

627



# A Microcomputer-Based Approach to Data Management: An Essential Tool in the Assessment of Trends and Results in Liver Transplantation

ROBERT D. GORDON, M.D. and THOMAS E. STARZL, M.D., Ph.D.

The recent National Institutes of Health Consensus Conference on Liver Transplantation concluded that orthotopic hepatic transplantation had passed beyond the stage of an experimental procedure, thereby opening the door to considerable expansion of the number of transplant programs in the United States. The Conference nevertheless emphasized the need for careful evaluation of future results, in order optimally to define the indication, prognostic factors, and other important variables for this complex and expensive procedure.

In 1985, microcomputer technology provides a highly effective means of storing and examining the results of a transplant program in a way that meets this stipulation of the Consensus Conference, without requiring major computer programming expertise on the part of the users. Since 1983 a microcomputer-based data registry has been maintained by the University of Pittsburgh Liver Transplant Program. This registry has permitted us to monitor continuously the activities of the liver transplant program and generate accurate, detailed life-table survival information on many groups of patients. The system has also been used to produce printed or projected presentation graphics suitable for publication in journals or books or for use during lectures. Much of the material presented in this issue of *Seminars* was produced using the system to be described in this article.

Because a microcomputer-based data registry similar to the one developed in Pittsburgh would seem to be a requirement for any future transplant

program, as well as having obvious applications for utilization of any patient data set, the University of Pittsburgh system will be described in some detail.

## SYSTEM DESCRIPTION

The hardware used to maintain the liver transplant registry is centered around a COMPAQ portable microcomputer (COMPAQ Computer Corporation, Houston, TX), an IBM-PC compatible machine based on the Intel 8088 microprocessor, which is a 16-bit CPU with a 20-bit program counter able to address one megabyte of RAM, and has an 8-bit data bus and a high resolution monochrome screen capable of displaying graphics. The system is equipped with 640 kilobytes of memory, the maximum amount of user-addressable memory permitted by the current version of the operating system (IBM DOS 2.10). The main unit houses two 360K 5-inch double-sided, double-density IBM format disk drives. An internal 1200 baud modem (Hayes 1200B, Hayes Manufacturing, Norcross, GA) is used for communication with main-frame systems and time-sharing services. In addition, the unit is connected to a high-capacity dual 8-inch 10 megabyte cartridge floppy disk drive unit (IOMEGA Corporation, Roy, UT). Although this storage device uses removable floppy disk cartridges, its performance characteristics are similar to a hard disk drive.

An Epson LQ1500 24-pin head dot matrix printer (Epson America, Torrance, CA) provides near letter quality hard copy and a Hewlett-Packard 7475A 6-pen plotter produces publication quality graphics. In addition, a dedicated color graphics display computer, the VideoShow 150 (General Parametrics Corporation, Berkeley, CA) is used

*From the Department of Surgery, University Health Center of Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania.*

Reprint requests: Dr. Gordon, Dept. of Surgery, 3601 Fifth Ave., Room 103, Falk Clinic, Pittsburgh, PA 15213.

with the microcomputer to produce presentation quality color slides.

## SOFTWARE CONCEPTS

### Data Base Management

A data base contains a file or group of files and is structured to allow easy access to and modification of specific items of data and production of reports and listings of any desired portion of the data.<sup>1</sup> The smallest unit of data maintained in a data base is called a "field" (also referred to as an item or attribute). For example, a patient's last name, first name, diagnosis, and data of surgery each constitute one field in a data base record. A "record" is a set of related fields that are stored as a unit. Thus, the set of fields representing the data for a particular patient constitutes the "data set" record for that patient.

A "data set" is a collection of records having the same fields. The structure of all the records in the data set is uniform; the order, type, and length of the fields in each record are the same. The content of each field in a record, however, may be different for each individual record. A "file" is a collection of data stored by the computer with a name for referencing it. One or more data sets may be stored in a file. A "data base" is one or more data sets on one or more files. It includes all the data stored in data sets that relate in any way to the activities of the user. A data base has different records for different parts (data sets) of its structure.

## SPREADSHEET

A spreadsheet, or visible calculator, is an electronic grid of intersecting rows and columns. Each intersection of row and column defines a "cell." A cell may contain a label (text), a number, a date, a time, or a formula. Formulas may be based on a wide range of mathematical or statistical functions and the results of the formula contained in one cell may derive from information contained in one or more other cells. If the data in the cells from which the formula is derived change, so does the data in the cell defined by the formula.

## SOFTWARE DESCRIPTION

The liver transplant data base in use at the University of Pittsburgh uses a relational data base manager, an integrated spreadsheet program, and several graphics packages to create a flexible, comprehensive reporting system. The system can send data to the University mainframe computer for further processing by advanced statistical programs, such as Statistical Package for the Social Sciences (SPSS), when necessary. The relationship between the program systems used for the registry is shown in Figure 1.

Dataease (Software Solutions, Milford, CT) is a menu-driven relational data base management program that can be highly customized by the user without the need to use programming code ("command language"). It has a powerful, but easy to use query language and versatile formatting

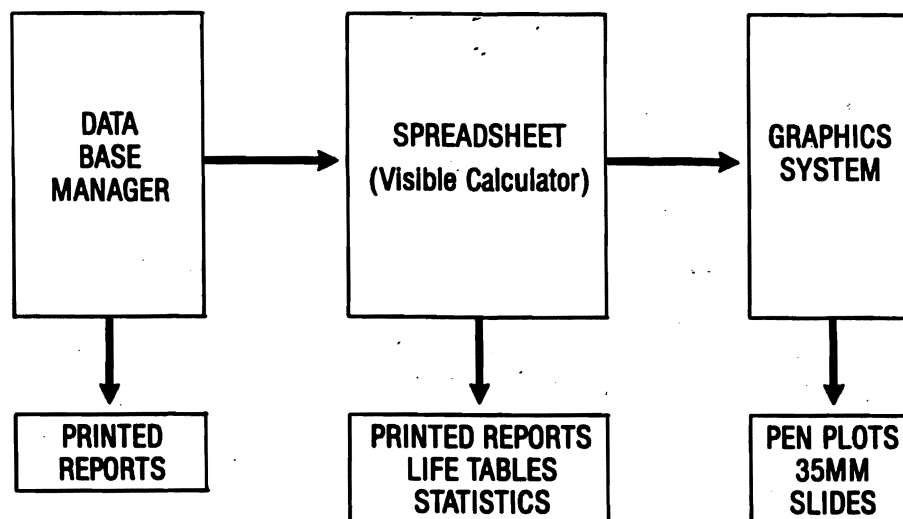


FIG. 1. Software integration. A relational data base manager is the foundation of the information management system. Data from the files created by this program can be transferred to a spreadsheet program for mathematical analysis, including generation of life-tables. The results of spreadsheet analysis can be exported to graphics software for creation of presentation quality plots or slides.

system for report generation. Dataease can export data directly to other programs, including spreadsheets or mainframe systems. It provides the foundation of the data management system.

Dataease organizes the data base into forms. Each data set is defined by a form and relationships between data sets can be defined by identifying to the system the fields that link different forms. A form "Itx recipients" contains the fields for each new patient admitted to the registry. A second form "Itx retransplants" contains the records for any patients requiring one or more retransplants. Each patient entered into the system is assigned an orthotopic transplant (OT) number. The OT number is used to link the "Itx recipient" data set to the "Itx retransplant" data set. The fields in the two data sets are described in Table 1.

Survival statistics in the form of life-tables can be generated from the data base registry using spreadsheet software. The report generator of Dataease is used to produce a Data Interchange Format (DIF) file that can be imported by the spreadsheet

program. Table 2 illustrates the query used in Dataease to generate the DIF file.

Symphony (Lotus Development Corporation, Cambridge, MA) is an integrated spreadsheet program that has an extensive set of query, mathematical, and statistical functions. It also can be programmed to perform a complex set of calculations automatically using a feature called a "macro." A macro is a learned sequence of keystrokes that permits a program to execute a defined sequence of program commands whenever desired by the user. The DIF file generated from Dataease is translated into a Symphony spreadsheet and a set of macros has been developed that permits rapid, automatic generation of life-tables on any group of patients in the registry.

Figure 2 illustrates the organization of the spreadsheet. The entire Symphony worksheet contains thousands of cells, but only a modest portion of the available worksheet is actually utilized. The spreadsheet has been organized into three principal reservations. The first is the *data base reservation*

TABLE 1. Data Set Fields\*

Field Name	Type	Description
<b>FORM: LTX RECIPIENTS</b>		
OT#	Numeric string	Primary patient sequence number
PATIENT	Text	Last name, first name
AGE	Numeric	Age in years
RACE	Text	White, black, Oriental, etc.
SEX	Choice	M(ale) or F(emale)
ABO	Choice	Recipient ABO blood group
DX	Choice	Predefined diagnosis codes
SUBCLASS	Text	Diagnostic subclassification
TX#	Numeric	Number of transplants
TX DATE	Date	Date of primary transplant
DONOR	Text	Primary transplant organ donor
DONOR ABO	Choice	Primary donor ABO blood group
STATUS	Choice	Alive or dead
SURV DATE	Date	Date to which patient survived
DAYS SURV	Calculated	Survival date - Transplant date
<b>FORM: LTX RETRANSPLANTS</b>		
OT#	Numeric string	Primary patient sequence number
PATIENT	Text	Last name, first name
TRANSPLANT #	Number	2 = second transplant, 3 = third, etc.
DONOR	Text	Name of donor
DONOR ABO	Choice	Donor ABO blood group
RETX DATE	Date	Date of retransplantation
RETX DX	Choice	Indication for retransplantation (rejection, technical failure, primary nonfunction)

\*In Dataease, each data is defined by a form. In the liver transplant data base, there are two forms, LTX RECIPIENTS and LTX RETRANSPLANTS. Every patient has a record in the form LTX RECIPIENTS and is assigned a sequence identification number (OT#) by the computer. There is only one record per patient in LTX RECIPIENTS. If a patient is retransplanted, a record is entered in LTX RETRANSPLANTS. The total number of transplants is updated in the TX# field in LTX RECIPIENTS. A patient may have none, one, or several records in LTX RETRANSPLANTS depending on the number of times a particular patient is transplanted. All the records in each data set for a given patient have the same OT#. This field is used to establish the relationship among records in the data sets for each patient.

**TABLE 2. Data Base Query to Produce Data Interchange Format File\*****for LTX RECIPIENTS**

with TX DATE between data-entry start date to data-entry end date ;

list records

OT# in order ;

PATIENT ;

AGE ;

RACE ;

SEX ;

TX# ;

TX DATE ;

SURV DATE ;

DX ;

SUBCLASS ;

any RELTX named "second" with ( TRANSPLANT # = 2 ) RETX DATE ;

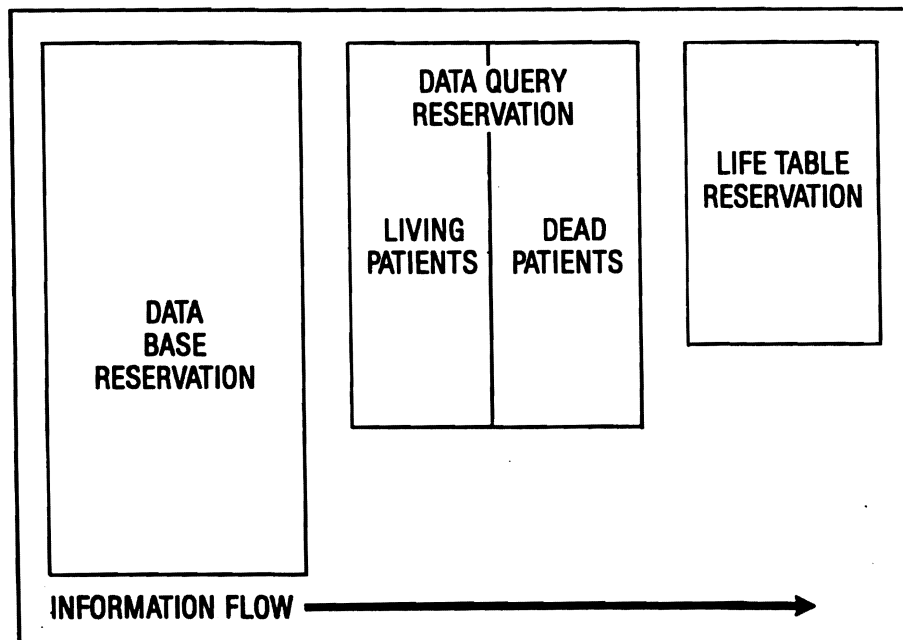
any second RETX DX ;

any RELTX named "third" with ( TRANSPLANT # = 3 ) RETX DATE ;

any third RETX DX ;

DAYS SURV.

\*The query language in Dataease is English-like and permits searches to link data contained in several forms. The query that produces the fields for construction of the data base reservation in the spreadsheet is shown above. The report is produced for patients receiving their first transplant between dates specified by the user ("data-entry start date" and "data-entry end date"). RELTX signifies the relationship between records in the LTX RECIPIENTS form and the LTX RETRANSPLANTS form. This relationship is defined by the field OT#, the primary patient sequence number that is assigned to every patient and is the same for that patient in every record of every data set. Thus, the query, in addition to data pertinent to a first transplant, will also produce the date and indication for any second or third transplant that a patient might have received. After completion of the query, the user can specify the format of the report. In this case, an export format (DIF) is chosen and a file is created on a disk containing the data in DIF format.



**FIG. 2. Spreadsheet organization.** The data base reservation is created from records imported from the relational data base manager. Any desired subset of these records can be selected for the data query reservation and forms the basis for computation of a life-table using formulas stored in the life-table reservation. The entire process is automated using a macro command, a memorized set of keystrokes that performs a complex set of instructions whenever needed.

which is created from the Dataease generated DIF file. The first row in this reservation contains the field names from the Dataease data sets. Each subsequent row contains one patient record from the registry.

The second reservation is the *data query reservation*. Using the query commands available in Symphony, subsets of patients can be selected from the data base reservation for analysis. Usually, two subsets are generated, one of patients living and the other of patients dead. For example, to prepare a life-table of patients receiving liver transplants for biliary atresia, a subset of all living patients and a second subset of all dead patients would be extracted from the data base reservation.

The third reservation is the *life-table reservation*. The cells in this reservation contain the numbers and formulas for calculation of a life-table. Symphony has a frequency distribution function that can analyze the survival intervals for each patient subset and create survival frequency distributions for living and dead patients over any range of user-specified intervals. Simple mathematical calculations then complete the life-table.

A macro has been developed that automates the transfer of data among the three data reservations. The user has only to specify the search parameters desired for the selection of subsets. Query of the data base reservation, listing of the subsets, generation of the frequency distribution of surviving and dead patients, and calculation of the life-table are performed automatically by execution of a simple macro command. A second macro produces a printed copy of the life-table.

If desired, Symphony can produce a life-table curve on the screen, printer, or plotter or the data can be supplied to a dedicated graphics program to produce illustrations of the highest quality. Graphwriter (Graphic Communications, Inc., Watham, MA) and Microsoft CHART (Microsoft Corp., Bellevue, WA) have been used for this purpose.

## CONCLUSIONS

The combination of a relational data base management program, a state of the art integrated spreadsheet, and dedicated graphics software have been used to develop a liver transplant registry on an IBM compatible microcomputer system. The software, readily available from retail suppliers, is versatile and can be customized by the user to perform a great variety of tasks without learning complex program code and without the services of a professional programmer. Such a system greatly enhances the ability of a transplant program to review its experience critically, to identify favorable or unfavorable trends early, and to modify its procedures accordingly.

*Acknowledgment.* Supported by research grants from the Veterans Administration and by Grant No. AM-29961 from the National Institutes of Health, Bethesda, Maryland.

## REFERENCE

1. Gordon RD: Basic information for microcomputer data base management. J Vasc Surg 1:585-589, 1984.