S·I·M·U·L·A·T·I·N·G ·S·O·C·I·E·T·Y

WILL CONTRACTOR LAND AS A STRACT STRACT

EPIAID

BY ANDREW G. DEAN

This series of programs helps construct models of disease from real-world epidemics

EPIDEMIOLOGY, the study of disease and its determinants in a population. is the discipline underlying publichealth disease control. Ideally, the epidemiologist, as portrayed in an exciting series of articles by Berton Roueche (see reference 1), flies into town from a distant point, calms the populace, and then, like Sherlock Holmes, unerringly gathers all the vital clues and a few spurious ones until the cause of the epidemic becomes clear. Although luck and skill can sometimes produce this result, the actual work of modern public-health epidemiology involves more person-to-person interviewing, paperwork, statistical processing, and library research than the popular image contains.

In technical terms, an epidemic is not necessarily an urgent, lifethreatening. or even widespread threat to health, but it is "the occurrence in a community or region of cases of an illness, specific healthrelated behavior, or other healthrelated events clearly in excess of normal expectancy." (See reference 2.) By this definition, an epidemic can include any condition the epidemiologist regards as unusual and worthy of investigation—including cardiovascular disease over the past halfcentury, the consequences of smoking, the sinking of the Titanic, famine in Africa, family violence, snowmobile injuries, or warts seen by family physicians.

The process of epidemic investigation consists of entering a real-life setting, constructing a mental model of a disease and its hypothesized causes, gathering information from ill and well or exposed and unexposed persons, testing the model to see if it is "correct." and repeating this process until either the model is adequate or resources are exhausted. A slightly more detailed version of the process is shown in figure 1.

Figure 1 is also the main menu of EPIAID, the set of experimental microcomputer programs being developed at the Centers for Disease Control (CDC) in Atlanta, Georgia. The CDC, as the national center for publichealth epidemiology, investigates major health problems and provides coordination for epidemiologists in states and counties and training for the more than 60 new epidemiologists who enter its Epidemic Intelligence Service (EIS) each year.

EPIAID was designed to provide better tools for field investigation by CDC's several hundred epidemiologists and their colleagues in state and local health departments. It is an experiment to see whether or not an expert system on a microcomputer is useful in epidemic investigation. Development of the system is also providing opportunities for examining how epidemiology is actually practiced. EPIAID offers a vehicle for storage, distribution, and refinement of commonly used epidemiologic tools such as questionnaires and case definitions for diseases. If field testing shows the system to be successful, it may be made available for both investigational and educational use to a wide variety of public-health agencies.

The partnership between the epidemiologist and the computer began with the availability of punch cards and mainframe computers. For research purposes, these resources and (continued)

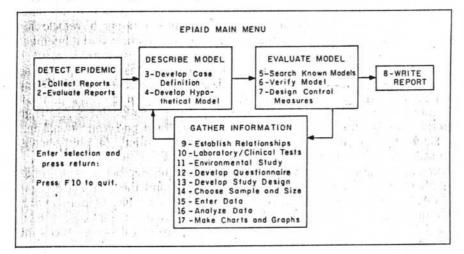
Andrew G. Dean, M.D., is a medical epidemiologist in the Epidemiology Program Office at the Centers for Disease Control (Building 1, Room 5115, Centers for Disease Control, Atlanta, GA 30333). later, more interactive systems have been invaluable, and most large epidemiologic research studies have been analyzed with the aid of computers. The tasks public-health disease control must accomplish, however, overlap only poorly with the services offered by mainframes. A typical acute-disease investigation occurs on short notice and requires travel to an unfamiliar site by an individual who will be busy with many tasks other than computer interaction.

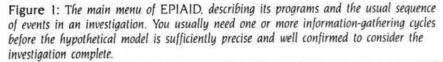
As an epidemiologist, you often operate from makeshift headquarters and work long hours while being harassed by worried citizens, the press, and myriad details of the investigation. You must devise a unique questionnaire appropriate for the situation, choose samples of ill and well or exposed and unexposed individuals, perform interviews and/or supervise those who do, and tabulate the data from a few dozen or a few hundred interviews to reach conclusions. If all goes well, you identify the factors leading to the epidemic and use them to design control measures such as improved food handling or water chlorination, changes in personal or occupational behavior, isolation of patients, or immunization of susceptible persons.

Typically, the newly devised ques-

tionnaire is hurriedly typed on plain paper, photocopied or mimeographed, and administered as quickly as possible-while local interest, the supply of interviewers, and the memories of the subjects remain at their peak. Preliminary tabulation is usually done by hand, although some investigations enter a later phase at your home base, in which you do more careful tabulation with the aid of a spreadsheet, a database or statistics package, or a data-processing department. Until recently, most epidemic investigations were tabulated by hand because the number of interviews did not exceed a few hundred and the necessary preparation for computer entry was not worth the steps involved. Lately, spreadsheets have proved to be a useful tool for managing questionnaire data, but setup time is still significant in a field setting.

My own experience includes interviewing 25 victims of insecticide poisoning, scribbling the results on 3 by 5 cards, and inspecting an African tea shop' in two hours before sundown because the airplane that brought us could not fly at night. One of my research investigations in the mid-1960s used 18,000 punch cards and a card sorter because the available computer was dedicated to





financial work.

An expert system for field epidemiologists must therefore be portable, 'easy to use even under distracting conditions, and conducive to setting up new data forms easily and quickly. Since the number of interviews seldom exceeds a few hundred, speed in data entry is less important than accuracy.

Since epidemic investigation, unlike billing or word processing, is not an everyday activity, programs must be easy to use the first time. The type of program that requires days or weeks of practice is not helpful to an investigator under distracting conditions. These limitations have kept even the most computer-eager epidemiologists from using computers in field investigations. They pose severe problems in the implementation, if not the design, of an expert system for epidemiologic investigation. Menus and prompts on the screen must take precedence over documentation or conditioned reflexes as a guide to what to do next.

Three previous programs written specifically for epidemiologists stimulated our thinking about EPIAID. Two of these. Socrates by Richard Curtis and the Epidemiologic Analysis System by Anthony Burton, were developed at CDC. They ask for definition of data items in a questionnaire, allow entry of data, and perform tabulations such as frequencies and crosstabulation, with simple statistical testing by chi-square or Fischer exact tests. The third, EPISTAT by Tracy Gustafson of the Texas State Department of Health, adds more statistical capabilities but focuses more on numbers than on textual information.

WHAT KIND OF EXPERT SYSTEM?

The burgeoning literature on expert systems makes it tempting to emulate the expert epidemiologist in software. Further thought suggests, however, that the last thing an epidemiologist wants to take on in an investigation is a somewhat rigid, less intelligent clone of himself. The most useful ex-(continued) pert system would be an *expert team* consisting of the following:

• a skilled secretary who captures ideas. text, and data items collected by the epidemiologist and arranges them neatly on paper in labeled file folders for use later when things are less hectic

• a professor of epidemiology who coaches gently when needed and asks whether something has been left out at times but avoids imposing a straitjacket on the investigator

• a medical librarian to look up information on previous epidemics and other items from the literature of medicine and epidemiology

 a statistician to process data and help interpret the results

• a writer to do the preliminary and final reports

• an artist to produce the graphs and charts

EPIAID provides elements of each of these. linked together by a menu so that the novice user can follow the programs in order and be led through the process of investigation, and the expert can pick and choose among the various epidemiologic tools offered. The programs offer

• interactive protocols for epidemic investigation

 access to information from previous investigations and from the literature of epidemiology and medicine

• tools for storing, structuring, and processing data, preliminary thoughts, and elements of the final report

THE DESIGN OF EPIAID

Compared to such other subjects for expert systems as spectrographic analysis. geologic exploration. or choosing an antibiotic for treatment of infectious disease. epidemiology operates in a broader. less-predictable environment. Decision making must incorporate elements of politics. personnel availability. ethics. and other parameters that are more easily handled by a human brain than by a computer. Even the more technical aspects of epidemiology are not easily distilled into rules. This element of fuzzy logic in EPIAID is handled by presenting options to the epidemiologist for refinement or even reversal at key points.

Epidemiology makes extensive use of information from medical literature, descriptions of previous epidemics. and tools from other investigations in the form of questionnaires. At present, the databases are scattered in both the filing cabinets and the brains of several thousand practicing epidemiologists in the United States, as well as in computerized bibliographic databases at the National Library of Medicine, CDC, and elsewhere. Even at CDC, the process of generating final reports about epidemic investigations has become slow and burdensome to the point that some reports are not completed before their authors leave for other employment.

EPIAID views the investigation as extracting information from the situation via systematic data-collection techniques, blending this information with data from past epidemiologic experience, and constructing an epidemic model that describes disease and its causes in the population being studied. During each round of data collection, EPIAID refines the model and then evaluates it to assess its suitability. The tests used are internal consistency, consistency with the medical and epidemiologic literature (matching known patterns), and the effectiveness of control measures designed from the model. When all three are judged satisfactory, the investigation is over, and the final report can be written.

IMPLEMENTATION

EPIAID operates under PC-DOS or MS-DOS. Although you can run it on a portable computer with 192K bytes of memory and two double-density 360K-byte floppy disks, a hard disk provides such advantages in convenience and speed as to be almost mandatory. EPIAID's programs are entirely menu-driven because they are intended for use under distracting conditions by occasional users who are interested in accuracy and organization rather than high-volume processing.

When you begin to run EPIAID, a banner is displayed while the program opens files; then you decide whether to continue with the current investigation or to begin a new one. The main menu (figure 1) appears, and you can select programs in the order suggested or you can pick and choose.

Since the specifics of the programs are mainly of interest to epidemiologists, the text box titled "EPIAID's Facilities" on page 230 looks at the kinds of facilities provided rather than dwelling on their contents.

PRESENT STATUS

EPIAID is an experimental program and is still in development. As of May 1, 1985, about 80 percent of the programming was complete and the necessary databases of previous epidemic models, questionnaires, and laboratory test information consisted of only a few examples. Once we have completed the program and entered larger amounts of background data, we plan to evaluate EPIAID first through review by epidemiologic colleagues and then through field trials in epidemics.

One of the most difficult hurdles in developing a system for use by publichealth epidemiologists is making the programs friendly enough for use in the field under distracting conditions. Field testing and the degree that epidemiologists adopt EPIAID will determine whether we have met this objective. Then, we can proceed to the more theoretical problems of database content and program logic flow.

Unlike expert systems in the physical sciences where you can define rules for most situations. EPIAID and the epidemiologist together must preside over a dynamic and complex process using information and previously useful tools from databases to collect new information from individuals or records in a real community. The new information must first be defined from an essentially infinite set of collectible items about individuals, and the results (continued)

com Simperin and Syst pow sequ reco rem 1.5 :

com

the

they

pc

SE(

PO

dist

simu

Av

EPIAID'S FACILITIES

E PIAID provides three basic kinds of facilities: entry, storage, and retrieval of structured records; access to databases of previous epidemics and tools for investigation; and interactive creation of text (phrase processing). Some of the modules in EPIAID combine all three functions.

RECORD HANDLING

The program to log disease reports stores entries from screen records resembling 3- by 5-inch index cards. One screen accepts the entry of cases reported by telephone or mail with details on name, address, date of onset, type of disease, etc. A second screen records the source of the report, usually a physician or hospital. Reports of groups of cases without individual details and of requests for information from individuals or from the news media are handled in alternate record formats. Function keys save, delete, or search for records, page through the file. or print lists of records.

Most investigations use questionnaires to record data from interviews. clinical examinations, laboratory tests, or hospital records. The usual dataentry program requires a data specification for each new questionnaire specifying the data type, name, length, and other characteristics of each fielda questionnaire blank. To bypass this step, a program called ENTER info was written in Turbo Pascal. ENTER info uses the original questionnaire, created as a plain ASCII text file on a word processor, as the data-specification file. When you run the program, it scans the questionnaire for underline characters. A continuous underline defines an alphanumeric field. A series of number signs (###) of the desired length indicates a numeric field. Number signs with a decimal point (##.#) describe real numbers. You define date fields as <date> and uppercase fields as < A >. With these exceptions, the questionnaire looks like one designed for manual entry.

ENTER *info* reads a questionnaire containing up to 80 fields and displays it on the screen ready for data entry within 25 seconds. Paging up and

down is handled automatically for any length of form; inserting, deleting, backspacing, and movement from field to field are accomplished with the appropriate keys on the IBM-style keyboard. The program saves each record after the last field is entered, and two Escapes provide exit. A utility program called CONVERT produces files for use in other database or statistics systems. CONVERT writes the data-specification statement for the Statistical Analysis Package (SAS), for example, and converts the data into 80-column "cards" with appropriate record and card numbers. You can send ENTER info data files over a network to another system and/or upload it to a mainframe computer for more extensive processing. CONVERT also produces files that you can use in the Epidemiologic Analysis System (EAS). written in BASIC at CDC. EAS does file handling, cross-tabulations, frequencies, sorting, and selecting. CONVERT will also produce files for StatPac, Lotus 1-2-3, SPSS, dBASE II, dBASE III, and other systems.

ACCESS TO DATABASES

EPIAID includes programs for creating and accessing databases of questionnaires from various investigations, case definitions of common or rare diseases (up to several pages long), information on laboratory tests available, and descriptions of past epidemics that the system can search for patterns that match the model of facts and hypotheses currently being used. Figure A contains a sample model, the screen for the program that searches a database of previous epidemics.

At present, all databases are contained on the same disk as the programs, but future plans include a program to create a search file for a remote on-line bibliographic retrieval system such as MEDLINE (maintained by the National Library of Medicine and available on a number of commercial database services). Medical librarians often perform such searches during an investigation. but our goal is to allow the epidemiologist to perform a search directly and download references and/or abstracts from the world literature into a disk file.

INTERACTIVE TEXT CREATION

Reports of epidemic investigations, whether published or unpublished, have a reasonably well-defined structure. Biomedical papers contain sections for introduction, methods, results, discussion, and references. EPIAID creates an English text draft of each section interactively. Evaluate Reports produces the introduction; Develop Study Design, the methods section; Analyze Data, the conclusions; and Verify Model, the discussion. The programs function as intelligent word processors or phrase processors. The system presents a suggested outline of the text and/or fragments of sentences on the screen for you to respond to and amplify in normal English sentences.

In the introductory memo, for example, Evaluate Reports constructs a heading with the time, date, investigation number, and organizational information. It asks for the subject and author of the memo. The program begins the text with "On" and prompts you to enter the "Date of first report." It places the date you enter in the text and adds a comma. The prompt window then asks for "Person or organization who made first report." You might enter "Dr. Steven J. Smith, Health Officer of Jones County." Your response can be up to two lines of text. which is added to that already created and left-justified. After each entry, the program checks a table within it for any special procedures to be executed at that time. A procedure may analyze the entry for content, save it for later use. perform mathematics, branch if certain conditions are met, or do anything that can be written into Q-Pro 4 code.

The programs that perform these procedures are built around a common core of Q-Pro 4 routines driven by three tables containing lines of text to be incorporated into the document, instructions to be presented on the screen, and the names of special procedures to be called after- you have made an entry. Each line has a number, and the program processes them in EPIAID

numerical order, somewhat like a program in BASIC. A procedure may center text, analyze its content, store a key word for later use, or jump to a subsequent line number if certain conditions are met. This rather simple facility can be used to create "intelligent" responses. Since the program displays the developing text and lets you edit as you create, you can provide responses in grammatical context. Therefore, the program does not need to try to adapt to the vagaries of English grammar and syntax. Figure B shows an example of the interaction that occurs. The final result is one or more pages of text in a dated and timestamped sequential file, which can be incorporated into the final report and/or edited later.

COMBINED PROGRAMS

The program for developing questionnaires uses the interactive text-generating facility, providing headings and prompting you to enter questions and entry fields for ENTER info in sections such as "Introduction." "Informed Consent." "Identifying Information." "Disease Symptoms," "Factors Influencing Disease" (known to epidemiologists as exposures), etc. At appropriate times, or in response to a function key, the program displays the current epidemic model at the bottom of the screen for ready reference as questions are phrased to test each of the hypotheses in the model.

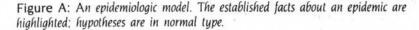
FI (See Files) brings up a list of questionnaires available from a questionnaire resource file. Any questionnaire can be brought into a window that overlies the screen temporarily. You can examine it using the Page Up and Page Down keys. Pressing the F4 key (Get File) appends the questionnaire from the resource file. line by line, to the text being constructed. You can accept. discard. or edit each line, and you can terminate the process by a command or by reaching the end of the resource questionnaire block. The combination of word processing, artificial intelligence, database access, and prompted data entry produces a questionnaire as a sequential text file that can be fed into ENTER info to start data entry.

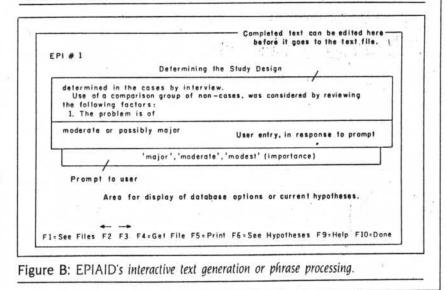
The program Develop Study Design uses the interactive text mode to produce a methods section for the final report. During the process, the epidemic model is available for reference, and the program creates blocks of text summarizing the facts and hypotheses developed so far. The previously developed case definition is brought in from an external file and included in the text. Intelligence is built into the program through the manipulation of key words such as "cases," "disease," "exposed persons," etc.

We usually design epidemiologic studies around persons known to be ill (cases) or exposed to some factor of interest (those involved with a chemical or a behavior like smoking or jogging). Epidemiologic studies may employ a comparison group (an analytic study) or merely describe the frequency of various factors in the

study group (a descriptive study). A comparison study based on known cases is called a case-control study; one based on exposure is called a cohort study. Given this information, the tables can be set up so that the system can create the text in reasonably useful form through its interaction with the epidemiologist. The system can systematically address major questions, such as whether to use a comparison group, and record the answers in the text. It leaves major decisions to the epidemiologist to make after reviewing the considerations the program suggests as relevant to the decision. If the study is to be descriptive, the program automatically skips those sections of the interaction that deal with comparison groups and matching.

2 FEW WEEKS Year(s) Control Measures Risk Factors	Location TRENTON, PA	Disease Name 1 KAWASAKI DISEASE	Route of Entry 1 ORAL 2 RESPIRATORY
13 1 FEW DAYS 1 CARPET 2 FEW WEEKS 2 FEW WEEKS Year(s) Control Measures Risk Factors		1 VIRAL AGENT .	
		1 FEW DAYS	
	Year(s) 1985	Control Measures 1 REMOVE AGENT	Risk Factors 1 JAPANESE ANCESTRY





OCTOBER 1985 . BYTE 231

must be interpreted with the complexities of the community situation in mind. The process is one of synthesizing a model of the current epidemic by merging data from the current situation with data from past epidemics and medical literature. The thinking process relies heavily on data from the collective experience of the discipline. Algorithms and computation, although important, are less prominent than in the physical sciences and are often simple (or complex but nonmathematical) enough to be supplied by the investigator in the field.

Many expert systems, particularly those available for microcomputers,



seem to be either mechanical or limited to narrower domains than the task of epidemiologic investigation. EPIAID leaves more control to the investigator than most expert systems do and relies much less on rules and probabilities. Present-day epidemiology does not have the necessary probabilities or rules available for incorporation into an expert system. Hopefully, more progress in this direction will occur when systems like EPIAID become available.

One characteristic of an expert system is its ability to "learn." EPIAID learns by accumulating new and more useful tools in its databases and by improving its logic. The database approach lets EPIAID learn a new geographic location. since disease patterns differ radically in different parts of the world—a database suitable for Botswana, for example, is nearly useless in Colorado.

One distinctive feature of EPIAID is that it produces blocks of English text describing and evaluating the hypotheses, methods, results, and conclusions of the model-building process. This can improve epidemiologic record keeping and offer a common ground for discussion of various record-keeping methods. The process of keeping written records of investigations badly needs the same kind of attention given to that process in individual medical cases (see reference 3). EPIAID may focus the discussion by offering a format for criticism and revision of case definitions, questionnaires, and the flow of logic in designing investigations.

However, interactive text creation, if not carefully constructed, can churn out hundreds of reports containing similar words in the same order, having little impact because of monotony. The idea that a computer can "write reports" is likely to arouse considerable emotion. It will be interesting to evaluate investigators' reactions to interactive text generation as we try to achieve the right balance of guidance and freedom in the text-creation tables so that the program assists rather than constrains the writing process. One of EPIAID's major limita-

和国家的名称是他国家的名称的人的意义。

232 BYTE • OCTOBER 1985

- MARTINE CONTRACTOR OF THE STATE

Inquiry 38

EPIAID

tions is its dependence on the investigator's typing skill; this will remain a problem until the development of practical systems for voice-toscreen dictation.

Another limitation is the value of the databases that are incorporated into EPIAID. Access to questionnaires used in previous investigations is only useful if the questionnaires are well constructed and evaluated. Searching a database of previous outbreaks for similar patterns is not useful if the database is incomplete. Frequently, the fact that a particular combination of factors has not been reported in the literature is as important as the reverse, and you cannot determine this from a partial database. You would need considerable resources to create and maintain a credible database of previous outbreaks in the structured form proposed by EPIAID's current design or a later enhancement of it. The epidemic model composed of 12 items, shown in figure A in the text box, was developed for EPIAID and is therefore a proposal, not a component of routine epidemiology, although most introductory courses include related material.

Currently, the literature of epidemics, both at the National Library of Medicine and at CDC, is indexed by key words and topics: however, it was not designed with the parameters of an epidemic model in mind. Systematic coding of negative as well as positive results from an investigation would be a new development and would require extra resources and agreement on the format. It would offer as an extra benefit the possibility of systematically evaluating the success of investigations in terms of the number of elements from the general model that were known at the end of the investigation. We need an agreedupon model to allow the "epidemiology of epidemics"-rather than of a particular epidemic-to develop to the point where we can include probabilistic thinking in an epidemiologist's expert system.

None of these problems is unique to computerized systems. If you're an epidemiologist relying on traditional

resources, you probably either maintain a manila folder of favorite questionnaires or go down the hall to ask for a colleague's. You search previous outbreaks through several trial runs with a medical librarian or call another epidemiologist who specializes in a particular subject. All of these processes are still necessary with computer assistance, but EPIAID aims to make the manila folder very large and convenient and the questionnaires in it suitable for data entry immediately after editing. With EPIAID the epidemiologist is one step closer to bibliographic information and can edit and incorporate the reference citations in a report rather than retyping them.

THE FUTURE

The role of the expert epidemiologist in the future may be partially to see that other epidemiologists have access to well-constructed and complete database materials and effective tools that have been tested in the field. EPIAID, by providing a medium for storing and using such materials, may stimulate interesting questions about how epidemiology is or should be practiced and how the worldwide epidemiologic database can or should be maintained. ■

REFERENCES

1. Roueche, Berton. The Medical Detectives. New York: Times Books, 1980.

2. Last. John M., ed. A Dictionary of Epidemiology. New York: Oxford University Press. 1983.

3. Lindberg, D. A. B., and M. F. Collen. AAMSI Congress 84 : Proceedings of the Congress on Medical Informatics. volume 2. Bethesda, MD: American Association of Medical Systems and Informatics, 1984.

ACKNOWLEDGMENT

EPIAID's menus and many of the programs and database searches were written in Q-Pro 4 by Richard McCorkle of Professional Software Inc. and the author. Q-Pro 4 from QNE International is a database language that allows rapid generation of data-entry screens and searches of indexed files. Programs for automatic database specification and data entry were written in Turbo Pascal by Jeffrey Dean, currently a student at Paidea School in Atlanta, Georgia.





A MCGRAW-HILL

10000

SIMULATING SOCIETY

FEATURES:

SB180 Microcomputer, Part 2 Skycam

REVIEWS:

Ericsson PC TLC-LISP

