46832

WHO/PEP/GEENET/93.19
DISTR: LIMITED
ORIGINAL: ENGLISH

ENVIRONMENTAL
OCCUPATIONAL
EPIDEMIOLOGY
SERIES

A COURSE IN MICROCOMPUTER USE FOR EPIDEMIOLOGISTS

BIBLIOTHEQUE OMS
20 SEP. 1993
WHO LIBRARY

WHO
World Health Organization
Geneva, 1993

In support of the International Programme on Chemical Safety

A Course in Microcomputer Use for Epidemiologists

Using Version 5 of Epi Info

Andrew G. Dean, M.D., M.P.H.

Centers for Disease Control
Atlanta, Georgia 30333, U.S.A.

Microcomputers have become an important tool for epidemiology studies. Data collection, editing and analysis -- for either small- or large-scale studies -- can be carried out more quickly and accurately through the use of microcomputer and user-friendly software.

This document is therefore being made available to members of the WHO Global Environmental Epidemiology Network (GEENET) in order to encourage and facilitate the use of microcomputers in environmental epidemiology studies. It was prepared by Andrew Dean, M.D., of the Centers for Disease Control, Atlanta, U.S.A. and contains an introduction to microcomputer use, together with examples from the Epi Info epidemiology software devised by the Centers for Disease Control. It also contains notes for instructors, as an appendix, for those organizing a short introductory training course on microcomputing in epidemiology.

WHO gratefully acknowledges the support of the Centers for Disease Control in making this material available for GEENET. As pointed out in the Introductory Notes, readers are encouraged to make copies for friends and colleagues. Comments or suggestions for improvements should be sent to Andrew G. Dean, M.D., (see Introductory Notes for address and contact numbers) with a copy to the GEENET office at WHO, Division of Environmental Health, CH-1211 Geneva, Switzerland.

Introductory Notes

These materials were produced by the Centers for Disease Control and are provided for use by the public health community. You are encouraged to give copies to your friends and colleagues. No warranty is made or implied for use of the materials for any particular purpose.

Trade names are used for identification or for examples; no endorsement of particular products by the Public Health Service or the U.S. Department of Health and Human Services is intended or implied.

Wordstar is a trademark of Micropro Corporation; Multimate, of Multimate/Multimate International, an Ashton-Tate company; dBase, of Ashton-Tate Corporation; Wordstar of Micropro International; Microsoft Word of Microsoft Corporation; and IBM, of International Business Machines, Inc.

Comments or suggestions are welcomed. Address comments or questions to: Andrew G. Dean, M.D., Epidemiology Program Office, Mailstop G34, Centers for Disease Control, Atlanta, Georgia 30333, U.S.A. Telephone (404) 639-1326, or (home) (404) 458-2271.

Contents

Introduction	How to Use This Course Manual	7
Chapter 1	A Short History of Microcomputing	11
Chapter 2	IBM-Compatible Microcomputer Anatomy . . .	19
Chapter 3	The Disk Operating System: Facilities at the Brainstem Level	33
Chapter 4	Using EPED, a Simplified Word Processor . .	57
Chapter 5	Making a Questionnaire and Entering and Editing Data with Epi Info	63
Chapter 6	The CHECK Program: Refining the Data-Entry Process	73
Chapter 7	Analyzing and Transforming Data	77
Appendix	Notes for Instructors	83

Introduction

How To Use This Course Manual

The CDC Course

This manual was designed for use in a course in microcomputing for epidemiologists at the Centers for Disease Control (CDC). It provides a minimal introduction to IBM-compatible microcomputers for public health professionals with no prior computing experience. Much of the course is built around Epi Info, Version 5, a public domain database and statistics system produced at CDC. Copies of Epi Info and its manual are required for the course.

At the end of the course, participants should be able to:

- Find, copy, and back up files
- Produce a memo, letter, or report in the computer
- Make a questionnaire and enter data from an outbreak or survey
- Add range checking, skip patterns, and other error-checking features to the data entry process
- Analyze and transform data, producing lists, frequencies, tables, epidemiologic statistics, graphs, and printed or file output
- Import and export files

Chapters 1 and 2 provide background material that should be read before beginning the class. Chapter 2 will be reviewed in the first class with microcomputers present so that the important parts of the computer become familiar. It is helpful to have at least one computer with the cover removed so that the internal components are visible.

At CDC, Chapters 1 through 4 are taught in a single 4-hour session. Epi Info is presented in another 3- to 4-hour session. It is later used in a field exercise that involves data collection by telephone or personal interview. Each student enters

data from the exercise and helps to analyze the data collected by the class.

Equipment

At least one IBM-compatible microcomputer is required for each pair of students, with DOS version 2.12 or a later version. The DOS manual that came with the computer will sometimes be helpful, and each student should have several blank floppy disks.

Epi Info, Version 5, and its manual are also required. It may be obtained from:

USD Inc.
2075 A West Park Place
Stone Mountain, Georgia 30087, U.S.A.
Tel (404) 469-4098
Fax (404) 469-0681

The current price is US$35 (US$45 outside North America) for four disks, a printed manual, and shipping. Quantity prices are available.

Notation

A few simple rules apply to reading the manual. Keys on the IBM-compatible keyboard are indicated by "< >." Thus, <ENTER> is the key labeled "ENTER." and <F1> is the function key labelled "F1." Material that you type is shown in boldface. Thus, if you are instructed to type:

A:\>**COPY A:*.* B:**

the computer will have supplied "A:\>" (or "A>"). You should type "COPY A:*.* B:" and press the <ENTER> key. The <ENTER> key is almost always needed to mark the end of an entry, although this will not be shown in the manual each time.

Items found between double lines like this are exercises, to be performed as they are encountered. They may be done in a class or as self-teaching exercises.

Educational Philosophy

There are many approaches to providing instruction in the use of computers. This course uses hands-on laboratory sessions and encourages use of the Epi Info manual. Enough DOS skills are taught so that the student can safely use other application software, but the emphasis is on accomplishing epidemiologic tasks with inexpensive or public-domain software.

People vary a great deal in their previous experience and psychological approach to computers. We have found it helpful to divide classes into sections based on previous experience with DOS-based software and to proceed at a faster pace for the more advanced students. Students who have never touched a computer generally need many more hours of practice than those who have mastered the basic skills.

Some students in this course may feel that the instruction is too sparse and the task too simple to be worthy of effort. This is particularly true of those who know computers already. If the simpler commands in the manual are not challenging, however, try to use the DOS and Epi Info manuals to progress beyond the scope of this manual and learn more about what can be done with DOS and available programs. Even an expert can learn something by writing extra DOS batch files or experimenting with new word processing techniques or Epi Info commands.

Others in the class, particularly those who have never used a computer, may feel that they are trapped in an alphabet soup of DOS commands and will never be able to sort them out. They should take comfort in the knowledge that no software command or miscommand can do damage to the computer or the operator (although FORMAT can erase a hard disk and ERASE or DELete can erase files). The exercises provided will prepare you to do a remarkable range of tasks in word processing, data entry and analysis, communication, and

bibliographic searching. At the end of the course you should be ready to "solo" with the computer, various manuals and books, and advice from telephone resources, supervisors, and friends.

Use the computer as much as possible before the first important outbreak investigation, typed report, or telecommunication. You will find that the computer will contribute to the accuracy, degree of documentation, and possibly even the speed of your work almost from the first time you use it.

Typing skills will develop, if necessary, as you work. If you have never had a typing course, you may want to consult a typing text or use one of the many computer typing instruction courses. so that you can practice with the fingers on the correct keys as you do the exercises.

Chapter 1

A Short History of Microcomputing

Cecil S. R. Duncan and Andrew G. Dean

The mystique that surrounded computers in the 1950s and 1960s is rapidly disappearing. No longer are they thought of as "thinking machines" or "giant brains" capable of being used only by computer scientists or specialists. Professionals everywhere are beginning to find them essential in their daily work, and there are said to be nearly half a million personal computers in Atlanta, Georgia, alone.

The Large Computer As Unapproachable Shrine

From the mid 1960s to the late 1970s, to get anything done by computer, a professional with a problem to be solved had to approach the "computer group" with hat in hand and plenty of money. On the way to the computer group's office, he or she had to walk past glass windows, behind which were the impressive gray cabinets of flashing lights, tape drives, card readers, and printers. By the time the office was reached, the impression had been gained that a programming cost of many thousands of dollars and several person-months or years was justified. Not included in that cost was the time required to educate the computer specialists in the problem to be solved. A change in software that had already been written required almost a repetition of the original effort.

The First Microcomputer

During the 1960s the aerospace industry boomed, and by 1975, its insatiable need for small, light, controlling devices had given rise to the main processing unit of a computer in a single small chip, a microprocessor. In that year an article appeared in *Popular Electronics* magazine that changed the world. The article described a

real computer, the Altair 8800, available in kit form for $397. Consisting of circuit boards and a bag of computer chips and switches with an optional cabinet, the computer was a real challenge to assemble and even more difficult to program. Quality control was a problem because the company received orders beyond its most optimistic expectations. There was no software; the user programmed the computer in binary language by use of switches on the front panel. Even adding two numbers required time, and there was no way to store programs or data. When the machine was turned off, all programs and data disappeared. But it opened the era of microcomputers.

At first, the Altair had no operating system, only a primitive set of instructions to the main chip to do such things as send a character to a given memory location, jump to another location in the program, or "halt." Such a system is needed to help the operator to communicate with the machine. It requires a complex set of instructions within the operating system, for example, to get characters from a keyboard and display them on a video screen--a big improvement over flipping eight switches on the front panel to indicate a character and seeing the result in small red lights above the switches. In effect, the user communicates with the operating system to run programs and save or print results. The operating system, in turn, runs the computer to produce the desired results. The operating system and other programs are called "software." The machine itself is called "hardware."

The Late 1970s

A year or so after Altair's appearance, an operating system for microcomputers appeared. It was called CP/M and is still used on some computers today, albeit in improved form. This software allowed the user to create programs and enter data, to save both on magnetic floppy disks or cassette tapes, and to load the desired programs and data back into the computer. No longer did the user have to flip switches for every bit of data. More importantly, computer users could exchange programs, because software written for one machine could be used on another with the same operating system. The programming language, BASIC, also appeared about this time. Originally designed for student use at Dartmouth, it was written for microcomputers by Bill Gates--now the billionaire President of Microsoft Corporation, but then a student dropout from Harvard University.

Soon another new microcomputer came on the market called the Apple. It featured a color display with graphics, and it came routinely with a floppy disk drive and an operating system.

IBM-Compatible Microcomputers

In 1982, IBM, the largest computer firm in the world, introduced a microcomputer. It incorporated most of the best features of previous microcomputers, and IBM sold hundreds of thousands, and then millions, of microcomputers. One of the most important products that IBM sold was standardization. Because of the IBM reputation in the business community, the disk format, screen display, and other features of the IBM microcomputer became universal standards, and subsequent computers produced by other firms were designed to be "IBM-compatible." This course is designed for IBM-compatible microcomputers because such machines can be found in the most remote parts of the world, and they offer a way of running the same software on computers produced by a number of different manufacturers.

The Macintosh

The Macintosh computer, introduced by Apple Computer, Inc., in 1984, used a "mouse" for pointing to objects on the screen, commands contained in pulldown menus, and "windows" to make application programs easy to use. All application programs following the Macintosh guidelines presented the same commands to the user. Pulling down a menu at the top left corner of the screen will always let you "Quit" a program, rather than having to guess whether to type "QUIT," "QU," "BYE," "DONE," or "END," press the <F10> key, press the <Esc> key, or use <Ctrl-C>, as may be necessary in different applications on IBM-compatible machines. This uniformity reduces the time needed to learn different applications on the Macintosh. Unfortunately, the Macintosh is considerably harder to program than the IBM and has not become as widely available. We therefore focus this course on IBM-compatible computers, while recognizing the many advantages of the Macintosh in "user-friendliness" and for complex graphic applications such as desktop publishing.

Connectivity

Developments of the past few years include methods of connecting many different kinds of computers together and providing software and hardware facilities for exchanging files. Local area networks (LANs) connect computers within the same office or building, and many programs and hardware devices are available to solve parts of the "Tower of Babel" that industry competition has spawned. Knowledgeable users today are more concerned with connecting various incompatible computers than with choosing THE best one for their own use and excluding all others. This manual, as an example, was partially written on an IBM computer and the files were converted and edited on a Macintosh for the present incarnation. It could also have been transported to the Wang system at CDC or even to the mainframe if desired.

Software

Although the first microcomputer users had to write all their own programs in BASIC or assembly language, soon programs for accounting and other purposes appeared. Such programs are called "applications" software, and they differ from operating systems in that applications software is designed to solve a specific problem. The user runs the applications software, which runs the operating system, that--in turn--runs the computer.

Until the very late 1970s, the microcomputer was still not used by most professionals. Applications software had to be written or purchased (if available) to do a specific job. Accounting programs did only accounting; statistical programs did only statistics; and if you wanted to do something else, you either had to write your own program (a time-consuming task) or had to find one already written. Many computers purchased by professionals were set aside because the time required to write or find programs was greater than the time required to do the job by hand, or because frequent hardware failures made them unreliable.

Word Processors

At the end of the 1970s, several new types of software appeared that have resulted in microcomputers becoming daily tools for professionals. The first was a word processing program. Here was a program that could be used for any letter or document that the professional wanted to write. One early word processing program was called "Electric Pencil." It was slow and tended to drop the last character on a line if the typist was fast. Also, it used a cassette for storage, which reduced the pace of saving and loading documents to a crawl, but it worked. Later came Wordstar, and word processing on microcomputers then became a serious and successful enterprise. Now one can buy Multimate, Word Perfect, Microsoft Word, and many other word processing programs. Sophisticated graphics and page-layout features are offered in "desktop publishing" packages like Aldus Pagemaker and Ventura Publisher.

Spreadsheets

The second software "breakthrough" was the spreadsheet. This type of program was invented by Dan Bricklin, a graduate student at Harvard Business School who grew tired of solving similar problems over and over; most involved rows and columns of numbers that were added, subtracted, multiplied, divided, averaged, or counted. He and a programmer friend wrote the first spreadsheet program, Visicalc. It allowed entry of items in row and column formation. At any location, a formula could be entered instructing the computer to perform mathematics, copy a value, or do other operations. The program automatically performed all these operations whenever data items were entered, providing a sort of magical slate that was always current.

The original Visicalc provided no capability for graphics (e.g., bar charts, line charts, pie charts), but an ancillary program called Visiplot allowed the user to plot Visicalc data on an Apple computer. Lotus 1-2-3 included integrated graphics and is still the number-one best seller among software programs because of its usefulness in business environments. Others, such as Excel and Quattro, offer the same or improved facilities.

Data Base Programs

The third important software development was the data base. A data base is a collection of related information, such as a list of names and addresses. Although spreadsheets can be used for data bases of limited size, other programs, called data bases, are available for that purpose. The largest-selling data base programs are dBASE III and III-Plus, although other good data base programs exist. The more capable programs offer ways to link data items between different files, for example, an individual interview form with a form collected for the household in which the individual resides.

Statistics Programs

Statistics programs were first developed on large mainframe computers for use by university research projects. Two of the most popular programs, SAS and SPSS, have become available as microcomputer programs. They offer powerful and diverse statistical facilities but occupy large amounts of disk space and require considerable time to learn to use. Because they offer statistics to a wide range of professions, the array of choices can be overwhelming, and many parts of the programs will rarely be used by most epidemiologists. They offer environments similar to those of their mainframe counterparts and minimize learning time for mainframe SAS and SPSS programmers.

Epidemiologic Software

Software designed specifically for epidemiology has proliferated in the last few years. Several commercial programs that combine data base and statistical functions are available, including EpiLog, True Epistat, and CRISP. In this course we will use a public-domain program called Epi Info that was written at CDC. It allows rapid construction of questionnaires, immediate data entry, and analysis of data under field conditions. It is also being used to maintain reportable disease records in at least 35 state health departments.

Communications and Bibliographic Searching

Communication programs grew from a public-domain program called Xmodem that was popular in amateur computer circles. We will use Crosstalk, but Smartcom, Procomm, and many others are also available. Most allow the details of interacting with a particular telecommuncations network to be recorded in a file ("script") so that passwords, telephone numbers, and answers to repetitive questions are contributed automatically by the program. A specialized communications program called Grateful Med allows direct access to the National Library of Medicine's MEDLARS data bases and enables anyone to perform a reasonable literature search within a few minutes and store the references on a disk for later perusal.

The Microcomputer As a Tool for Solving Real Problems

By using an operating system, a word processor, an epidemiologic statistics program, and perhaps a spreadsheet and a data base program, a professional can solve almost any problem that can be solved by a small computer, and with a communications program access a bibliographic data base or send the results to another computer--all without learning a programming language.

Yes, it takes time to learn anything, but these software packages can be quickly learned and put to use by the professional who knows the problem best. Once competent in these applications, the professional must take hat in hand and visit the computer specialists only when the problem is so big that the microcomputer cannot handle the job: processing a file of over 20,000 records, for example, or setting up a permanent data base to be used by many people. Such problems belong in the hands of computer specialists and large, mainframe computers.

The solution to the professional's need for computer tools is at hand; both the hardware and software exist. There are still some major barriers, however. One barrier is the time it takes to learn a particular program and to set it up for a given task. One aim of this course is to reduce that time, to give you a head start. Another barrier is the difficulty in recognizing that the computer can be used to solve a real-world problem. One of the authors has observed a professional who was preparing a list of names in alphabetical order by writing the names on slips

of paper and moving the slips around, while sitting within 6 feet of a microcomputer with Lotus 1-2-3 and dBASE software, either of which is capable of sorting a sizable list in alphabetical or numerical order in seconds.

This course has been constructed as much as possible around real-world examples of the type encountered in epidemiologic work. The word processor and general-purpose epidemiologic program are in the public domain so that they may be copied for friends and colleagues or used for classes without fear of copyright infringement.

For those who go on to take courses in SAS or SPSS for mainframe processing of large files, the microcomputer will serve as an introduction. Others will gradually acquire skills with spreadsheets, data bases, graphics programs, or other applications through experimentation or courses in whatever environment they will be working. Now that the microcomputer has made computing accessible to essentially any professional, professionals of the future will be handicapped if they do not learn to use this valuable tool.

Chapter 2

IBM-Compatible Microcomputer Anatomy

Key Concepts--Bits, Bytes, and Files

Computer logic is built up of "bits." A bit is either on or off, zero or one, true or false, and is the smallest unit of storage in the computer. It is equivalent to a switch set either "on" or "off."

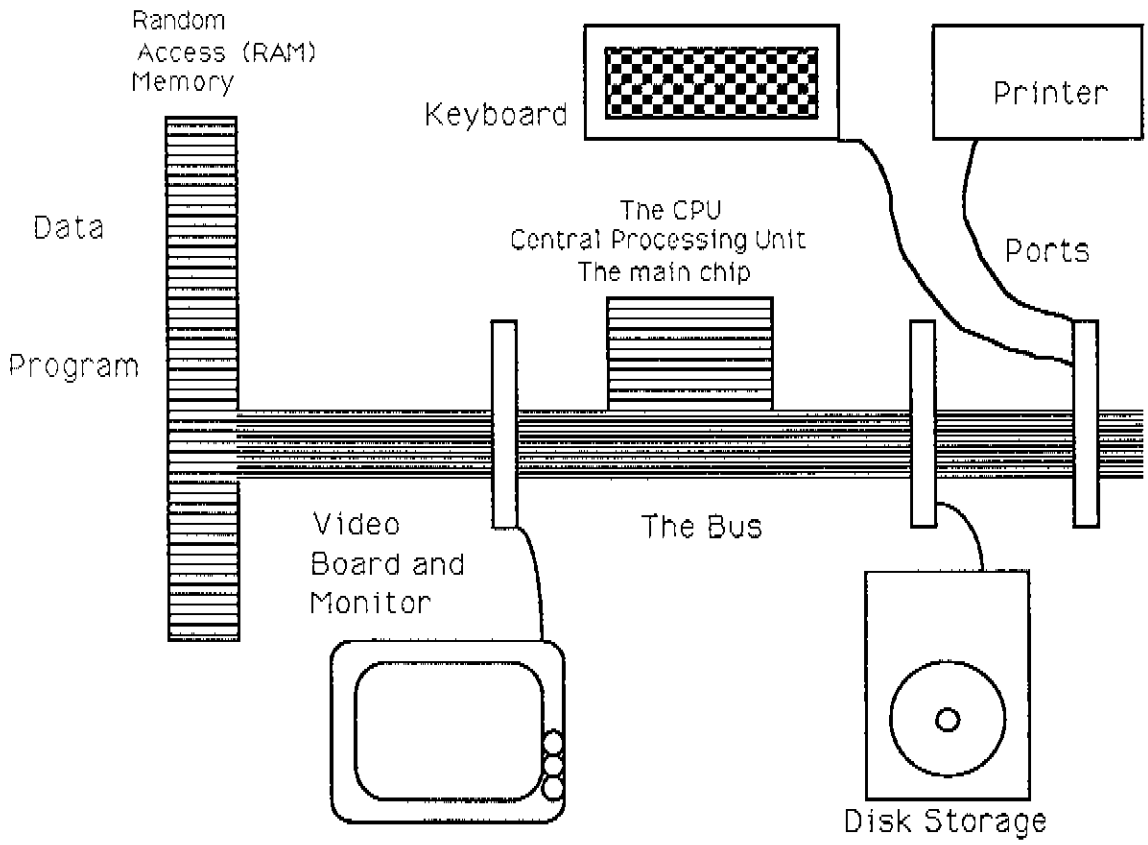
A "byte" consists of eight bits. Using binary numbers, a byte can represent the numbers 0 through 255. By international convention called the ASCII code ("as-key"), the letters of the alphabet, digits, and many punctuation symbols are represented as bytes. The byte 01000001, or 65 decimal, represents a capital "A," for example. A byte can therefore be thought of as a single character. In computer programs, one or several bytes represent an instruction to the computer's central processing unit to perform a particular operation such as "fetch a character from the memory location whose address is stored in register H."

Files are collections of bytes that are stored as a single unit, usually on a floppy or hard disk, and are loaded into memory when needed. They may contain either programs (instructions) or data. In the early days of computing, Von Neumann recognized that programs and data could be handled in the same way--placed in files, loaded into the same kind of memory. This was one of the key concepts in the development of today's computers.

Components of a Microcomputer

The computer system hardware includes the monitor, system unit, and keyboard, as shown in Figure 1. If a printer or modem is included, these items are also considered hardware.

FUNCTIONAL PARTS OF A MICROCOMPUTER



Functionally, the computer may be thought of as consisting of input devices and output devices ("peripherals"), and in between, the central processing unit ("main chip") and storage facilities (memory and disk storage).

Input	**Processing**	**Memory/ Storage**	**Output**
Keyboard	CPU	RAM	Monitor
Modem		Disks	Printer
		Tape	Modem

Instructions and data are received through the input devices, processed and perhaps stored in the computer, and then the results are sent to the output device.

Within the computer, a standard series of connectors on a device called the "bus" allows boards from different manufacturers to be plugged in and to function with the rest of the computer. Although the details are complicated, the entire computer is built on the concept of the transistor, which is an on/off switch, and in a sense, a microcomputer can be thought of as a collection of 10 or 20 million switches all set to on or off, some of them having the ability to control the others.

The Power Cord, Main Switch, and Power Supply

The power cord has a third prong on its plug for grounding purposes. It is important that the outlet you use has a properly grounded, three-prong receptacle. If not, you can lead a wire from the ground connection (round prong) to a water pipe or radiator for safety purposes.

Plug your computer into an AC outlet. If necessary, locate the connection for the power cord on the computer and insert the female end of the power cord. Set up the keyboard, if necessary, and then turn on the computer's power switch. After a short wait, you should see something on the screen.

If a computer is available with the top off, locate and examine the power supply, which converts AC current into the 5- and 12-volt DC current required by most computer components. The video tube is an exception and may contain THOUSANDS OF VOLTS, even for some time after the computer has been turned off. KEEP FINGERS AND CONDUCTIVE OBJECTS AWAY FROM THE REAR OF THE VIDEO TUBE.

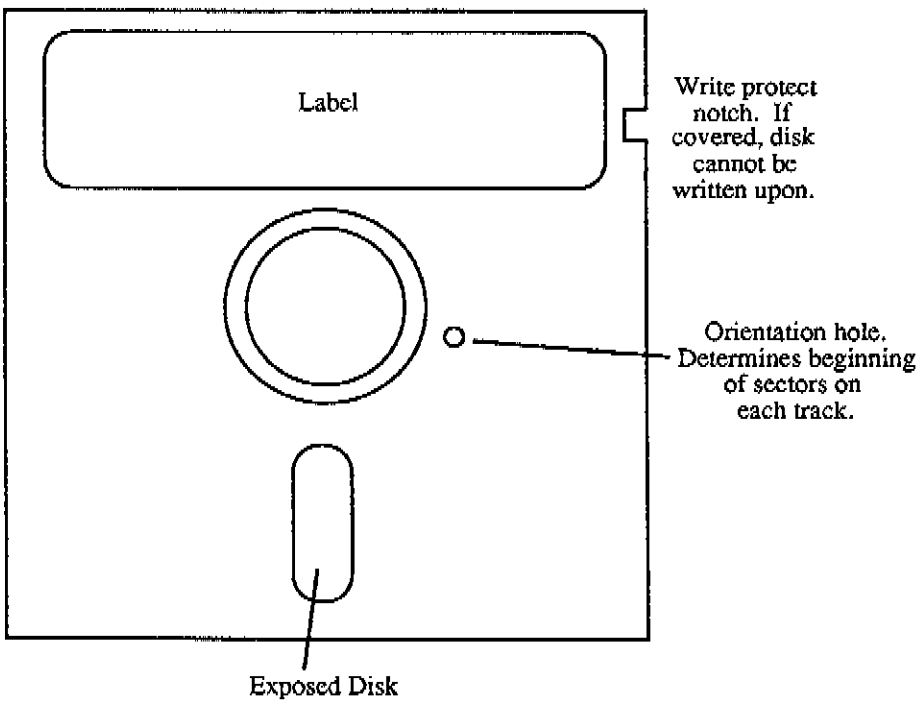
Floppy Disks, Hard Disks, and Files

Computer software consists of programs that cause the computer hardware to perform useful work. Software and data items or documents are stored as files, usually on some kind of magnetic recording media.

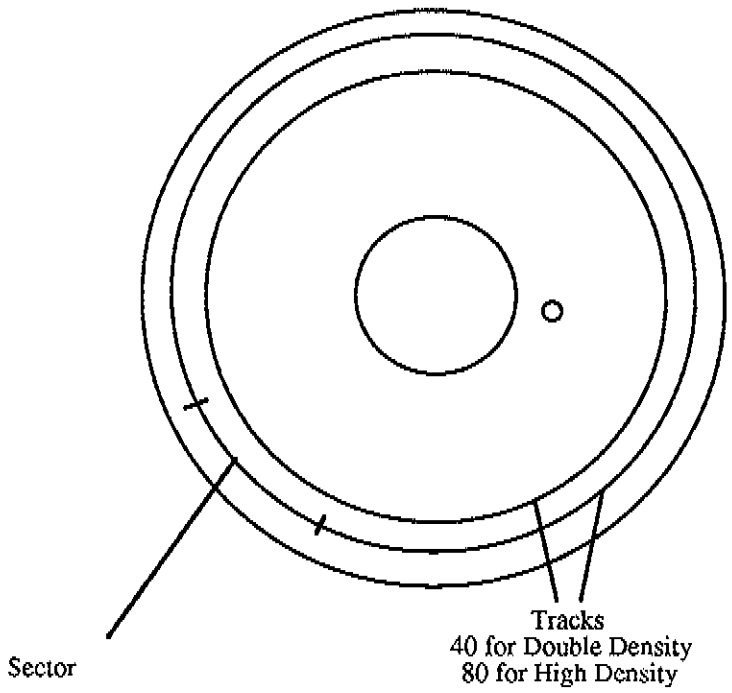
"Floppy disks" are a common storage medium, consisting of a flexible plastic disk in a protective jacket. The disk is covered with ferric oxide of the same type used on recording tape. "Bits" (a one or zero, on or off) and "bytes" (eight bits or one character) are stored as magnetized regions on the disk. The disk is arranged into *tracks* and *sectors* so that individual locations can be located by the operating system and the disk drive. To use a disk, insert it in one of the floppy disk "drives" on the front of the system unit. The label must face up, or to the right or to the front, depending on how the disk drive is oriented. A lever or button on the drive closes the door to the drive and tells the computer that a disk has been inserted.

Examine a floppy disk. Rotate the disk inside its jacket and locate the hole that marks the beginning of the sectors. Do not touch the recording surfaces.

A 5 1/4 Inch Diskette



A 5 1/4 Inch Diskette Without Its Protective Jacket



The floppy disk drive turns the disk and recording heads read the tracks and sectors, loading a file into the internal memory of the computer for processing. When software is "saved," the recording head records the data or program back on the floppy disk to be stored until needed in the future. Floppy disk drives are referred to by name, which is a letter followed by a colon, such as A:. Usually the first drive is on the left or top and is called drive A, with a second drive, if any, being called B.

Disks for IBM-compatible computers come in several types:

5 1/4 inch Double-Sided, Double-Density (DS/DD)	360 K
5 1/4 inch Double-Sided, High Density (DS/HD)	1200 K or 1.2 megabytes
3 1/2 inch microfloppy	720 K
3 1/2 inch microfloppy	1440 K or 1.44 megabytes

Unfortunately, the high-density diskettes are the same size, shape, and color as double-density diskettes, and you must look closely at the labels to see the telltale "HD" or "DD."

Examine several types of floppy disks. Determine from the manual or the instructor which types your computer uses.

Floppy disks should be handled carefully. Never touch the exposed recording surfaces. Avoid bending them, and always keep them stored in their envelopes and protected from dust and moisture as much as possible. When writing on the disk label, use a felt-tip pen, because ballpoints will make dents in the underlying disk and destroy the alignment of the tracks and sectors. Magnets and magnetized metal, including electric motors and possibly AC electrical cords, will erase parts of the disks. Airport X-ray machines will usually not cause problems unless the disk is passed through many times.

The computer may also have a hard disk, which is inside the machine. The hard disk can store more information than a floppy disk but cannot be removed from the machine. It is constructed on the same principle as the floppy disk, but the rigid metal disk and dust-free atmosphere inside the drive allow data to be recorded more densely on the hard disk. Hard disks hold from 10 to several hundred megabytes (million bytes) of data or programs. The hard disk on most machines is drive C. Hard disks in portable computers are reasonably safe from airport X-ray machines since they are enclosed in metal housings.

Does your computer have a hard disk? What is its capacity? How many floppy disks will it take to back up (transfer the contents of) the hard disk to floppies? Consult the manual or the instructor for further information.

The Video Monitor

There are many different standards for video displays for IBM-compatible computers. You may encounter monochrome Hercules Graphics Boards, the IBM Color Graphics Adaptor (CGA), the Enhanced Graphics Adaptor (EGA), and the Video Graphics Array. Most XT compatibles have either the Hercules or CGA displays, and most AT-compatibles (286 processors) have an EGA display that has higher resolution and is therefore more pleasant to the eye. VGA boards have still higher resolution and are often used with computers having 386 processors. Most laptop and portable computers, including the Compaq, have a monochrome display that simulates the CGA, EGA, or VGA in various shades of its single color.

Identify the type of video display your computer has. Is it color or monochrome, and does it have a CGA, EGA, or Hercules graphics board, or possibly no graphics board? Ask the instructor or consult the manual to find out.

When you turn on the computer system, a small line (or a small highlighted square) will appear on the screen. That line is called the "cursor," and it is very important. Anything that you type on the keyboard will begin to appear at the cursor position.

The Keyboard

The keyboard is still the most common way of "speaking" to a microcomputer. If you have not used a computer before, several of the important keys may be new.

The most important "different" key is the **<carriage return>** or **<enter>** key. This key is not named on some keyboards, but may have an arrow pointing down and to the left, representing a carriage return and a line feed. On still other keyboards this key is labeled "Enter." The key is used to tell the computer to take action or to enter that which has been typed, such as a command or an item of data.

Programs are not consistent in how they require use of the carriage return. In some cases, simply typing the command is sufficient; in others the command must be followed by a carriage return before anything happens. In this course, we have designated the carriage return as "<Enter>". If at any time when a command or data entry is given nothing happens, press the carriage return.

Another important key is the **control** key. This key is marked on most keyboards as "Ctrl". In use, this key is held down while another key is pressed, very much as the shift key is used for upper case. To reduce space in computer documents, the control key is represented by the "^" symbol or by Ctrl. Thus control C can be represented as ^C or <Ctrl-C>. The control key does not print a letter, again like the shift key. To give an example, <Ctrl-C> interrupts most of the DOS commands like COPY in case you want to stop in the middle of an operation.

The **Alt** (alternate) key is used like the control key, to change the function of other keys. Its use is less common than the control key, and there is no convention for referring to it in the literature. In this document, the alternate key is referred to as the <Alt> key.

The **Esc** (escape) key is used to escape a condition. In some cases, prompt messages on the screen inform the user to press the key. In other cases, the user may try it to get out of a situation that may have been created unintentionally or sometimes to return to a previous menu or part of a program.

The function keys, **<F1>** - **<F10>**, on the left side of the keyboard have purposes assigned by the programmer. Therefore, their purpose depends on the software in use. Usually they are used to perform actions such as saving a file or leaving the program. In many programs, <F1> is the "Help" key and may display a message containing information about how the program works.

On the right side of the keyboard is the numeric keypad. These keys have dual use. In one mode, they do not enter numbers as marked, but the **arrow keys** instead move the cursor up, down, left, right, or to other places as designated. **PgUp** and **PgDn** move up and down a page (screen), and **Home** and **End** move to the beginning and end of documents. In the other mode, the keys represent a 10-key calculator. The mode is changed by the Num Lock key at the top of the keyboard. If you get numbers on the screen when you meant to move the cursor

(or vice versa), change the mode of the keypad with the Num Lock key. Unless you are an accountant accustomed to entering numbers rapidly on the keypad, Num Lock is more of a nuisance than a help.

Below the numeric keypad are the **<Ins>** (insert) and **<Del>** (delete) keys. The <Ins> key causes characters that you type to be inserted, spreading apart the existing text to make room for the newly typed characters. When the <Ins> key is pressed again, the insert mode is cancelled and anything you type will "type over" the previous text. The <Del> key causes characters already typed to be deleted.

The **<PrtSc>** (print screen) key does just what it says--sends a copy of whatever is on the screen to the printer. On some computers you must hold down the shift key to activate <PrtSc>.

The center area of the keyboard generally follows the layout and function of typewriter keys. Like most modern typewriters, any key that is held down will repeat until released.

Examine your keyboard. Find the keys described above. Press the Num Lock key. On some keyboards, a light indicates that Num Lock has been pressed. Press Num Lock again to return to the original setting.

The Central Processing Unit (CPU or Microprocessor)

There are several models of IBM-compatible microcomputers. The original IBM PC was based on a microprocessor called the Intel 8088 and had floppy disks only. The IBM PC/XT used the same "chip" but had a hard disk. In the past few years, the faster IBM PC/AT with the 80286 chip has been popular. Now users who want greater speed are buying computers containing the 80386 chip which is about twice as fast as the 80286, and the 80486 offers still more speed.

The speed of processing is affected not only by the type of chip but also by the "clock speed" (4.77 megahertz in the PC, up to 33 megahertz for the 80386).

Some kinds of programs containing many numeric calculations run much faster with a "numeric coprocessor," an optional accessory that speeds up calculations.

If you have access to a computer with the top removed, find the "motherboard," which usually lies on the bottom of the computer, and the "bus" connectors that contain the accessory boards, such as the video graphics board or accessory memory boards. Have someone point out the CPU if it is visible. What type is it?

Random Access Memory (RAM)

The CPU can process only one or two items (usually characters) at a time and has no place to store a program. Data and programs are therefore kept in MEMORY. The contents of memory can be retrieved by giving the location or address of a byte to be retrieved. In today's microcomputers, the contents of any address can be retrieved and in any order; memory is therefore called **Random Access Memory** or RAM.

RAM consists of transistors and/or capacitors packed densely into chips; it remains active only when supplied with power. When the computer is turned off, the RAM "forgets" whatever it contained. Thus if you have typed a long document into a word processor but have not yet saved it as a disk file, there is apt to be weeping and gnashing of teeth if the power goes off.

The size of RAM is measured in units of 1024 bytes called kilobytes or K or in larger units, roughly a million bytes, 1000 K or 1 megabyte. Most IBM-compatible microcomputers now have 640 K of usable RAM but many have 1 or 2 megabytes that can be used by some programs or as "RAM Disk," in which memory is used to simulate a hard disk that operates much faster than a normal hard disk.

How much RAM does your computer have? How many typewritten pages of 300 words each will it hold? The instructor can help determine this, or, if you already know how to run programs, the program called CHKDSK, a DOS utility, will display the amount of RAM as "total memory."

Read Only Memory (ROM)

When the computer is first turned on, some instructions must be present for the CPU to load DOS from a disk and run the operating system. These are stored in another kind of memory called Read Only Memory, or ROM. ROM is programmed at the factory and can only be read, not written upon. It retains its contents without electrical power, however, and is therefore useful for storing the startup or "bootstrap" instructions. Games in video parlors often store their programs in ROM, since there is no need to change the program or store new data.

External Connections

The rear panel of the system unit provides connections for the hardware components. The computer power cord connects here, and there may be a power outlet for the monitor cord. The keyboard plugs into a special connector. In the back (IBM) or on the right (Compaq) are the ends or the "boards" that are seated on the "bus" inside the computer. Some of the boards will have connector plugs or jacks. These are the input and output ports through which the computer communicates with the monitor, printers, and modems. In one of the slots is a small plug that should fit the video cable from your monitor.

The printer connector will be the most difficult to locate because the location of that plug is not standardized in location. It is important to know if your printer is a "serial" or a "parallel" printer and to connect it to the corresponding port. If it is a serial printer, you will also need to issue a "mode" command to tell DOS through which port and how fast to send printer instructions. The DOS and printer manuals should explain how to set up a printer, but you may need help

from your dealer or a knowledgeable friend or instructor the first time you do this.

Locate the serial and parallel port connections for your computer with the help of an instructor or the manual. Serial connectors may be 9-pin or 25-pin, male or female; parallel ports may have "Centronics" connectors or 25-pin connectors (unfortunately) resembling those for serial ports, but usually of different "gender." Cables for modems and printers are usually different in the way the internal wiring is arranged. The moral of this tale is that when buying a cable, be very specific about the computer or peripheral for each end, and consult someone who is knowledgeable. When going on an outbreak investigation, be sure to take along all the proper cables and test them well in advance.

General Care and Maintenance

You can relax in the knowledge that you cannot break the computer by typing something on the keyboard. You may end up somewhere in the program other than the place intended, but the computer will not be harmed. In fact there are very few ways to harm the computer short of dropping it or spilling liquids into it. One way to cause harm is to open a disk drive while the drive light is on. That can cause loss of data, and in some cases, harm the disk.

You should also be careful not to use the computer during thunderstorms because voltage surges often occur that can damage the computer. You can buy a voltage-spike protector to connect into the power cord, which provides some protection, but the effectiveness of these devices varies. Compaq computers and many others have some voltage spike protection built in, but this, of course, will not help protect the contents of the RAM during a power outage.

Reasonable care is required to keep the computer operating properly. Handle it gently, and before moving the desk model machines (particularly the XT with the hard disk), use the special program for parking the hard disk recording head so that it will not damage the disk surface. If you remove the cover of the computer,

be sure to unplug the power cord first. Given reasonable care, the machine should give good service for years.

The first thing to check when a computer doesn't work is the "connectors." Be sure that all the cable connections are well seated and, if possible, secured with screws. If you have access to a computer with the top off during this course, examine the boards and observe how they can come loose from their connections with the motherboard (bus) during shipping. Experiment with pushing the boards down with a finger until they are properly seated in their sockets. This is one of the more common computer "repairs," sustaining a minimum charge of $30-$50.

Chapter 3

The Disk Operating System: Facilities at the Brainstem Level

MS-DOS and PC-DOS

The disk operating system is the software "closest to" the hardware of the computer. It receives commands from the user or from other programs and reinterprets them in a form that causes the computer to do useful work.

The operating system used in IBM-compatible computers is called DOS for Disk Operating System. The version sold by IBM is called PC-DOS; the one from Microsoft is MS-DOS. The differences are small and not important in this course. We will refer to the operating system as DOS from now on.

Applications software, such as word processing, spreadsheets, data bases, and Epi Info, must communicate with the computer hardware through the operating system. The applications software sends requests to the operating system, which in turn sends instructions to the central processing chip of the computer.

Because of its key and constant role in the functioning of any computer, DOS must always be loaded into the computer first. When you turn your computer on, a ROM in the computer directs the loading of the DOS system from drive A. If drive A is empty, DOS is loaded from the hard disk if possible.

On a computer with a hard disk, if no disk is present in drive A, DOS will be loaded from the hard disk. If, however, drive A contains a disk, but the disk does not have the DOS files, an error message will result. You should therefore be sure that drive A is empty or not latched when you turn your hard-disk computer on.

The essential part of DOS consists of three programs, COMMAND.COM and two hidden files (called IO.SYS and MSDOS.SYS, but never visible in directories).

They are the ones that are loaded into memory when the computer is turned on. The most visible program in DOS is COMMAND.COM. This program interprets instructions for loading other programs, saving programs, getting directories (DIR), copying files (COPY), and other basic activities. Other DOS programs can be loaded when needed but are not required for routine operation. These programs include those for formatting disks (FORMAT), "setting up" the output ports (MODE), and other chores.

The following exercises assume that your computer has a hard disk set up with a root directory and directories for DOS and Epi Info. If your computer is new and has never been set up before, see the manual and, if possible, consult a knowledgeable friend for the setup procedures.

LET'S POWER UP, turn it on, and boot up. Be sure that the computer is plugged in as in the previous chapter. If you have a floppy disk machine (no hard disk), place a disk containing DOS in drive A and close the latch; if you have a hard disk, be sure that drive A is empty. Now turn the power switch on. If you have a separate video monitor, you may also have to turn on the monitor. Soon you should see some numbers at the top of the screen. These show the progress of a memory check. If you have a bad memory chip or loose board, an error message will be given. The last number given (perhaps 640 K) indicates the amount of RAM memory in your computer.

Now you will see the indicator light for one or more drives become active while the DOS files are loaded into RAM. You may now be asked for the date and time. If so, type these in the format specified; if not, your computer probably has a battery-powered clock that supplies DOS with the date and time automatically. The computer may also load programs or perform other chores at the behest of a file called AUTOEXEC.BAT (more on this near the end of this chapter). When the housekeeping is all done, however, you should see a prompt:

C:\>
or
A:\>

(Perhaps without the ":\", depending on how your machine is set up.)

This means you are on (LOGGED ONTO) drive C or drive A and that DOS is ready to receive a command. In the rest of this chapter, we will assume that C is the primary drive where DOS resides. If this is not the case, substitute "A" or another appropriate drive letter in the commands.

Because not all applications programs completely eliminate the use of DOS, you need to know at least the essential DOS commands and how to name files and navigate through various parts of floppy and hard disks. It is also important to know how to make backup disks and to adopt a regular routine for doing so. This provide insurances against losing valuable files in case of hardware or software disaster.

Changing the Logged Disk Drive

To change the logged drive, the command is (new drive letter):

If your prompt is C:\>, type "A:" (without the quotation marks) and then press <Enter>. The prompt should change to A:\>. Try changing the logged drive several times. Note what happens if no disk is in the drive or if you type "D:". If you get the message "Abort, Retry, Fail," type "F" and then at the next prompt give the letter of the desired drive and a colon, as in "C:".

DOS File Names--A Key Concept

The names of files in DOS are constructed in a special way of:

8 or fewer characters, an optional period, and an optional 1 to 3 extra characters

AAAAAAAA.BBB
Filename Period Extension

They must begin with an alphabetic or numeric character, but can contain some embedded special characters such as hyphens (no spaces). Upper and lower case letters have the same meaning.

Legal DOS file names include:

COMMAND.COM
command.com
Letter1.txt
ANALYSIS.EXE
My-Words.TXT
1.BAT

Illegal names would be:

ThisIsMyFile (Too long)
This.Is.It (Only one period allowed, unlike the rules on many other computers)
*Text (Comma, *, and others not allowed, especially at beginning)
1,2,3.Fil (No commas allowed)

The final three letters (the "extension") are often used to indicate the kind of program or file being named. DOS identifies *programs* by their .COM, .EXE, and .BAT extensions. Extensions often used for convenience are .DAT for DATA, .TXT for TEXT, and .BAK for backup files. Epi Info uses .QES for questionnaire files and .REC for RECORD or data files.

The DIR Command: What Files Are Available?

DIR is a DOS command standing for DIRECTORY. It displays the names of all files in the current directory that match the file name (if any) that you supplied. For each file it also gives the size of the file in bytes, and the date and time when it was created.

Type DIR at the > prompt and then press <Enter>. You should see a list of the files in the current disk or directory (much more on directories later). Now type:

C:\>**DIR *.BAT <Enter>**
(Remember that C:\> is already there, since it is not in bold type.)

The space after DIR or any command is important, since this is the only way DOS distinquishes one word from another. Also note that you must press <Enter> to tell DOS to execute the command. We will assume this from now on and not print <Enter> after each command.

The "*" is a "wildcard" character, meaning "any file." We have just requested the names of all files ending in ".BAT". Try some other combinations. DIR can be followed by: *.* meaning all files as in:

C:\>**DIR** *.* or simply **DIR** (all the files)
or
C:\>**DIR A*.*** (all those beginning with A)

Of course you may have to alter the commands to match the files available on your disk. Adding "/P' causes DIR to pause after each screen and wait for you to press a key. "/W" displays the filenames in a Wide format so that more will fit on the screen, for example:

C:\>**DIR /P**
C:\>**DIR /W**

Seeing What's in a File: the TYPE Command

The TYPE command displays the contents of a file.

At this point you must change directories. Directories will be explained later, but for now, just type:

C:\>**CD EPI5**

The prompt will change to show that you are in the EPI5 directory:

C:\EPI5>

Now give the TYPE command:

C:\EPI5\>**TYPE OSWEGO.REC**

Remember that the "C:\>" is already on the screen and that you must end a command by pressing the <Enter> key. You should see some text that represents the contents of this file.

Note that some of the text disappeared from the top of the screen as it whizzed ("scrolled") by. You can cause TYPE to pause by adding | MORE (the "|" consists of two lines, one above the other on the DOS keyboard). Try this command:

C:\EPI5\>**TYPE OSWEGO.REC | MORE**

When the MORE prompt appears, press <Enter> to see the next screen.

If you TYPE a long file, you may want to escape before the end is reached. <Ctrl-Break> will accomplish this. Try it with OSWEGO.REC.

Printing the Screen or the Contents of a File

To send the contents of the screen to your printer, press the <PrtSc> key. On some keyboards, <PrtSc> may be in the upper-case position and require pressing the shift key at the same time. To turn the printer on and print all subsequent screen displays, press <Ctrl-P>. A second <Ctrl-P> will turn the printer off again.

If you don't have a printer, move to another computer that has one. Be sure that the printer is connected to the computer and to a power outlet. Load the printer with paper, being sure that the sprocket holes on each side of the paper are engaged on the sprockets evenly. The lever for the platen should be in the released position so that the paper can slide between the platen and the other rollers. The other position is for plain (non-computer, no sprocket-holes) paper, or friction feed.

Turn on the power switch for the printer. If there is a "select" button, press it until the indicator light goes out, showing that "local" is selected. Now use the line-feed control, or if there is none, the knob on the platen, to move the paper so that the top of a page is directly under the printhead. Press the "select" button so that "on line" is selected.

Press the <PrtSc> key (with <shift> if necessary).

The screen should be printed.

Now turn on the printer by pressing <Ctrl-P>.

Then:

C:\>**TYPE OSWEGO.QES**

The contents of OSWEGO.QES should be printed.

Now press <Ctrl-P> again to turn the printer off. This command "toggles" the printer so that the output will again go to the screen

only. There is another DOS command for printing.

C:\EPI5>**PRINT OSWEGO.QES OSWEGO.REC**
will print both files. Answer the question about output with <Enter>.

Running Programs

To run a program from DOS, type the name of the program, without the extension or the period, after the DOS prompt. This causes DOS to find the program on the disk, load it into RAM, and begin running the program. From that point until you have left the program, whatever the computer does is controlled by that program, working through the operating system.

There are three kinds of programs that can be run by DOS, with three different file name extensions. A program called EPI, for example, could exist in three forms:

EPI.BAT	(A text file containing DOS commands)
EPI.COM	(Two formats for programs, .COM being the older
EPI.EXE	and simpler of the two.)

Any of the three would be run by typing EPI at the DOS prompt.

Use the DIR command to look for EPI.* You should find a program called EPI.EXE.

Now run the program by typing:

C:\EPI5> **EPI**

(Depending on how DOS is set up, the prompt may be C> only.)

You should see a menu, which is actually a menu written in the programming language called Pascal and contained in a file called EPI.EXE. DOS has loaded the file into random-access memory and

executed its commands one by one to produce the menu and allow interaction from the keyboard. For now, you should merely press <F10> to leave the menu.

Then type:

C:\EPI5>**CD **

to return to the "root" directory, as will be explained later.

As soon as the program stops running, the memory it occupies becomes available to DOS for loading the next program.

FORMATting a Disk

Before a new floppy disk can be used, it must have sectors and tracks recorded on the surface as landmarks for DOS to use in storing files. The DOS command FORMAT--actually a separate program--does this job. It will also take care of putting a new copy of DOS, including the hidden files, on the new disk if you wish.

Unfortunately, **FORMAT is dangerous**, since it is possible to format (erase) your entire *hard* disk if you fail to specify which drive is to be formatted or indicate the wrong drive. **It should be used with full attention and you should be sure to indicate which drive you want to format, or the program may assume you want to format the "current drive," which could be your hard disk!!**

Insert a new disk into drive A. Give the command:

C:\>FORMAT A: **(Use FORMAT A:/4 to format a low- or double-density disk in a high-density drive.)**

A: indicates the disk drive. FORMAT will ask if you are ready to proceed; reply "Y" for "yes." The disk drive will now become active

for some time, and the number of tracks formatted will appear on the screen. For a double-density disk this is 40; high-density disks have 80. When the DOS prompt appears, type DIR A: to see the contents of the disk. No files will be present.

If you have a Compaq Portable III, IBM PC/AT, or another computer with a high-density disk drive, things are more complicated. Since the tracks on the high-density disks are narrower, there are some incompatibilities. High-density disks can be read or written on only in high-density drives; that much is simple. Double- (industry hype for low-) density disks can be read by either type of drive. **But you should be careful never to write on the same disk with both high- and low-density drives. Doing so can make the disk unreadable.**

If you want to format a low-density disk for transfer of programs to an older computer, for example, use the low-density drive if you have one, and if not (as on the Compaq Portable III), use the command **FORMAT A:/4**, which will result in a double-density disk format. According to some authors, however, the resulting disk is not reliable when used in a double-density drive, and it is better to format the disk on a computer with a double- (low-) density drive.

Many computers are now equipped with 3 1/2-inch drives. These also come in double-density and high-density varieties, holding 720K or 1.44 megabytes (1440K) respectively. Consult the DOS manual for methods of formatting a double-density disk in a high-density drive, although it is generally safer to format a disk in a drive designed for the correct density.

Copying Files

Making a copy of a file is one of the commonest computer operations. The DOS command COPY will accomplish this. Another command, DISKCOPY, is sometimes useful for copying an entire disk, but COPY is adequate most of the time. COPY requires two extra pieces of information: the name and location of the program to copy, and the name and location of the new file that will result.

Note that the new file produced by COPY can have a name either the same as or different from the original. The format for the COPY command is:

COPY	Disk location & File name (The Original or Source)	Disk location & File name (The Copy To Be Made or Destination)

The space between the two file names can be understood as meaning TO, but an error message will result if you actually type the word rather than the required space.

COPY understands several different wordings; a few examples with the file names SOURCE and DESTIN will illustrate.

A:\>COPY A:SOURCE C:DESTIN
Copy SOURCE from drive A to drive C and call it DESTIN.
(The "C:" may have to be replaced with the letter for another drive on your system.)

A:\>COPY A:*.* C:
Copies all files on drive A to drive C and gives them the same names.

If the drive letter is omitted, the drive where we are currently located ("logged") is understood.
A:\>COPY *.* B: is the same as A:\>COPY A:*.* B:

If only one file name is given, it is understood to mean the source; the destination files will be given the same name.

A:\> COPY B:SOURCE means "Go and get the file B:SOURCE and put the copy on the current drive (A) with the same name (A:SOURCE)."

As with the DIR command, the asterisk (*) may be used as the "wildcard" indicator.

C:\>COPY A:*.EXE means COPY all the .EXE files on A to drive C.

If the destination file already exists, it will be replaced by the new copy.

Place the disk you have formatted in drive A.

Then give the command:

C:\>CD \EPI5

Now copy a file from drive C to drive A. First do DIR C: (same as DIR C:*.*) to see what files are present.

Then give the command:

C:\EPI5\>COPY C:HOUSE.QES A:

Now do DIR A: to verify that the file was actually copied to drive A. The date, time, and file size should be identical in the original and its copy. If no name is given for the destination file, it will be assumed to be the same as that of the source file.

Make another copy and give it a different name:

C:\EPI5\>COPY C:HOUSE.QES A:HOUSE1.QES

Now DIR A: should show that you have two files on drive A:. Try the command:

C:\EPI5\>COPY *.BAT A:

What did it do?

Practice using the copy command until you can easily make copies to and from any drive. What happens if the destination file you give already exists? You can find out by copying PERSON.QES to HOUSE1.QES, examining HOUSE1.QES before and after with the TYPE command.

Renaming Files

The DOS command RENAME will change the name of a file. The command syntax is very similar to that for COPY:

RENAME Original File Renamed File

Type:

C:\EPI5\>**RENAME A:HOUSE1.QES HOUSE1.OLD**

(Note that the second name cannot have the drive letter specified. This is reasonable, since it must always be the same as that of the original file.)

Use DIR to see the results

Deleting (Erasing) Files

The commands ERASE and DEL (for DELete) are identical, and you can use the one that is most natural. Deleting a file is normally irreversible, but the actual process does not erase the recorded data on the disk, and there are programs like Norton Utilities that will resuscitate (unerase) a file that has been deleted inadvertently. If you do delete an important file inadvertently, be sure to avoid copying or creating any more files until you have found someone who can recover it, since space occupied by deleted files on a disk is written over by creation of other files.

Type:

C:\EPI5\>**DEL A:HOUSE1.OLD** **(Or ERASE A:HOUSE1.OLD)**

Use DIR to see the results

Directories, Paths, and the Hard Disk

A floppy disk is like an in-basket, small enough so that objects (files) can be found within it despite their lack of structure. Hard disks, however, hold hundreds of files, and DOS provides subdivisions called "directories," which might be thought of as equivalent to file drawers or folders.

Every disk (even a floppy disk) has a "root" directory, referred to as "\" (backslash). The backslash is entirely different from "/," the forward slash, and both are located in different places on various keyboards, but such is life in the DOS world.

Directories are arranged in a "tree," like the Linnean classification of the biological world, so that all subdirectories "belong" to the root, and may in turn have other subdirectories that belong to them. The location of any given file can be described by giving the drive letter followed by a list of the directories and subdirectories that lie on the "path" of the file.

C:\EPI5\OSWEGO.QES tells us that the file OSWEGO.QES is in the directory called EPI5 that is connected to the root directory on disk drive C. It is like saying that in the country of C and the state of EPI5 there dwells a file called OSWEGO.QES. Directories can contain subdirectories, so that a complete "path" describing a file might be "C:\EPI5\MYDATA\OUTBREAK.REC".

Floppy disks can also have subdirectories other than the root, but usually do not because dealing with the subdirectories is not worth the trouble for a limited number of files. In this section we will first learn how to see the directory structure on a disk, then how to navigate among directories, and finally how to make or remove directories. If you do not have a hard disk, you may skip this section.

To see the directories contained in the root (subdirectories of the root), use the DIR command and look for all the files with no extension:

C:\EPI5\>**CD ** (Change to the root directory)
C:\>**DIR *.**

This will list several files with "<DIR>" after the name. They are the directories (actually subdirectories, but usually this term is only used to emphasize that a directory is within another directory).

Now let's move to the EPI5 directory:

C:\>CD \EPI5

CD stands for "CHANGE DIRECTORY" and can be followed by the complete path of the destination directory. If the directory being requested is a subdirectory of the current directory, the complete path is not required. We could have used CD EPI5 instead, since we were already located in the root.

Use DIR to verify that you are indeed in a different directory. Now return to the root directory:

C:\EPI5\>CD \

Check again with DIR.

Finally, let's make and remove a new directory:

C:\EPI5\>MD TEMP (or MKDIR TEMP)

MD stands for MAKE DIRECTORY. It will make the directory as a subdirectory of the current directory, so your location at the time the command is given is important. Although we have now made a new directory, you are still in the root. To enter the new directory use:

C:\EPI5\>CD TEMP

Use DIR to see that C:\EPI5\TEMP is empty except for the two standard directory files that help DOS maintain the directory structure.

In order to remove the directory, it must be EMPTY and we must be in the directory above it, in this case in \EPI5. Move \EPI5 with:

C:\EPI5\>CD \EPI5

Check to see the TEMP <DIR> entry in \EPI5:

DIR *.

And then remove TEMP with:

C:\EPI5\>RD TEMP
(for Remove Directory)

Use DIR again to verify that TEMP<DIR> is gone:

C:\EPI5\>DIR *.

CD TEMP will also let you know that the directory is gone.

You can use the "Path name" with almost any DOS command to work with files that are not in the current directory. To see the file names in \RSURV, use the command:

C:\EPI5\>DIR \RSURV\

\RSURV\ (actually C:\RSURV\) is the path name, meaning "the directory RSURV that is a subdirectory of the root on drive C."

Now run a program in \RSURV called RSURV (actually RSURV.BAT):

C:\EPI5\>\RSURV\RSURV

Press <F10> to exit from the program.

There is a DOS command called PATH that sets up a series of directories so that DOS automatically looks for programs to be run in all these directories. You have in fact been using this feature because the DOS commands like FORMAT are not in the root directory but neatly stowed in their own directory called \DOS. You were able to run them because a PATH command had been set up by your AUTOEXEC.BAT file. AUTOEXEC.BAT will be explained in a later section. To see the PATH that is set up in your computer, type:

C:\EPI5\> **PATH**

Backing Up the Hard Disk

A hard disk has a metal and ceramic reading and recording head riding at high speed only microns above a soft ferric oxide layer on a metal platter. When a speck of dust lodges between the head and the disk, tearing a hole in the coating, the disk "crashes." Most professionals with experience of more than a few years in computing have had personal encounters with hard disk failure or crashes. **It is important to adopt a regular policy of making backup disks and storing these in a safe place.** The contents of the hard disk can then be replaced if necessary.

Partial backups, containing only new material, can be made using the copy command, copying only the current working file to a floppy disk. This can be done as often as once an hour. Set the interval for the amount of work you are willing to repeat if a crash or lesser disaster occurs. In the class we will describe how to make a complete backup of the hard disk so that incremental or partial backups can be made later to keep it up to date. At intervals, perhaps weekly or monthly, you will want to repeat the complete backup procedure, keeping the last backup for extra security. Most experts recommend that two or more sets of disks be kept in circulation as backups, including the current one, and the next-to-current set. Disks older than that can be recycled in most situations.

More elaborate backups are needed if you maintain large data sets for others. One copy of such backups is often stored in a vault or other location remote from the building housing the computer in case of fire or water damage. If the computer is at the office, for example, you may want to store one backup at home.

We will begin to make a complete backup of hard disk C onto floppy disks in drive A. If you do not need the backup and time is short, pressing <Ctrl-C> will interrupt the process.

Format at least one high-density floppy disk as described earlier in this chapter.

Type

C:\>BACKUP C:*.* A:

If you add "/S", all subdirectories of the "\" or root directory will be included, meaning the entire hard disk. The backup would probably require many floppy disks. In DOS version 3.3, you can add "/F" so that disks are formatted automatically if necessary.

Insert the first disk in drive A when requested to do so. Be sure to label the disks as they are used with the date, time, computer, and the command used to create them ("BACKUP C:*.* A:"). The program will guide you to insert disks as necessary until the process is complete. If you do not wish to complete it, use <Ctrl-C> to interrupt. If your computer is new, however, and has important software on the hard disk, this is a good chance to make the all-important first backup.

To make an INCREMENTAL backup, use the /D parameter. Only files that have been created or altered since the last backup will be backed up. If you add /D:07/01/88, BACKUP will back up only the files created or changed since July 1, 1988. Since this depends on the date and time recorded for each file, it is important to keep your DOS date and time correct by responding to the date and time questions (if any) that are displayed when you turn the computer on.

Although in this class we will not actually restore files from the backup disk, it is important to know how to do this. If your disk crashes and you are starting again after repair or replacement, the new hard disk would first need to be formatted and

set up with the DOS system. You would then need to create directories that match those on the original disk by using the MKDIR command. This might be done as follows:

C:\>**MD DOS**
C:\>**MD EPI5**
C:\>**MD CDC**

You would then use the following command to restore the files contained on your backup disks:

C:\>**RESTORE A: C:*.*/S**

If you wanted to restore only selected files, the *.* could be made more specific or preceded by a single directory name, as in RESTORE A: C:\EPI5*.REC.

Other options for backing up hard disks, such as XCOPY (for DOS 3.3) and commercial programs such as FASTBACK may be discussed in class.

DOS Batch Files, AUTOEXEC.BAT, and CONFIG.SYS

A batch file is a file containing DOS commands. It can be made in a word processor such as EPED. When the name of the file is typed at the DOS prompt, DOS "runs" the file, executing each of the commands in the file just as though you had typed them at the keyboard.

AUTOEXEC.BAT is a special kind of batch file that DOS runs automatically each time the operating system is loaded, if a file by this name is present.

Examine the contents of your AUTOEXEC.BAT by doing:

C:\>**TYPE \AUTOEXEC.BAT**

Some of the commands will be familiar, and others may be new. The DOS manual explains each command, a few of which are used only in

batch files. One useful command is PROMPT PG, which causes DOS to display the directory along with the disk letter at every prompt. After you have made a change in AUTOEXEC.BAT, the change will not take effect until you run AUTOEXEC or reboot the computer.

There is another file, called CONFIG.SYS, that DOS consults whenever the operating system is loaded. It contains instructions to the operating system about the number of files that can be open at one time and other details.

Look at your CONFIG.SYS file with:

C:\>**TYPE \CONFIG.SYS**

It probably contains the statements such as:

FILES = 20
BUFFERS = 20

If not, or if there is no CONFIG.SYS in the root directory, you may need to create one with these commands. Without them, the number of files open at one time is limited to 4. The hard disk will be noticeably slower without a buffer setting of 15 or 20, although the "buffers" do occupy extra RAM space.

Rebooting the System

After changing something in AUTOEXEC.BAT, or if a program has gone hopelessly wrong, it is necessary to "reboot" the computer, starting from the beginning by loading a fresh copy of DOS into memory. This can be done by turning the computer off and then on again after a brief wait, but there is an easier way.

Press the following keys in order, holding down the first two:

<Ctrl><Alt><Del>

This is called a "warm boot" (or the Paganini maneuver because it requires long and agile fingers); it causes DOS to reload from the disk. Programs and data that are in memory will be lost.

DOS Review Exercise and an Alternative Approach to Backing Up

Having a good backup strategy is insurance against major frustration and/or disaster in using a computer. There are commercial programs that do a good job of backing up an entire hard disk in compressed form onto a number of floppy disks. Since one of these has not been supplied with your computer, you need to develop an alternative strategy. This means trusting CDC resources to replace the programs on the hard disk if necessary, but backing up any new files that you create on floppy disks. You can do this by using the floppy drive for text files and smaller data files. For greater processing speed, you will probably want to keep larger data files on the hard disk, but transfer them at intervals (after every one to several hours of work, depending upon how much you are willing to lose) to a floppy disk for safe storage.

Pretend that you have just created a series of programs to process surveillance data. Their names all end in .PGM and they are located in a directory called \RSURV on your hard disk.

1. **Format a floppy disk.**
2. **Change to the \RSURV directory, copy the .PGM files in \RSURV to the disk you have formatted, label it, and put it aside for backup.**

Now pretend that you want to "restore" all these files to a directory that you will create called \TEMP.

1. Create a directory called \TEMP on drive C.
2. Copy the files with names ending in .PGM to this directory.
3. Use the directory command to confirm that they have been copied.
4. Delete (erase) all these files and then remove the directory to prepare the computer for the next class.
5. Change to the \EPI5 directory and run the EPI.EXE program.

Summary

This completes the DOS section of this course. Although we have only scratched the surface, you now should know how to:

- Set up and turn on the computer and the printer
- Find files with DIR
- Display the contents of a file with TYPE
- Print the screen or the results of any DOS command
- Identify or construct a DOS file name
- FORMAT a disk
- COPY files
- DELete or RENAME files
- Move around among directories on a hard disk
- Create or remove directories
- Make backups of the hard disk
- Examine and change the AUTOEXEC.BAT and CONFIG.SYS files
- "Reboot" the computer

Further progress requires time spent at the keyboard, real tasks to be accomplished, and frequent reference to the DOS manual and other books. Several such references are listed below, but most bookstores have other selections that may match your own style of learning better.

References

1. Schildt, H. *DOS Made Easy*. Osborne McGraw-Hill, Berkeley, California, 1988.

2. DeVoney, C. *Using PC DOS*. QUE Corporation, Indianapolis, Indiana, 1986.

3. Levy, S. *Hackers: Heroes of the Computer Revolution*. Anchor Press/ Doubleday, Garden City, New York, 1984.

Chapter 4

Using EPED, a Simplified Word Processor

Word Processing Programs

For most professionals, word processing is the most frequently used function in a personal computer. Composing a letter, memo, report, or questionnaire is central to many kinds of activities.

The first computer programs for processing text were called editors. They allowed text to be typed on the screen and later brought back and edited. Their big advantage over the typewriter was error correction before printing, so that a perfect page could be produced without the aid of either a talented typist or gallons of white-out. To recall a line in an editor, one often had to refer to the line number, and justifying the text after a deletion or insertion often required a separate command and a long wait.

"Word processors" generally show text on the screen as it will be printed, allow you to move about on the screen with the arrow keys or some other device, and make justification easy or completely automatic. Most allow plain, **bold**, or *italic* type, and show it as such on the screen. Word processing programs commonly used at CDC include Word Perfect, Microsoft Word, Wordstar, and Multimate (because it is most compatible with Wang word processors).

Many key concepts are common to all word processors. These include opening and saving files, entering and correcting text on the screen, selecting and moving or deleting blocks of text, searching for and replacing words and phrases, and making backup files.

EPED is a public domain program included in the Epi Info programs for epidemiologists. It may be considered a "high-end text editor" or "low-end word processor," since it has full-screen editing and semi-automatic justification. It is

used in this course because it offers an easy and inexpensive way to begin word processing. If your work requires letters, memos, and other documents without embedded graphics, you may not need to use another word processor. If you do decide to move to one of the programs in the $200-$600 range, however, the concepts learned in EPED will still apply to most other word processors.

Word Processing Using EPED

To save time, the Epi Info programs have already been loaded onto the hard disk in the computer you are using. Instructions for this step are in Chapter 4 in the Epi Info manual, but you will not need to perform it today if you are using these materials in a formal class setting.

In this part of the course, we assume that you have Epi Info, including EPED, installed in a directory called EPI5 on hard disk C. If you are using the D or E drive or working from floppy disks on drive A or B, substitute the appropriate disk letter and/or directory in the instructions below.

Turn on the printer, monitor, and computer and allow DOS to load (remember drive A must be empty or open if you have a hard disk). When you see the C:\> prompt, change to the Epi Info directory:

C:\>**CD \EPI5**

You should see a new prompt:

C:\EPI5>

Now type EPI to run the main menu of Epi Info. You should see a menu that looks like this:

```
Epi Info Version 5

Public Domain Software for Epidemiology and Disease Surveillance

Centers for Disease Control                World Health Organization
Epidemiology Program Office                Global Programme on AIDS
Atlanta, Georgia                           Geneva, Switzerland

Programs                                   Description

EPED      Text editor                      Use EPED to:
ENTER     Data entry                     2 Create questionnaires
ANALYSIS  Data analysis                  2 Write memos or letters
                                         2 Receive guidance in investigations
CHECK     Data validation                2 Write or edit Epi Info programs
STATCALC  Epi calculator
CONVERT   Export data                      EPED is a word processor for creating Epi Info
IMPORT    Import data                      questionnaires or for general text processing.
MERGE     Merge data files                 It contains the EPIAID system for thinking about
VALIDATE  Compare files                    epidemic investigations and producing drafts of
                                           investigation reports.
About Epi Info
Quit Epi Info                              Params: EPED <Textfile or EPIAID program name>

EPED
St , letter, or space to select            F9 for Setup, F10 to Quit
```

Note that the highlight bar is on the first item, "EPED - Text editor." To select others, you would move the bar with the up or down arrow keys, but since we want the first, merely pressing <Enter> (or typing the hot key, "E") will select it. Press <Enter> and you will soon see a banner announcing EPED and then a screen with a series of menu choices at the top. This is the EPED word processing screen.

Running the Tutorials in EPED

You will now see a full screen, like a blank sheet of paper, on which you can create a document. The commands available in EPED are displayed across the top of the screen.

EPED contains several tutorials that present the main features interactively. The tutorials are contained in the portion of EPED called EPIAID.

At the top of the screen major types of EPED commands or menus are displayed. Note that EPIAID is reached by pressing the function key labelled "F3."

Press <F3>. A menu will appear. Use the down-arrow key to move the cursor bar to "Word Processing Tutorial." Press <Enter>. A brief introduction to the tutorial will appear. Press <Esc> as you finish reading each screen until you come to another menu called "The EPED Word Processing Tutorial." The menu offers five small tutorials that display the important features of EPED. Choose the first by pressing <Enter>. Follow the instructions in the tutorial until you have completed "The Basics of Word Processing" and returned to this same menu. Now choose the second topic using the down-arrow key and <Enter>. Continue in similar fashion until you have completed each of the five tutorials.

You may see a "File not found" message at the beginning of one or more tutorials. The tutorial programs attempt to erase the work of previous students to avoid confusion, and the message indicates that you are the first one to do this tutorial.

When you have completed all five tutorials, choose "Return to the EPIAID Main Menu," and then press <F10> to return to EPED itself.

You should now see a blank screen with the function key options listed at the top, ready for typing or retrieving a document.

Printing Documents

You can print a document from EPED, even if the file has not been opened for word processing. To print properly the printer must be turned on, connected, and set so that the printhead is just below the perforation between two pages.

Check the printer, making sure that it is on and supplied with paper. Turn the select light off and move the paper with the linefeed button so that the printhead lies just below the perforation between two pages.

Press <F5> PRINT and move the cursor to "Name of File." Press <Enter> and then enter "MYFILE.TXT." Choose "Print File Now," and the document should be printed. If not, or if the document is not formatted correctly, ask the instructor to help you set the page length and margins, or install a printer setup (.PDF) file to initialize your printer. Setting up EPED for a particular printer is discussed on pages 261-262 of the Epi Info manual.

Practice and Further Assistance

If time allows, practice preparing letters or other documents in EPED. When you need further information, use the manual or help file to learn about additional features. An introduction to the use of EPED is contained in Chapter 6 of the Epi Info manual. Reference material is provided in Chapter 33.

At any time in EPED, the help file, containing all the information in Chapter 33, can be examined on the screen by pressing <F1>. To find the help section for a particular command, first find the appropriate menu and press the corresponding function key. When the menu is displayed, move the cursor bar to the desired command and press <F1>.

Display the instructions for the "Add File Other Window" command by pressing <F2> to display the file menu, moving the cursor bar to the "Add File Other Window" command, and pressing <F1>.

Now remove all menus using the <Esc> key and press <F1>. Choose an item from the menu of help topics and display the help text by pressing <Enter>.

Summary of Topics Covered

At this point you should be equipped for the following tasks:

- Creating a document
- Retrieving and editing a document
- Finding and replacing text
- Moving, copying, or deleting blocks of text
- Printing documents

With additional practice and use of the manual, you should become familiar enough with EPED to use it comfortably for most word processing tasks.

Chapter 5

Making a Questionnaire and Entering and Editing Data with Epi Info

Overview

In many branches of epidemiology, the computer has become an investigative tool as important as laboratory work. An epidemiologist needs to understand all the steps in transferring data from an affected population to questionnaires, to data files, and finally to lists, frequencies, cross tabulations, and reports.

Epi Info is a series of programs for IBM-compatible microcomputers that allows rapid construction of a questionnaire, automatic production of a data file from the questionnaire, entry of data, and processing to produce epidemiologic statistics. It has been used in offices, hotel rooms, laboratories, on cruise ships, and in rural Africa for rapid data entry and processing. Because it is a programmable system, it has been further enhanced for use in maintaining notifiable disease reports, and is used for that purpose in 35 US state and territorial health departments.

There are other statistics programs with a wider range of statistical operations. SAS is frequently used at CDC, and SPSS and BMDP elsewhere. Epi Info occupies less disk space than the larger systems and is readily available. It is used in this class as an introduction to microcomputer data processing. Many epidemiologists use it as their main processing system; others use Epi Info for data entry and preliminary processing and then convert the data files to another format such as SAS for more advanced statistics.

The objective of these exercises is to equip you to use Epi Info for outbreak investigation, surveys, or other questionnaire processing. In this chapter, you will:

1. Construct a questionnaire in the computer in Epi Info format
2. Make a data file from the questionnaire automatically
3. Enter data

Each chapter in the Epi Info manual contains a tutorial in a box at the beginning of the chapter. If you complete the exercises in the course manual quickly, choose a chapter in the manual that you have not covered and try the tutorial, or practice what you have learned by working with a data set of your own making.

Running Epi Info

Epi Info is located in a directory on the hard disk that is a subdirectory of the root called EPI5.

Turn on the printer, monitor, and computer and allow DOS to load (remember drive A must be empty or open if you have a hard disk). When you see the C:\> prompt, change to the Epi Info directory:

C:\>**CD \EPI5**

You should see a new prompt:

C:\EPI5>

Now type EPI to run the main menu of Epi Info. It should look like this:

```
Epi Info Version 5

Public Domain Software for Epidemiology and Disease Surveillance

Centers for Disease Control                World Health Organization
Epidemiology Program Office                Global Programme on AIDS
Atlanta, Georgia                              Geneva, Switzerland

Programs                                   Description

EPED      Text editor                      Use EPED to:
ENTER     Data entry                     2 Create questionnaires
ANALYSIS  Data analysis                  2 Write memos or letters
                                         2 Receive guidance in investigations
CHECK     Data validation                2 Write or edit Epi Info programs
STATCALC  Epi calculator
CONVERT   Export data                      EPED is a word processor for creating Epi Info
IMPORT    Import data                      questionnaires or for general text processing.
MERGE     Merge data files                 It contains the EPIAID system for thinking about
VALIDATE  Compare files                    epidemic investigations and producing drafts of
                                           investigation reports.
About Epi Info
Quit Epi Info                              Params: EPED <Textfile or EPIAID program name>

EPED
St , letter, or space to select            F9 for Setup, F10 to Quit
```

Making a Questionnaire

Making a questionnaire is very similar to typing a memo, except that word wrap should be turned **off**.

From the main menu, select EPED.

Bring up the SET menu with <F6> and turn word wrap OFF. Now enter the following questionnaire on the screen:

```
Centers for Disease Control
```

(To center a line, first type the line, then place the cursor on the line. Hold down the <Ctrl> key and then press O and C in succession--lightly and just once each. If you hold down the O or C rather than giving them a quick tap, the repeat function of the keyboard will send a succession of "Os" to the computer rather than the <Ctrl-O-C> that you want.)

```
EIS Course
Computer Skills Survey

SSN: Last 4 digits ####
Last name _________________ First name ____________
Today's date <mm/dd/yy>          Hour of class  ##


Previous computer experience (working knowledge)
     IBM-compatible microcomputer <Y>
     Macintosh                    <Y>
     Other microcomputer          <Y>
         (Please specify _________________)
     SAS/SPSS/BMDP                <Y>
     Mainframe/minicomputer       <Y>

Do you own your own computer?          <Y>

If so, what kind? ___________________________

Typing Skills
     1. None
     2. Intermediate
     3. Comfortable or advanced
     Typing #
```

Check each of the places where data will be entered against the field formats on the back of the Epi Info manual. Items like <mm/dd/yy> and <Y> are symbols for data entry FIELDS or VARIABLES and must be entered exactly as shown.

When everything is correct, press <F9> to save the file. Give it a name ending in ".QES" and beginning with your first name (up to eight letters), such as SUSAN.QES. When the file has been saved, press <F10> to leave EPED.

Making a Data File

Once you have created a questionnaire and saved it with a file name ending in .QES, the ENTER program will create a data file in a few seconds so that you can begin to enter data. The data or .REC file is made automatically by the ENTER program.

ENTER constructs a new description of the questionnaire and the data to be entered, using your .QES file as a template, and puts both in a new data file that is given the extension .REC. Although we will use a short questionnaire, the process is the same with questionnaires of up to 500 lines.

Choose ENTER from the main menu of Epi Info.

In response to "Data file:," enter the name of your questionnaire and data set, e.g., SUSAN.

In response to "Questionnaire file," enter the name again, e.g,. SUSAN. Now press <F4> to make a new .REC file from your .QES file. You should see your questionnaire on the screen. The data file has been made by Epi Info, and you are now ready to enter data.

Before microcomputers and rapid questionnaire-processing programs, the work you have done so far required weeks of negotiating with the data-processing department, conversations with a programmer, the writing of "specs," and finally of a data entry program that would allow entry of data by a keypuncher, but not usually by the epidemiologist. Now this phase can be completed in a matter of minutes.

If one or more blanks is not displayed properly, you may need to go back to EPED and correct errors in the format of the .QES file.

When you make a new data file after editing the .QES file, be sure to use a NEW AND DIFFERENT name for the .REC file, like SUSAN1. Otherwise, ENTER will bring back and display the first data file you made rather than making a new one. New files are made only if no data file exists by the name specified.

Entering Data

Respond to the questions on the screen and fill in the blanks. After the last entry, you will see the message "Write data to disk?"

Answer "Y", and you will be ready to enter the next questionnaire. Make four different questionnaire records, skipping some of the blanks if you wish.

Try the arrow keys and the <PgUp>, <PgDn>, <Home>, and <End> keys. Note the number of the current questionnaire in the lower right corner of the screen.

Editing Data

Be sure you have saved the last questionnaire, and then use <F7> (previous record) and <F8> (next record) to review your work. Watch the record numbers on the lower right as you do so.

Edit one of the entries by placing the cursor in any field and changing the contents. Then press the <End> key and "Y" to save the edited copy.

Hold down the <Ctrl> key and press <N> for "New record." This will clear the screen and bring up the next blank questionnaire record.

Finding Records with Particular Characteristics

The data entry blanks that appear on the screen are called **fields** or **variables**.

The ENTER program allows you to find records based on the value of any field. After you have located an existing record, perhaps by searching for the name, Jones, the information in the record can be changed and then saved in its new form.

There are several function keys that allow you to search for particular records. For information on these keys, hold down the <Ctrl> key. This changes the prompt at the bottom of the screen to display the commands for finding records.

To find a record, first enter the Find mode by pressing <Ctrl-F> as above. This will clear the entry blanks. Enter the search items in the proper fields, using items that you know will match at least one record. Then press <F1> to find the first instance of these search items. The first matching record should appear.

To see ALL the records that contain "Y" for a particular question, put the cursor on one of the Yes/No fields and enter "Y". Bring the cursor back to this field with the up arrow key. Now press <F1> to find the first record with a "Y" in this field, and <F4> to see the next record containing a "Y" at this location. Continue to press <F4> until the end of the file has been reached.

To find a particular record by number, press <F2> and enter the number of the record. Try it.

Printing the Questionnaire or the Records

The <F5> key in ENTER will print a copy of the questionnaire and, if desired, print the records in a compressed format suitable for proofreading.

If you do not have a printer, skip this step. Be sure that your printer is hooked up and ready to print. Press <F5>. Answer "Y" to "Print a copy of the questionnaire?" Give 1 and 2 as the first and last records to be printed, and "N" to "Single sheets?" The questionnaire and first two records should be printed.

Revising the Structure of a Data File

Now that you have entered several records, it is possible to add or delete questions in the questionnaire and still use the data you have entered. ENTER will automatically make a new .REC file and merge the old data into the new format.

Press <F10> to return to the main EPI menu. Choose EPED and bring back your questionnaire by pressing <F2> for the FILE menu and then choosing "Open File This Window." Give the file name for your questionnaire, e.g., SUSAN.QES.

Now add a question or revise or delete one or more questions. Press <F10> to leave EPED, and answer "Y" to save the changes.

Choose ENTER from the EPI menu. Enter the name of your data (.REC) file, but do not press <Enter> yet. Instead, press the <F6> key to revise the file and then read the instructions that appear on the

screen. Follow these instructions to revise the file. When the new file appears on the screen, verify that the changes occurred by paging back over the records with the <F7> key.

Summary of Topics Covered

This chapter is intended to provide you with the following skills:

- Designing and typing a questionnaire in the Epi Info .QES format
- Entering data
- Revising the questionnaire and, if necessary, transferring data already entered to the new format
- Finding records with particular characteristics or record numbers
- Printing the questionnaire and data

In order to become proficient, take every opportunity to use Epi Info for processing forms or questionnaires, even if the number of entries is small. You will be happy you did so when you are confronted with a large outbreak investigation where the pressure of time makes it more difficult to practice the rudiments of data processing.

Additional Help in Making Questionnaires

In EPED, you can make a questionnaire by simply typing it on the screen as we have done. The questionnaire function in EPIAID provides a longer tutorial and produces a preformed questionnaire suitable for a foodborne epidemic. If you have extra time, you may want to perform the following exercise.

Run EPED from the main menu. Press <F3> to see the EPIAID menu, and choose "Make Epi Info Questionnaire" by moving the cursor bar with the down arrow key and pressing <Enter> to make the choice.

Follow the instructions in the tutorial carefully and make a questionnaire. The items in parentheses in the questionnaire are meant to be edited and replaced with your own words. For example, "Breakfast (date and day of week)" might be changed to "Breakfast on Tuesday, May 15."

When you are asked for a file name, use the first eight letters of your last name followed by a period and .QES, as in ROCKWELL.QES

Chapter 6

The CHECK Program: Refining the Data-Entry Process

Overview

Often it is useful to have the computer check for errors during the data entry process, to do automatic coding of entries, and to skip over parts of the questionnaire if certain conditions are met. The CHECK program makes it possible to instruct ENTER to perform such operations automatically. By using CHECK and investing a little extra time in setting up conditions, you can protect your data against many common types of error and also make data entry easier and more automatic. This is particularly helpful if several people or those with minimal computer experience will be entering the data.

CHECK makes a file with a name ending in .CHK. The .CHK file contains instructions for ENTER to restrict the data entered in specified fields. When ENTER is run, it automatically looks for a file with the same name as the .REC file but ending in .CHK.

Using CHECK is optional, and the ENTER program will function just as well if you decide not to make a .CHK file.

Run the CHECK program from the EPI menu.

Enter the file name OSWEGO and respond that you are "Ready" by typing "Y" and pressing <Enter>.

You will see the OSWEGO questionnaire on a screen that looks like the ENTER screen but that has a different series of prompts at the bottom.

Enter "1" in the AGE field. Now press <F1> to make this the minimum age allowed in the field. Note the change in the line just above the function key prompts. Now enter "90" in this field and press <F2> to define the maximum of the range.

Move the cursor to the field labelled SEX. Enter "M" and press <F6> to make this a "legal value." Do the same for "F" and for "U" (Unknown). Press <F4> to make this a "Must Enter" field. The result is that only M, F, and U will be accepted in this field, and a missing value is not allowed, since we have decided to use U for this purpose.

Now let's install a conditional jump or skip pattern so that date and time of onset are skipped if there is no illness. Place the cursor in the ILL field, enter "N", and press <F7>. Following the instructions on the screen move the cursor to BAKED HAM and press <F7> again. The cursor will now jump automatically to BAKED HAM when N is entered for ILL. If Y or blank is entered, the cursor will proceed normally to Date of Onset.

Press <F10> and affirm that you wish to save the results. The EPI menu will appear.

Run the ENTER program with the file OSWEGO. Enter a record to verify that the range, legal values, and jump work properly. Do not save the results, but press <F10> and "N" to return to the menu.

Other features of CHECK that were not demonstrated include automatic coding, repeat fields, file indexing, and changing the names of variables. These are described in Chapters 10, 22, and 23.

CHECK contains a programming language, and advanced users can place specific program statements in the .CHK file that perform complex functions such as displaying a help window if subtracting one date from another produces a negative result. Many examples are contained in the RSURV surveillance system described in Chapter 26 of the manual.

Change directories to \RSURV and type RSURV to see the main menu of the surveillance system. Choose the first item and enter one or more records to see how the functions work.

Enter "10" for disease code and see that an entirely different questionnaire appears. This is a hepatitis form that has been linked to the main file by instructions in the .CHK file. In the COUNTY field, press <F9> to see the available codes and then type "O" for "Orange"; note that the cursor jumps to the first "O" in the list, which happens to be "Orange." The county and FIPS codes are provided automatically by codes set up in the .CHK file. Try to enter an INVEST. REPORT date that is before the INVEST. SENT date. The help window that pops up is another one of the functions that can be programmed in the .CHK file.

Examine the GEPI.CHK file that performs these functions in EPED. Although some of the programming is moderately complex, you should be able to alter some of the expressions to accomplish your own goals. If you do so, be sure to save a backup copy of the file first and restore it later so that other students will begin with the original file.

Summary of Topics Covered

With the CHECK program, you should be able to do the following:

- Set up range checking and legal codes
- Install skip patterns in your questionnaire with the "Jumps" command
- Understand automatic coding and know where to find further information
- Be aware of file linking, cross-checking between fields, help windows, and other advanced features of .CHK files

Chapter 7

Analyzing and Transforming Data

Overview

The program called ANALYSIS produces lists, frequencies, tables, statistics, and graphs from Epi Info or dBASE files. Simple commands will cause ANALYSIS to select records using specified criteria, sort or list records, do frequencies or cross tabulations, do logical or mathematical operations on a variable, put the results in a new variable, and direct the results to the screen, a printer, or a disk file. Many of these operations are illustrated in the tutorial called TUTOR1 that comes with the Epi Info system.

The commands form a mini programming language, and they may be entered one by one from the keyboard or placed in a program file that is then "run" from ANALYSIS. Commands for creating variables, selecting records, and recoding variables and the concept of programming are presented the the TUTOR2 tutorial.

The tutorials use a data file called OSWEGO.REC, which consists of information from a famous epidemic of staphylococcal food poisoning associated with ice cream in upstate New York that is often used in teaching epidemiology.

Sometimes it is useful to obtain statistics by entering summary data directly on the screen rather than into a data file. The STATCALC program does calculations from tables, provides sample size estimates, and does statistics for dose-response determinations (chi square for trend).

To use Epi Info data in other programs like dBASE, Lotus 1 2 3, SAS, or SPSS, Epi Info files must be converted into the formats needed by these programs. This is the function of the CONVERT program. CONVERT produces 12 different file formats from an Epi Info file.

ANALYSIS can use either Epi Info or dBASE files directly, but other files must be transformed into the Epi Info format. This is the function of the program called IMPORT.

Be sure that you are in the \EPI5 directory, using CD \EPI5 if necessary. Run EPI. From the main menu of Epi Info choose the ANALYSIS program.

When the prompt "EPI>" appears, type:

RUN TUTOR1

Enjoy the tutorial.

Now READ your own data file, and try LISTing the data. To see the variable names, type VARIABLES. Do:

FREQ *

Choose two variables for cross tabulation and do TABLES.

You already know enough to process a simple survey or case-control study, but ANALYSIS offers many more possibilities for selecting, altering, combining, and otherwise massaging the records in a file.

Do the second tutorial by giving the command:

RUN TUTOR2

If you need more information at any time (except during the tutorial), press <F1> to access the help file or type the name of a command and then press <F1> to see the section for that command. For more advanced work do the exercises in Chapter 12.

Experiment with various commands using your own data, OSWEGO.REC, or the file called EPIDEMIC.REC described in Chapter 12.

Doing Calculations on the Screen in STATCALC

STATCALC offers three kinds of statistical resources. The tables option allows you to enter numbers on the screen and obtain quick results with the same kind of statistics obtained in ANALYSIS with the TABLES command. It offers an additional kind of statistic, called "exact confidence limits," that can be used to check the Cornfield confidence limits produced by ANALYSIS or to replace them for final publication, since they are more accurate when the odds ratio is very different from 1.0.

The sample size calculations offer a method of estimating the theoretical sample sizes needed for a study of proportions. Although these are only a guide, they can be helpful as one factor in planning studies.

The Mantel-Haenszel test for trend is used for measuring dose-response and other relationships involving response to an ordered series of exposures.

Run STATCALC from the main menu. Choose "Tables" from the STATCALC menu.

Enter four numbers, with an <Enter> after each. Then press <Enter> again.

After looking at the results, press <Enter>. Now press "E" for exact confidence limits. These are more accurate than the Cornfield confidence limits. Press <Enter> to begin another table.

Enter six or more numbers in the new table before pressing an extra <Enter>.

Enter a new 2-by-2 table, and after the statistics appear, press <F2> for another "stratum" or table. Enter four more numbers, an extra <Enter>, and after the statistics appear, a final <Enter>. The stratified analysis will appear.

Press <Enter> again and then "E" for exact confidence limits on the odds ratio for the stratified analysis.

Press <F10> to return to the STATCALC menu.

Choose Sample Size and experiment with the calculations offered. Try to design a study and guess how many subjects will be required, then check the results with STATCALC.

If you have time and the interest in statistics, experiment with the Mantel-Haenszel test for trend. A numerical example is given in Chapter 15 of the Epi Info manual and in the STATCALC help file.

Explore and read the help files by pressing <F1> from various points on the menus.

Converting Files to Other Formats for Export

Epi Info will make files in 12 different formats for use in other systems. We will make a SAS file as an illustration.

Run CONVERT from the menu.

Enter OSWEGO as the first file name.

Enter the number 2 for "Choice" to make a SAS file.

Press <Enter> to accept the new file name.

Answer "Y" to "Ready?" and press <Enter>.

Examine the resulting file by choosing EPED from the menu and opening the file OSWEGO.SAS. Compare it with the original Epi Info file, OSWEGO.REC.

It is common practice to "upload" such files to the mainframe at CDC and perform further analysis in the SAS system.

Advanced Work

Chapters 24, 25, and 26 describe the REPORT command, the handling of several related files at one time, and the notifiable disease surveillance system that is distributed as a demonstration with Epi Info. If you have mastered the basic functions of Epi Info, you may wish to explore these chapters and the associated tutorials.

You may also wish to examine the EPIAID system described in Chapters 20 and 21 and experiment with initiating an investigation or designing a study with EPIAID's help.

Summary of Analytic Topics Covered

This chapter has been brief, but the interactive tutorials cover considerable material. You should now be prepared for the following tasks in analyzing data:

- READing a file
- Producing lists, frequencies, cross tabulations, means, and graphs
- Selecting particular subsets of records
- Writing and running an ANALYSIS program
- Recoding variables

- Using IF statements to carry out operations IF certain conditions are true
- Sending results to a printer or to a file
- Using STATCALC to do statistics for tables or calculate sample sizes
- Being aware of the uses for chi square for trend in STATCALC
- CONVERTing files to other formats for export
- IMPORTing files from other programs

With additional practice, you should be able to use ANALYSIS to produce results for most epidemiologic studies. Although it does not contain advanced statistical functions like logistic regression, most experts recommend that you first examine the table statistics, such as the Mantel-Haenszel stratified analyses, that are easily done in ANALYSIS. Once these have been studied and understood, the data files can be prepared for use in other programs that offer more complex statistics. If the conclusions are not similar, however, serious thought is in order before relying exclusively on "advanced" analyses.

Appendix

Notes for Instructors (Prepared for WHO Use)

1. Title: A Course in Microcomputer Use for Epidemiologists, Using Version 5 of Epi Info.

2. Description: This is a short course for professionals in public health or other fields in the use of IBM-compatible microcomputers, the DOS operating system, and Epi Info, a software system for word processing, questionnaire tabulation, data base management and statistical analysis.

3. Objectives: Participants will be able to turn on and operate an IBM-compatible microcomputer, find and copy files, change directories, and perform other simple DOS operations. They will be able to type a letter or memorandum in the Epi Info word processor, make a questionnaire, enter data in the questionnaire, and search for and edit records. They will be able to produce lists, frequencies, and cross tabulations from the data entered and to do basic data "cleaning" operations such as selecting records and recoding values. They will be able to write simple programs for analysis, perform statistical calculations on the screen, and export files from Epi Info.

4. Basic points: IBM-compatible computers are widely available in public health facilities. Epi Info is a series of programs for performing common public health tasks on such computers, developed jointly by the Centers for Disease Control (CDC) in the United States and the Global Programme on AIDS, World Health Organization (WHO), Geneva. It may be freely copied and distributed. Spanish and French translations are planned.

The course covers the basics of operating a computer with the DOS operating system for persons who have not previously operated an IBM-compatible microcomputer, and then offers tutorial instruction in the operation of Epi Info. The course is designed to be carried out on a computer equipped with Epi Info. It uses the Epi Info manual and the tutorial files that come with Epi Info as the core material for the course.

At CDC, the course is carried out in a computer laboratory with 10 to 15 microcomputers and one or two students per computer, but the course manual might also be used for self-instruction. It is helpful to have instructors experienced in the use of DOS and Epi Info available, but the course is detailed step-by-step in the course manual so that the instructor need provide only backup assistance. Some instructors may wish to introduce and amplify points or give demonstrations. It is highly recommended that the instructor provide additional exercises from a particular field of study at the end of the course to solidify the points covered.

5. Target group: This course was originally designed for Epidemic Intelligence Service Officers being trained at CDC, most of whom are physicians or have doctorates in another field. Previous experience with computers varies from none to extensive. The class is divided into three groups of about 20 each on the basis of a questionnaire asking about previous experience with the DOS operating system, roughly corresponding with beginning, intermediate, and advanced computer users. The "beginning" group is not expected to progress as far through the manual as the others, and some parts of the course may be skipped if time is short. Students with previous DOS experience may skim rapidly through Chapters 1 and 2 and begin work in Chapter 3, or even skip the first two chapters entirely.

6. Timing: The time allotted for the CDC course is about 4 hours for DOS and Word Processing and another 4 hours for Epi Info. An additional 4 hours for more practice using Epi Info is recommended. Practical exercises in data processing should be provided for the students to carry out with Epi Info, since the tutorial approach in the class manual needs further practice to consolidate and apply what the students have seen in the tutorials.

Sessions of more than 4 hours in a single day should be avoided, particularly for students who have no experience with typing, and it may be better to limit the sessions to 2 hours for some groups. Students without previous computer experience, or who are taking the course in a language other than their own, may need additional time during the course or free access to computers for extra practice.

7. Facilities and equipment: Each student should have access to an IBM-compatible microcomputer with 640 K of random access memory. Except in unusual circumstances, the computers should have a hard disk and a color monitor. Epi Info should be previously installed on each computer, although instructors may prefer to have the

students install Epi Info so that they will see how this is accomplished. A large monitor or projector connected with the instructor's computer for demonstrating selected points is helpful but not essential.

In Chapter 2, several exercises require materials such as common types of floppy diskettes with and without the jackets removed, keyboards, computer chips, and at least one computer with the top removed so that the various parts can be pointed out.

8. Materials: Each student should have a copy of the course manual, the Epi Info manual, and the Epi Info programs. These materials may be copied locally or obtained as follows:

Epi Info manual and programs:	USD, Inc. 2075 A West Park Place Stone Mountain, Georgia 30087, U.S.A. Tel (404) 469-4098 Fax (404) 469-0681 The cost is US$35 in North America or US$45, including air shipment, elsewhere.

Quantity discounts are usually available for classes, and the manual is available separately for US$25.

Single copies of this manual, "A Course in Microcomputer Use for Epidemiologists, Using Version 5 of Epi Info," are available from the author, and may be freely reproduced.

9. Procedure: The students are divided into beginning, intermediate, and advanced groups on the basis of previous experience with DOS. They should be encouraged to read Chapters 1 and 2 before coming to the first class, particularly if they are in the beginning group.

The instructor may choose to give a brief orientation to the class and an introduction to computing. In small classes, it is helpful to have each student give a few details about his or her computing background, expectations for the class, and intended applications for Epi Info.

Attention is then directed to Chapter 2 of the course manual. The instructor explains that the class should follow the directions in bold type between the sets of double lines. The class then follows the exercises, e.g., turning on the computer, viewing the various kinds of diskettes, and exploring the keyboard. If possible, the cover should be removed from one computer and the instructor should point out the motherboard, CPU, RAM chips, bus, disk controller and drives, and video board. A discussion of power surges, 110-volt vs. 230-volt supplies, and travel difficulties may follow.

Chapter 3 provides an introduction to the most important DOS commands: DIR, TYPE, FORMAT, COPY, RENAME, ERASE (DEL), and CD. The concepts of file names, paths, directories, the logged drive and directory, backing up files, AUTOEXEC.BAT, CONFIG.SYS, and rebooting the system are introduced. The instructor may choose to omit any of the topics if desired. If a better backup program than BACKUP is available, for example, this could be introduced, or students could be advised to use the COPY command for backup purposes. The DOS Review Exercise near the end of the chapter may be used for review for students who start the course with some knowledg of DOS, or as a test for those who have progressed to this point. It offers an alternative method of backing up hard disks, that of backing up most files immediately onto floppy disks without using a special backup program. This is particularly useful for students who will work mostly with word processing and short data files. If this approach is chosen, it is probably best to skip the section on using DOS BACKUP earlier in the chapter.

Chapter 4 teaches basic word processing through use of EPED. If the students are already familiar with word processing, this lesson can be abbreviated or modified. The tutorials in EPED are most useful for students fluent in the same language. In teaching a Spanish-speaking class to use the English version of Epi Info, for example, we chose to write instructions on the blackboard in Spanish rather than use the tutorials, since the tutorials contain many explanatory screens in English.

Chapter 5 teaches the construction of a simple questionnaire with a variety of different types of fields. For beginning students or those working in other than their native language, simply typing the questionnaire on the EPED screen and saving the file as detailed in the course manual is the easiest way to understand the concepts. If the students are more advanced computer users or have used Epi Info before, the tutorial mentioned at the end of the chapter is more challenging. It teaches more advanced concepts of questionnaire design and allows the students to construct a questionnaire suitable for an enteric disease outbreak interactively.

No matter how the text of the questionnaire is created, the resulting questionnaire is accessed through the ENTER program that automatically creates a data file and presents the questionnaire on the screen ready for data entry. One of the high points of the class occurs as the students see the completed questionnaire, ready for data entry, a few seconds after they have finished typing the .QES "template."

The course manual provides instruction for data entry, finding and editing records, and revising the structure of the questionnaire. The latter is almost always requested by the students to correct the inevitable typing mistakes that occur in creating the questionnaire.

Chapter 6 introduces the CHECK program for adding range checking, skip patterns, automatic coding, legal values and other constraints to the data entry process. It may be included in the course or skipped if time is short.

In Chapter 7, analysis is taught through two tutorials, TUTOR1 and TUTOR2. Usually the class is silent for a considerable period as the students work their way through the various parts of the ANALYSIS program. Additional concepts, particularly the use of the SAVE command and the construction of programs, may be stressed by the instructor.

STATCALC provides statistical resources for 2-by-2 to 2-by-9 tables, sample size calculation, and chi square for trend. The exercise provides a brief introduction, but additional exercises are contained in the help files for STATCALC, accessed through the <F1> key, if desired.

The course manual ends with an introduction to the CONVERT program. CONVERT allows Epi Info files to be exported in 12 different formats for use in other systems. It is important to point out that tutorials for other programs in Epi Info are found at the beginning of each chapter in the Epi Info manual and that the IMPORT, MERGE, and VERIFY programs can be explored when needed through these tutorials.

10. Evaluation and Practical Exercise: As mentioned above, it is important to follow the tutorials with an exercise that solidifies the points covered in less passive form, since the tutorials provide only an overview for most students. At CDC, the current exercise is based on two case-control studies of eosinophilia-myalgia syndrome (EMS). The students design and construct a questionnaire, enter data, and analyze one case-control study. They are given data files for both studies for more advanced analysis, including matched analysis. Working in groups, they prepare 5-minute presentations on their

findings at the end of a 3- or 4-hour period, presumably for the "Commissioner of Health" or for a "Press Conference."

The subject matter of the exercise should be as close as possible to the field in which the students will use their computer skills. If disease surveillance is the subject, the exercise could be modification or construction of a disease surveillance system, starting from the RSURV or SURV examples on the Epi Info disks. Environmental health students could be asked to devise a system for processing restaurant inspections or toxic substance investigations. Occupational health students might process health hazard evaluations. The epidemic investigation in Chapter 12 of the Epi Info manual could form the basis for an exercise in an epidemiology class; basic questionnaire construction and data entry should be added, however, to solidify basic Epi Info skills.

The results of the practical exercise not only provide an important part of the learning experience, but they also give the instructor an indication of how much the class has learned and what skills remain to be acquired. All of the classes should be followed by a standard anonymous evaluation by the students of the teaching methods, the instruction, and the materials of the class, using a simple questionnaire. The results may be easily tabulated by the instructor using Epi Info.

THE GLOBAL ENVIRONMENTAL EPIDEMIOLOGY NETWORK

G E E N E T

GEENET

The Network was established in 1987 as a means for the World Health Organization to strengthen education, training and research in institutions involved in epidemiological teaching and research on the health effects of environmental hazards, and other epidemiological applications in environmental and occupational health.

The Network aims at improved communication and collaboration between institutions in this field in developed and developing countries. A series of documents with information of value for training and research development is prepared for the Network and lists of Network members are distributed on a regular basis. Training and research promotion workshops are organized in collaboration with national and international agencies.

For further information write to:

Environmental Epidemiology
Division of Environmental Health
World Health Organization
1211 Geneva 27
Switzerland

Telephone: (022) 791 37 60
Telefax: (022) 791 07 46

WHO