Pan American Health Organization

PAHO/ACMR 9/P/6

NINTH MEETING OF THE
ADVISORY COMMITTEE ON MEDICAL RESEARCH

Washington, D.C. 15-19 June 1970

PANEL ON MEDICAL EDUCATION

OPPORTUNITIES FOR EVALUATION OF MEDICAL SCHOOL PROGRAMS

Edwin B. Hutchins, Ph.D.

and

Horst Noack, M. D.

OPPORTUNITIES FOR EVALUATION OF
MEDICAL SCHOOL PROGRAMS

by

Edwin B. Hutchins, Ph.D.
and
Horst Noack, M. D.
Iowa State University

Over the past century the progress of medical education has been steadfast. There exists an accumulated wisdom in the education and training of physicians that has incorporated not only the strides made in the biomedical sciences but the advances in the philosophy and technique of education as well. While it is not difficult to engage a medical man in a heated discussion of the relative importance of art and science in the practice of medicine, it is less likely that the advocate of science in medicine will take up the torch for science in the practice of teaching. Yet the importance of this is implicitly recognized by Campbell and Stanley (1963, page) who indicate their commitment to the experiment as "the only means for settling disputes regarding educational practice, as the only way of verifying educational improvements, and as the only way of establishing a cumulative tradition in which improvements can be introduced without the danger of fadish discard of old wisdom in favor of inferior novelties."

Research in medical education has paralleled to some extent developments in the behavioral sciences. It has been dependent upon methodological developments, the development of measurement devices and upon knowledge about the appropriateness of various designs. It has only been in the last 15 years or so that a general concern for the multivariate character of educational research has been a pervasive factor in the study of medical education. It will be our purpose here to illustrate with available data from the Association of American Medical Colleges (AAMC) longitudinal study some of the problems involved in drawing inferences about the relationships obtained among a few selected variables. Although the examples we have chosen might not be of great concern to those dealing with the pressing problems surrounding the delivery of health care we think they demonstrate typical difficulties in the evaluation of medical school programs.

The AAMC longitudinal study represents one of the most extensive data collection efforts undertaken in the field of medical education. This study began in 1956 and at this point in time spans some 14 years although the last data collection was in 1965. The study was primarily a descriptive and predictive study aimed at answering three broad questions: (1) What kinds of measures of student characteristics could contribute to prediction of success in medical school? (2) What kinds of changes take place in the personality characteristics of medical students as a function of their medical education? (3) What personal characteristics were associated with various career groupings within medicine? In the course of the study, an additional question was posed: (4) What characteristics of medical schools relate to personality, interest and outcome variables of the students they select and educate?

Design and Data Collection

From the 78 member schools of the AAMC in 1956 a sample of 28 medical schools was drawn according to three criteria: public or private, geographic region, and ability level of the student body. Figure 1 shows some of the measures obtained during the course of the study. Three major classes of variables were considered: (1) Historical variables. Examples are the estimated expenditures of the medical schools per year in 1910 (Flexner,), Faculty-student ratio in 1910 and percentage of a school's graduates over a period of 25 years who later entered academic medicine. (2) Institutional variables, included current estimated annual expenditures of the schools, current faculty-student ratio, and student's perception of their environment as assessed by the Medical School Environment Inventory, an instrument developed for this particular purpose (Hutchins,). (3) Student variables. Of interest for our purpose is the Medical College Admission Test (MCAT) which measures scholastic aptitude. This instrument consists of four scales (Verbal Ability, Quantitative Ability, Modern Society, and Science) and has been widely used as a criterion for admission to medical school. Three psychological instruments, the Allport-Vernon-Lindzey Study of Values (AVL), Edwards Personal Preference Schedule (EPPS), and the Strong Vocational Interest Blank (SVIB) were administered to the total Freshman class in the 28 sample schools in 1956 and again when they were seniors in 1960 for the purpose of measuring students' values, personality characteristics and interests. In addition, performance on the examinations of the National Board of Medical Examiners (NBME) and peer ratings of professional competence were obtained at the end of the Sophomore and Senior years. The students' plans regarding their career choices were also assessed prior to graduation from medical school.

The time axis in Figure 1 illustrates the longitudinal character of the study. The arrows indicate roughly the hypothesized relationship between the classes of variables with respect to time. It is easy to think of arrows between all of the variables which gives us a picture of the complexity of the situation. Figure 1 is neither a complete description of the study nor a valid model of all significant variables but it does suggest the potential problems inherent in dealing with a few variables. The difficulties that confound the analysis of complex models are alluded to and will serve as a basis for discussion of some of the major problems involved in drawing inferences from the data.

In our model there is reason to assume that the history of a medical school has some relationship to the present educational program, or that ability and personality characteristics of the students entering a medical school affect the career choices of the graduating students four years later. This kind of reasoning usually implies a cause-effect relationship between different variables.

The ultimate purpose of our research effort is to learn something of the causes affecting our criterion measures. Yet the difficulties involved in drawing causal inferences from data obtained in natural settings are paramount. How do we know that the variables included in the study and not other related yet unknown or omitted factors account for the results? How shall we define an experimental treatment in our example? If the 28 medical schools we studied represent 28 different treatment conditions, are the most important specific treatment variables estimated expenditures, faculty-student ratio, or are they other characteristics like structure of the curriculum, research orientation of the faculty and the like?

Since we are not dealing with a controlled experiment, several extraneous variables may be present "which might produce effects confounded with the experimental stimulus" (Campbell and Stanley, 1963, p. 5), in our case the specific treatment variable under consideration. The generation of quasi-experimental designs in large part involves the elimination of the effects of extraneous variables on the dependent measure. Campbell and Stanley (1963) list eight classes of extraneous variables which might jeopardize the inferences which can be drawn from the experiment. Although all of these are relevant to our discussion, we will limit ourselves to a few.

1. Because of differing applicant pools and selection criteria, the entering classes of medical schools differ with respect to scholastic aptitude ( ), personality characteristics ( ) and other factors. Thus we begin any study of programs with serious <u>selection biases</u> when comparing schools. In a situation like this we cannot say whether, for example, differences in the proportion of graduates planning to go into general practice is due to differential selection, differential treatment, or both.

2. Preselection of extreme groups can introduce <u>statistical regression</u> artifacts. For example, if groups of subjects for study were selected on the basis of extreme scores on the MCAT, a group with high scores would tend to score lower on a later aptitude or achievement test simply due to unreliability of measurement.

3. <u>Maturation</u> is likely to be involved. As students grow older they will develop and change without regard to our educational treatment. Different groups of students might also change or mature at different rates.

4. <u>Experimental mortality</u> is a serious problem caused by differential drop-out rates as shown by Johnson and Hutchins ( ).

5. The effects of the administration of a test upon later testing can also affect the measures taken. Thus in the longitudinal study there is some evidence that the participants in the study viewed their schools more positively simply as a matter of participation in the testing effort.

The possibility that these types of extraneous variables may cause the observed results makes the analysis and interpretation of data collected in a natural setting extremely difficult. In order to illustrate this we will deal with a measure of the career plans of students. In these studies we first distinguish between a student's concern for whether he should seek a career involving a full-time general or specialty clinical practice or a career holding some balance between clinical practice and academic medicine. A second consideration is the specific choice of speciality whether pediatrics, obstetrics, psychiatry, and so on. While differences between students choosing the various specialities have been studied, these results we will leave for another time.

The general measure of career choice is of some social significance since by availing ourselves of information on the career plans of our graduating students we can have some vision of what the face of medicine will be in the future. This leads as well to other questions. Thus, given limited manpower resources, does Africa need neuro-surgeons, internists, public health specialists, researchers and teachers, nutritionists or all of these and more in some optimal combination? Similarly, what demands do the topography of the land and the geographic distribution of people make for the delivery of health care in Canada, or Peru or Mexico? If we can answer these and similar questions, perhaps we can then move to the next query: "How do we design an educational system to produce the desired health manpower resources?" Given a system in which the student freely elects his type of career, we would need to know something about both our students and their environment for learning in order to answer such a question. While a number of articles on student characteristics related to choice of a career have been published we are interested here in the development and evaluation of educational programs and therefore will draw from research on institutional determinants related to the career choices of the school's graduates in order to illustrate certain methodological considerations.

The 28 schools represented in our studytherefore become the units of investigation in the examples we will prexent here. As such they represent 28 "treatments" in the study design. The criterion variable is the proportion of the graduating seniors who elect to go into straight clinical medicine as opposed to a career which would involve at least some teaching and research. The schools in the stratified random sample varied from a low of 18% of their graduates seeking straight clinical practice to a high of 88% anticipating this type of career. Since the schools do vary greatly with regard to this measure of their output, we can appropriately ask why. A pervasive general hypothesis has been that the schools were selective in their choice of students (or conversely students may be selective in choice of school) and differential selection factors ultimately made a difference in the proportion of the class choosing straight practice careers four years after their admittance. The layman's lore regarding medicine has it that faculty members must take a vow of poverty while full-time practitioners join country clubs and drive Cadillacs. The hypothesis follows that the economic values of students prevail and that this in large part determines their choice even upon entrance to medical school. The evidence for this is presented in Figure 2. The scatter plot indicates that indeed a correlation of .74 does exist between these two variables.

Thus the apparent inference to be drawn is that schools whose incoming students tend to value highly economic considerations in their life style tend also to produce higher proportions of students interested in full-time straight clinical practice, whereas schools whose incoming students do not represent the stereotype of the businessman tend to produce higher proportions of their graduates seeking careers in academic medicine.

There are, of course, many obvious alternative hypotheses. One might postulate, for example, that the ability level of the class would be as important for the election of academic medicine as the economic value is for the election of a straight practice career. Figure 3 presents the relevant data. Here the average science subtest score of the MCAT is plotted against the same career choice criterion. This relationship is negative, the correlation being -.75.

Other rival hypotheses intrude upon our inference as well. Thus, if educational programs have any impact on our medical graduates we should need to account for institutional differences also.

A major factor in many programs is the amount of contact faculty have with students, a factor easily measured by the faculty-student ratio. This, too, varies considerably across schools from a ratio of almost 6 students per full-time faculty member to a ratio of nearly one to one. Figure 4 relates this measure to the school's output of clinical practitioners. While the correlation of -.60 is not quite as high as those obtained from measured student characteristics, it is indicative of the importance of institutional variance and invites alternative kinds of inference. One envisions a school with faculty resources approaching a tutorial setting in which a close personal relationship between faculty member and student leading to the emulation of the faculty member by the student as he enters a career in academic medicine. Conversely, a school with limited faculty resources must necessarily program differently and its students, left more to their own devices, direct their attention to the practice of medicine with little concern for the promulgation of knowledge and its transmission to the next generation.

Enticing as such inferences are, we are obligated to look further in our effort to eliminate rival hypotheses.

For example, the budget of an educational institution is often a primary determinant of factors such as the faculty-student ratio, the ability of the institution to be innovative or indeed to differentially attract students of one characteristic or another. In Figure 5 we have the total expenditures of our schools plotted against the career choice measure. We continue to find some degree of relationship. The correlation of -.55 indicates that some 30% of the variance in choice of career by the senior class can be accounted for by simply knowing the general level of expenditures for the schools. In addition, this ordering of schools in terms of size of budget has been quite stable over a long period of time.

This gives rise in turn to another rival hypothesis concerning the history of the school. In looking back over a half century we find that we can do almost as well in our prediction by using the data from the Flexner report of 1910. Figure 6 relates data on general expenditures of the schools in 1910 to the career choice output variable 50 years later. The correlation of -.65 is as interesting as that obtained with data that was collected concurrently. Finally, the appendix to the Flexner report of 1910 also contained data on the ratio of faculty to students at that time. The relationship to the 1960 career choice information is presented in Figure 7. Here the correlation of -.34 is somewhat lower and accounts for a relatively small proportion of the variance.

These illustrative scatter plots presenting the actual data for 28 schools relating six variables to the schools output of straight clinical practitioners are further summarized in the diagram given in Figure 8. Presented here are additional relationships among the causal variables. Science Achievement correlated -.66 with the economic value measure while school expenditures and faculty-student ratio correlate .05 and .72 respectively for the 1910 and 1959 data. The interrelationships among the variables in this nomothetic network indicate the complexity of the analysis if causal inferences are to be made from the data. In this task the ability to eliminate rival hypotheses will depend on the design of data collection, the nature of the sampling, the choice of independent, dependent and control variables and the ingenuity of the investigator to simultaneously attend to a wide variety of factors.

In summary these include the fact that medical education operates in a broad social setting so that the support of the school, its history, the character of its student body, the innovativeness of its administrative leadership and the characteristics of its faculty along with numerous other variables confound the process of drawing inferences from research in medical education. This will follow whether we are concerned with a specific classroom experiment, an effort to predict success in a given school or an understanding of the national forces that shape and mold the total enterprise. Emulation of the univariate physical science model, which in the past has characterized behavioral and educational research, will no longer serve if we are to make significant strides in research in medical education. It will be incumbent on the researcher in medical education to make use of the best possible counsel in the design, execution and analysis of his study if he is to make good use of the limited resources available for the task at hand. This task involves the generation of information which can effectively guide the training and development of health care personnel in a world which has before it a model of high level health care but a model which today is realized by a very small proportion of the total population.

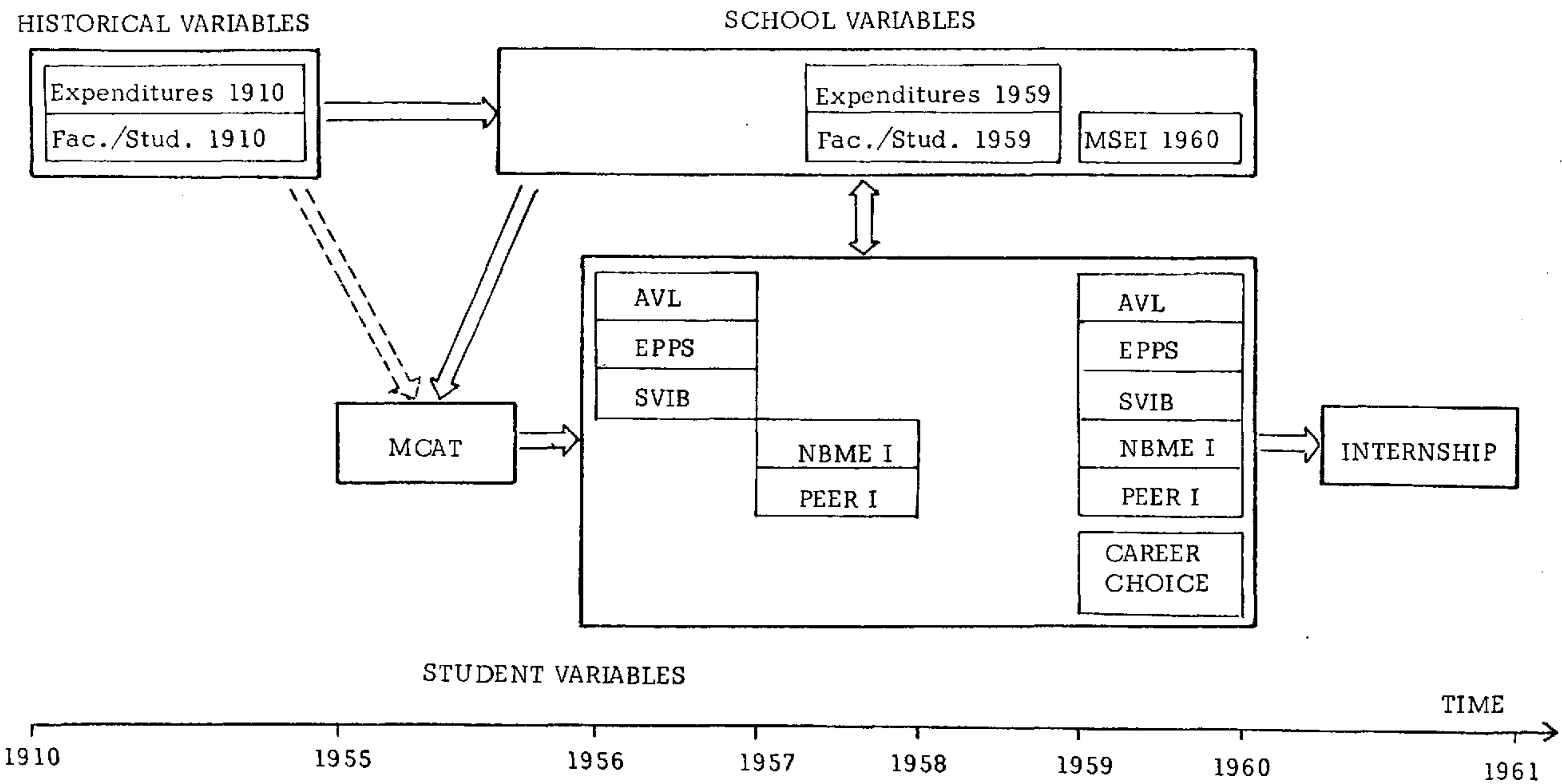

Figure 1. Association of American Medical Colleges Longitudinal Study. Time relationship of some selected variables.

Fac./Stud. = Faculty-Student ratio; MSEI = Medical School Environment Inventory; MCAT = Medical College Admission Test; AVL = Allport-Vernon-Lindzey Study of Values; EPPS = Edwards Personal Preference Schedule; SVIB = Strong Vocational Interest Blank; NBME = Examination administered by the National Board of Medical Examiners; PEER = Peer ratings of professional competence.

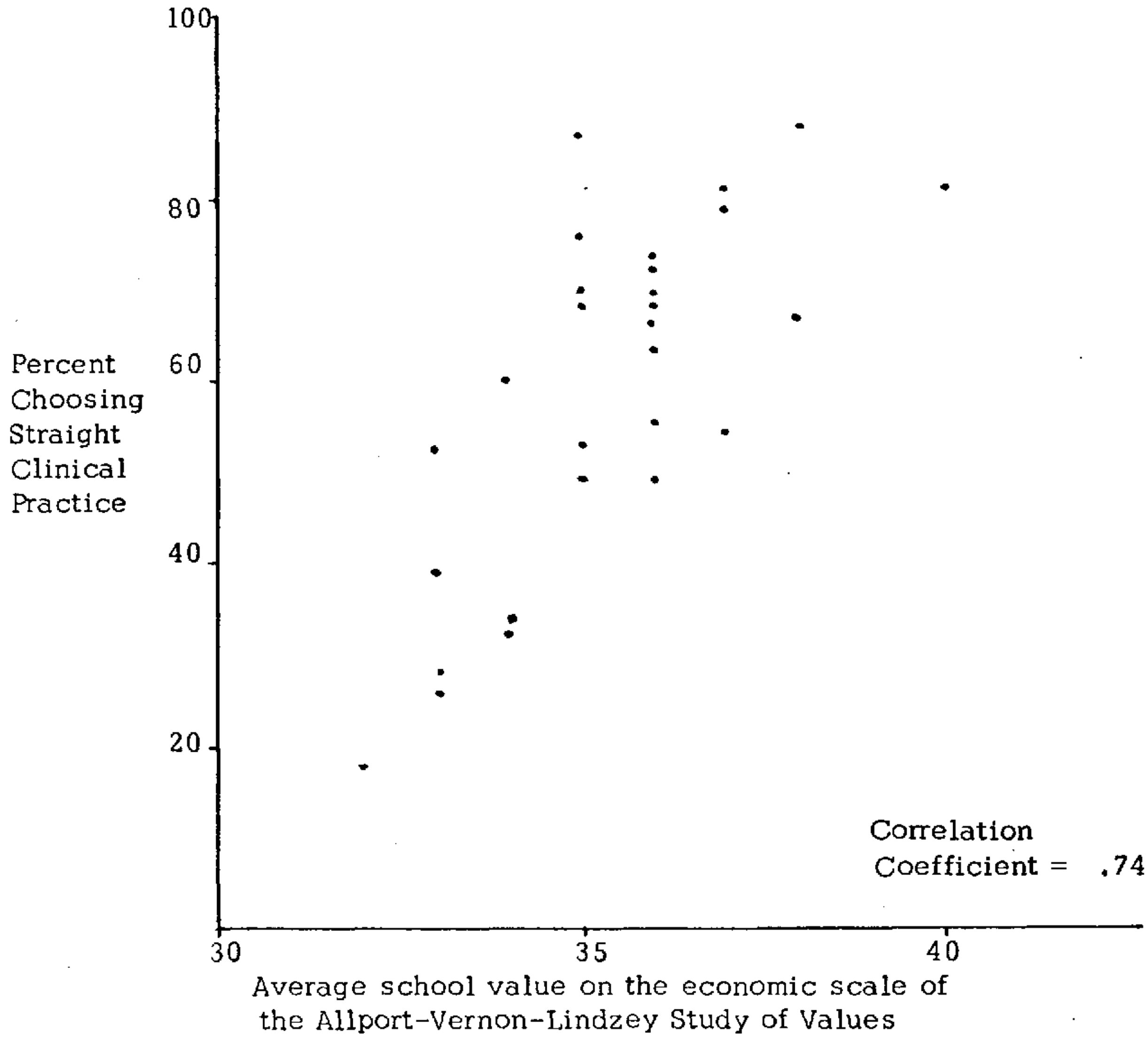

Figure 2. Scatter plot illustrating the relationship between the average economic values of incoming freshmen medical school classes and their later proportionate choice of straight practice careers as graduating seniors. N = 28 medical schools.

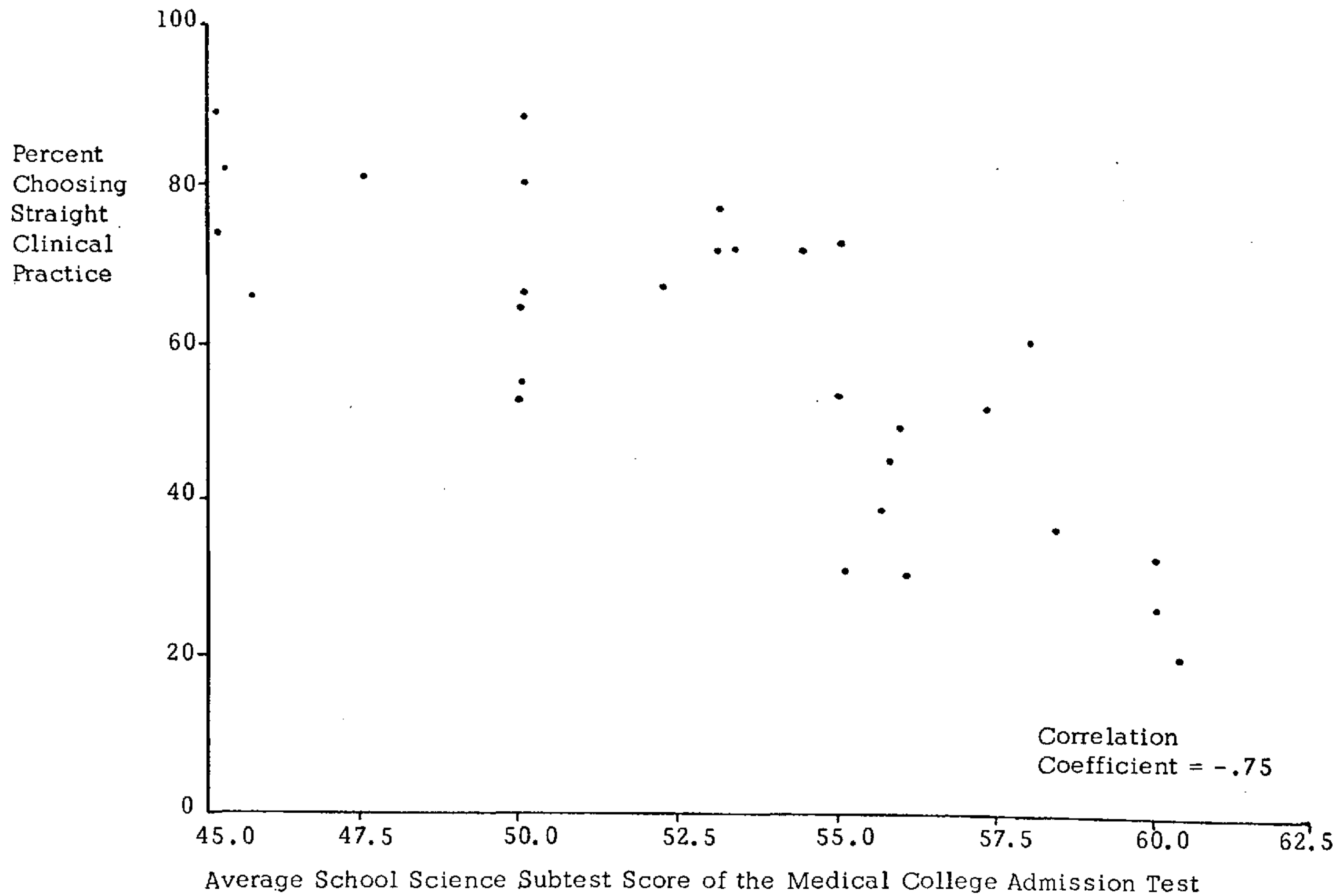

Figure 3. Scatter plot illustrating the relationship between the average science subtest scores of incoming medical school classes and their later proportionate choice of straight practice careers as graduating seniors. N = 28 medical schools.

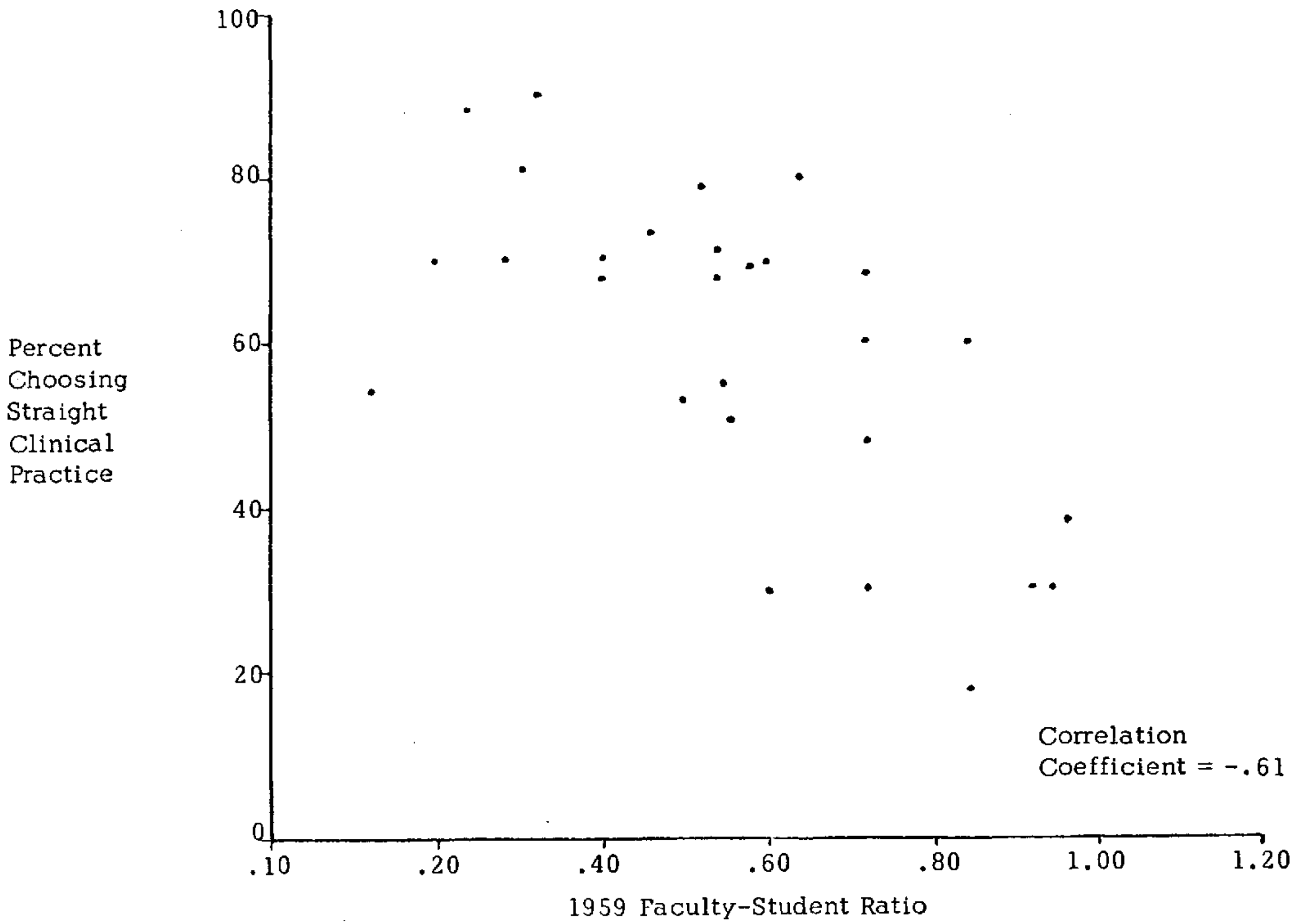

Figure 4. Scatter plot illustrating the relationship between faculty-student ratio in 1959 and proportionate choice of straight practice careers as graduating seniors. N = 28 medical schools.

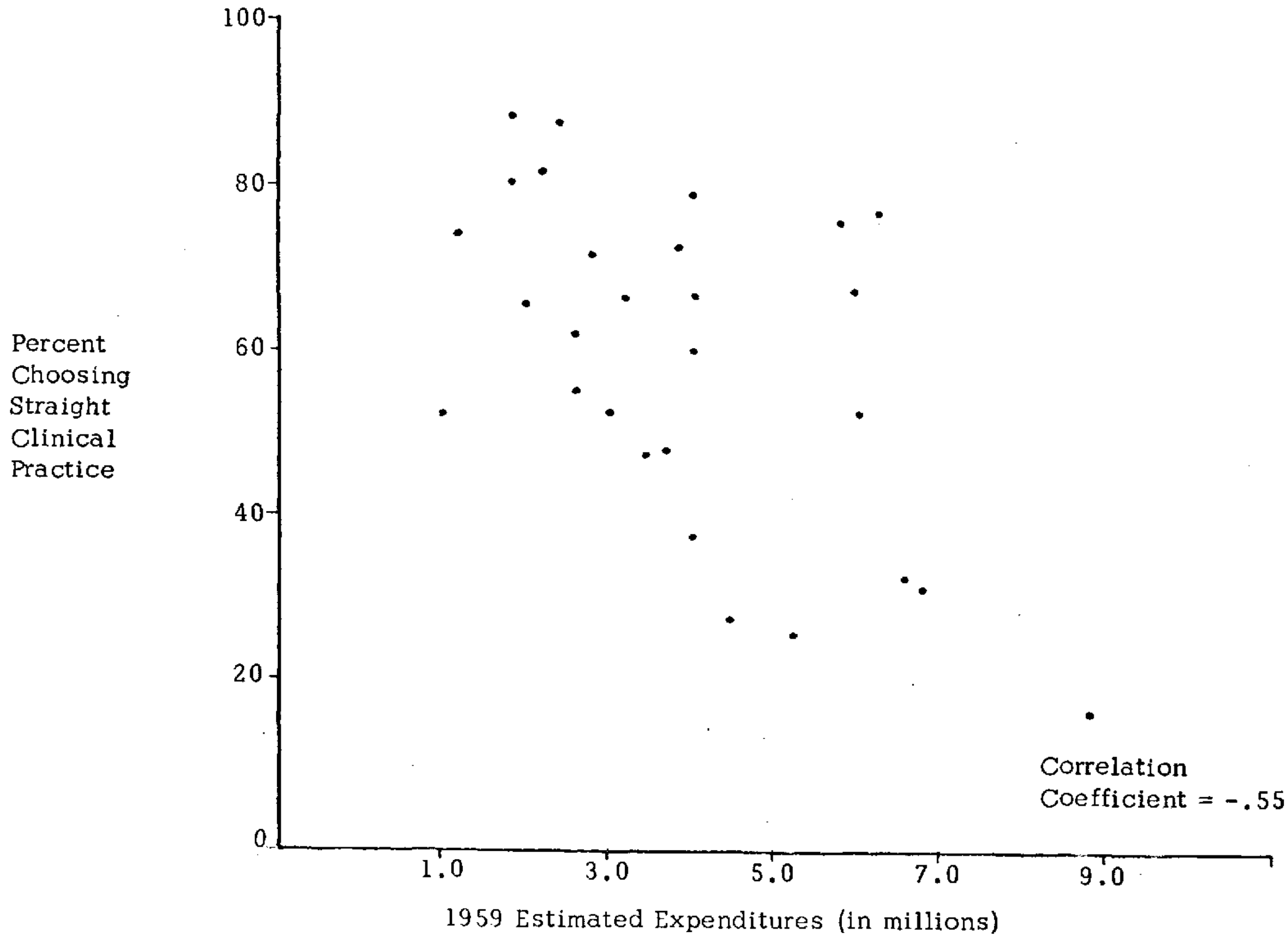

Figure 5. Scatter plot illustrating the relationship between estimated medical school expenditures (in millions) in 1959 and proportionate choice of straight practice careers as graduating seniors. N = 28 medical schools.

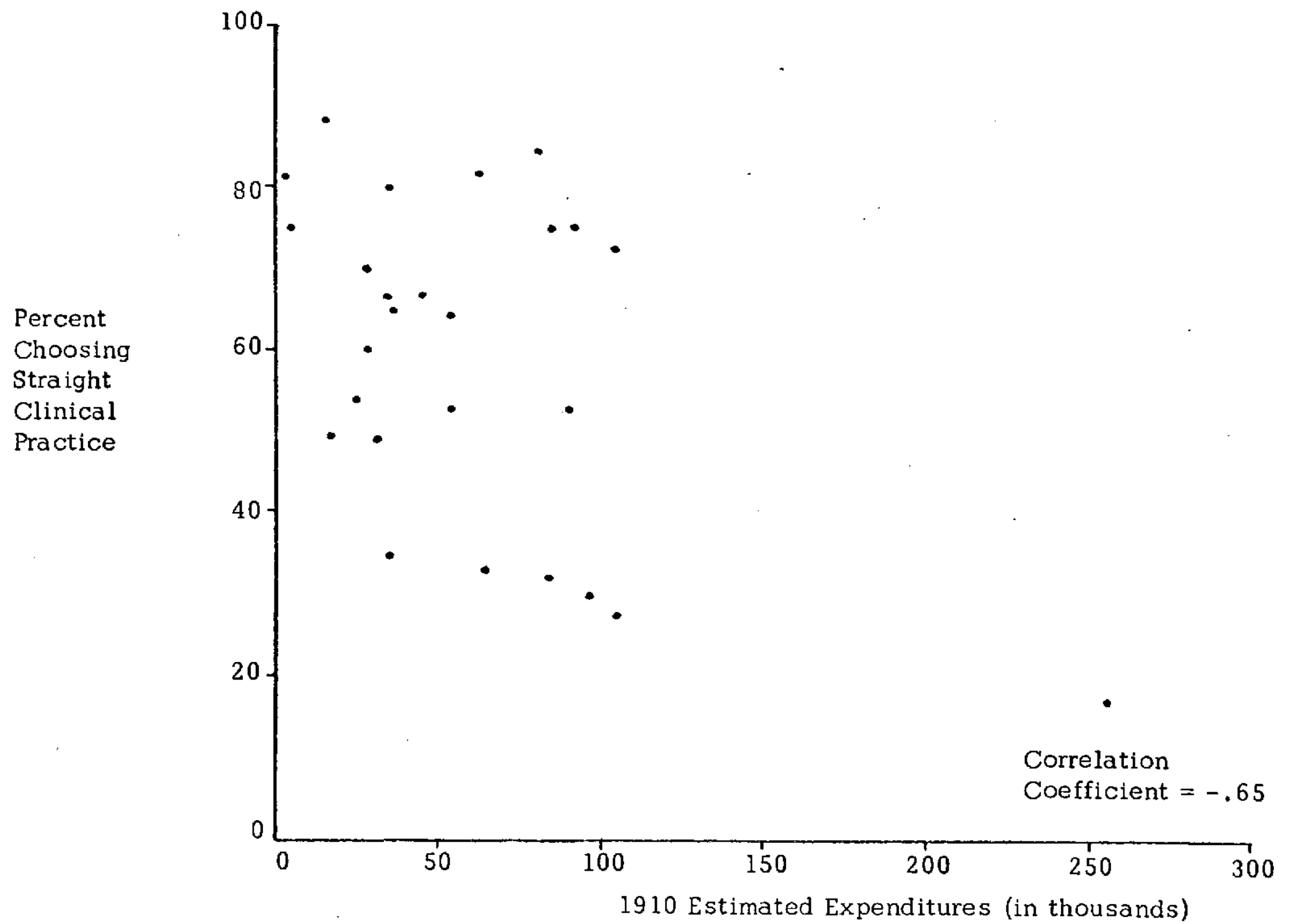

Figure 6. Scatter plot illustrating the relationship between estimated expenditures (in thousands) in 1910 and proportionate choice of straight practice careers as graduating seniors. N = 28 medical schools.

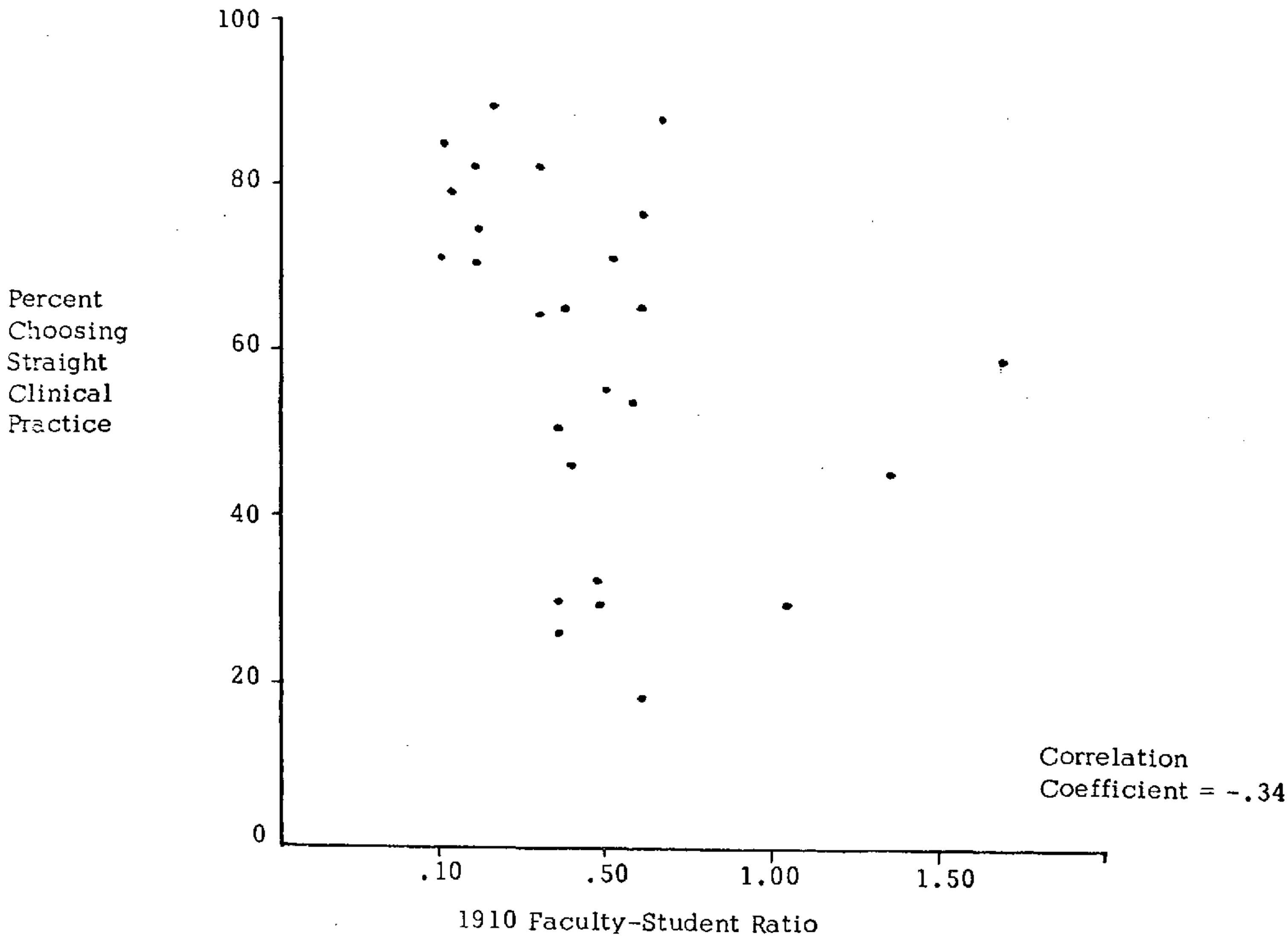

Figure 7. Scatter plot illustrating the relationship between faculty-student ratio in 1910 and proportionate choice of straight practice careers as graduating seniors. N = 28 medical schools.

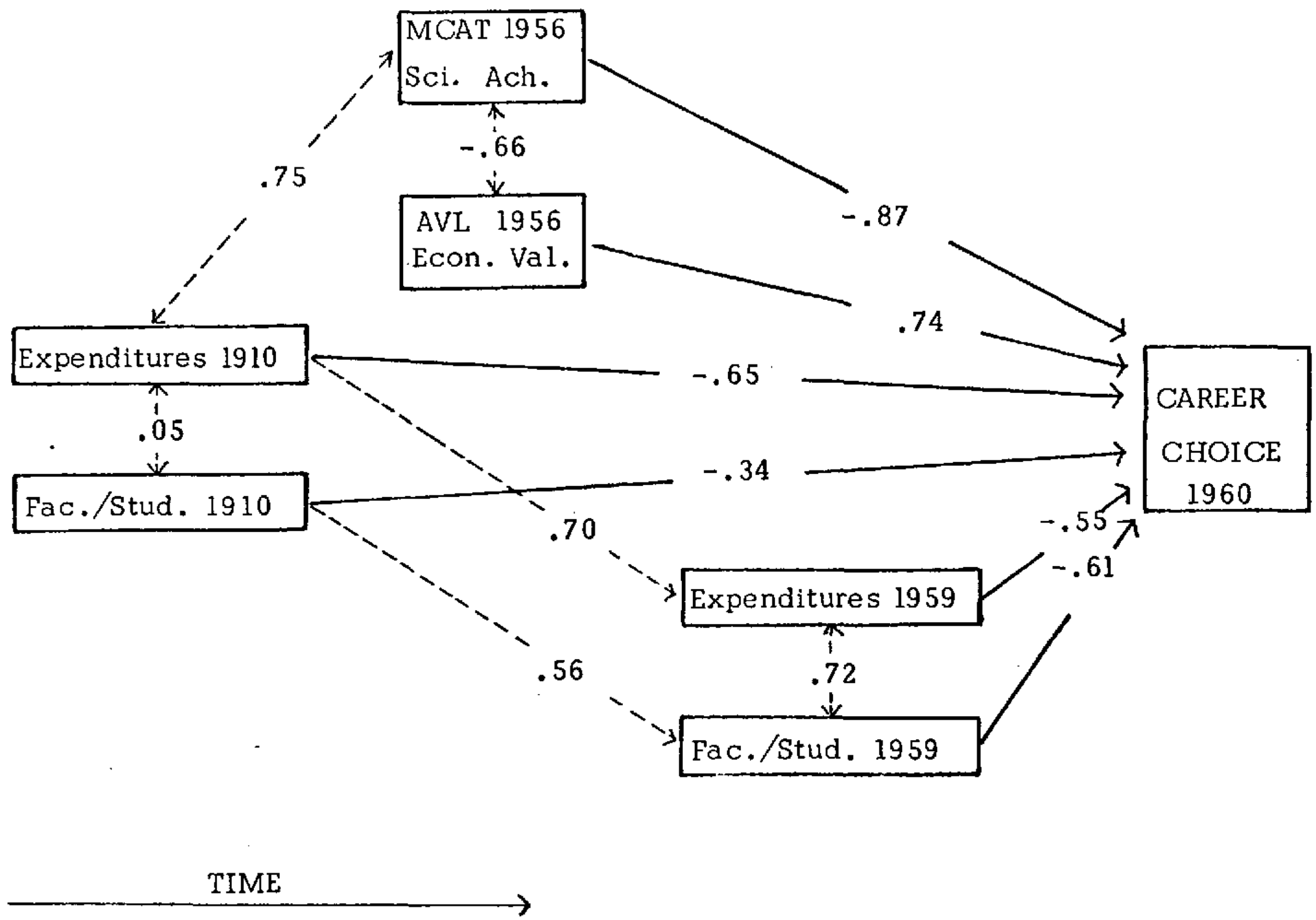

Figure 8. Correlation coefficients between selected independent variables and proportionate choice of straight practice careers as graduating seniors in 1960. N = 28 medical schools.

Fac./Stud. = Faculty-Student ratio; MCAT 1956 Sci. Ach. = average school science subtest score of the Medical College Admission Test; AVL 1956 Econ. Val. = average school value on the economic scale of the Allport-Vernon-Lindzey Study of Values.