Epidemiological Bulletin

Pan American Health Organization:
Celebrating 100 Years of Health

Vol. 25, No. 4 December 2004

Software Programs for Mapping and Spatial Analysis in Epidemiology and Public Health

Introduction

In recent years, there has been growing interest among ministries of health and other health sector institutions in the use of Geographic Information Systems (GIS) as a tool to strengthen the analytical, management, monitoring, and decision-making capacity in public health, as well as a tool for advocacy and communication between technical personnel, policymakers, and the general public. This interest is the result of: i) recognition of the GIS capacity for managing geographical dimensions, integrating health-related data from various sources, helping to discover and visualize new patterns and geographical relations in data that would otherwise be difficult to identify, and displaying these on maps that constitute a more expressive and visual representation; ii) technical meetings and congresses devoted to GIS in public health, as well as the inclusion of GIS as a topic in scientific events in epidemiology and public health; iii) publication of a significant number of articles that show the potential and uses of GIS in several areas, including epidemiological studies, public health management, and improvement of community health;[1,2] and iv) the growing number of health studies and projects that are being developed by academic teams and health service professionals that include the use of GIS as a tool for analysis and results communication.

Notwithstanding these developments, the adoption of GIS in the health sector at local levels has been limited, primarily due to low level of access to commercial GIS programs, their cost, complexity, and limited availability of analytical techniques and methods for problem solving in epidemiology and public health. In response, efforts have been made to overcome these problems through the development of GIS software packages, mapping tools and methods for statistical spatial data analysis in different environments.

The purpose of this article is to inform public health professionals about the availability and current status of various GIS, mapping, and epidemiological analysis software, highlighting advantages and limitations in terms of their potential use in solving public health problems.

Geographic Information Systems in Public Health

An acknowledged definition of GIS refers to an organized set of computer technology (computer hardware, software packages, geographic and non-geographic data in digital format), methods, procedures, and personnel, designed for the capture, storage, retrieval, manipulation, display, and analysis of geographically referenced data; whose purpose is to support decision-making for the solution of problems that arise in a given geographical space.[3]

GIS software packages are sets of algorithms, methods, and automated procedures, implemented in a program with specific functions to guarantee these processes. The functions usually include the following:

- Entry, retrieval, and edition of cartographic and attribute data in different standardized formats and sources;
- Display and handling of multiple cartographic layers;
- Visualization of data in the form of maps, tables, charts, and layout;
- Management of multiple scales and projection systems, in layers and on maps;
- Creation of different types of thematic maps;
- Spatial data selection and queries, including attributes or geographical operations on one or more cartographic layers;
- Performing geographical functions, such as proximity, distance, etc.;
- Creation of buffer zones;
- Measurement of distance on maps and between geographical objects;
- Availability of a database management system that includes operations for relating nonspatial with spatial data;
- Geo-referencing and geo-coding of data;
- Capacity for development of new functions and customization of the software user interface using script languages;
- Capacity for managing and maintaining work or project sessions.

A detailed description of GIS generic functions can be found in other publications,[4] as it is not the objective of this article.

GIS programs have evolved significantly in recent years, particularly in the improvement of user interfaces from command-line-based ones to graphic interfaces that are simpler and more user-friendly. The number of functions available for data management and processing has substantially increased, and operations previously considered to be highly complex have been integrated in simple and user-friendly ways. Of the commercial GIS packages currently available, the most prominent are ArcView™, ArcGIS™, MapInfo™, Maptitude™, Idrisi™, and Geomedia™. From these, ArcView™ and MapInfo™ are the most frequently used in epidemiological studies and public health applications.

From the public health perspective, the concept of GIS includes the design, development, and use of statistical and epidemiological methods linked with GIS. It also comprises the technology in for the description and study of the magnitude and distribution of health problems in populations, health situation analysis, surveillance of health events, epidemiological analysis, in addition to planning and evaluation of interventions, management, and decision-making. Given their multidisciplinary nature, GIS in public health is closely related to disciplines such as epidemiology, biostatistics, geography, and information technology in order to achieve their objectives.

According to the above concept and in addition to the generic, GIS software in public health must include specific functions in health-related data analysis, focusing on public health problems and under its specific context and terminology. However, commercial GIS software is deficient in public health or epidemiological methods, which are the foundation for health data analysis and solving public health problems. This is because these programs development has taken a general multipurpose approach, using concepts and terminology from geography, cartography, and information sciences. Although it is possible to carry out many health data analysis operations using current GIS applications, in most cases it requires complex processes involving multiple functions, along with knowledge of conceptual and methodological frameworks of GIS, epidemiology, and public health. This could demand highly skilled users. This issue, coupled with the low levels of access to commercial GIS software owing to cost, limited availability of staff trained in GIS, and restrictions on access to cartographic data in digital formats, constitute the principal barriers to wider adoption of GIS in public health.

There have been many efforts to develop GIS in public health in an attempt to overcome these barriers. The Public Health Mapping Program, in the Communicable Disease Surveillance and Response Department (CSR), of the World Health Organization (WHO), has developed HealthMapper to respond to critical information needs for surveillance in public health and programs for the prevention and control of communicable diseases.[5] The Health Analysis and Information Systems (AIS) area, of the Pan American Health Organization (PAHO), has developed the SIGEpi software package, which is an integral part of its cooperative initiative to strengthen the analytical capacity in epidemiology and health.[3,6,7] Also, the Division of Public Health Surveillance and Informatics of the Centers for Disease Control and Prevention (CDC) in Atlanta, United States, has developed the disease-mapping tool EpiMap as part of the epidemiological analysis package EpiInfo.[8,9] These tools have played a key role at global, regional, national, and local levels.

In response to a request from the Assembly of the University Consortium for Geographic Information Science, held in the summer of 1999, Rushton et al. presented a series of proposals for improving GIS in public health, emphasizing the key themes in research on GIS in health and the need for using analytical epidemiology methods in GIS to effectively assist in determining the relationship between geographical patterns of disease distribution and environmental and social conditions.[10]

As a result, the interest of epidemiology and public health professionals in studying and analyzing geographical distribution of diseases and their relation to potential risk factors, linked to the possibility of managing spatial dimensions of epidemiological data through GIS, has stimulated the development of relevant statistical methods in both disciplines, particularly spatial statistical methods. These methods have been integrated into software packages such as DispmapWin,[11] SpaceStat,[12] DMAP,[13] SaTScan,[14] CrimeStat,[15] GeoDa,[16] ClusterSeer,[17] EpiAnalyst, and ResearchAnalyst,[18] some of which will be briefly commented on below. These developments have been the result of the intense activity aimed at developing spatial data analytical functions integrated into GIS.

Disease Mapping Tools and GIS software in Public Health

EpiInfo/EpiMap

EpiInfo and EpiMap were developed with the object of giving public health services, epidemiologists, and other health professionals an inexpensive and easy-to-use tool for epidemiological analysis, capable of managing data, analyzing epidemiological health data, and displaying results in the form of maps.

The design and development of the EpiInfo/EpiMap software system was guided by the need to cover the main functions and procedures for disease surveillance and epidemiological studies, in response to the needs of CDC and WHO. It generally consists of a group of closely related programs: MakeView permits the creation of questionnaires and databases; it offers functions for data entry, editing, validation, consistency check up of data based on the designed questionnaire. Once the data are entered and validated, the Analysis program allows the user to perform statistical and epidemiological analysis, presenting results in the form of tables and graphs. EpiReport is for combining analysis´ results and other data processing, making possible professional presentations that can be stored in HTML format for publication and distribution on the Internet.

EpiMap is the program for displaying the results of analyses carried out in EpiInfo, as well as other data and indicators, on maps. It contains a small number of GIS functions. It allows data geo-referencing from individuals and health events, based on geographical variables such as residence, work address, or other geographical references. It also provides tools for the creation of thematic maps using intervals, graduated symbols, single values, or dot-density methods. However, it does not provide functions for operations on geographic data or statistical methods, offering basically the capability of visual inspection of the geographic relations displayed on the map.

EpiInfo/EpiMap is an indispensable analytical tool for any health team or health unit engaged in epidemiological research and field studies. It offers a wide range of methods for meeting most needs in epidemiological studies, such as outbreaks or other public health analyses. It does not include methods for statistical spatial data analysis. It runs on the Microsoft Windows operating system platform (95/NT/98/2000/Me/XP) and it has good user documentation, with examples that facilitate learning the program. The software package, documentation, cartographic data, and examples, along with other related resources, can be downloaded from the CDC website (http:/www.cdc.gov/epiinfo/downloads.htm). Its online distribution is free. If manuals and other printed materials are required, the cost is low.

HealthMapper

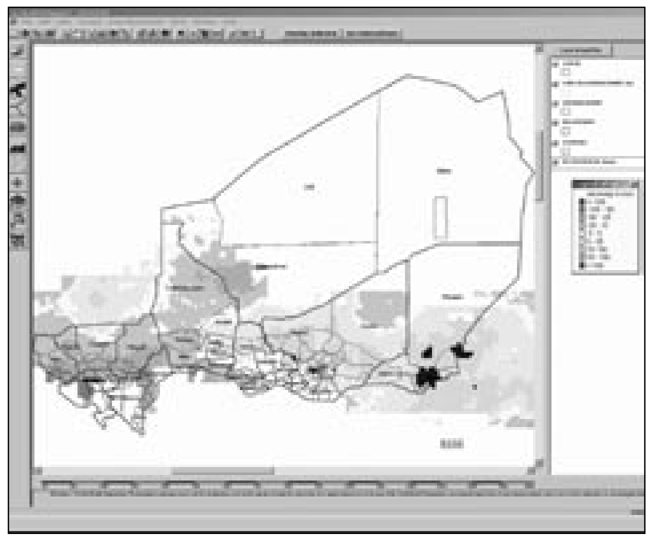
HealthMapper was developed by CSR/WHO with the initial objective of building a system for mapping Guinea Worm Disease, strengthening epidemiological surveillance of communicable diseases, through the creation of a basic database containing cartographic data and indicators commonly requested by other priority disease control programs. Its particular focus is on the African countries. The original objectives were later expanded for meeting the needs of other control programs such as malaria, HIV/AIDS, lymphatic filariasis, onchocerciasis, leprosy, epidemic diseases, and tuberculosis. It has also expanded its application in other regions of the world.

The design and development of HealthMapper was based on the following premises: i) a ready to use cartographic digital database, with administrative boundaries maps, and environmental factors such as rivers, lakes, elevation; as well as basic health data and indicators, like schools, health infrastructure services, and drinking water supply; ii) a simple data management system allowing the entry and updating of health indicators related to cartographic data in a previously standardized form; iii) an accessible and user-friendly interface with automated functions for creating maps, tables, and graphs; iv) a system that operates from local up to global levels, mainly under the conditions of countries in Africa; and v) an alternative tool that can be used free of charge or at low cost.

HealthMapper is aimed at simplifying the processes of collection, storage, updating, retrieval, and analysis of data for epidemiological surveillance, in particular, and public health, in general. It provides a basic set of geographic information system functions, mainly those for mapping and displaying thematic maps. Its main users are public health professionals and policymakers working at national and local level.

The package includes three main components: database, data manager, and mapping interface. The database component, essential to HealthMapper, is a collection of standardized cartographic data that has been developed through direct collaboration with countries' authorities and integrates data from health and other sectors. It currently contains cartographic data for most African and Southeast Asian countries, with a process for continuous and systematic data collection, updating, and standardization. It uses the MS Access database management system for data, making it possible to import and export tables in other formats. Moreover, it has a function for importing geographic data directly from Global Positioning System (GPS) receivers.

The Data Manager component has an interface that allows the user to link indicators to cartographic data for mapping and analysis. This component has three basic functions: 1) to serve as an interface for updating and maintaining the cartographic data; 2) to allow the linkage of health indicators with cartographic data; 3) to facilitate the transfer of indicators from one level of aggregation to another.



The mapping interface provides a series of functions used most frequently in mapping and basic spatial analysis in public health. This allows the user to visualize and analyze the data through thematic maps and graphics. The key functions of this component include: generation of thematic maps with intervals, graduated symbols, and dot-density format; location and selection of geographical units, measurement of distances, zoom-in and zoom-out (change of map scale); overlay of multiple layers on the map such as topographic relief, roads, rivers, health units, schools, towns, and drinking water supply points; creation of areas of influence; calculation of rates taking into account the population and the number of cases within a given radius of a selected point; creation of graphs based on selection of geographical units to evaluate trends or compare indicators in time; creation, storage, and retrieval of the most frequently used maps.

This software is a tool to support health surveillance and disease prevention and control programs. Its use in countries not included in the database requires at the first stage, the preparation of the cartographic data, health indicators and other data of interest. The indicators' database has a predetermined structure that is managed by the data manager component. In a second stage, it requires the definition of relationships between indicators and cartographic layers, all by units of analysis and levels of aggregation. Once the two stages are completed, the creation of thematic maps, health and other data included in the database is straightforward and easy to carry out. HealthMapper has a user-friendly interface that facilitates its use, with most of the functions presented in simplified form.

HealthMapper is very useful for people who need a tool to maintain a set of aggregate indicators organized by different administrative units (e.g., states/departments/provinces, municipalities, localities) and to have the ability to generate and display thematic maps. It is important to note that these functions are the ones most frequently used and requested by most health surveillance units, as well as the disease prevention and control programs. The epidemiological analysis capacities that HealthMapper offers are limited, as they are oriented more to the descriptive component of the distribution and magnitude of health risks and their determinants. It is stand-alone software, it is independent of other application platforms and requires the Windows operating system (Windows 98/NT/2000/Me/XP). Its distribution is free or inexpensive and is based on institutional agreements with CRS/WHO.

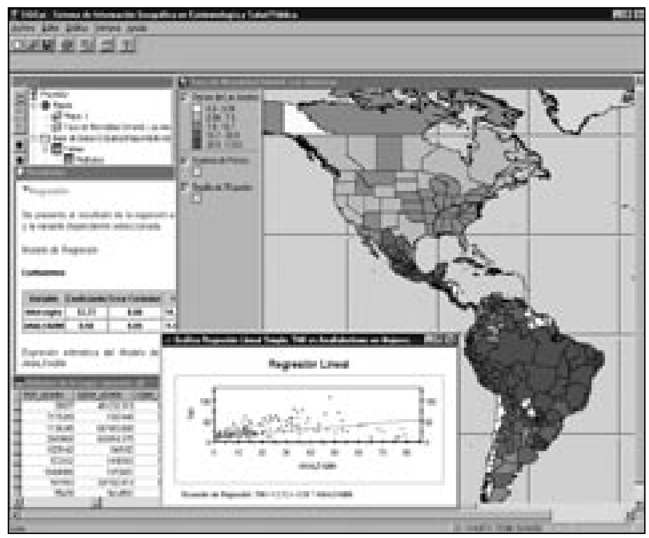
SIGEpi

SIGEpi was developed by AIS at PAHO with the purpose of contributing to the attainment of the objectives of strengthening analytical capacities in epidemiology and public health in the Region of the Americas and overcoming some barriers restricting the use of GIS in public health. Its design and development was guided by the premise of offering public health users a versatile platform-independent GIS tool that does not rely on other commercial GIS software. It includes: i) most GIS functions and some additional simplified functions, ii) methods for analyzing data on health and its determinants, including descriptive and exploratory techniques and techniques for estimates of risk measures, clustering detection, and identification of health needs and priorities, making it possible to effectively support evidence-based decision-making; iii) a friendly GIS environment, both easy to use and oriented to public health; and iv) a low-cost alternative GIS tool that can be used by health services at the local, national, regional, and global levels.

The audience and potential users for SIGEpi are public health professionals, technical personnel, and health units and services managers at different decision-making levels of the countries: as well as researchers and other professionals who need to analyze health data with geographical references. It is a platform independent of other programs, developed to run on the Windows operating system family (Windows 98/NT/2000/Me/XP).

The graphic user interface for SIGEpi contains multiple windows and types of documents, including Projects, Maps, Tables, Graphs, Results, and Layouts, each with its own specific functions, menus, and tools.

The SIGEpi program has a Relational Database Management System (RDBMS) that uses MS Access native format (.mdb) = and permits data interchange in other formats such as Dbase, Excel, Btrieve, EpiInfo, and delimited ASCII text, as well as data tables from EpiInfo version 6.x (.rec). It is an open system from the data management perspective, allowing the user to include and manipulate data without a structure established a priori. With this model, the user can create the database, create the structure for tables and/or import existing data tables produced by other standard information systems or study-specific data. Its RDBMS is integrated into the system, allowing processing of non-spatial data without the need to resort to other programs. Providing specific dialogue boxes for this purpose makes it possible for the user to interact with the database and create queries for table generation containing new variables, measures, and indicators. This approach does not require the user to know the Relational Database Model Structured Query



Language (SQL)[2] to process data, facilitating its use by less experienced professionals. Database tables can be linked with cartographic layers on the map, permitting visualization of the variables and their use in geographical operations and other analytical procedures.

The Map window is the specific interface for managing maps and cartographic layers. This offers a significant number of GIS functions, beginning with the simplest ones including the opening of layers and the definition of their graphic properties, changes in scale, interaction with the map and the extraction of information, and graphic selection of units from layers, among other things. It also has other complex functions, such as the selection of geographical units from a layer based on another layer (spatial query), and the creation of geographical layers based on geographic data from GPS receivers, as well as the raster images geo-referencing. The system for managing the geographic bases has the capacity to handle and process files in the most common formats, such as ESRI Shapefile and other standardized formats including various image formats. It also manages border files from EpiMap version 2 (.bnd). The map's window, the attribute tables for cartographic layers and graphs are dynamically linked, so a selection made in one of them is reflected in the others. The common GIS functions implemented in SIGEpi are not detailed here, since they are not the focus of this article and can be found in other publications .

SIGEpi provides the following methods for analysis of health data:

- Descriptive statistics: This includes the frequency distribution calculation, in addition to measures of central tendency and dispersion, all useful during data exploration. Contains also correlation analysis functions, which allow the identification of associations among variables and determination of co-linearity between indicators, an important step for selecting indicators to include in a model. Also has available linear regression analysis, which offers the capacity to build simple and multiple linear regression models, supporting ecological data analysis.

- Risk estimates: This includes crude and specific rates, ratios, and proportions calculation; rates standardization by direct and indirect methods, which is important for adjusting risk estimates for confounding factors; spatial rates smoothing and spatial estimates for standardized of mortality and morbidity rates based on Bayesian methods, which are useful for eliminating variability in data when working with small areas, with uncommon diseases, and generally speaking, diseases with small numbers. These methods are also useful when the purpose is to improve the risk estimate power and help visualize the spatial trend of an estimate.

- Identification of critical areas and population groups: This simple interface assists in the construction of a complex conditional expression with several variables and indicators, facilitating identification of geographical units and population groups sharing the worst conditions. The use of different classification methods facilitates the selection of cut-off values for each element in the expression. The identification of critical areas can be complemented with geographical conditions applying the cartographic layer's queries functionality.

- Construction of Composite Health Index: This helps the prioritization of health needs and interventions through a method that constructs a composite index by standardizing indicators in different measurement units. A simple interface permits the selection of indicators to include in the model, defining the direction of each with respect to the index being constructed. This is useful in determining unmet health needs and priority areas and population groups in greater need. It is also possible to give standardized weights to the indicators included in the model. The Composite Health Index makes it possible to synthesize information on a health problem from several dimensions, and they has diverse uses.

- Spatial Clustering. These methods are relevant for epidemiological surveillance and public health, particularly for detecting clusters of geographic units in which a parameter under observation deviates from the expected value and the identification of risk factors and contextual determinants for etiological hypotheses formulation. For identifying the presence of global spatial clusters, the Moran's I and Geary's c global spatial autocorrelation indexes are applied. Calculations of local indices of spatial autocorrelation (LISA) are also included to identify the location of clusters of high and low values in aggregate data. Also included in this section is a method for detecting spatial-temporal clustering using the Knox test, which permits measurement of association in space and time for disaggregated or individual health events.

- Measurement of association between environmental exposure factors and health effects, applied to epidemiological studies at the individual level. This method uses the locations on the map of both individual case data and environmental or ecological factors in the area studied and permits calculation of different epidemiological measures of association between exposure to the environmental factor and the health effect. The results include confidence levels, contingency tables, and stratified analysis, if required by the user. The measures of association are useful in epidemiological cohort and case–control studies.

All the functions and methods have interfaces whose design employs the perspective, language, and conceptual framework of epidemiology and public health. SIGEpi provides a significant group of GIS functions, although it has limitations in editing geographic databases. Nevertheless, it offers functions for the generation of cartographic point or lines layers from tables with variables of coordinates of latitude and longitude, along with data derived from GPS receivers.

SIGEpi is useful for anyone needing health related GIS applications aimed at: 1) studying the number and spatial distribution of health events in the population; 2) identifying critical population groups and geographical units; 3) estimating the risk of becoming ill or dying; 4) conducting analyses of the association between risk factors and determinants with possible health effects; 5) identifying and laying out etiological hypotheses; 6) carrying out analyses of health status; 7) monitoring health events and their determinants; 8) prioritizing and targeting health actions and interventions; and 9) monitoring and evaluating health interventions. Although SIGEpi is designed for applications in epidemiology and public health, it can be used as a GIS tool in other disciplines and sectors.

Its distribution is based on an inexpensive license. The software package and other related materials can be downloaded from the Internet at http:/ais.paho.org/sigepi/index.htm?xml=sigepi/soporte.htm. The related materials include a user's manual, an on-line help system, and case studies with companion data sets that describe step by step how to carry out procedures for solving public health problems.

AIS at PAHO will continue to provide support for the development of SIGEpi, given its significance as a technical cooperation instrument for strengthening analytical capacity in epidemiology and public health along with evidence-based decision-making.

DispmapWin

DismapWin is a program for advanced statistical analysis of epidemiological data, developed by Schlattmann of the Benjamín Franklin University Clinic, Free University of Berlin. This program permits the mapping of health data and provides measures of risk such as crude rates and ratios. The analytical methods it offers include the analysis of unobserved heterogeneity in epidemiological data using mixed models, also called hierarchical or multi-level analysis. It also offers the capacity to carry out ecological analysis using the Poisson regression model, adjusting the spatial dependency of the independent variables through a mixed regression model.[20, 21]

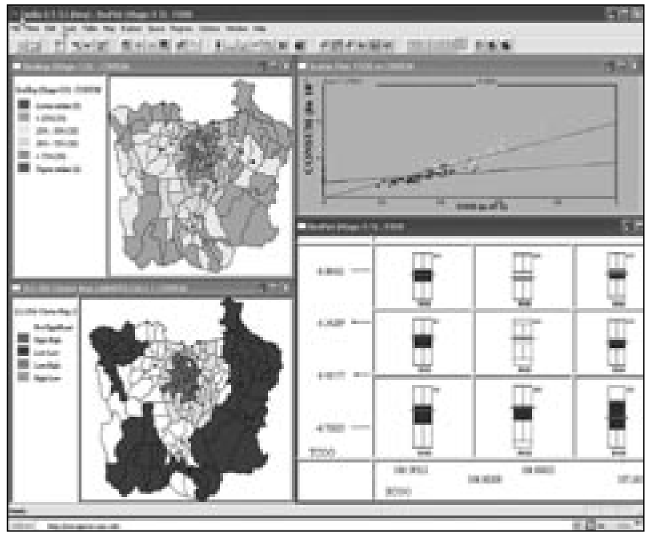
The analytical functions of this program can be grouped into two types: i) descriptive spatial statistical measures; and ii) statistical modeling. In the first category, methods are presented for identification of the presence of clusters and spatial heterogeneity through the Moran and Ohno-Aoki tests. These methods allow detecting the existence of dependency or spatial heterogeneity in measures of effects and health risks, providing indications on the existence of exposure factors in specific population groups and geographical units. For statistical modeling, the program provides mixed regression models, which make it possible to adjust for the effects of dependencies and spatial heterogeneity of data when modeling the relations of possible risk factors and a measure of effect.[20]

DismapWin reads data files in ASCII and dBase III format and uses the boundary file format of EpiMap version 2 (.bnd), it can handle basically geographic data in the form of areas or polygons. It includes limited GIS functions, restricting them to the display of thematic maps with intervals using percentile classification. Risk estimates based on mixed regression models are displayed in thematic maps, including maps of levels of significance. This program is platform-independent, not requiring other programs. It is free and can be downloaded from its website.[11] From the standpoint of application and use, it is relevant to processes for spatial statistical modeling of health data.

GeoDa

GeoDa is a program designed for exploratory analysis of discrete spatial data in the form of points and polygons, offering an important group of statistical methods for spatial data analysis. Its main objective is to provide a user-friendly graphic environment with a natural route for empirical spatial data analysis, beginning with mapping and simple visualization, continuing to exploration and analysis of spatial autocorrelation, and ending with spatial regression. It was developed by Anselin and co-workers at the Spatial Analysis Laboratory of the Department of Agricultural and Consumer Economics at the University of Illinois, Urbana-Champaign.[16]

This software has its origin in the first efforts to bridge the statistical software packages with ArcInfo GIS program. These efforts led to the development of SpaceStat and



its extension for ArcView,[23] and DynESDA and its extension for ArcView, which introduced the concept of dynamically linked windows (“linking and brushing”) in a GIS environment. These solutions required ArcView for construction of the neighborhood matrices and spatial weights..

GeoDa’s design consists of an interactive environment that combines maps with statistical graphs, using the technology of dynamically linked and brushed windows. In general terms, its functions can be divided into six categories: 1) utilities and manipulation of spatial data: data input, retrieval, and conversion; 2) transformation of data: transformation of variables and creation of new variables; 3) mapping: creation of thematic maps, cartograms, and animated maps; 4) exploratory data analysis: creation of several kinds of statistical graphs, such as histograms, box diagrams, and scatter diagrams; 5) spatial autocorrelation: statistics of global and local spatial autocorrelation with inference and visualization; and 6) spatial regression: diagnosis and estimates of parameters for spatial regression models. More details on the design and functions of GeoDa can be found in other publications.

GeoDa has capacities for generation and visualization of maps, although its GIS functions are very limited. It is a tool that meets several analytical needs in epidemiology and public health. However, the terminology and orientation are focused on professionals in disciplines such as statistics, spatial analysis, and econometrics in general.

This is an independent platform program that runs on the Microsoft Windows operating system (Windows 98/NT/2000/Me/XP), as well as on virtual emulators for Windows on Mac operating systems (MacOS 9 and MacOS X). It is free and it can be downloaded from the Internet. It has a good documentation; its user’s manual adequately describes the use of the functions and details the implemented methods, uses, and results interpretation.

Programs for Statistical Analysis of Spatial Data Requiring GIS

DMAP

The Disease Mapping and Analysis Program (DMAP) estimates disease or death rates as a measure of epidemiological risk in a geographical area of study. It was developed by the University of Iowa's Department of Geography . The risk estimates are done using the concept of spatial filters and statistical tests using Monte Carlo simulation to determine the significance of the estimates. It uses the approach of constructing continuous risk distributions in a geographical area based on discrete observations.

DMAP calculates rates for a given area using individual and aggregate data. Individual data are data from cases referenced to a geographical location, using a pair of geographic coordinates as variables. Aggregated data refers to the number of cases and population per geographical unit, such as cities, census tracts, municipalities, or other administrative areas. When data refer to areas, the coordinate pair of the polygon's centroid is used.

In general terms, the method consists of overlapping a grid of equidistant points uniformly distributed over the entire geographical area and counting, for each point, the number of cases and population at risk within a given radius. With these values, the measure of risk and its statistical significance is estimated using the Monte Carlo simulation method. The results are placed in three files in ASCII text format, one for the grid, one for the risk estimates, and a third for the statistical significance of the estimates.

In order to estimate the continuous distribution of risk over a geographical area, it is necessary to process the results produced by DMAP in a GIS and apply interpolation techniques, using the Kriging method, and finally, construct contour lines to represent risk areas. This process requires use of the GIS software ArcView, with the Spatial Analyst and 3D Analyst extensions. This is a constraint for many users, due to the cost of the required software and the need for familiarity with geo-statistical techniques.

DMAP is also an independent platform program and runs on the Windows operating system. It is free and can be downloaded from the website of the University of Iowa's Department of Geography.

SaTScan

The SaTScan program was developed to analyze health event data in time, space, and space-time using scan statistics. Both the software and the statistical method were developed by Kulldorf. Its purposes are i) to support epidemiological disease surveillance by detecting clusters of health events in space, time, or time and space, and to confirm their statistical significance; ii) to test whether a disease is randomly distributed in space, time, or space and time; iii) to evaluate the statistical significance of clusters of high-risk diseases, low-risk diseases, and combinations thereof; and iv) to perform procedures for early detection of epidemics in epidemiological surveillance on a periodic and repetitive basis.

SaTScan uses two types of models, one based on Poisson, where the number of events in an area is assumed to follow a Poisson statistical distribution according to a known population at risk. In this case the data are aggregated by geographical units; and another model based on Bernoulli for individual event data or for individuals with values of 1 and 0 to identify cases and controls.

The program uses different types of data, depending on the required analysis. The formats used by the SaTScan program for data files for import and export are ASCII text and dBase. Files with the resulting identified clusters can be imported to a GIS for display on maps. Management of the SaTScan results files by a GIS requires the import and display process to be implemented in terms of thematic maps, with the production of cartographic files for the circles representing the clusters.

SaTScan is an independent program platform that runs on the Windows operating system. It is free of charge and it can be downloaded from the Internet. The user documentation is of good quality, facilitating the learning of the software, including methods and interpretation of results. It also has a long list of bibliographic references on the methods used and several examples of applications in health.

CrimeStat

CrimeStat® is a spatial statistical program specifically designed to analyze the occurrence of criminal incidents. It was developed by Levine under the auspices of the National Institute of Justice of the United States of America. Its purpose is to provide a set of complementary statistical tools to help law enforcement agencies and criminal justice investigators in their efforts to map criminal activity.[15, 28]

CrimeStat uses several types of data entry files, a primary file and a secondary file. Both contain the location of incidents in the form of pairs of coordinates, as well as the date each incident occurred. The secondary file contains data associated with the primary one and can be used for comparison purposes in methods for detecting clusters of nearest neighbors adjusted for risk and with dual kernel interpolation. It also uses a file that defines a regular or irregular grid which can be overlaid on the area of study. This grid can be created by CrimeStat or by a GIS.

The spatial statistical methods used in CrimeStat are: 1) spatial distribution: a set of methods for describing the spatial distribution of incidents, such as the mean center, center of minimum distance, standard deviational ellipse, and Moran's I spatial autocorrelation index; 2) distance analysis: a set of statistical techniques to describe the properties of distances between incidents, including nearest neighbor analysis, linear nearest neighbor analysis, and Ripley's K statistic; 3) 'hot spot' analysis: includes routines for conducting 'hot spot' analysis, including the mode, the fuzzy mode, detection of hierarchical nearest neighbor clustering and risk-adjusted hierarchical nearest neighbor clustering, as well as other routines for the Spatial and Temporal Analysis of Crimes (STAC), including detection of K-means clustering and the Local Indicators of Spatial Autocorrelation (LISA) proposed by Anselin; 4) spatial modeling, which includes interpolation methods using single-variable kernel density estimation (e.g. thefts in dwellings) to produce a surface or contour estimate of the density of incidents, as well as dual-variable kernel density estimation for comparing the density of incidents to the density of a baseline (e.g. thefts in dwellings with respect to the total number of dwellings) and other techniques for spatial-temporal analysis, such as the Knox index and the Mantel indices, which make it possible to detect association among incidents in time and space, as well as other methods with greater application in analysis of crimes.

The program uses the location of criminal incidents (e.g., location of thefts) as data entered, in dBase ('dbf'), Shapefile cartographic ('shp') or ASCII text format files. Based on these data, it makes it possible to apply spatial statistical methods, yielding results in cartographic format files that can be used directly by different GIS programs such as ArcViewTM, MapInfo®, Atlas*GISTM, SurferTM for Windows, and ArcView Spatial AnalystTM. The program does not have the capacity for visualization in maps or for GIS functions. Its design is centered on methods for spatial data analysis, allowing the results to be presented through a GIS. An important characteristic from the programming standpoint is the availability of an Application Programming Interface (API) that allows other programs to invoke its procedures and functions.

Most of the methods in the CrimeStat program are applicable to the analysis of spatial health data and public health surveillance. The spatial distribution functions can be used to describe the occurrence of disease or death at the individual level, helping to visualize and characterize these health events. They can answer questions such as whether the cases are scattered or concentrated, where the events are concentrated, and whether the cases are occurring in proximity to a source of contamination or an environmental risk factor. The functions for distance analysis help describe the parameters and properties of distances between health events, providing measures such as minimum and maximum distance, mean distance, and mean distance from nearest neighbors. The analysis of hot spots, also called cluster analysis, makes it possible to identify where there are concentrations of health events greater than one would expect by chance. When one has aggregate data by geographical units, the Local Indices of Spatial Autocorrelation method serves the same purpose. Spatial modeling methods also have application to public health, particularly when the concern is to estimate continuous surfaces for the risk of becoming ill or dying based on discrete measurements.

CrimeStat is used extensively by police departments in the United States as well as by other criminal justice agencies. It is an independent platform program that runs on the Windows operating system (Windows 98, NT, 2000, XP). It is free of charge and can be downloaded from the Internet. Its functions are well documented in its user's manual, and the program comes with data sets that facilitate its learning and the interpretation of results.

EpiAnalyst

EpiAnalyst is an extension of the ArcView 3.x software, which allows use of statistical spatial data analysis methods within the ArcView environment. To do this, it provides links with programs like DMAP, SaTScan, CrimeStat, S-PLUS, EpiInfo, and other ArcView extensions, such as Arc-SDM (Spatial Data Modeller), Spatial Analyst and 3D Analyst. It also offers other support utilities for analysis such as spatial union between cartographic layers, interpolation of variable values in geographical units in the form of areas, and creation of Thiessen polygons.

The EpiAnalyst extension was developed by the Public Health Research Laboratory.[18] It offers the advantage of providing access to most spatial data analysis methods from within the ArcView environment. However, the drawbacks to its use are its cost and the license requirements for commercial software packages such as ArcView and/or ArcMap and the Arc-SDM, Spatial Analyst, and 3D Analyst extensions, as well as the S-PLUS software package. These costs are prohibitive for the majority of users at the community and local levels in the health sector.

Conclusions

There is currently a variety of software and tools for spatial data analysis that can be used in public health, ranging from GIS programs to programs specifically for analysis. Although there is still much to be done for the adoption and effective use of these programs for solving public health problems, the achievements to date offer a range of alternatives for technical teams in the health sector.

EpiMap and HealthMapper are feasible alternatives for technical teams that require basic GIS functions. If more complex functions and/or analytical methods are needed, other alternatives should be chosen. SIGEpi stands out for offering a broad GIS platform with functions for statistical spatial data analysis integrated into the program. It is notable for providing specific methods for different analytical approaches in public health and its low cost. It is the alternative for analysis and support for public health decision-making with the greatest potential for adoption in the health sector. The GeoDa program is a feasible alternative that offers a significant set of exploratory and analytic methods and functions for the display of geographic data in an independent environment platform.

From the set of the specific programs for spatial data analysis described in this article, some require the use of separate GIS software to some extent. DMAP has greater dependency on ArcView and Spatial Analyst for full use of its analytical methodology. However, in the cases of SaTScan and CrimeStat, results can be displayed with any GIS software, providing the opportunity for using a free or low-cost GIS. This gives them an edge over others that require a commercial GIS.

Technical teams in public health should evaluate the validity of tools available, considering their own needs. They should take into account that, in general, a set of tools and programs will be needed, since it is unlikely that a single tool will offer the complete solution for their analytical needs.

Bibliographic references

1. Issue Focus: Geographic Information Systems in Public Health, Part 1. J. Public Health Management and Practice. Vol 4. No. 1, 1999.
2. Issue Focus: Geographic Information Systems in Public Health, Part 1. J. Public Health Management and Practice. Vol 5. No. 2, 1999.
3. Organización Panamericana de la Salud (OPS). Cooperación técnica de la OPS en sistemas de información geográfica aplicados en epidemiología (SIG-EPI) en las Américas. Boletín Epidemiológico/ OPS, 17(2):8-10, 1996.
4. Castillo-Salgado C., et al. Sistemas de Información Geográfica en Salud: Conceptos Básicos. Washington, DC.: PAHO, 2003.
5. Thrall, SE. Geographic Information System (GIS). Hardware and Software. J Public Health Management Practice, 5(2), 82-90, 1999.
6. World Health Organization. Herramienta HealthMapper. Sitio en Internet: http://www.who.int/csr/mapping/tools/healthmapper/healthmapper/en/ , revisado el 10 de Diciembre del 2004.
7. Martinez Piedra R., et al. SIGEpi: Sistema de Información Geográfica en Epidemiología y Salud Pública. Boletín Epidemiológico/OPS, 22(3):4-5, 2001.
8. Dean, A.G. EpiInfo and EpiMap: Current Status and Plans for EpiInfo 2000. Journal of Public Health Management Practice, 5(4), 54-57, 1999.
9. CDC. EpiInfo. Sitio en Internet http://www.cdc.gov/epiinfo/ , revisado el 10 de Diciembre del 2004.
10. Rushton G, Eles G, McMaster R. Considerations for Improving Geographic Information System Reseca in Public Health. URISA Journal, Vol 12, No. 2: 31-49, Spring 2000.
11. DismapWin. Sitio en internet http://ftp.ukbf.fu-berlin.de/sozmed/DismapWin.html , revisado el 10 de Diciembre del 2004.
12. TerraSeer. SpaceStat. Sitio en Internet http://www.terraseer.com/products/spacestat.html , revisado el 10 de Diciembre del 2004.
13. Deparment of Geography, University of Iowa. Disease Mapping and Análisis Programa (DMAP). Disponible en sitio en Internet [http://www.uiowa.edu/~geog/research/dmap/index.htm] , revisado el 10 de Diciembre del 2004.
14. SaTScan. Spatial and Temporal Scan Tests Program (SaTScan). Disponible en sitio en Internet [http://www.satscan.org/] , revisado el 10 de Diciembre del 2004.a
15. National Archives of Criminal Justice Data. CrimeStat Spatial Statistics Program: Version 2.0. Disponible en [http://www.icpsr.umich.edu/NACJD/crimestat.html] , revisado el 10 de Diciembre del 2004.
16. Spatial Análisis Laboratory, Department of Agricultural and Consumer Economics, University of Ilinois at Urbana Champaigne. Programa GeoDa. Disponible en sitio en Internet [http://sal.agecon.uiuc.edu/geoda_main.php] , revisado el 10 de Diciembre del 2004.
17. TerraSeer. ClusterSeer. Disponible en sitio en Internet [http://www.terraseer.com/products/clusterseer.html] , revisado el 10 de Diciembre del 2004.
18. Public Health Research Laboratories. Disponible en sitio en Internet [http://www.phrl.org/] , revisado el 10 de Diciembre del 2004.
19. Anselin L. Computing environments for spatial data análisis. J Geograph Syst, 2:201-220, 2000.
20. Schlattmann P; Dietz E; Böhning D. Covariate adjusted mixture models and disease mapping with the program DismapWin. Stat Med15(7-9):919-29, 1996.
21. Schlattmann P. The Computer Package DismapWin. Comments in: Stat Med 1996; 15(7-9):931.
22. Anselin, L. SpaceStat, a Software Program for Analysis of Spatial Data. National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara, California. 1992.
23. Anselin L and Bao S. Exploratory Spatial Data Análisis linking SpaceStat and ArcView. In: Fisher M, Getis A (eds). Recent development in spatial análisis. Springer. Berlin, Heidelberg, New Cork. 1997.
24. Anselin L, Syabri I and Youngihn Kho. GeoDa: An Introduction to Spatial Data Analysis, Geographical Analysis, [En prensa] Disponible en sitio en Internet [http://sal.agecom.uiuc.edu/pdf/geodaGA.pdf] , revisado el 10 de Diciembre del 2004.
25. M.A. Weinstock. “A generalized scan statistic test for the detection of clusters.” International Journal of Epidemiology 10:289-293, 1981.
26. Kulldorff M. and Information Management Services, Inc. SaTScanTM v6.0: Software for the spatial and space-time scan statistics. Sitio en Internet http://www.satscan.org , revisado el 10 de Diciembre del 2004.
27. Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods, 26:1481-1496, 1997.
28. Ned Levine. CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 2.0). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. May 2002.
29. Levine, N. (2004). The CrimeStat program: Characteristics, use and audience. In: Proceedings of the SCISS Specialist Meeting “New Tools for Spatial Data Analysis”. Santa Barbara, California, USA. May 10-11, 2002. Disponible en sitio en Internet: [http://www.dpi.inpe.br/gilberto/csiss/papers/levine.pdf] , revisado el 10 de Diciembre del 2004.

Note: The Pan American Health Organization does not endorse, neither have any affiliation with the following GIS programs: ArcView and ArcGIS (trademarks of ESRI), MapInfo (trademark of MapInfo Corp.), Maptitude (trademark of Caliper Corp.), Idrisi (trademark of Clark Labs, Clark University) and Geomedia (trademark of Intergraph).

This paper has been prepare by Ramón Martínez-Piedra, Enrique Loyola-Elizondo, Manuel Vidaurre-Arenas, and Patricia Nájera Aguilar, members of the Health Analysis and Information Systems Area of Pan American Health Organization.

Health Inequalities

The two core values on which the work of the Pan American Health Organization (PAHO) revolves are equity and Pan Americanism. These values form the basis for cooperation with countries, and measuring and monitoring inequalities are therefore fundamental activities for decision-making. Technical cooperation with countries should focus on the identification of inequities and on the formulation of effective strategies for reducing and, ultimately, eliminating them.

Although important gains have been made in the overall health situation in the Region, the Americas remains the region with the greatest inequity in income distribution in the world. There are still large disparities in health status between different countries and social groups. It is known that groups with lower socioeconomic status not only suffer a greater burden of disease, but, in addition, they tend to develop chronic diseases and disabilities at earlier ages; they have less access to health services, and the services they do receive are of poor quality.

The measurement of health inequalities is an indispensable condition for moving forward in the effort to improve the health situation in the Region, where the analysis of average values is no longer sufficient. This type of analysis is a key tool for action aimed at achieving greater equity in health. In general, health information systems and health situation analyses do not include evaluation of inequalities. Measuring inequalities in health and living conditions constitutes the first step toward the identification of inequity in the field of health. According to Whitehead (1991) and Schneider (2002), inequality is not synonymous with inequity. Inequity is an unfair and avoidable inequality, and therein lies its importance for decision-makers. Classifying an inequality as inequity implies knowing its causes and being able to substantiate a judgment as to the unfairness of those causes.

In light of the importance of the matter, this issue of the *Epidemiological Bulletin* presents a summary of the special issue on measuring health inequalities published by the *Pan American Journal of Public Health* in December 2002 (Vol. 12, No. 6). In addition, it revisits one of the articles, Methods for Measuring Inequalities in Health, by Schneider et al., which presents a simple explanation of the various methodologies for examining health inequalities. The article by Schneider et al. provides a guide for calculating indicators and discusses the advantages and disadvantages of each of them. This issue of the *Epidemiological Bulletin* includes a summary of the introduction and the type of indicators and their characteristics; subsequent issues will present the methodology for calculating the indicators most frequently used to measure inequalities in the field of health, including rate ratio and rate difference, effect index, population attributable risk, index of dissimilarity, slope index of inequality and relative index of inequality, Gini coefficient, and concentration index.

Outline of the Inequality Journal of Health

Inequities in health are the differences in the levels of health among distinct socioeconomics groups that are considered unfair, on the basis of a detailed judgment of their causes.[1] Similarly, Kawachi, Subramanian & Almeida-Filho[2] define inequities in health as the inequalities in health that are considered unfair or stemming from some form of injustice. Common to the different published definitions of inequities in health, is the underlying assumption of injustice.[3] Then, to evaluate inequities in health, there is the need for making possible the measurement of fairness in a given society. Although some methodologies have been proposed, all are based in a necessary judgment of value and are dependent of the adopted theory of justice and the accepted explanations for the etiology of the observed inequalities, involving therefore a political conception.[2,3] Consequently, the International Society for Equity in Health adopts, as operational definition, that inequities are "systematic [and potentially remediable] differences in one or more aspects of health status across socially, economically, demographically or geographically defined populations or population sub-groups".[4] A readily measurable dimension of this concept, that has allowed an approach of the iniquities in health in a society, is that of inequalities in health. These are understood as a generic term that involves population differences, variations and disparities in the health achievements of individuals and groups that need not imply moral judgment of these differentials, nor strict considerations on their remediation.[2]

Recently, a great interest in the subject of inequities and inequalities in health has been noticed. Several studies have pointed out the important magnitude of health inequalities in countries with varying degrees of development. In addition, and more disturbing, is that regardless of a general improvement of the average population health conditions, an increasing trend of health inequalities occurred in recent years in many regions. The importance of health policies as instruments to correct and decrease those inequalities in highlighted each time more frequently; however, when existing inequalities are ignored, policies may influence the health sector in such a way that the sector could actually constitute another determinant of the amplification of health inequalities in a society. In this regard, the concept of equity in health services must also be mentioned, which implies that there should be no differences in health services where health needs are equal (horizontal equity) or that improved health services be provided where greater health needs are present (vertical equity).[5] Obviously, the issue of inequalities in health and their marked socio-economic determination is not a subject that only concerns the health sector, both with respect to its conceptual frame, as well as from the perspective of the identification and development of effective interventions for their solution. Due to the variety of determinant factors involved in the causal net of health inequalities, a muti-sectoral approach, with policies, programs and interventions leading to reduction or limitation of the negatives effects on health, are required.

The subject of the inequalities in health has been an important object of attention of the Pan American Health Organization (PAHO), integrating the concept in its mission for strengthening the use and analysis of information on inequalities for public health management in the Americas. In addition, several initiatives to stimulate the theoretical-conceptual debate on this subject have been adopted and a great effort devoted to promote the institutionalization of evidence-based practices that aim at equity in health among population groups of our societies. Among these practices, the need for monitoring and surveillance of health inequalities and the assessment of their etiology, which frequently show important local specificities, are highlighted. Accordingly, PAHO has made available technical information, methodologies and other materials to facilitate the approach of health inequalities within the practice of the health services, and for the sensitization of the health professionals, in particular decision makers, regarding this theme.

One of PAHO's recent specific initiatives in this direction has been the publication of a special issue of the Pan American Journal of Public Health on "Measuring Health Inequalities" that was distributed in December of 2002. This issue included six original articles and three current topics. Moreover, it contained a special guideline on "Methodologies for measurement of health inequalities". The following section presents a brief commented summary of these articles.

The articles by **Barbosa, J. & Barros, M.B. ["Epidemiologia e desigualdades: notas sobre a teoria e a história" – article in portuguese] and by Starfield, B. ["Equity in health: perspective on non-random distribution of health in the population" - current topic in english]** present a conceptual historical framework on the subject. Barbosa & Barros introduce the history of the precursory thoughts that lead to the current conceptualization of the health inequities that is amply discussed on the basis of a relevant bibliographic review. The article includes historical explanations of disease causation in human populations and its inherent health inequalities, as well as the development of the role of Epidemiology in this context. The article by Starfield, starts by indicating that heath inequalities do not occur randomly, and discusses the determinations of clustering of health-related problems identified in human populations. The author proposes a complex net of causality of health status and determinant risk factors that serves as the basis for understanding that health events aggregate in a systematic way (and not at random). She also points out that it is very unlikely that this complex causal net acts in the same way in all areas/regions and calls attention on to the need for further studies on the genesis of health inequalities. She concludes suggesting that the information and knowledge originated from such studies be translated into effective actions for problem solving.

Another article adds a reflection on the situation of inequalities in health in the Americas, including examples of different methodological approaches used for their measurement, presented by **Alleyne, G.A.O, et al. ["Overview of social health inequalities in the region of the americas using different methodological approaches" - article in english]**. In this work the authors carry out a descriptive analysis of the health situation in the region of the Americas, followed by an exploratory approach of the social inequalities in health among countries in this Region. Highlighting the great disparities in health status and in some of their determinants, the authors indicate that the search for equity in health is not only a theoretical matter, but a concrete reference for technical cooperation among countries, once that equity is recognized as an imperative assumption required to achieve an integral human development in the Region.

As examples of analyses of health inequalities in countries of the Americas, the journal includes the articles by **Duarte, E.C., Et al. ["Expectativa de vida ao nascer e mortalidade no brasil em 1999: uma análise exploratória dos diferenciais regionais" – article in portuguese] and Gattini, C; Sanderson, C. & Castillo-Salgado, C. ["Variaciones de indicadores de mortalidad evitable como aproximación a desigualdades en salud en comunas chilenas" – article in spanish]**. In the work carried out in Brazil (Duarte et al), the inequalities in health among the Brazilian regions and states were studied, with emphasis on life expectancy at birth, infant mortality and child mortality (children less than 5 years of age) according to specific causes (diarrheal and acute respiratory diseases), and mortality rates due to external causes (homicides and traffic accidents). Ecological associations between the selected health indicators and socio-economic and demographic indicators were analyzed. The study results indicated a pattern of health inequalities in Brazil with a marked intra-national and intra-regional polarization, as well as juxtaposition of diseases related to both underdevelopment and development. This complex situation demands more specific actions from the health sector aiming to minimize these inequalities. The article by Gattini, Sanderson and Castillo-Salgado analyzes the variation of preventable mortality indicators, as an approach to assess the magnitude of health inequalities, among distinct geographical areas. They used for this methodological approach, health data from small geographic areas in Chile. The authors observed significant inverse (negative) associations between preventable mortality indicators (potential years of life lost, preventable mortality rates and infant mortality rates) with those of socio-economic development and discussed the implications of their findings for policies on prioritization to those areas with the worst situation and on increased coverage to prevent the highest number of deaths.

The special issue of the journal includes, in addition, some useful methodological experiences for the study of health inequalities that are described as follows.

In the article on 'methodologies for the measurement of health inequalities' **[Schneider, MC et al. "Metodologias para la medición de desigualdades en salud"– report in spanish]** the authors discuss appropriate indicators of health situation, health services and socio-economic status for studies on health inequalities. In addition, they describe selected basic methodologies for measuring health inequalities, which may be useful for the health services settings at their various levels. The circumstances and levels of application as well as advantages and limitations of the different measurement methods are also discussed.

Using a theoretical and practical approach, **Bacallao, J. Et al ["Indicadores basados en la noción de entropía para la medición de las desigualdades sociales en salud" – current topic in spanish]**, describe the advantages for measuring inequalities based on the notion of entropy. This concept has its origin in the Physics, the Statistics and the Information Theory. In this article, the authors started with a review of the classic indicators used for measurement of health inequalities, and then describe the indices based on the notion of entropy. Regarding this last aspect, the definitions and the properties of such indices are discussed and examples of their use for measuring health inequalities included. The authors' conclusions are in favor of the usefulness of these indices in the field of inequalities, considering that some of their properties are unique to this end.

The article by **Metzger, X ["Agregación de datos en la medición de desigualdades e inequidades en la salud de las poblaciones" – article in spanish]** had as its objective to determine the consequences and adequacy of using different levels of data aggregation for measuring of inequalities in the health situation of populations. The author exemplified the calculation of the most frequently used measures, having the infant mortality rate from Costa Rica as health indicator. Metzger discusses the need for considering the benefit vs. loss trade-off when opting between a higher or a lower level of aggregation of geographical units used in health inequality studies. With the Costa Rica example, the effects on the consistency of results obtained from studies that use larger aggregates are described. The author concludes that some measures generate important result discrepancies according to the different levels of aggregation utilized and recommends considering the study objectives to guide an adequate selection of methodologies to be used.

In another article, **Loyola-Elizondo E., Et al. ["Los sistemas de información geográfica (sig) como herramienta para el monitoreo de las desigualdades en salud" – article in spanish]**, discuss the usefulness of Geographic Information Systems (GIS) for monitoring health inequalities, highlighting their capability for data and information integration from different sources and types and their subsequent processing. This allows simplifying, improving speed and automation of epidemiological analyses at diverse levels of aggregation. To illustrate the properties of GIS, the authors used as example the inequalities on the infant mortality rates (because it is one of the indicators with broadest coverage) in the countries of the region of the Americas, according to socio-economic indicators, analyzed in three different levels of aggregation (regional, national and local). They conclude that, based on the definition of the magnitude and distribution of health events and their determinants, the adequate use of GIS in the study of health inequalities contributes to facilitate public health management. In addition, GIS enables focalizing and planning of interventions in the high priority areas and groups, among other applications.

The compilation of these articles in the special issue of the Pan American Journal of Public Health constitutes a useful instrument to initiate the debate on the need to include the subject of health inequalities in the agenda of decision makers in the health sector.

The articles may be reviewed in full in their electronic version at the following Internet address: http://www.scielosp.org/scielo.php?script=sci_issuetoc&pid=1020-498920020012&lng=en&nrm=iso

References:

(1) Kunst AE & Mackenbach JP. Measuring socioeconomic inequalities in health. Copenhagen: WHO Regional Office for Europe, 1994. 115 p.
(2) Kawachi I, Subramanian SV & Almeida-Filho N. A glossary for health inequalities. J. Epidemiol Community Health. 2002; 56:647-652.
(3) Macinko J & Starfield B. Annotated Bibliography on Equity in Health, 1980-2001. International Journal for Equity in Health. 2002; 1(1).
(4) International Society for Equity in Health. The Toronto Declaration on Equity in Heath. Conference Statement: 2nd International Conference. Toronto, June 2002.
(5) Starfield B. Improving equity in health: A research agenda. International Journal of Health Services. 2001; 31(3):545-566.

Methods for measuring health inequalities (Part I)

Maria Cristina Schneider, Carlos Castillo-Salgado, Jorge Bacallao, Enrique Loyola, Oscar J. Mujica, Manuel Vidaurre, and Anne Roca.

Measuring health inequalities is essential for analyzing the determinants of health and advancing a theory, which, in turn, is fundamental for action. Nevertheless, how to carry out these measurements is a subject of debate. There are various measurement methods, with differing levels of complexity, and choosing one rather than another depends on the objective of the study. The purpose of this article is to familiarize health professionals and decision-makers with the methodological aspects of measurement and the simple analysis of health inequalities, utilizing basic data that are reported regularly (for example, mortality, morbidity, and resources), aggregated by geopolitical unit (for example, a country or a state). However, the methods presented are applicable to the measurement of various types of inequalities and at different levels of analysis.

TYPES OF INDICATORS

Methodological considerations

Two areas for analysis of inequalities can be identified: health status and health services. Indicators for the measurement of health status basically use morbidity and mortality data; many of the studies published to date have used secondary mortality data or surveys. Measurements of inequalities in the area of health services use mainly survey data and incorporate concepts such as need, access, efficacy, effectiveness, and others that require a somewhat more complex methodology. This article looks only at measurement of inequalities in health status.

Studies that measure inequalities can be classified according to two factors: time and unit of analysis. With regard to the former, they may be cross-sectional or longitudinal, and with regard to the latter, individual or ecological.

In cross-sectional studies all the observations are carried out at a single point in time; although there may be several replications of each observation, they all refer to one moment in time. These studies tend to use vital statistics that contain information on social status, occupation, schooling, and other individual attributes, although they may also use secondary data from surveys conducted for various purposes, such as the Demographic and Health Surveys, which are carried out in 13 countries of the Region. Alternatively, a specific survey may be conducted to study inequalities. In longitudinal studies, on the other hand, the observations are made over a period of time, prospectively or retrospectively.

In individual studies, the unit of observation and analysis is the subject (all variables are recorded as individual attributes), while in ecological studies the unit of analysis is a conglomerate of individuals who are grouped together according to geo-demographic, socioeconomic, or other criteria. These studies are usually based on secondary data aggregated by geopolitical unit.

The principal drawback of analyses of aggregate data is the risk of assuming that the results found in populations (aggregates) are applicable or reproducible equally among individuals (ecological fallacy).[1] Nevertheless, their great advantage is that they take into account social, geographic, and community factors of a contextual nature that cannot be analyzed in individual studies and that act as factors that confound or modify the effect of other proxy variables.

When individual data are used, the variables employed establish a ranking, both among and within the groups. This is the case, for example, with social position, schooling, and income level. In the ecological approach, on the other hand, the only ranking possible is among groups, since the attributes that are used (GNP, percentage of poverty, percentage of literacy, unmet basic needs, income ratio, unemployment rate, and others) lack meaning at the individual level.

This article assumes the use of secondary data already existing in countries, aggregated by geopolitical unit, such as those provided by the "Basic Indicators" initiative, which relies on subnational data already available in various countries of the Region. Analyzing health inequalities through these data will be very useful for public health policy-makers.

Because of the regional nature of the analyses carried out by PAHO/WHO, the examples given here use aggregate information by country, but the methods presented can be used for analysis in smaller geopolitical units (state, municipality, community, or neighborhood), depending on the objectives of the study.

The majority of traditional health indicators, such as mortality or morbidity from infrequent diseases, have very large standard errors and are therefore unstable when applied to small populations (under 100 000 inhabitants). Classical statistical techniques, both descriptive and inferential, are not applicable in these cases, and it is necessary to rely on weighting and the use of distributions appropriate to very infrequent events, such as the Poisson distribution.

In the examples given in this document, each geopolitical unit is considered an observation. These units can, in turn, be aggregated in socioeconomic groups according to the number of units studied, the indicator used, and the type of comparison to be drawn. In general, when observations are grouped together, information is lost and biases tend to appear (among them the ecological bias) when effects or associations are estimated. Whether or not to group geopolitical units is a decision made by the investigator.

There are various options for defining socioeconomic groups. One of them[2] consists of using per capita GNP to form groups, such that internal homogeneity is maximized. An example of aggregation of geographic units based on a socioeconomic indicator is the use of quintiles, which is one of the simplest ways of creating groups.

When analyzing social inequalities in the field of health, selection of the socioeconomic indicator is crucial because this variable defines the groups and the ranking within and between groups. A poor selection of indicator or of the categories created can bias the study. Obviously, when a single variable is used to define the socioeconomic status of the geopolitical units, as occurs in the examples in this article, the results cannot be extrapolated to the other factors that define socioeconomic status. The generalizations in some examples presented in this article should not be taken literally as their intent is purely didactic. Selection of the wrong socioeconomic indicator or an inappropriate definition of the categories of this indicator is one of the difficulties with aggregate studies.

Not all health inequalities are of social origin, but this article focuses on the type most frequently found in the literature. Social inequalities in health are health differences between groups of people categorized a priori on the basis of some significant characteristic relating to their socioeconomic position.[1]

Characteristics of indicators

Several major reviews of methodology for measuring inequality in health status exist. Those of Mackenbach and Kunst[3] and Wagstaff et al.[4] have been taken as the basis for this article.

Each indicator has advantages and disadvantages and each serves different purposes. The selection of the indicator should be consistent with the theoretical framework and objectives of the study. Good indicators of inequality should: 1) reflect the socioeconomic dimension of health inequalities, 2) use information on all population groups defined by the indicator, and 3) be sensitive to changes in the distribution and size of the population across socioeconomic groups.[4]

Regardless of the type of indicator used, it is extremely important that the information be of good quality and can be validated. Any method used should include a descriptive analysis of variation among groups in the phenomenon studied.

Indicators differ in complexity depending on the objectives of the study. Mackenbach and Kunst[3] recommend that decision-makers use simple methods, but that investigators confirm the results using more complex methods.

Measurements can be expressed as relative differences (e.g., rate ratio) or absolute difference (e.g., rate difference); both are important and tend to have complementary value. Relative measures are more stable and readily understood. In some cases, however, absolute measures are more useful for decision-makers, especially when goals have been set, because they enable a better appraisal of the magnitude of the public health problem. Absolute measures can be derived from relative ones and vice versa.[5]

Another methodological option is to use measures of the effect or impact of socioeconomic status on health. The essential difference between the two options is that impact measures take into account the actual socioeconomic situation and measure changes in health conditions that can be expected as a result of potential interventions. For this reason, impact measures are especially important for decision-making and the formulation of public policies designed to achieve equity.[2]

Measures of effect are based on fixed categories of the socioeconomic variable (e.g., primary schooling versus university education). Measures of impact, in contrast, use categories defined by a socioeconomic indicator quantifiable in population terms (e.g., highest income quintile versus lowest income quintile), such that if the distribution of the indicator varies, it also modifies the measurement of inequality.

Among measures of effect, rate ratio and rate difference are two of the most frequently used indicators. Another is regression- based effect index. One of the best-known indicators of total impact in the health field is population attributable risk (PAR), adapted from epidemiology. This indicator can also be estimated through regression analysis. Regression is also used to estimate the slope index of inequality (SII) and the relative index of inequality (RII). The index of dissimilarity is another example of an impact measure.[3,4]

Indicators from the field of economics are also used to measure inequalities in health, such as the Gini coefficient, with its corresponding Lorenz curve, and variants of both, such as the concentration index and curve, which combine indicators with graphic representations.

Notes:

A) Braveman P. Challenges in monitoring social inequalities in health: examples from a few continents (draft). Rockefeller Foundation Global Health Equity Initiative, 1999.

B) The relationship between the concepts of effect and impact can be compared to the one that exists between the well-known epidemiological concepts of relative risk and attributable risk.

REFERENCES (Part I)

1) Greenland S, Morgenstern H. Ecological bias, confounding and effect modification. *Int J Epidemiol* 1989;18:269–274.
2) Pan American Health Organization. The health situation. In: *Annual Report of the Director*. 1996 Edition. Washington, DC: PAHO; 1997. (Official Document No. 283).
3) Mackenbach JP, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Soc Sci Med 1997;44:757–771.
4) Wagstaff A, Paci P, Van Doorslaer E. On the measurement of inequalities in health. Soc Sci Med 1991;33:545–557.
5) Kunst AE, Mackenbach JP. Measuring socioeconomic inequalities in health. WHO Regional Office for Europe, 1994. (Document EUR/ICP/RPD 416). http://www .who.dk/Document/PAE/Measrpd416.pdf. 12 November 2002.

Source: Originally published with the title "Métodos de medición de las desigualdades de salud" in Pan American Journal of Public Health 12(6), 2002.

Lymphatic filariasis

Rationale for surveillance

Lymphatic filariasis remains a major cause of overt or hidden clinical disease in much of Asia, Africa, the Western Pacific and certain parts of the Americas. It is the second leading cause of permanent long-term disability. The prevalence is increasing world-wide, with at least 120 million people affected at different stages of the disease. Both diethylcarbamazine (DEC) and ivermectin, given as single doses, have been shown to be very effective in reducing microfilaraemia.

Filariasis was identified by the International Task Force on Disease Eradication as one of 6 potentially eradicable diseases. Current WHO policy is to achieve elimination of infection in humans mainly through combination drugs in large populations, complemented by antivector activities. Surveillance is therefore essential.

Recommended case definition

Clinical case definition

Hydrocoele, lymphoedema, elephantiasis or chyluria in a resident of an endemic area for which other causes of these findings have been excluded.

Laboratory criteria for diagnosis

Microfilaria positive o sign of "dance" of the adult filarial on ultrasound in the male genital area, antigen positive.

Case classification

Suspected: Not applicable.

Probable: A case that meets the clinical case definition.

Confirmed: A case with laboratory confirmation.

Recommended types of surveillance

There are currently three options and the choice will depend on the local situation:

• Routine monthly reporting of aggregated data on probable and confirmed cases from periphery to intermediate level and to central level or

• Sentinel population surveys (standardised and periodical) or

• Active case finding through surveys of selected groups or through mass surveys.

International: Annual reporting from central level to WHO (for a limited number of countries).

Recommended minimum data elements
Case-based data at peripheral level
Case classification (probable / confirmed).
Unique identifier.
Geographical information (location).
Laboratory result.

Aggregated data for reporting
Number of new cases.
Number of laboratory-confirmed cases.
Number of chronic conditions (hydrocoele, lymphoedema, elephantiasis or chyluria)

Recommended data analyses, presentation, reports
• Number of cases by geographical area and by year
• Monthly and yearly incidence, point prevalence (if active case detection), by geographic origin, by sex, by parasitological diagnosis

In areas where a program based on universal treatment was initiated:
- Prevalence of microfilaremia/antigenemia in sentinel communities

Principal uses of data for decision-making
• Estimate the magnitude of the problem and define populations at risk
• Improve and focus the elimination activities
• Improve the management and follow-up of filariasis-infected patients
• Identify technical and operational difficulties

Summer Courses in Epidemiology, 2005

Johns Hopkins University Bloomberg School of Public Health - XXIII Graduate Summer Institute of Epidemiology and Biostatistics

Place: Baltimore, MD, USA
Dates: 20 June to 8 July 2005

The courses offered are:
Three-week courses
- Principles of Epidemiology
- Intermediate Epidemiology
- Statistical Reasoning in Public Health I
- Statistical Reasoning in Public Health II
- Epidemiologic Basis for Tuberculosis Control
- Infectious Disease Epidemiology

Two-week courses
- Regression Analysis in Public Health Research
- Regression Analysis in Public Health Research Laboratory

One-week courses
- Introduction to the SAS Statistical Package
- Conducting Epidemiological Research
- Molecular Biology for Genetic Epidemiology
- Genetic Epidemiology in Populations
- Family Based Genetic Epidemiology
- Gene Expression Data Analysis
- New Perspectives on Management of Epidemiologic Studies
- Data Analysis Workshop I
- Data Analysis Workshop II
- Epidemiology in Evidence-Based Decisions
- Ethics Issues in Human Subjects Research in Developing Countries
- Nutritional Epidemiology
- Epidemiologic Applications of GIS
- Pharmacoepidemiology
- Survival Analysis
- Causal Inference from Epidemiologic Data
- Epidemiologic Methods for Planning and Evaluating Health Services
- Public Health and Human Rights: Methods and Approaches
- Epidemiology of HIV/AIDS
- Applications of the Case-Control Method
- Multilevel Models
- Design and Conduct of Clinical Trials
- Methods and Applications of Cohort Studies
- Social Epidemiology
- Tobacco Control: National and International Approaches

Weekend course
- Clinical Research: An Introduction to Design and Analysis

For more information, please contact: Ayesha Khan, Program Coordinator, Graduate Summer Institute of Epidemiology and Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD, 21205, USA.
tel: (410) 955.7158 / fax: (410) 955.0898.
email: akhan@jhsph.edu
website: www.jhsph.edu/summerepi

University of Michigan School of Public Health - XL Graduate Summer Session in Epidemiology

Place: Ann Arbor, Michigan, USA
Dates: 10 to 29 July 2005

The courses offered are:
One and three-week courses:

- Fundamentals of Biostatistics and Epidemiology
- Infectious Diseases
- Public Health Practice
- Cancer
- Clinical Trials
- Epidemiologic Measures
- Logistic Model
- Environmental and Occupational Epidemiology
- Behavioral Change, Social Epidemiology
- Longitudinal Studies
- Sample Survey Data Analysis in SAS & SPSS
- SUDAAN
- Global Health
- Genetics
- Bioterrorism
- Pharmacoepidemiology
- Nutritional Epidemiology and Geographic Information Systems
- Aging
- Meta-Analysis
- Health Policy
- Reproductive & Perinatal Epidemiology
- Research Ethics.

Special Weekend Courses:
- Introduction to Genetics in Epidemiology.

CME Credit Available

For more information, please contact: Jody Gray, Graduate Summer Session in Epidemiology, The University of Michigan School of Public Health, 109 Observatory St., Ann Arbor, MI 48109-2029.
tel: (734) 764.5454 / fax: (734) 764.3192
email: umichgss@umich.edu
website: www.sph.umich.edu/epid/GSS

National Institute of Health Public of Mexico - 10º. Program for Updating in Public Health and Epidemiology

Place: Cuernavaca, Morelos, México
Dates: August 1-26, 2005

Will be offered different courses in the following areas:
- Public Health
- Epidemiology
- Environmental Health
- Health Systems
- Biostatistics
- Infectious Diseases

For more information contact: Teresa Téllez Allende, Coordinadora Ejecutiva, Programa de Actualización en Salud Pública y Epidemiología, Instituto Nacional de Salud Pública, Av. Universidad 655, Col. Sta. María Ahuacatitlan, C.P. 62508, Cuernavaca, Morelos, México.
tel: +52(777).3293.010 y 3293.000 Exts. 3116, 3117 y 3119 / fax: +52(777).3293.080 y 3111.148
email: verano@insp.mx
website: http://www.insp.mx/verano/

Epidemiology General Directorate of the Secretariat of Health of Mexico - XXII A Diploma Course in Applied Epidemiology

Place: México City, D.F., México
Dates: July 4-29, 2005

The courses and workshops include:
- Epidemiology
- Biostatistics
- Methodology of research
- Introduction to the designs of epidemiological studies
- Case-Control Study
- Cross-sectional Studies
- Studies of outbreaks
- Evidence-based Decision-making
- Management in public health.

Applications must be received before June 17, 2005. For further information, please contact: Dra. Arminda Torrscano Granillo, Coordinadora operativa, Dirección General de Epidemiología, Francisco P. Miranda No. 177, Col. Unidad Lomas de Plateros, Delegación A. Obregón, C.P. 01480, México, D.F., México.
tel: +52(55).5593.3661 / fax: +52(55).5651.6286
email: internacionaldge@dgepi.gob.mx
website: http://www.dgepi.salud.gob.mx

PAHO's Epidemiological Bulletin is published quarterly in English and Spanish. Catalogued and indexed by the United States National Library of Medicine. Printed on acid-free paper.

http://www.paho.org/English/DD/AIS/beindexe.htm

PAN AMERICAN HEALTH ORGANIZATION
Pan American Sanitary Bureau, Regional Office of the
WORLD HEALTH ORGANIZATION
525 Twenty-Third Street, N.W.
Washington, D.C. 20037