgery.

Whether a given woman should have a hysterectomy is an individual decision. Whether elective hysterectomies should be financed by third parties is a policy decision which is more likely to be at least partially influenced by direct cost considerations. If such a decision resulted in an increase in the net cost of care, such costs would be borne by the public through higher insurance premiums. If the decision resulted in a decrease in the net costs of medical care, the public might be expected to benefit through decreased premiums. The findings of this study, and the companion analyses, suggest that a policy of permitting elective hysterectomies probably would not be cost-beneficial when direct costs are considered. However, financial criteria are necessary, though not sufficient, for making policy decisions on a societal level.

With the increasing costs of medical services, and the increasing role of government and third parties, limits probably will be set on the amount of money the country will spend on health services, and on the types of services that can be reimbursed. Cost-benefit analyses such as this can provide information which is useful in making such decisions; even so, many other factors will need to be considered, regardless of whether the decision is being made by an individual, an underwriter, or the government.

References and Notes

1 Mead, S. *Surgical Operations in Short Stay Hospitals, United States, 1968* (DHEW Publication No. (HSM) 73-1762) Rockville, Md. Health Service and Mental Health Administrations, National Center for Health Statistics (January 1973).

2 Bunker, J. P. "Surgical Manpower: A Comparison of Operations and Surgeons in the United States and in England and Wales." *New England Journal of Medicine*, 282:135–144 (January 15, 1970).

3 Wennberg, J. and Gittelsohn, A. "Small Area Variations in Health Care Delivery." *Science*, 182:1102–1108 (December 7, 1973).

4 Rice, D. P. "Estimating the Cost of Illness." *American Journal of Public Health*, 57:424–440 (March 1967).

5 Krizay, John and Wilson, Andrew. *The Patient as Consumer: Health Care Financing in the United States*. Lexington, Mass., Lexington Books, 1974, p. 4.

6 Bunker, J. P. et al. "Elective Hysterectomy: Pro and Con." *New England Journal of Medicine*, 295:264–268 (July 29, 1976).

7 Bunker, J. P., Barnes, B. A. and Mosteller, F. *Costs, Risks and Benefits of Surgery*. Oxford University Press, New York, N.Y. 1977. Chapter 17.

8 Deane, R. T. and Ulene, A. "Hysterectomy or Tubal Ligation for Sterilization: A Cost-Effectiveness Analysis." *Inquiry*, 14:73–86 (March 1977).

9 Cole, P. and Berlin, J. "Elective Hysterectomy." *American Journal of Obstetrics and Gynecology* 129(2):117–23 (September 1977).

10 Polivy, J. "Psychological Reactions to Hysterectomy: A Critical Review." *American Journal of Obstetrics and Gynecology*, 118:417–426 (February 1974).

11 Raphael, B. "The Crisis of Hysterectomy." *Australia and New Zealand Journal of Psychiatry*, 6:106–115 (June 1972).

12 Richards, D. H. "Depression after Hysterectomy." *Lancet*, 2:430–433 (August 25, 1973).

13 Barker, M. G. "Psychiatric Illness after Hysterectomy." *British Medical Journal*, 2:91–95 (April 13, 1968).

14 Richardson, W. C. et al. "Seattle Prepaid Health Care Project: Final Report." Available NTIS, Autumn 1978.

15 Eighth Revision International Classification of Diseases, adapted for use in the United States. Department of H.E.W., Washington, D.C.: U.S. Government Printing Office.

16 *Length of Stay in PAS Hospitals, United States, 1972*. Commission on Professional and Hospital Activities, Ann Arbor, Mich., 1973.

17 *Social Security Bulletin*, 39:12, 1976. Table M-42.

18 *Vital Statistics of the United States 1970*. DHEW, Public Health Service, N.C.H.S., Rockville, Md. 1975. Vol. 1, Sec. 1. Table 1-55, pp. 1–73.

19 Mundell, R. A. "Inflation and Real Interest." *Journal of Political Economy*, 71:280–283 (June 1963).

TALLER SOBRE ENSEÑANZA DE LA INVESTIGACION OPERATIVA Y EL ANALISIS DE SISTEMAS EN PROGRAMAS DE ADMINISTRACION DE SALUD

El Taller se celebró en Caracas, Venezuela, del 8 al 12 de marzo de 1982, con los participantes que se mencionan en el Apéndice. Fue una actividad del Programa de la OPS/Fundación W. K. Kellogg de Educación en Administración de Salud y tuvo por objeto tratar sobre la incorporación de las técnicas de la investigación operativa y el análisis de sistemas[1] en los programas regulares de capacitación y en los de educación continua de administradores de sistemas de salud.

Los objetivos específicos del Taller fueron:

- Intercambiar experiencias entre los participantes en relación con la enseñanza de la investigación operativa y el análisis de sistemas aplicados a los problemas de atención de salud.
- Desarrollar un plan básico para la enseñanza en esos campos.
- Seleccionar una bibliografía básica para ser usada en la actualización de educadores y administradores de los servicios de salud así como en el proceso educacional.
- Diseñar un programa básico de educación continua en investigación operativa y análisis de sistemas para la actualización de profesores, investigadores y administradores en el sector salud.

En 1970 no se conocía en América Latina la aplicación de la investigación operativa a los servicios de salud. En 1971 la OPS organizó un simposio sobre el análisis de sistemas en el que se recomendó el empleo prioritario de la investigación operativa a nivel hospitalario.

Ese mismo año se inició un programa de verano en el Instituto Tecnológico de Estudios Superiores de Monterrey, México, en el cual estudiantes de ingeniería industrial y profesores participaron junto con administradores de salud en la definición y solución de problemas hospitalarios. Desde entonces, diferentes modalidades de este programa se han desarrollado en otras universidades de Colombia, Costa Rica, Chile, Perú, México y Brasil.

En la actualidad se tiene conocimiento de unos 11 programas de ingeniería industrial y de sistemas en los cuales estudiantes de dicha especialidad han participado en investigaciones de servicios de salud, y 54 programas que ofrecen capacitación en administración de salud.

[1] El análisis de sistemas y la investigación operativa se deben concebir de modo amplio para que comprendan las técnicas cuantitativas de la ciencia de la administración, cuya aplicación a la solución de problemas reales puede ser de gran utilidad. Dichas técnicas pueden ser de diferente grado de complejidad, desde la ingeniería industrial clásica hasta la programación matemática.

Problemas identificados

Durante el Taller los participantes examinaron los diversos problemas[2] que se observan en la América Latina en la capacitación de administradores de salud en técnicas científicas de gestión. Dichos problemas se resumen seguidamente:

Problemas en el sector salud

- Los médicos, que son los responsables de la toma de decisiones en el sector, por lo general no conocen las técnicas del análisis de sistemas y la investigación operativa. Por consiguiente, las decisiones se toman con base en criterios subjetivos y políticos, sin el beneficio de metodologías cuantitativas y de información pertinente.
- No se cuenta con información adecuada a nivel nacional, regional o local para tomar las decisiones. La solución de problemas en los sistemas de salud exige el desarrollo y la utilización de sistemas de información.
- Los administradores de nivel intermedio no tienen los conocimientos adecuados de estadística descriptiva, análisis de sistemas y técnicas de evaluación requeridos en un enfoque racional del proceso decisorio.
- Algunos administradores desconfían de los profesionales capacitados en técnicas científicas de gestión, lo que dificulta las relaciones interinstitucionales y perjudica el proceso educativo.
- En el proceso de capacitación de administradores en métodos cuantitativos (especialmente de médicos) se corre el riesgo de que estos, con el poco conocimiento adquirido, se consideren "expertos" en la materia.
- En las instituciones de salud se han creado pocas plazas para personas capaces de aplicar las técnicas científicas de gestión de problemas, por lo que en el sector salud las oportunidades de trabajo para profesionales en investigación operativa son limitadas.
- Los especialistas en investigación operativa muestran una tendencia a concentrarse en los problemas pequeños, ignorando otros más importantes que tienen importancia nacional.

Problemas en el sector educativo

- La mayoría de los profesores universitarios no posee la experiencia adecuada en la aplicación del análisis de sistemas y la investigación operativa en el sector salud para motivar y enseñar eficazmente a estudiantes de administración de salud.
- La enseñanza no debe consistir simplemente en disertaciones académicas. Para que sea eficaz, debe mostrar a los administradores de salud cómo aplicar las metodologías y técnicas a la solución de los problemas. Para ello se requiere de una adecuada preparación por parte de los docentes y de ayudas didácticas tales como casos de estudio y de un laboratorio sobre toma de decisiones.
- En el diseño de programas de administración de salud es deseable la integración de las escuelas de ingeniería industrial y de sistemas, administración y medicina. Un problema fundamental de esta integración consiste en la orientación de las escuelas de administración e ingeniería hacia el sector privado o industrial; en general se tiene poca experiencia o poco interés en el sector público.
- La mayoría de los profesores de salud pública en América Latina son médicos y, por lo general, muestran resistencia hacia la participación de otras disciplinas en el proceso de capacitación de personal de salud.

---

[2]Para un ejemplo específico véase: Pérez, C. E. *Improving the Managerial Capability of the Colombian Health System,* trabajo presentado en la reunión conjunta de ORSA/TIMS, Toronto, mayo de 1981.

• Como las universidades latinoamericanas por lo general no poseen los recursos financieros para mantener personal docente de calidad, existe un continuo éxodo de profesores hacia el sector privado y el extranjero. Quizás el problema más serio con que se enfrentan las instituciones educativas en América Latina sea precisamente el de contratar y mantener recursos humanos calificados.

• Existen obstáculos institucionales y políticos que dificultan el establecimiento de relaciones formales entre las escuelas de salud pública, ingeniería y administración. Sin embargo, sin estas relaciones resulta difícil desarrollar programas educativos integrados. Estos obstáculos solo podrán superarse si las autoridades responsables de la administración de las escuelas y de las universidades reconocen este problema y la necesidad de encontrarle solución.

Recomendaciones

En el Taller se formularon una serie de recomendaciones orientadas al desarrollo de programas de educación, tanto continua como regular, para administradores de salud en materia de análisis de sistemas e investigación operativa, así como también de programas para preparar expertos en análisis de sistemas para trabajar en el sector salud. Hubo consenso en que la aplicación exitosa de estas técnicas en la solución de problemas de salud requiere de un enfoque multidisciplinario de los problemas que no pueden ser resueltos por individuos, ya sean administradores o analistas, trabajando independientemente. Por lo tanto, hay que considerar simultáneamente la educación de administradores y expertos. Más aún, para ser más efectivos, los programas educativos deben superponerse, en la mayor medida posible, de modo que los alumnos—futuros administradores y analistas—puedan asistir a los mismos cursos, trabajar juntos en proyectos de grupo y así fomentar el interés y la estima recíproca por la otra disciplina y desarrollar actitudes que faciliten la tarea conjunta en el futuro.

En cada una de las 22 recomendaciones se indican las personas o instituciones a las que incumbe la función principal en su ejecución, usando los siguientes códigos:

OPS: (Organización Panamericana de la Salud). A pesar de sus limitados recursos financieros, la OPS debe seguir desempeñando las importantes funciones de obtener y diseminar información, propiciar la reunión de personas e instituciones para lograr la adopción de las recomendaciones formuladas, y el desarrollo continuo de recursos financieros tanto de fundaciones como de los gobiernos interesados.

IE: (Instituciones educativas). Se refiere a las instituciones educativas en los diferentes países que ofrecerán directamente los programas o que contratarán a otras instituciones para desarrollar los programas de estudios y cursos.

Organismos: Se refiere a organismos nacionales, regionales o locales que proveerán los recursos financieros, los participantes y alumnos para las actividades propuestas; son también fuentes de datos y señalan áreas problema para su estudio.

IO/AS: (Investigadores operacionales y analistas de sistemas). Se refiere a los participantes en el taller y a otros individuos en los países latinoamericanos que participarán en la ejecución de los programas y cursos que se recomiendan.

Las recomendaciones son las siguientes:

Programas educativos[3]

1. Se recomienda desarrollar tres tipos de programas para enseñar investigación operativa y análisis de sistemas a estudiantes de administración de salud y analistas de sistemas en América Latina *(OPS, IE, IO/AS)*:

- *Cursos cortos,* principalmente para administradores y otros profesionales del sector salud. El curso típico durará de 16 a 40 horas y tendrá como objetivo enseñar al estudiante el empleo de metodologías y técnicas particulares en la solución de problemas reales de importancia. Se recomienda desarrollar una serie de estos cursos ya que es ilusorio pensar que un solo curso es suficiente. No se exigirán conocimientos previos de matemática. El curso permitirá al estudiante entender la perspectiva del enfoque de sistemas, apreciar la necesidad de utilizar información en el proceso de toma de decisiones y reconocer que existen individuos capacitados que pueden aplicar las técnicas de la administración científica en la solución de problemas del sector salud. Deberán planificarse cursos cortos parecidos para investigadores operacionales y analistas de sistemas a fin de lograr que su función en el sector salud sea más eficaz.
- *Programa de certificación,* con una duración aproximada de 240 horas, para administradores de salud en ejercicio. El programa ofrecerá al administrador un conocimiento más profundo de las técnicas científicas de gestión. Debe ofrecerse, de preferencia, en la noche y fines de semana, para permitir al administrador/estudiante continuar con su trabajo regular. Como requisito en el campo de las matemáticas solo se deberá exigir conocimientos básicos de algebra.
- *Programa de maestría,* de dos años, para administradores que deseen adquirir un conocimiento más profundo sobre investigación operativa y análisis de sistemas. Como requisito de admisión se deberá exigir conocimientos de cálculo. Se deberá ofrecer un segundo programa a nivel de maestría para preparar individuos con orientación técnica (ingenieros industriales y de sistemas, matemáticos, físicos) en la solución de problemas del sector salud. Con base en la experiencia de los Estados Unidos de América, estos individuos podrán asumir en períodos relativamente cortos puestos administrativos en el sector salud y se convertirán en una fuente secundaria de administradores con formación cuantitativa.

2. Los estudios de investigación operativa y análisis de sistemas dependen de datos. La toma de decisiones óptimas se dificulta si no se posee información válida y confiable. Sin embargo, en la mayoría de los programas de administración de salud se presta muy poca importancia a este aspecto. Se considera que la OPS debería llevar a cabo un seminario regional sobre "datos e informática en sistemas de salud", que podría incluir temas tales como: fuente y acopio de datos, diseño y desarrollo de sistemas de información, procedimientos de entrada, procesamiento, control de calidad, clasificación y análisis de datos, intercambio de información y otros temas afines (*OPS*).

3. En cada país se deberá realizar una serie de talleres que reúnan a las autoridades del sector salud y técnicos competentes en el área de sistemas e investigación operativa. Dichos talleres deben estar orientados a la solución de problemas concretos de la prestación de servicios de salud *(OPS, Organismos, IE)*.

4. Se deberá organizar un taller ulterior para analizar las dificultades que se encuentren en la ejecución y actualización de las recomendaciones aquí for-

[3] La descripción de los programas educativos recomendados aparece en la pág. 12.

muladas. En este taller se discutirían los problemas relacionados con la ejecución de programas y proyectos, prestando especial atención a los obstáculos políticos, culturales y sociales. Deberían participar los asistentes en el presente Taller y otros individuos involucrados en proyectos de investigación operativa y análisis de sistemas a nivel nacional o regional en los países latinoamericanos *(OPS).*

Recursos y material didáctico auxiliar

Antes del Taller, se realizaron dos estudios bibliográficos (español y portugués, e inglés) que sirvieron de referencia para el mismo y que también podrán utilizarse en los cursos que resulten de sus recomendaciones. Los estudios se concentraron, en primer lugar, en artículos y documentos que pudieran servir de base para la forn. 'ación de casos de estudios, útiles en la enseñanza de análisis de sistemas e investigación operativa para administradores de salud, y en segundo, en artículos que ilustraran la problemática de la prestación de servicios de salud.

Se identificaron 60 artículos en español y portugués y se presentó un resumen de 43 de ellos.

El examen de las publicaciones en inglés (adelantado con la ayuda de la computadora) reveló miles de artículos relacionados con el análisis de sistemas y la investigación operativa en el sector salud. Se seleccionaron 116 artículos y 24 libros divididos en ocho subcategorías: evaluación de proyectos, planificación y programación de recursos humanos, distribución de recursos, estimación de demanda, análisis de costo-beneficio, control de inventarios, evaluación de tecnologías y análisis de áreas de capacitación. En la selección se tuvo presente el nivel matemático de los artículos, teniendo en cuenta que se utilizarán para el adiestramiento de administradores, y se eliminaron aquellos artículos puramente teóricos y los que exigían conocimientos de matemática superior. Aunque se insistió principalmente en publicaciones recientes, se incluyeron varios artículos "clásicos" y otros que tratan temas de importancia a nivel de los países.

5. Después de examinar los 116 artículos en inglés, los participantes seleccionaron 38 para traducción al español. Estos artículos servirían como fuente y como base para el desarrollo de casos de estudio a ser usados en los tres tipos de cursos recomendados.[4]

En la serie final de publicaciones que se recomiendan para los programas de administración en salud también se deberán incluir 12 de los 43 artículos en español y 10 documentos distribuidos previamente por la OPS, así como la lista de los 24 libros en inglés y la bibliografía complementaria (artículos no seleccionados para traducción); de esta forma, los instructores contarán con una serie importante de recursos bibliográficos. Si bien los libros y los artículos complementarios están escritos en inglés, resultarán provechosos en los programas de

[4] En la pág. 14 se incluye la lista de los artículos.

maestría, donde se supone que los alumnos poseen conocimientos básicos de ese idioma *(OPS, IE)*.

6. A fin de mejorar la calidad de la enseñanza y la pertinencia del contenido de los cursos, se debe facilitar el desarrollo de casos de estudio y de programas "paquetes" de computadoras como material didáctico auxiliar *(OPS)*.

- Se deberán desarrollar varios casos de estudio, que varíen en complejidad desde simples versiones de material disponible actualmente en publicaciones, hasta casos más complejos que pueden requerir 8-16 horas de clase y laboratorio para resolverlos. Los problemas seleccionados deben ser representativos de los que enfrentan los administradores en salud latinoamericanos. Los organismos nacionales o los analistas que hayan participado en la solución de problemas en el sector deben proveer el material básico para los casos. Otros casos de estudio pueden obtenerse de la expansión de algunos artículos incluidos en los estudios bibliográficos.
- Para mejorar las clases (y algunos casos de estudio) se deben preparar programas de computadoras para sesiones de laboratorio. También deben obtenerse series de datos reales que permitan a los alumnos el uso de los programas a fin de lograr una mejor comprensión de las técnicas particulares y su empleo en la solución de problemas reales.

Como parte de este esfuerzo, la OPS debe fomentar el intercambio de este tipo de información entre universidades de América Latina y el Caribe.

Deben circularse resúmenes de proyectos y tesis sobre problemas de los sistemas de salud. Estos proyectos pueden servir como casos de estudio y material para los "laboratorios" *(OPS, IO/AS, IE, Organismos)*.

7. Durante el Taller se trató de identificar lagunas en la literatura. Se consideró que existen varias áreas deficientes, tanto en inglés como en español (cuadro 1). Estas deficiencias deberán identificarse con más detalle por representantes de los países latinoamericanos. Posteriormente, la OPS deberá comisionar publicaciones o iniciar estudios que puedan reducir estas deficiencias *(OPS, IO/AS, Organismos)*.

Cuadro 1. Areas en que existen lagunas en la literatura.

Area	Tipo de curso					
	Curso corto		Programa de certificación		Programa de maestría	
	Español	Inglés	Español	Inglés	Español	Inglés
Análisis de sistemas	D	D	S	D	D	D
Informática	D	D	D	D	D	D
Análisis de decisiones	D	E	D	E	D	E
Investigación operativa	D	E	D	E	D	E
Conceptos generales	D	D	D	D	D	D

D: Deficiente.
S: Satisfactorio.
E: Excelente.

8. Los propios participantes y las instituciones docentes deben considerar la posibilidad de colaborar en la elaboración de programas "paquetes" para microcomputadoras, que se puedan usar para fines didácticos y para la solución de problemas reales. Lo anterior implicaría que los participantes de las diferentes instituciones utilicen máquinas similares y un lenguaje común. Los últimos desarrollos tecnológicos en el campo de las microprocesadoras pone a la disposición del usuario computadoras capaces de resolver problemas de gran complejidad a un costo relativamente bajo. En general, el costo de desarrollar los programas será mayor que el costo de la máquina; de ahí la gran importancia de la colaboración *(IE, IO/AS).*

9. Teniendo en cuenta la considerable superposición de los temas considerados en los talleres sobre administración de salud organizados por la OPS, esta deberá comunicar a futuros talleres los resultados de los ya realizados, en especial cuando el tema lo justifique *(OPS).*

10. La OPS debe explorar la posibilidad de obtener fondos para la publicación de un libro en español sobre investigación operativa aplicada a los problemas de los servicios de salud y dirigido a profesionales latinoamericanos; se deberá prestar especial atención a problemas regionales y nacionales. El libro podría incluir algunos casos de estudio *(OPS, IO/AS).*

11. La OPS debe facilitar la preparación de un artículo que describa las labores del presente Taller y sus resultados. El artículo se publicará en una revista apropiada y serviría para fomentar más actividades de este género y mejorar la comunicación entre especialistas de los diversos países *(OPS).*

12. La OPS debe facilitar la distribución de documentos de trabajo y tesis preparados en los Estados Unidos de América a educadores latinoamericanos. En especial se recomienda circular periódicamente el *Catalog of Hospital Management Engineering Technical Papers* publicado por el centro de distribución de ingeniería de gestión hospitalaria de la Asociación Americana de Hospitales; citas selectas de *Hospital Management Abstracts,* y resúmenes de tesis de maestría sobre problemas de los servicios de salud (que podrían obtenerse por medio de la Asociación de Programas Universitarios de Administración de Salud *(OPS, IE).*

Programas de intercambio de información

13. La OPS debe organizar la revisión de manera regular de las publicaciones pertinentes sobre servicios de salud (en español e inglés) y mantenerse enterada de las principales reuniones nacionales e internacionales de las asociaciones de investigación operativa y de salud pública. La información resultante debe enviarse sistemáticamente a los programas de administración de salud en América Latina. También se deberá incluir una lista de cursos y reuniones pertinentes a nivel profesional en los Estados Unidos de América y en países latinoamericanos *(OPS, IO/AS).*

14. El intercambio de información en los talleres puede mejorarse mediante la

presentación formal de informes sobre proyectos o actividades de investigación por parte de algunos de los participantes. Estas presentaciones podrían programarse antes de la inauguración del taller *(OPS)*.

Recomendaciones generales

15. Para lograr una presentación más eficaz de las metodologías de investigación operativa y análisis de sistemas a los administradores de salud, formando simultáneamente un grupo de analistas competentes, se debe dar prioridad al desarrollo de los cursos cortos y al programa de maestría de dos años; aunque es importante, el programa de certificación tiene menos prioridad. Según se indicó, se deberán desarrollar varios cursos cortos usando varios formatos y contenidos para satisfacer a grupos específicos *(IE, IO/AS)*.

16. Todos los participantes deben promover el desarrollo de programas de investigación operativa y análisis de sistemas con énfasis en el sector salud en los departamentos y escuelas de ingeniería industrial, medicina, administración y administración de salud. Debe insistirse en la creación de programas interdisciplinarios, en particular cuando solo una universidad está involucrada *(IE, IO/AS)*.

17. Un componente importante de los programas de certificación y de maestría deberá ser (por lo menos) un curso de investigación operativa en salud, conjuntamente entre las escuelas de ingeniería, salud pública y medicina. En ese curso debe incluirse, como mínimo, un proyecto en el que los alumnos trabajen como equipo interdisciplinario *(IE)*.

18. Hasta que las universidades latinoamericanas no se conviertan en los principales centros de formación de profesionales a nivel de maestría en estos campos, las universidades en los Estados Unidos de América y Canadá continuarán siendo los centros de formación a ese nivel en la Región. En consecuencia, es conveniente que las universidades latinoamericanas establezcan acuerdos institucionales con programas y escuelas de esos países. Las universidades latinoamericanas deben procurar que los programas de Norteamérica sean relevantes y tengan en cuenta las necesidades educativas de los estudiantes latinoamericanos. Si es posible, las tesis de grado deben tratar sobre problemas que sean también importantes en América Latina *(IE)*.

19. A medida que los programas de administración en salud se desarrollen en América Latina, es conveniente que se establezcan vínculos firmes y formales con los programas de ingeniería industrial, investigación operativa y de administración. Sin estos vínculos, las técnicas científicas de gestión no se integrarán eficazmente en los planes de estudio de administración en salud *(IE)*.

20. La OPS puede facilitar aún más la realización de proyectos de investigación y programas de cooperación entre profesionales de los Estados Unidos de América y América Latina si se informa a los participantes y a otros individuos interesados sobre problemas prioritarios y posibles fuentes de financiamiento *(OPS, IE, Organismos)*.

21. Es preciso reiterar que la investigación operativa y el análisis de sistemas exigen grupos interdisciplinarios para la solución satisfactoria de los problemas de prestación de servicios de salud. Esta realidad debe reflejarse cuidadosamente en la elaboración y ejecución de todos los programas y planes de estudio *(OPS, IE, Organismos, IO/AS).*

22. Para que los estudios y proyectos sobre investigación operativa se traduzcan en acciones exitosas, es preciso que los administradores de los sistemas involucrados participen en todos los niveles y puedan reconocer el valor de emplear técnicas científicas de gestión *(OPS, IE, Organismos, IO/AS).*

Descripción de los programas educativos recomendados por el Taller

Como ya se señaló, se recomendaron tres tipos de cursos: cursos cortos, programas de certificación y programas de maestría. Se convino en que el énfasis debe estar primero en los cursos cortos, después en los programas de maestría y finalmente en los de certificación.

Cursos cortos. Los cursos (de 16 a 40 horas) se ofrecerán intensivamente durante un período de dos a cinco días o espaciados durante varias semanas o un semestre. Como en un curso solo se puede presentar material introductorio, se deben programar una serie de cursos afines que se complementen mutuamente y que ofrezcan a los participantes una experiencia educativa más completa. Los cursos deberán incluir problemas a ser resueltos por los alumnos, individualmente y en grupo. Estas sesiones de laboratorio podrían usarse para familiarizar a los alumnos con la computadora, ya sea por medio de un juego, por ejemplo, la simulación de un sistema regional de salud o de un hospital, o mediante un modelo sencillo de asignación de recursos o de pronósticos.

A continuación se presentan ejemplos de cursos cortos:

- *Curso para administradores de hospitales y sistemas de salud.* No se requieren conocimientos de matemáticas, aunque una preparación básica en esa materia es aconsejable. En el curso se presenta una introducción a las técnicas de la investigación operativa y del análisis de sistemas, insistiendo en la comprensión de las posibilidades y limitaciones de los mismos. Los alumnos adquieren conocimientos sobre el enfoque de sistemas, análisis de costo-beneficio, sistemas de información, fuentes de datos y sobre trabajo en grupos interdisciplinarios[5]. Asimismo, aprenden a reconocer los problemas que pueden solucionarse con estas metodologías, a evaluar el esfuerzo y tiempo aproximados que se requieren para la solución y a reconocer posibles barreras políticas y posibles fuentes de ayuda para la realización de estos estudios. Los temas específicos a incluirse en este curso pueden referirse a problemas simplificados de decisión, problemas de asignación de recursos, problemas de planificación (CPM/PERT). El curso debe incluir un proyecto de grupo en el que se formule un problema no estructurado.

[5]Este sería un curso ideal, pero se reconoce que en un solo curso no se podrán presentar adecuadamente todos esos temas.

• *Curso para ingenieros industriales, investigadores operacionales y personal de otras disciplinas técnicas* interesados en participar en la solución de problemas del sector salud. Los alumnos se familiarizan con los sistemas de salud y con los tipos de problemas comunes en el sector. Se presentan y discuten técnicas particulares para solucionar algunos de los problemas.

Programa de certificación (por lo general, de 240 horas). El programa estaría diseñado para el administrador de salud en ejercicio que desea un conocimiento más profundo que el ofrecido a través de una serie de cursos cortos, y también para el profesional que aspira a un cambio de profesión. La participación de los administradores de nivel intermedio en el programa puede significar para estos un ascenso de categoría. Se requiere preparación matemática en materia de álgebra.

El egresado del programa estará en condiciones de plantear y solucionar problemas cuantitativos sencillos (programación lineal, teoría de colas), realizar análisis básicos de costo-beneficio; recopilar y reducir datos (estadística descriptiva); usar programas "paquetes" de computadora para análisis; formular (pero no solucionar) problemas más complejos. Asimismo, tendrá un conocimiento básico de los costos de la prestación de servicios de salud. Debe enfatizarse el concepto de equipo interdisciplinario. El curso debe incluir un proyecto de grupo de 2-3 meses de duración.

Los cursos del programa deben incluir una introducción a métodos cuantitativos; un curso más riguroso en la aplicación de métodos cuantitativos; cursos introductorios en economía de la salud, bioestadística, y un curso en sistemas de información. Pueden considerarse cursos opcionales sobre teoría de planificación y evaluación en salud; métodos cuantitativos de planificación y evaluación, y comportamiento de las organizaciones en el sector.

Programas de maestría (dos años). Se recomendaron dos tipos de programa de maestría. El primero estaría dirigido a administradores de hospitales y sistemas de salud que necesitan un adiestramiento más intenso que el ofrecido en el programa de certificación. Aunque los alumnos conserven sus trabajos regulares, se espera que residan en una institución académica por un período adecuado. Se deben exigir conocimientos de matemática a nivel universitario. El programa debe incluir todos los cursos del programa de certificación, un curso avanzado de estadística, un segundo curso de investigación operativa (modelos probabilísticos) y un curso introductorio en econometría. Los alumnos deben participar en dos o tres proyectos de grupo durante el programa; estos proyectos deben realizarse como parte de un curso interdisciplinario.

El segundo programa de maestría estaría diseñado para profesionales de ingeniería que desean solucionar problemas de los sistemas de salud. El programa tiene una orientación muy fuerte hacia la metodología y técnicas de la investigación operativa. Además, los alumnos deben familiarizarse con los conceptos de contabilidad de costos, principios de economía y sistemas de información. Se espera que los egresados del programa puedan desempeñar varias de las siguientes funciones:

a) Solucionar problemas de programación lineal y de inventario.
b) Formular y analizar modelos estadísticos.
c) Realizar análisis estadísticos avanzados (ANOVA y regresión múltiple).
d) Diseñar estudios para recopilación de datos.
e) Programar en un lenguaje de alto nivel (FORTRAN, PASCAL o PL-1).
f) Diseñar la evaluación de programas.
g) Realizar simulaciones mediante el uso de la computadora.
h) Presentar informes técnicos.
i) Interactuar con las autoridades del sector y participar en equipos interdisciplinarios.
j) Tomar un problema no estructurado, resumirlo, formular un modelo matemático, identificar soluciones e implantar la mejor y más aceptable de ellas.
k) Realizar estudios de planificación a largo plazo.

Artículos seleccionados para traducir al español

Los artículos que los participantes recomendaron se tradujeran al español son los siguientes:

1. Berkson, D., I. Whipple y cols. Evaluation of an automated blood pressure measuring device intended for general public use. *Am J Public Health* 69(5), May, 1979.
2. Escudero, J. C. On lies and health statistics: Some Latin American examples. *Int J Health Serv* 10(3):421-434, 1980.
3. Fetter, R. B., Y. Shin y cols. Case mix definition by diagnosis-related groups. *Med Care* 18(2) Supplement:1-52, 1980.
4. Frerichs, R. y J. Prawda. A computer simulation model for the control of rabies in an urban area of Colombia. *Management Science* 22(4):411-421, 1975.
5. Greenland, S., E. Watson y R. Neutra. The case-control method in medical care evaluation. *Med Care* 19(8), August, 1981.
6. Hartunian, N., Ch. Smart y M. Thompson. The incidence and economic costs of cancer, motor vehicle injuries, coronary heart disease, and stroke: A comparative analysis. *Am J Public Health* 70(12), December, 1980.
7. Lev, B., G. Revesz y cols. Patient flow analysis and the delivery of radiology service. *Socio-Econ Plan Sci* 10:159-166, 1976.
8. Meredith, J. Program evaluation techniques in the health services. *Am J Public Health* 66(11):1069-1073, 1976.
9. Nutting, P., G. Shorr y B. Burkhalter. Assessing the performance of medical care systems: A method and its application. *Med Care* 19(3), March, 1981.
10. O'Connor, R. W. y G. L. Urban. Using a model as a practical management tool for family planning programs. *Am J Public Health* 62:1493-1500, 1972.
11. Reisman, A., J. Mello da Silva y J. B. Mantell. Systems and procedures of patient and information flow. *Hosp Health Serv Admin* Winter: 42-71, 1978.
12. Schoenbaum, S. C., B. J. McNeil y J. Kavet. The swine-influenza decision. *N Engl J Med* 295(14):759-765, 1976.
13. Shuman, L. J., H. Wolfe y R. Dixon Speas, Jr. The role of operations research in regional health planning. *Operations Research* 22:234-248, 1974.
14. Vraciu, R. Programming, budgeting, and control in health care organizations: the state of the art. *Health Serv Res* 14(2), Summer, 1979.
15. Duran, L. y A. Reisman. Design of alternative provider team configurations: experience in both developed and developing countries. Technical memo 947, Department of Operations Research, Case Western Reserve University, Cleveland, Ohio, 1980.

16. Hancock, W., D. Magerlein y cols. Parameters affecting hospital occupancy and implications for facility sizing. *Health Serv Res* 13(3), Fall, 1978.

17. Hindl, A., N. Dierckman y cols. Estimating the need for additional primary care physicians. *Health Serv Res* 13(3), Fall, 1978.

18. Reisman, A., B. V. Dean y cols. Physician supply and surgical demand forecasting: a regional manpower study. *Management Science* 19(12):1345-1354, 1973.

19. Abernathy, W. J. y J. C. Hershey. A spatial allocation model for regional health-services planning. *Operations Research* 20(3):629-642, 1972.

20. Goldman, J. y H. A. Knappenberger. How to determine the optimum number of operating rooms. *Modern Hospital* 111:114-116, 1968.

21. Revelle, C., D. Bigman y cols. Facility location: A review of context-free and EMS models. *Health Serv Res* 12(22):129-146, 1977.

22. Brodheim, E. y G. P. Prastacos. The Long Island blood distribution system as a prototype for regional blood management. *Interfaces* 9(5):3-20, 1979.

23. Duraiswamy, N., R. Welton y A. Reisman. Using computed simulation to predict ICU staffing needs. *J Nur Admin*, February, 1981, pp. 39-44.

24. Harrington, M. B. Forecasting area-wide demand for health care services: A critical review of major techniques and their application. *Inquiry* 14:254-268, 1977.

25. Centerwall, B. S. y M. H. Criqui. Prevention of the Wernicke-Korsakoff syndrome. *N Engl J Med* 299(6):285-289, 1978.

26. Couch, N. P., N. L. Tilney y cols. The high cost of low-frequency events: The anatomy and economics of surgical mishaps. *N Engl J Med* 304(11):634-637, 1981.

27. Eisenberg, J. M. y A. J. Rosoff. Physician responsibility for the cost of unnecessary medical services. *N Engl J Med* 299(2):76-80, 1978.

28. Henry, J. B. y R. L. Roenfeldt. Cost analysis of leasing hospital equipment. *Inquiry* 15(1):33-37, 1978.

29. Klarman, H. E. Application of cost-benefit analysis to the health services and the special case of technologic innovation. *Int J Health Serv* 4(2):325-352, 1974.

30. McGregor, M. y G. Pelletier. Planning of specialized health facilities: size vs. cost and effectiveness in health surgery. *N Engl J Med* 299(4):179-181, 1978.

31. Schwartz, W. B. Decision analysis: A look at the chief complaints. *N Engl J Med* 300(10):556-559, 1979.

32. Willems, J. S., C. R. Sanders y cols. Cost effectiveness of vaccination against pneumococcal pneumonia. *N Engl J Med* 303(10):553-559, 1980.

33. Warner, K. E. y R. C. Hutton. Cost-benefit and cost effectiveness analysis in health care. *Med Care* 18(11):1069-1084, 1980.

34. Kendall, K. y S. Lee. Formulating blood rotation policies with multiple objectives. *Management Science* 26(11), November, 1980.

35. Jackson, M. N., J. P. LoGerfo y cols. Elective hysterectomy: a cost benefit analysis. *Inquiry* 15(3):275-280, 1978.

36. Evans, J. R., K. Lashman Hall y J. Warford. Shattuck Lecture—Health care in the developing world: Problems of scarcity and choice. *N Engl J Med* 305(19):1117-1127, 1981.

37. McNeil, B. J. y S. J. Adelstein. Measures of clinical efficacy: The value of case finding in hypertensive renovascular disease. *N Engl J Med* 293(5):221-226, 1975.

38. McNeil, B. J., E. Keller y S. J. Adelstein. Primer on certain elements of medical decision making. *N Engl J Med* 293(5):211-215, 1975.

En el cuadro 2 se indica para qué tipo de programa es aplicable cada uno de los 38 artículos seleccionados para traducir al español.

Cuadro 2. Tipos de programa en los que resultarán más provechosos los artículos seleccionados para traducir al español.

Artículo No.	Programa de maestría	Programa de certificación	Curso corto
1.		X	X
2.	X	X	X
3.	X		
4.	X		
5.	X	X[a]	
6.	X	X	X
7.	X		
8.	X	X	X
9.	X	X	X
10.	X	X	
11.	X	X	
12.	X	X	X
13.	X	X	X
14.	X	X	
15.	X	X	X
16.	X	X	X
17.	X	X	X
18.	X		
19.	X		
20.			X
21.	X	X	
22.	X	X	X[a]
23.	X	X	X
24.	X	X	X[a]
25.	X	X	X
26.	X	X	X
27.	X	X	X
28.	X	X	X
29.	X	X	X[a]
30.	X	X	X
31.	X	X	X[a]
32.	X	X	X
33.	X	X	
34.	X	X	
35.	X	X	X
36.	X	X	X
37 y 38	X	X	X

[a]Opcional.

Artículos seleccionados en español

Se seleccionaron los siguientes artículos, que se recomendó que la OPS distribuyera, al igual que sus propias publicaciones pertinentes a este campo.

1. Ackoff, R. L. Posibilidades actuales de la investigación operativa. *Administración de empresas* 5(50):125-135, mayo, 1974.

2. Barrenechea, J. J. La selección de prioridades como integrante del proceso de decisión. *Medicina sanitaria y administración de salud* (Tomo II, Parte 3: Atención de la salud), págs. 206-214.

3. Dunia, W. A. de, C. Carmendia y cols. *Servicio de emergencia: Consideraciones conceptuales sobre su funcionamiento y organización especial (Centros ambulatorios de salud y hospital general de 200 camas.* Nota técnica 76-NTe-13). Ministerio de Obras Públicas de Venezuela, Dirección General de Desarrollo Urbanístico, Secretaría Técnica, Unidad de Investigación, 1976.

4. Facultad de Ingeniería, Universidad de Costa Rica, y OPS, estudio de ingeniería industrial en el subsistema "Procesar y distribuir alimentos" del Hospital Nacional de niños. 1974.

5. Grundy, F. y W. A. Reinke. *Investigaciones de práctica sanitaria y métodos matemáticos de gestión:* Capítulos I, II, IV y VIII. OMS, Cuadernos de Salud Pública 51, 1974.

6. Novaro, S. Asignación de camas de servicios de hospitalización: una técnica posible. *Atención médica* 2(1-2), 1973.

7. Rodríguez, R. J., L. C. Arcón y L. A. Almeida Pimentel. Estado actual del sistema de control de pacientes del hospital de clínicas. *Bol Of Sanit Panam* 84(6):493-504, 1978.

8. Schmidt, L. *Consecuencias técnicas de investigación del sistema de información para la salud, Venezuela* (SIS-V/80). Ministerio de Sanidad y Asistencia Social, Dirección de Planificación, Presupuesto e Informática, Comité de Informática, 1980.

9. Schmidt, L., E. Zorilla y M. Quintero. Sistema automático de monitoreo de señales EKG para una unidad de cuidados coronarios. Trabajo presentado en el Congreso Internacional de Sistemas, Venezuela, julio de 1981.

10. OPS y Departamento de Ingeniería Industrial, Instituto Tecnológico y de Estudios Superiores de Monterrey, Nuevo León, México. Análisis de sistemas en lavandería de un hospital general. Washington, D.C., Documento HRR/13/2-B, 1976.

En el cuadro 3 se indican los cursos para los que son aplicables los artículos seleccionados.

Cuadro 3. Tipos de programa en los que resultarán más provechosos los artículos en español.

Artículo No.	Programa de maestría	Programa de certificación	Curso corto
1.	X	X	X
2.	X	X	
3.		X	X
4.	X	X	
5.	X	X	X
6.	X	X	
7.		X	X
8.		X	X
9.		X	
10.	X	X	

A continuación se incluyen como ejemplo los contenidos de varios cursos, que podrán adaptarse según las necesidades.

EJEMPLOS DE CONTENIDO DE LOS CURSOS CORTOS

Introducción a la investigación operativa y al análisis de sistemas para profesionales de salud (40 horas)

1. Introducción a los conceptos de análisis de sistemas, planificación, evaluación y toma de decisiones. Se debe presentar inicialmente una visión panorámica y un modelo conceptual del análisis de sistemas aplicado a los problemas de prestación de servicios de salud. Este modelo debe servir de referencia a medida que se presenten los temas con mayor profundidad.

2. Definición de un sistema de decisión, tipos de decisiones, tipos de modelos. Discusión sobre los niveles de toma de decisión.

3. Introducción a la formulación de modelos; el arte de modelaje, modelos típicos de problemas importantes en los sistemas de salud.

4. Presentación de un caso o de un microproyecto que los alumnos deben desarrollar durante el curso. El énfasis debe estar en la formulación del problema. La solución de una parte del problema puede lograrse usando programas "paquetes" para computadoras.

5. Introducción a métodos de planificación y evaluación, incluyendo la preparación de diagramas, CPM (método de la ruta crítica), PERT (técnica para la evaluación y control de programas), gráficas de Gantt. Se debe incluir la discusión de métodos para la evaluación de procesos y de resultados utilizando ejemplos apropiados.

6. Introducción a modelos de asignación de recursos. Presentación de un caso sencillo de programación lineal. Demostración de la técnica gráfica de solución. Estudio de supuestos en la formulación del problema. (Es deseable el uso de la computadora.)

7. Introducción a fuentes y al acopio de datos. Deben estudiarse los problemas de la administración y procesamiento de datos, la confiabilidad de los datos, sistemas de información y posibles áreas problema.

8. Introducción al análisis de costo-beneficio y a presupuestos. Deben examinarse el valor del dinero en el tiempo, cálculo de costo de proyectos, ejemplos sencillos de análisis de costo-beneficio, definición y cálculo de la tasa interna de retorno y relación costo-beneficio.

9. Análisis de costo-beneficio. Definición de beneficios y costos marginales, matriz de impacto-incidencia, ejemplo completo de costo-beneficio mediante cálculos manuales. (Es deseable un ejemplo utilizando la computadora.)

10. Mercadeo de los servicios de salud. Definiciones básicas, diferencia entre venta y mercadeo, planificación estratégica, aplicaciones específicas al sector.

11. Introducción al concepto de incertidumbre. Conceptos básicos de probabilidad, eventos, estimación y cálculo de probabilidades, el teorema de Bayes, distribuciones y valores esperados.

12. Análisis de decisiones, incluyendo árboles de decisión, alternativas, resultados, experimentos, valores y su obtención. Análisis e interpretaciones del problema de decisión, estudios de cosas sencillas. Las sesiones de laboratorio deben incluir un caso completo sobre análisis de decisión.

13. Técnicas de pronóstico. Concepto de promedio móvil, suavización exponencial, cuadrados mínimos. Interpretación de regresión simple. Estudio de caso; es deseable el uso de programa de computadora.

14. Administración de inventarios y materiales. Costos importantes en el sistema de inventarios. Modelos básicos de inventario, tamaño de lote y nivel de orden.

15. Ejecución de estudios de investigación operativa y análisis de sistemas. Grupos interdisciplinarios, trabajo con autoridades, validación de los resultados de los estudios, estudio de obstáculos en la implantación de los resultados.

Curso (8 a 16 horas) para profesionales de salud[6]

Tema	Curso de un día	Curso de dos días	Curso de dos días (con matemática)
Introducción al análisis de sistemas	1 hr.	1 hr.	1 hr.
Concepto de sistemas	1	1	1
Técnicas gráficas	2	5	5
Técnicas matemáticas	-	-	8
Acopio de datos	1	3	3
Ayudas formales para creatividad	0,5	1	1
Estrategias para implantación	0,5	1,5	1,5

Métodos cuantitativos aplicados a la administración de servicios de salud (45 horas)[7]

El curso está diseñado para administradores de salud (comúnmente médicos y enfermeras), con conocimientos limitados de matemática y estadística.

1. Elementos de matemática: coordinadas cartesianas, ecuaciones lineales, procedimientos elementales de solución gráfica.
2. Elementos de estadística: teoría de conjuntos, probabilidades, probabilidad Bayesiana, distribuciones discretas, distribuciones continuas, regresión lineal simple.
3. Elementos de investigación operativa: método de la ruta crítica, programación lineal, teoría de colas.

Técnicas de mejoramiento de métodos en hospitales (40 horas)

El curso está diseñado para directores de departamentos hospitalarios y personal administrativo auxiliar. Los participantes deben definir y solucionar un problema real que afecta a su departamento.

1. Introducción al mejoramiento de métodos; administración científica; enfoque sistemático para la solución de problemas.
2. Principios de simplificación del trabajo: extensión al marco hospitalario; identificación del problema; empleo del personal en la simplificación del trabajo; CPM/PERT.
3. Sesión de trabajo sobre definición del problema.
4. Análisis de procesos. Definiciones y significado simbólico; diagramas de flujo, evaluación de diagramas y acciones consecuentes.
5. Análisis de operaciones: preparación de diagramas de operaciones, economía de movimientos; técnica de micro y videomovimiento.
6. Simplificación del trabajo: estudio sistemático de los problemas; sistemas automáticos.
7. Sesión de trabajo sobre análisis del problema.
8. Muestreo de trabajo: determinación del tamaño de la muestra; procedimiento, interpretación de resultados.

---

[6] Se recomienda usar el texto de Reisman, A. *System Analysis in Health Care Delivery.* Lexington, Massachusetts, Lexington Books, 1979.

[7] El curso es ofrecido conjuntamente entre la Universidad de São Paulo y la Fundación Getulio Vargas, Brasil.

9. Otras técnicas de acopio de datos: estudio de tiempos; análisis de asociación, cuestionarios.
10. Sesión de trabajo sobre medición.
11. Análisis de costos: medidas de costos; cálculo de datos; criterios no monetarios; comparación de métodos.
12. Estadística elemental para análisis de diferencias; prueba de hipótesis; interpretación de datos.
13. Sesión de trabajo sobre evaluación de resultados.
14. Importancia de los resultados de los estudios sobre mejoramiento de métodos.
15. Sesión de trabajo: informes y críticas de proyectos.

Curso/Seminario sobre investigación de sistemas de salud (45 horas)

Este curso está programado para estudiantes de administración de salud y de ingeniería industrial/investigación operativa que desean profundizar sus conocimientos en la aplicación de métodos cuantitativos a los problemas de los sistemas de salud. Además de las discusiones académicas, los participantes formarán pequeños grupos interdisciplinarios para resolver un problema real de un hospital o del sistema de salud. Los temas a tratar en el seminario son:

1. Introducción a la ingeniería industrial/investigación operativa aplicadas al sector salud.
2. Evaluación del sistema de distribución de dosis unitarias de medicamentos.
3. Modelos de costos hospitalarios: distribución escalonada de costos; microcosteo.
4. Reembolso, modelo de incentivo: enfoque de ingeniería industrial.
5. Modelos de evaluación de utilización de servicios.
6. Modelos para la dotación de personal de enfermería.
7. Modelos de reembolsos predictivos.
8. Agrupación de hospitales para control de reembolsos y costos.
9. Sistemas de admisión y programación de cirugías.
10. Aplicaciones de la programación lineal: planificación de menús en hospitales, radiología terapéutica.
11. Planificación regional: localización de una red de centros de salud.
12. Servicios médicos de urgencia: sistema de información; simulación de sistemas de urgencia; evaluación del personal paramédico.
13. Análisis y modelos de economía de salud.
14. Evaluación de la eficacia de la atención médica.

Sistemas de información y evaluación de la calidad de la atención médica (30 horas)

Este curso está programado para planificadores y administradores de los servicios de salud. Presenta una visión panorámica del uso de la información y de la tecnología del procesamiento de datos para la planificación, administración y control de los sistemas de salud.

1. Origen, naturaleza y uso de la información en los sistemas de salud; documentación médica; registros administrativos.
2. Información para los procesos de toma de decisiones; importancia y objetivos de esa enseñanza de informática al personal de salud.
3. Objetivos de los usuarios de la información; niveles funcionales; núcleo de centralización o de distribución de los sistemas de información; impacto de la información en las organizaciones.

4. Necesidades de los administradores; información para administración y planificación.
5. Análisis de sistemas; proyectos sobre análisis de sistemas; planificación y control de los proyectos; acopio y distribución de datos; proceso para establecer sistemas de información.
6. Computadoras y automatización: conceptos básicos, procesamiento electrónico de datos; archivos y estructuras lógicas; bancos de datos; computadoras y sus componentes, centros de procesamiento de datos.
7. Opciones para el procesamiento de datos: sistemas centralizados, descentralizados y distribuidos.
8. Aplicaciones del procesamiento de datos en los sistemas de salud; ventajas e inconvenientes del uso de computadoras.
9. Sistemas de información para la evaluación y el control; métodos de evaluación del uso y de la calidad de los servicios, auditoría médica.

Métodos cuantitativos y analíticos en administración de salud (40 horas)

Este curso[8] está diseñado para: 1) desarrollar en el participante la apreciación y comprensión de los procesos básicos del análisis de sistemas, toma de decisiones y control; 2) identificar la influencia de estos procesos en las actividades y responsabilidades del administrador de salud. Se discuten algunos modelos básicos de decisión y control así como la metodología del análisis de sistemas, ciencia de gestión y economía. Se recalcan las ventajas y limitaciones de estos modelos para la toma más eficaz de decisiones en el sector. Entre otros, se requieren conocimientos de álgebra a nivel universitario, álgebra lineal básica, probabilidad y estadística básicas.

1. Revisión general de la literatura relacionada con el análisis de sistemas/investigación operativa y ciencia de gestión y el sector de la salud.
2. Concepto y análisis de sistemas: un prólogo para la toma de decisiones en el sector.
3. Modelos cuantitativos y sus funciones en la toma de decisiones y control en el sector salud.
4. Modelos elementales para el análisis de decisiones determinísticas: inventarios y administración de proyectos.
5. Modelos complejos de decisiones determinísticas: programación matemática.
6. Solución de problemas de programación lineal por computadoras; análisis de sensibilidad.
7. Acopio de datos apropiados para la toma de decisiones en el sector salud: técnicas de medidas.
8. Aplicaciones de la programación entera y programación con metas en el sector de la salud.
9. Análisis de costo-beneficio y costo-eficacia: un modo fundamental de pensar en la toma de decisiones en el sector de la salud.
10. Realimentación, control y evaluación de programas para una administración más eficaz de la atención en salud.
11. Temas especiales tratados por conferencistas invitados.

[8] Curso desarrollado por Barnett R. Parker, Ph.D., Escuela de Salud Pública, Universidad de Carolina del Norte, Chapel Hill.

Apéndice
Participantes en el Taller

Ing. Moisés Arteaga Cabrera, Ministerio de Planificación, La Paz, Bolivia
Carmela Calapinto de Rosales, Ministerio de Sanidad y Asistencia Social, Caracas, Venezuela
Ing. Antonio dal Fabbro, Escuela de Salud Pública, São Paulo, Brasil
Dr. Mariano Fernández, Ministerio de Sanidad y Asistencia Social, Caracas, Venezuela
Dr. David R. Gómez Cova, Escuela de Salud Pública, Universidad Central de Venezuela, Caracas, Venezuela
Dr. Pedro Guédez Lima, Dirección de Salud Pública, Ministerio de Sanidad y Asistencia Social, Caracas, Venezuela
George Kastner, Instituto de Estudios Superiores en Administración, Caracas, Venezuela
Ing. Carlos Enrique Pérez (Relator), Universidad de Pittsburgh, Pittsburgh, Pensilvania
Ing. Víctor L. Pérez, Universidad de Chile, Santiago, Chile
Ing. Diego Prieto, Universidad de Los Andes, Bogotá, Colombia
Ing. Carlos M. Quesada Solano, Escuela de Ingeniería Industrial, San José, Costa Rica
Arnold Reisman, Universidad Case Western Reserve, Cleveland, Ohio
Dr. Roberto Jaime Rodríguez, PROAHSA, Hospital das Clinicas, Facultad de Medicina, Universidad de São Paulo, São Paulo, Brasil
Dr. Richard H. Shachtman, Escuela de Salud Pública, Universidad de Carolina del Norte, Chapel Hill, Carolina del Norte
Dr. Larry J. Shuman (Relator), Universidad de Pittsburgh, Pittsburgh, Pensilvania

Organización Panamericana de la Salud

Dr. Humberto Moraes Novaes, Asesor Regional y Coordinador del Programa OPS/Kellogg de Educación en Administración de Salud, División de Recursos Humanos e Investigación, Washington, D.C.
Ing. Jorge Ortíz, Asesor Regional en Investigación de Servicios de Salud, División de Recursos Humanos e Investigación, Washington, D.C.
Ing. Jorge Peña Mohr, Asesor Regional en Administración Institucional, División de Servicios Integrados de Salud, Washington, D.C.

LISTA DE PARTICIPANTES

MOISES ARTEAGA	Subsecretario de Planificacion La Paz, Bolivia
DR. DAVID GOMEZ COVA	Escuela de Salud Publica Caracas, Venezuela
DR. ANTONIO DAL FABRO	PROAHSA - Hospital Das Clinicas Ave. 9 de Julh 2029 01313 Sao Paulo, Brasil
DR. CESAR HERMIDA	Curso de Postgrado en Administracion de Salud Facultad de Medicina Calle Iquique s/n Quito, Ecuador
DR. GEORGE KASTNER	Instituto de Estudios Superiores Final Calle Occidente San Bernandino Caracas, Venezuela
DR. CARLOS PEREZ	Insdustrial Engineering Dept. University of Pittsburgh U.S.A.
DR. VICTOR PEREZ	Departamento de Industrias Universidad de Chile Santiago, Chile
DR. DIEGO PRIETO	Escuela de Ingenieria Industrial Universidad de los Andes Bogota, Colombia
DR. CARLOS QUESADA	Escuela de Ingenieria Industrial Universidad de Costa Rica San Jose, Costa Rica

DR. ARNOLD REISMAN — *Case Western Reserve University*
Cleveland, Ohio

DR. ROBERTO J. RODRIGUEZ — *PROAHSA*
Ave. 9 de julho 2029
01313 Sao Paulo, Brasil

DR. LARRY SHUMAN — *Industrial Engineering Department*
University of Pittsburgh
U.S .A.

PROFESSOR RICHAR SHACHTMAN — *University of North Carolina*
Rosenan Hall 201H
Chapel Hill, N.C. 27514

DIVISION DE RECURSOS HUMANOS E INVESTIGACION

DR. HUMBERTO M. NOVAES — *Coordinador*

ING. JORGE ORTIZ — *Investigador Operacional*

DIVISION DE SERVICIOS INTEGRADOS DE SALUD

SR. JORGE PENA MOHR — *Asesor Regional en Administracion de Salud*

OFICINA SANITARIA PANAMERICANA- CARACAS

DR. BARRY W. WHALLEY — *RA-II Venezuela*