Wendy Gancher

The Computer: An Effective Research Assistant

SUMMARY

The development of software packages such as data management systems and statistical packages has made it possible to process large amounts of research data. Data management systems make the organization and manipulation of such data easier. Floppy disks ease the problem of storing and retrieving records. Patient information can be kept confidential by limiting access to computer passwords linked with research files, or by using floppy disks. These attributes make the microcomputer essential to modern primary care research. (Can Fam Physician 1984; 30:885-887).

SOMMAIRE

Le développement de progiciels et de logiciels tels les systèmes de traitement des données et d'analyses statistiques ont rendu possible le traitement de grandes quantités de données de recherche. Les systèmes de traitement des données facilitent l'organisation et la manipulation de telles données. Le disque mou simplifie le problème d'emmagasinage et d'extraction des dossiers. La confidentialité des informations concernant les patients est préservée en limitant l'accès aux mots de passe de l'ordinateur pour les dossiers de recherche ou en utilisant des disques mous. Ces attributs font du micro-ordinateur un outil essentiel à la recherche moderne en soins de première ligne.

Ms. Gancher recently completed graduate studies in the Department of Community Health and Epidemiology, and is a research assistant in the Family Medicine Centre, at Queen's University, Kingston, ON. Reprint requests to: Dr. J. E. Anderson, Hotel Dieu, Family Medicine Centre, Queen's University, 220 Bagot Street, P.O. Bag 8888, Kingston, ON. K7L 5E9.

A RESEARCH PROJECT has four major components. The first stage involves basic planning and the initiation of organizational ideas. At this stage, the investigator outlines how the project will be conducted, in an effort to resolve the problem or fulfill the purpose that initiated the project. Once this is completed, the investigator proceeds with the actual execution of the plan. It is at this stage that data are collected and organized into a set of useful information. The third component is analyzing this information to evaluate its contribution in answering the initial questions. The final stage involves presenting the results of the investigation and conclusions based on the interpretation of these results. Each stage is vital, and the organization within each stage ensures a smooth and logical progression.

Initially, most health care data exist in the cumbersome form of patient records, survey forms, patient statistics or test results. Most of these items are best described as piles of disorganized information sheets. The storage of such data is a problem, but organizing data into a form useful for research is an even greater task. The inability to organize and analyze larger data sets can be an impediment to research.

Recently, the microcomputer has made the organization and storage of research data much simpler. Organizational packages have been developed for microcomputers that make manipulating large sets of data easier. One type of package, generally referred to as database management systems, facilitates data entry, assists in organizing data under parameters of the investigator's choice, manipulates data on the basis of the information to be obtained and restructures data into meaningful output. These capabilities make the system invaluable to the researcher.

The purpose of this article is to briefly outline the use of a database management system and the employment of this system in the construction of our example—an age-sex register for a primary care practice. The particular package used is an actual system called DBase II (trademark of Ashton-Tate). Although this system is unique, it is representative of the other database management systems available for most microcomputers.[1]

Setting Up Data Files

We will look at a hypothetical data set created to establish the age-sex structure of a primary care practice. In order that data collection may be carried out efficiently, a decision should be made before collection begins about the type of data to be collected. In this way, a skeletal format for data entry can be set up on the computer. This ensures that all pertinent data are col-

lected and organized in the most useful manner.

Informational items that are directly related are organized into a file, much as they would be if they were organized manually. The structure of the database must be established in order to enter any information into this file. For example, suppose we wished to construct the age-sex structure of a particular primary care practice. Our file would consist of all patients in that practice and all pertinent information about them. For the purpose of our example, the relevant data items have been identified as the patients' names, ages, sex, and date of last visit to the physician. These items are referred to as 'fields'. Each patient and the accompanying information are a 'record'. Each record is composed of the fields established in the structure.

In our example, we will initially use ten records. This will help illustrate some of the major advantages of this system. However, this system cannot be fully appreciated until it is applied to a larger data set. Each field in a record is identified by a name, which is usually related to the information that is to be entered. The type of information, whether it is numbers or characters, and the maximum space required to record it, are also entered. The investigator can then sit at a terminal and create a structure in five minutes or less. With the help of Dbase the age:sex file is created (see Figure 1).

Fig. 1. Replication of the screen image of the file structure.

```
• CREATE AGE:SEX
ENTER RECORD STRUCTURE AS FOLLOWS:
  FIELD     NAME,TYPE,WIDTH,DECIMAL PLACES
  001       NAME,C,30
  002       AGE,N,2
  003       SEX,C,1
  004       LAT:VIST,N,6
  005
INPUT DATA NOW? N
```

Fig. 2. Replication of the screen image of a blank record.

```
RECORD # 00001
NAME        :                                    :
AGE         :    :
SEX         : :
LAT:VIST    :         :
```

Fig. 3. Replication of the screen image of a record creation.

```
RECORD # 00001
NAME        :CAMPBELL, JENNIFER                  :
AGE         :42:
SEX         :F:
LAT:VIST    :         :
```

Fig. 4. Replication of the screen image of editing a record.

```
• FIND LAWRENCE, CATHY
• DISPLAY
00002  LAWRENCE, CATHY                  13 F  820404

• FIND LAWRENCE, CATHY
• DISPLAY
00002  LAWRENCE, CATHY                  13 F  840112
```

Fig. 5. Replication of the sceen image of record addition.

```
• LIST
00007  ADAMS, WILLIAM                    5 M  821013
00001  CAMPBELL, JENNIFER                42 F  820502
00010  DAWSON, MARK                      2 M  831129
00011  FISHER, MARTHA                   24 F  840114
00005  GROOPER, KEN                      4 M  830109
00002  LAWRENCE, CATHY                  13 F  820404
00004  LEATTON, MAXWELL                 34 M  820208
00009  MATTHEWS, MOIRA                  37 F  830827
00006  MCNAB, CECILIA                   54 F  831111
00008  PRESTON, ALICE                   16 F  830923
00003  TREMBLAY, BRIAN                  51 M  830306
```

Entering and Editing Data

Once the structure of each record is established, the investigator or an assistant can easily enter the data, which in this case is the patient information. If the assistant does the data entry, the investigator is relieved of a tedious job (see Figures 2 and 3).

Any information within the file can be changed with little effort. Suppose we wish to update one patient's last visit. First, we must locate the patient record and then make the appropriate changes (see Figure 4). Records can be entered all at once or daily as new patients join the practice. If a patient is added to this indexed file, the new record is automatically inserted into its appropriate alphabetical position (see Figure 5). Records can be removed as simply as they are added. Once the file is in its final form the data can be processed.

Using the Data

Once all the data have been entered, they can be organized around any factor common to each patient record. For example, if the filing order for patients is normally alphabetical, the computer can index the file on the basis of the patient's surname (see Figure 6).

If the examination of the data requires no more than a tally of certain data items, then the database management system can be used. In our example, we must know how many individuals fall into the separate age-sex categories. For example, we must tally the number of males who fall in the zero to four year age group (see Figure 7). Similarly, all the age-sex groups can be tallied to complete the structure. The capabilities of this data manipulation system are wide-ranging, but its organizational capacity is its selling point.

Separate packages are available for the microcomputer if the data are to be analyzed statistically. Usually, this is a simple matter of converting the data into the statistical format and choosing the test to be performed. The test will display all conventional test results for your interpretation. A common feature of these packages is their ability to format data into tables, which can be used for presentation.

Storage of Data

Besides the organizational capabilities offered by the microcomputer in conjunction with its data packages, the computer provides ample storage space for data files. Most microcomputers have two distinct, possible storage methods.

The first storage area is within the main memory or hard disk of the computer. Discrete use of passwords can limit accessibility of files to protect confidential patient data. Unfortunately, most microcomputers have limited hard disk space. There are two ways to handle this problem. The first is to increase the main memory by replacing the smaller hard disk drive with one having a greater storage capacity, or by adding a second disk of the same size. This can be expensive.

The second option is to expand the capacity of the microcomputer to include a soft disk drive. The soft or floppy disk is a temporary disk that can be used as an alternative for data storage when inserted into a disk drive and formatted. The floppy disk is a separate entity and provides additional security for patient records. As well, any information stored on disk can be relocated in an area away from the main computer. If, by some chance, the main computer was damaged, the information on the floppy disk would act as a backup and could be restored on a similar machine. If the data being collected are not to be used immediately, floppy disk storage can be used to 'archive' data, leaving the hard disk clear for active files.

One further advantage of the floppy disk is its portability. Data can be entered at one site and sent anywhere in the world. In most cases, the computer receiving the data must be compatible with the system used to enter the data, but this is changing. Now there are conversion programs, which convert data recorded on one system so that they may be interpreted by a totally unrelated system. This aspect of portability makes the floppy disk essential to coordinating data collection in multicentre projects.

Fig. 6. Replication of the screen image of listing and ordering records.

```
• LIST
00001  CAMPBELL, JENNIFER          42  F  820502
00002  LAWRENCE, CATHY             13  F  820404
00003  TREMBLAY, BRIAN             51  M  830306
00004  LEATTON, MAXWELL            34  M  820208
00005  GROOPER, KEN                 4  M  830109
00006  MCNAB, CECILIA              54  F  831111
00007  ADAMS, WILLIAM               5  M  821013
00008  PRESTON, ALICE              16  F  830923
00009  MATTHEWS, MOIRA             37  F  830827
00010  DAWSON, MARK                 2  M  831129

• A_INDEX ON NAME TO ORDER
00010 RECORDS INDEXED
• USE REGISTER INDEX ORDER
• LIST
00007  ADAMS, WILLIAM               5  M  821013
00001  CAMPBELL, JENNIFER          42  F  820502
00010  DAWSON, MARK                 2  M  831129
00005  GROOPER, KEN                 4  M  830109
00002  LAWRENCE, CATHY             13  F  820404
00004  LEATTON, MAXWELL            34  M  820208
00009  MATTHEWS, MOIRA             37  F  830827
00006  MCNAB, CECILIA              54  F  831111
00008  PRESTON, ALICE              16  F  830923
00003  TREMBLAY, BRIAN             51  M  830306
```

Fig. 7. Replication of the screen image of selective tallying of records.

```
• DISPLAY STRUCTURE
STRUCTURE FOR FILE:        REGISTER.DBF
NUMBER OF RECORDS:         00011
DATE OF LAST UPDATE:       00/00/00
PRIMARY USE DATABASE
FLD      NAME        TYPE      WIDTH      DEC
001      NAME        C           030
002      AGE         N           002
003      SEX         C           001
004      LAT:VIST    N           006
** TOTAL **                     00040
• COUNT FOR SEX = "M" •AND• AGE < 04
COUNT = 00001
• COUNT FOR SEX = "F" •AND• AGE > 25 •AND• AGE < 44
COUNT = 00002
```

Conclusion

The advent of the microcomputer and the continual development of software packages have made research involving large data sets possible. The time and effort saved by using the microcomputer may be directly translated into the funds required for the initial set up of such a system.

It is important to keep abreast of changes in the area of data management and analysis, because they are occurring quite quickly. It may be wise to seek expert assistance initially, in order to determine which software packages would be most suitable for the type of research data you will be handling and for your methods of analysis. ◉

Reference

1. Byers RA: Everyman's Database Primer Featuring DBase II. Culver City, CA., Ashton-Tate, 1982.