AN INTRODUCTION TO THE CLINFO PROTOTYPE DATA MANAGEMENT AND ANALYSIS SYSTEM

PREPARED FOR THE NATIONAL INSTITUTES OF HEALTH

R-1541-NIH

DECEMBER 1977

G. F. GRONER, M. D. HOPWOOD, N. A. PALLEY, W. L. SIBLEY

This research was supported by the National Institutes of Health under Contract No. NO1-RR-5-2111. Views or conclusions expressed herein should not be interpreted as representing the official opinion or policy of the National Institutes of Health.

Published by The Rand Corporation

AN INTRODUCTION TO THE CLINFO PROTOTYPE DATA MANAGEMENT AND ANALYSIS SYSTEM

PREPARED FOR THE NATIONAL INSTITUTES OF HEALTH

G. F. GRONER, M. D. HOPWOOD, N. A. PALLEY, W. L. SIBLEY

R-1541-NIH
DECEMBER 1977

Rand
SANTA MONICA, CA. 90406

PREFACE

This report provides a description of the final release of the CLINFO prototype system for the management and analysis of clinical research data. With minor changes, it comprises the first three sections of the *CLINFO User's Guide.* The *User's Guide* includes a reference manual with complete details and is aimed at active users of the system.

The prototype described in this report has been developed as part of the CLINFO project, a scientific inquiry sponsored by the Division of Research Resources (DRR) of the National Institutes of Health (NIH). The goals of the project are to identify and characterize the information-analytic tasks and the information flows in clinical research, and to develop methods for facilitating these tasks and flows.

As implemented, the prototype provides data management and analysis capability for clinical researchers. In addition, it collects data that are used to evaluate its effectiveness in clinical research centers, where it is being implemented on an experimental basis. This release reflects many changes and improvements suggested by users of the system in earlier evaluations. It incorporates suggestions from these continuing evaluations to provide a set of functional specifications for a data processing tool that can be used throughout the clinical research community.

The CLINFO project is being conducted by clinical investigators at the Baylor College of Medicine, the University of Washington, and Vanderbilt University; by information scientists at The Rand Corporation; and by staff members of the DRR. A clinical investigator at the University of Oklahoma participated previously. The project has thus far

- Broadly characterized the information-related activities of the clinical research process.
- Identified research data management and analysis as major problems in clinical research.
- Carried out an extensive survey among clinical investigators which indicated that these problems are widespread.
- Examined a number of existing systems aimed at alleviating these problems.
- Prepared an initial functional description of a prototype data management and analysis system.
- Implemented this prototype in three clinical research centers, using standard minicomputer hardware, an operating system supplied by the manufacturer, and specialized computer programs written in the BASIC language.
- Evaluated the effectiveness of the prototype and developed functional specifications and diffusion plans for a widely available CLINFO system.

Project results to date are described in a number of articles and in the following Rand Corporation reports:

N. A. Palley and G. F. Groner, *A Survey of Clinical Investigators and Their Information Processing Activities,* R-1539-NIH, July 1974.

G. F. Groner, M. D. Hopwood, N. A. Palley, N. Z. Shapiro, and W. L. Sibley, *A Plan for the Development and Evaluation of a Data Management and Analysis System for Clinical Investigators,* R-1542-NIH, August 1974.

W. L. Sibley, M. D. Hopwood, G. F. Groner, and N. A. Palley, *A Prototype Data Management and Analysis System for Clinical Investigators: An Initial Functional Description,* R-1621-NIH, August 1974.

G. F. Groner, N. A. Palley, and N. Z. Shapiro, *A Structural Characterization of Clinical Research Participants and Their Activities,* R-1540-NIH, January 1975.

G. F. Groner, W. R. Baker, Jr., T. G. Christopher, M. D. Hopwood, N. A. Palley, W. L. Sibley, and H. K. Thompson, Jr., *The Design and Evaluation of a Prototype Data Management and Analysis System for Clinical Investigators,* P-5746, November 1976.

G. F. Groner, N. A. Palley, M. D. Hopwood, W. L. Sibley, and B. Fishman, *CLINFO User's Guide: Release Three,* R-1543-3-NIH, September 1977.

M. D. Hopwood, G. F. Groner, N. A. Palley, W. L. Sibley, J. C. Mabry, H. K. Thompson, Jr., T. G. Christopher, R. O. Ling, W. W. Lacy, and J. L. O'Connor, *An Evaluation of the CLINFO Data Management and Analysis System,* R-2260-NIH, forthcoming.

CLINFO prototypes are currently operating experimentally at three clinical research sites that have established suitable physical and scientific environments in cooperation with the CLINFO project investigators. The first prototype has been operating at the Baylor College of Medicine in Houston, Texas, since January 1976. The CLINFO project director there is Dr. Howard K. Thompson, Jr., and the system manager is John C. Mabry. The second prototype has been at the University of Washington School of Medicine in Seattle, Washington, since September 1976. T. Graham Christopher, M.D., is the CLINFO project director, and Ronald O. Ling is the system manager. The third prototype has been operating at the Vanderbilt University School of Medicine, Nashville, Tennessee, since February 1977. William W. Lacy, M.D., is the CLINFO project director, and Joel L. O'Connor is the system manager.

This report is primarily intended for those who will be actively employing the CLINFO prototype in their clinical research activities, but it should also be of interest to members of the NIH staff, to information scientists concerned with the information processing needs of clinical investigators, and to all clinical investigators interested in data management aids.

As CLINFO is an experiment in its own right, users are encouraged to be innovative and to attempt operations and manipulations which may not be explicitly described in the CLINFO documentation. We hope that the results of these innovative pursuits, both successes and failures, will be reported to us; such information provides insights into the needs of the system and can be of considerable value to the CLINFO project investigators.

ACKNOWLEDGMENTS

Although the authors accept responsibility for the contents of this report, many others deserve credit for their contributions to the CLINFO Prototype Data Management and Analysis System. Our former colleagues Steven S. Zucker and William H. Josephs participated in the system design and implementation. Barry Fishman designed and implemented analytic portions of the system and wrote the corresponding sections of the *User's Guide.* Marie Hall and JoAnn Lockett have previously been responsible for maintaining the computer hardware upon which we develop, test, and use the CLINFO software; Ms. Lockett has made other contributions as well.

Our co-contractors, Howard K. Thompson, Jr., of the Baylor College of Medicine (the first CLINFO site), T. Graham Christopher, of the University of Washington (the second CLINFO site), and Arthur W. Nunnery, of the University of Oklahoma, deserve special credit. They have collaborated with us since the beginning of the CLINFO project in July 1972, educating us about clinical research, participating in developing the project's research plan, ensuring that the design of the prototype system meets the needs of a substantial number of clinical investigators, and evaluating the prototype itself. John C. Mabry, of the Baylor College of Medicine, Ronald O. Ling, of the University of Washington, and William W. Lacy and Joel L. O'Connor, of Vanderbilt University, have also influenced the system design. The users of the prototype systems in Houston, Seattle and Nashville have been most helpful in uncovering deficiencies and problems and in keeping our attention focused on the practical needs of the clinical investigator.

The CLINFO project and system could not be successful without continued guidance, encouragement, and support from several staff members of the National Institutes of Health and from the group of NIH advisors for the CLINFO project. We especially want to thank our project officer, William R. Baker, Jr., whose foresight and persistence have made our endeavors both exciting and possible.

Our colleague Thomas L. Lincoln deserves credit for having educated all of us at Rand over the past several years about the need for such a system and for using the prototype, as we develop it, extensively and in many stressful and novel ways.

Finally, the authors acknowledge the assistance of Jacqueline Bowens, the project secretary, who typed successive drafts of this report, using the WYLBUR text editor. Ms. Bowens has also helped to operate the CLINFO computer hardware at Rand.

CONTENTS

I. INTRODUCTION

The CLINFO Prototype Data Management and Analysis System is a computing system for the storage, retrieval, and analysis of data (called clinical study data) pertaining to medical research on human subjects. It also provides a means for evaluating the efficacy of such systems in the environment of General Clinical Research Centers (GCRCs) funded by the GCRC branch of the Division of Research Resources, National Institutes of Health (NIH).

GCRCs are small inpatient/outpatient research units, generally with fewer than 20 beds and usually located within a university-affiliated hospital, that serve as institutional research resources for improving the quality of care that physicians offer. They provide patient care, special prescribed dietary services, biological sample collection, and specialized laboratory facilities. Approximately 80 GCRCs are funded by NIH. The CLINFO experiment is designed to determine the requirements for a GCRC data processing tool and its computational/statistical support staff.

The CLINFO prototype is based on the premise that clinical research data are a precious commodity representing considerable intellectual, organizational, and financial investment. They are derived from tests and measurements made on human subjects, and they hold the promise of new knowledge in the treatment or prevention of human disease. The unique set of data collected during the study of a particular group of patients, under a single protocol, is the focus of the CLINFO prototype.

The data for a number of studies may be physically stored within the CLINFO system, and several investigators may use the system simultaneously. However, each investigator should view the system as his personal data storage, manipulation, and analysis tool. The design of each study will probably be unique in terms of the number and identity of variables, the frequency of data collection, the number of subjects, and the treatment and control conditions under which the data are collected; nevertheless, there are characteristics and dimensions common to all collections of clinical research data.

WHAT CLINFO DOES

The CLINFO prototype was designed to support clinical research data processing activities by providing an interactive computer system that

- Allows the investigator to describe the contents of a computer-based *study data file.*
- Provides for data entry into the study data file and several forms of data screening and encoding.
- Protects the investigator's data from loss due to computer malfunction.
- Allows the investigator to extract from the list of all patients in the study those patients who have particular characteristics.

- Allows for the creation of *worksheets* of data from a variety of sources, in particular, from the investigator's computer-based study data file.
- Allows the investigator to perform any of a number of mathematical or statistical analyses upon these worksheets and produce simple reports, plots, and graphs from them.
- Allows for the transformation of CLINFO data files into a form compatible with a simple programming language, BASIC.
- Provides for collection of statistics to aid in performing the CLINFO experiments and managing the CLINFO prototype as a clinical research center resource.
- Provides an operating environment for the prototype system which allows the inclusion of additional system features that are not part of the original design.

The CLINFO capabilities are provided by a set of computer programs that have been developed and continuously refined at The Rand Corporation, under the sponsorship of the NIH Division of Research Resources. The prototype is implemented on a Data General minicomputer which, with all associated hardware except user terminals, costs about $100,000. Up to eight terminals may be connected to the minicomputer; however, it is expected that only four of them will be in use simultaneously.

The minicomputer is physically installed in or near the clinical research center, and its operation is overseen by a resident system manager responsible for day-to-day operation, and by a CLINFO project director with overall responsibility for the local CLINFO experiment. Designed for the active investigator-users of a GCRC, CLINFO provides facilities for about 30 current studies resident in the system.

To use the prototype, the user logs onto a free-standing, remote terminal with a password that has been assigned to him and then proceeds to describe his study, enter data, retrieve previously entered data, create patient subsets, or analyze and display data. The password is required both to maintain the privacy of the research data and to protect it against malicious or inadvertent tampering. Passwords are assigned by the system manager.

The system is highly integrated and allows the user to perform a variety of functions without the need to write programs or to reformat data when passing them from one function to another. The CLINFO user does not have to remember or formulate complex commands in order to use the system. Rather, CLINFO asks him about the next step to take while suggesting possible responses, and he responds by typing the name of a variable or worksheet or a short word such as *yes* or *r*. Whenever a choice among functions is called for, the prototype displays on a CRT (cathode ray tube) screen the current options available and requests a response. By typing in one-word commands beginning with "!", however, experienced users may move from function to function without the need to pass through intermediate menus of choices. On command, a small printer attached to the terminal will copy the contents of the CRT screen onto paper.

The design philosophy of the CLINFO prototype requires that it be very "forgiving"; thus, inexperienced users may learn to use the system by logging on and trying the options presented, without being concerned that they will harm the system or their data in any way.

The resident CLINFO system manager at each site is available to provide assistance to CLINFO users. The system manager holds formal training sessions for new users and instructs, assists, and advises both new and experienced users at their request. There are also certain system functions that only the system manager can carry out. These include, for example, maintaining backup copies of all user data, retrieving copies of files that may have been deleted inadvertently, preparing files of data for analysis on other systems, and assigning passwords to new users. In addition, the system manager oversees the maintenance of the CLINFO system, writes special-purpose application programs for users, and responds to users' problems; he also communicates with the system designers about user requests, actual or suspected errors in the system, and system utilization. In addition, each CLINFO site is visited periodically by a Rand Corporation CLINFO system representative, who learns about users' needs and problems first-hand, informs users about system changes that are planned or in progress, and assists users as necessary.

CLINFO TERMINOLOGY

Patients, Items, and Panels

The unit of analysis is usually the patient (or case or subject). The investigator assigns a unique *patient abbreviation* (e.g., initials or hospital number) to each subject. A number of different observations, measurements, and tests are made on each subject. We call these measurements, tests, and classifications *items* (e.g., systolic blood pressure, height, sex), and we refer to the specific quantities or qualities associated with the items as *values* (e.g., 140 (torr), 63 (inches), male).

The values for certain sets of items are usually collected at the same time (for example, height, weight, and other physical data or a number of measurements made on a single laboratory sample). Such collections of item values are termed *panels*. In the CLINFO system the panels and items are named by the user.

The data for a single study may include a number of *instances* of each panel, where each instance corresponds to a different data-collection time. For example, if blood were drawn every 15 minutes after a drug is given, each resultant set of blood chemistry values would correspond to an instance of the blood panel, and the data collection times for a particular research subject would be the times that his blood is drawn. At the time of data entry, each instance of a panel is uniquely identified by its *context*, i.e., by the panel name, patient abbreviation, and time of data collection.

Time and Events

The protocol defining the clinical research study usually specifies the times at which measurements are to be taken relative to some experimental procedure, but the data are actually collected at real clock times which are often different for different patients. For the purposes of accuracy and accountability, it is useful (but not necessary) for the data to be identified by the (clock) time of collection; but for review and analysis, it is necessary to align the data in time relative to a fixed

occurrence across all of the patients, or relative to the occurrence of a measurement whose time cannot be predicted.

The CLINFO system allows the user to specify *events* that may be used during retrieval to indicate a relative zero time, around which instances of panels (or items) may be arranged. The system always recognizes two events: *begin*, which is the time of the first data collection, and *end*, the time of the most recent instance of any panel. Other events may be specified to be the time of the first, last, maximum, or minimum value of any item, or the time of occurrence of a specific value of an item. Thus, whereas an item may have different values at different times, an event has only a single value which *is* a time; and whereas individual item values in the data base do not change, the time of an event may change as data are entered and the item associated with the event takes on new values.

Numbers, Codes, and Text

The error-free recording of study data is an obvious requirement for meaningful analysis but is not easily assured. The CLINFO system approaches this problem by providing user-specified error-checking procedures as part of the data-entry activity.

The data recorded by the CLINFO system are of two types, numeric and textual. Most of the data stored by CLINFO will be of the *numeric* type, which includes real numbers, times, dates, and codes. Short words or *character strings* may be used to represent predefined values of items, e.g., *male* and *female* might be used as values of the item *sex*, and *lung*, *breast*, *liver*, *colon*, etc., might be used as values of the item *primsite* (i.e., primary site). Entered character strings are translated into numeric *codes* for storage, and the numeric codes may be decoded into character strings for display.

The textual type of data provides for the recording of comments and observations. Textual data may be retrieved and displayed, but the system does not provide for analysis of them (e.g., it cannot search for a specific word or phrase), nor does any error checking take place. (The CLINFO hardware does support a text editor called Clem, not described here, which is useful for such applications.)

The Study-Specific Schema

The CLINFO prototype (1) "understands" the association among items in a panel, (2) recognizes the names of items, (3) screens values upon entry for conformity with user-specified ranges or codes, and (4) flags events, all in a manner specific to an individual study. A good deal of information must be supplied to the system, and this is done by a procedure called *describe*, which assists the user in constructing a data dictionary that describes the items to be included in his study. This data dictionary, called a *schema*, includes the names of panels, items, and events, the data types of the items, the kinds of screening to be performed, and the equivalence between the numeric codes and character strings.

CLINFO Activities

Describe is one of the major groups of functions in CLINFO called *activities*. Other activities provide for entering, retrieving, and analyzing data, creating pa-

tient subsets, and manipulating rectangular arrays of data. An individual function (e.g., listing the patients in a specified subset, or performing a chi-square test) is called a *subactivity* or *option.* A new CLINFO user may perform a function by selecting an activity from a menu (i.e., a list displayed on the CRT), selecting a subactivity, and then responding to a series of prompts. An experienced user may go directly to the subactivity by typing the activity name and the subactivity number or by using keys labeled with abbreviated activity names.

Subactivities are generally designed to perform commonly used, discrete functions. For example, Frequency Distribution, *analyze,* option 9, computes the numbers of times that distinct numbers or codes occur. Some more complex or less common functions require that several subactivities be used in combination. For example, a cumulative distribution is plotted by computing frequencies with *analyze,* option 9, obtaining the cumulative distribution by applying *calculate,* option 4 (Cumulative Sum) to the results, then plotting the distribution with *analyze,* option 7 (Scatter Plot). Such sequences are similar to those performed manually and are easily devised once the user is familiar with the system's capabilities. Other functions, particularly those involving data retrieval, that do not correspond to typical manual operations are more difficult to devise, and help from the system manager may be required. Helpful hints listed below indicate how to perform some of these more complex functions. If a particular sequence of steps comes to be used repeatedly, this sequence can be stored in a CLINFO *response file* (see below) to simplify the processing.

Patient Set

The user may find it useful to identify different groupings of the patients in the study and to examine the data for these groups separately. These groups are called *patient sets* or *subsets* and they are established by using the *subset* activity. Since sets are merely named, ordered lists of patient abbreviations, a patient may belong to any number of sets. When retrieving data, the user has the option of specifying the set named *all* (which contains all of the patients on a study) or any other set. Subsetting can also be used in the inverse manner by finding those patients who are (or are not) members of specified sets. A set refers to a list of patients, *not* to their data.

OVERVIEW OF CLINFO OPERATION

An ideal data analysis system would perform as shown in Fig. I-1. That is, data are entered into the system in some predefined order; they are then rearranged according to the requirements of the proposed analyses, and those analyses are carried out, producing the "results."

In the real world of clinical research, however, the data become available a little at a time, and often in unpredictable order. Furthermore, the organization of the data may be quite complex, with some items collected only once, other items collected repeatedly, and other items (or sets of items) missing for some patients. But most statistical analyses require that the data upon which they operate be in neat, two-dimensional arrays, and such arrays are also a natural form for human obser-

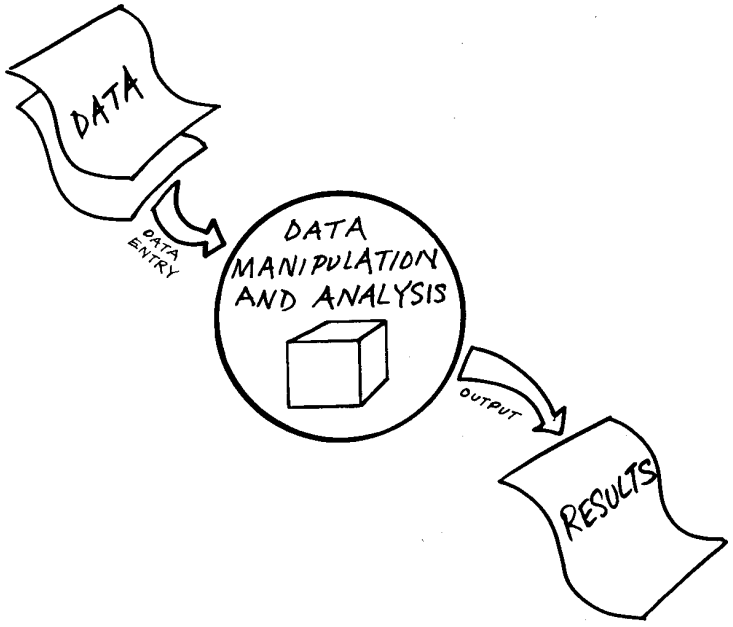

Fig. I-1—An ideal data-analysis system.

vation of data. In order to accommodate these conflicting needs (i.e., ease of data entry, manipulation, and analysis), the CLINFO system provides two complementary facilities for storing experimental data: the *Study Data File* (SDF) and *worksheets*. Data stored in the SDF are organized as shown in Fig. I-2. The SDF may be thought of as having three dimensions: patient, time, and item. The value of an item (or the group of values for items in the same panel) is represented by a point (or a line of points) in the three-dimensional space of the SDF. Because different patients may have different numbers of repetitions of some panels, the space may be very sparsely filled even when all the collected data are entered.

Worksheets are two-dimensional arrays which may be created (and destroyed) at the option of the user and which permit easy manipulation and analysis of data. Data may be entered directly into worksheets by a variety of methods, discussed below. When data are entered in these ways, the rows and columns of the worksheets may stand for whatever the user wishes. However, when the worksheets contain data retrieved from the SDF, they may be of only one of three major types, diagrammed in Fig. I-3. They are termed "patient," "item," or "time" worksheets, and each represents a different slice through the SDF. This relationship is indicated in Fig. I-3 by showing the "third" dimension as a dotted line. Since the slices are two-dimensional, they can contain values for only one "point" on the excluded dimension. Thus, in the patient worksheet, the rows can represent different items, and the columns, different times (or instances). Every value in that worksheet belongs to the same patient, as illustrated in Fig. I-4. A single patient worksheet might contain all of the data stored for that patient.

Similarly, the other two worksheet types result from the retrieval of (1) many values of a single item (for a group of patients at a number of times) or (2) many

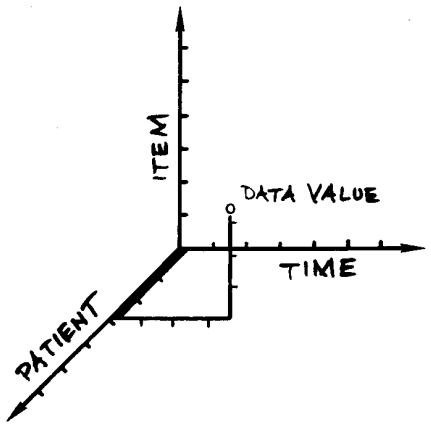

Fig. I-2—Organization of data stored in the SDF. Each data value is associated with a particular patient and a particular data collection time for a schema item. Multiple data values may be stored for a single item, each with a different data collection time.

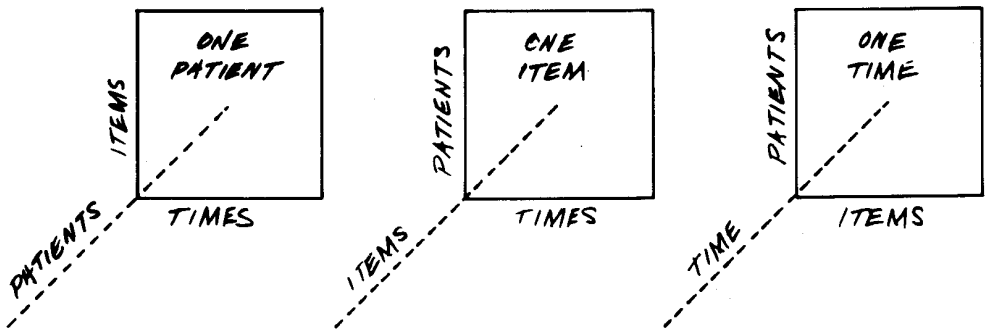

Fig. I-3—Types of worksheets retrieved from the SDF. Each of the three types represents a slice through the SDF. A patient worksheet contains values for multiple items at multiple times for a single patient.

values collected at a particular time (or relative time) for a group of patients and a number of different items.

The primary relationships among the SDF, the worksheets, the original data, and the "results" are diagrammed in Fig. I-5. The arrows represent CLINFO *activities;* that is, user-directed operations that transform the data as they pass through the CLINFO system. The four activities illustrated in this simplified diagram are:

1. *Enter* (data entry into the SDF).
2. *Retrieve* (retrieval of portions of the SDF into worksheets).
3. *Analyze* (the generation of statistical results or other reports, graphs, or charts).
4. *Worksheet* (the creation and direct entry of data into worksheets).

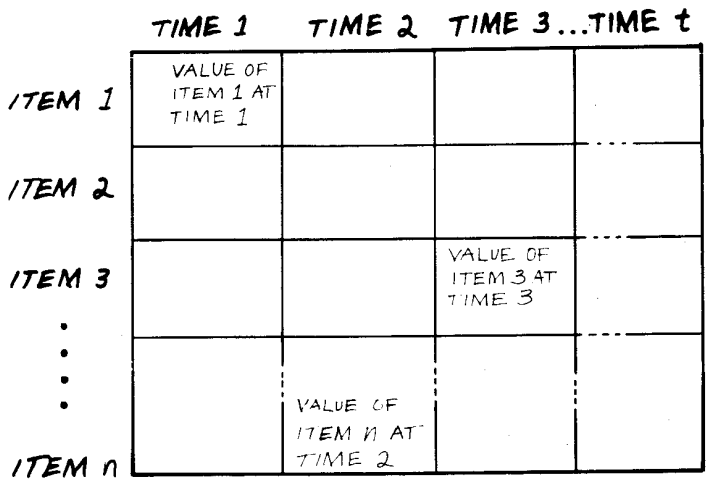

Fig. I-4—A patient worksheet. Each cell is the value of a particular (or relative) time.

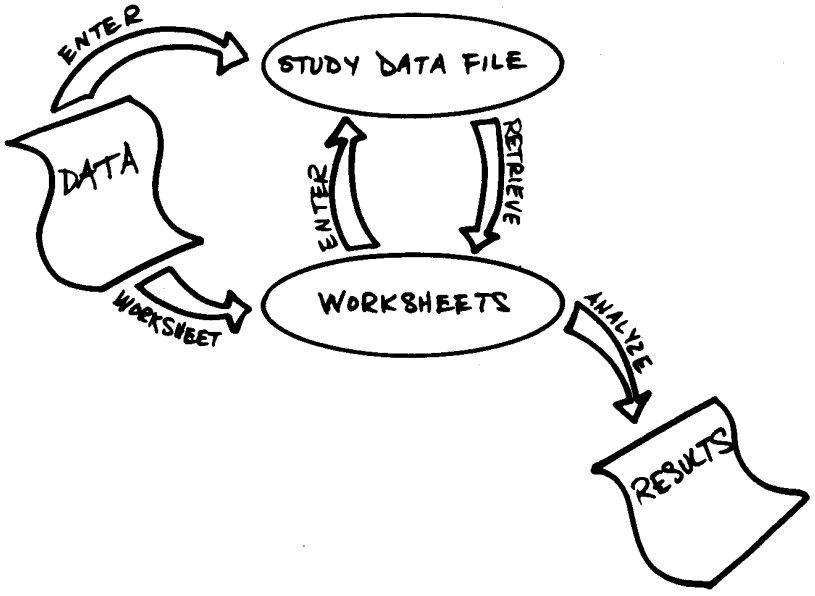

Fig. I-5—Schematic relationships among major CLINFO activities (arrows), data storage (ovals), and data input/output. Primary storage is in the SDF; all analyses are done on worksheets.

These activities may be used in several different sequences, including the following three:

1. Enter data into the SDF (*enter*); retrieve data from the SDF into worksheets (*retrieve*); then analyze or plot worksheet data (*analyze*).
2. Create, name, and label worksheets (*worksheet*); enter data into worksheets (*worksheet*); then analyze or plot the data (*analyze*).
3. Create worksheets and enter data into them (*worksheet*); analyze the data and store resultant derived data in new worksheets (*analyze*); discard the original worksheets (*worksheet*); enter the derived data into the SDF (*enter*); retrieve data from the SDF into worksheets that have different structures and contents than the original ones (*retrieve*); then analyze or sort the data in the new worksheets (*analyze*).

Each of the major activities has within it various options. For example, the *worksheet* activity provides for the naming and titling of new worksheets, the labeling of their rows and columns, the entry of data into the worksheet (by row, column, or cell), sorting of data within worksheets, and rearranging rows and columns within existing worksheets.

As described above, the SDF is a highly structured file into which user data are placed, and from which they are retrieved. The particular structure of an investigator's SDF depends on his particular study—the number of cases, the number of different items (variables), the time relationship among the items, the number of replications of groups of items, etc. This information (and more discussed below) is contained formally or informally in the investigator's experimental protocol. It must be communicated to the CLINFO system so that the SDF can be constructed to represent the data structure of the study, so that names may be associated with items, and so that values of items may be screened before they are entered into the SDF.

This requires the introduction of several more concepts, including two more activities and two more files. Figure I-6 illustrates these new concepts and presents a more complete functional picture of the CLINFO system.

One of the new files in Fig. I-6 is the *schema*, which contains a full description of the SDF; the investigator uses the schema to make CLINFO specific to his problem. All data added to the SDF are first checked against the schema to see that they conform to the investigator's specifications. The arrows representing the *enter* and *examine* activities are thus shown as passing through the schema. The activity that permits the user to construct the schema is called *describe*. This activity is used only at the start of a study; once constructed, the schema is "locked" to protect it against accidental changes and to protect the "meaning" of each stored item value from inadvertent changes. The Update File provides a buffer between various data entry procedures and the SDF. It holds data temporarily until they can be merged into the SDF. Thus, data entry is not slowed down by the rather complicated process of inserting new data into the proper locations in the SDF. Furthermore, the data in the Update File may be reviewed (using option 3 of the *enter* activity) and changes may be made before merging. The *examine* activity permits the user to view the contents of the SDF directly, to modify (or delete) the values of individual items (such changes are automatically flagged for data-auditing purposes), and to delete entire panels.

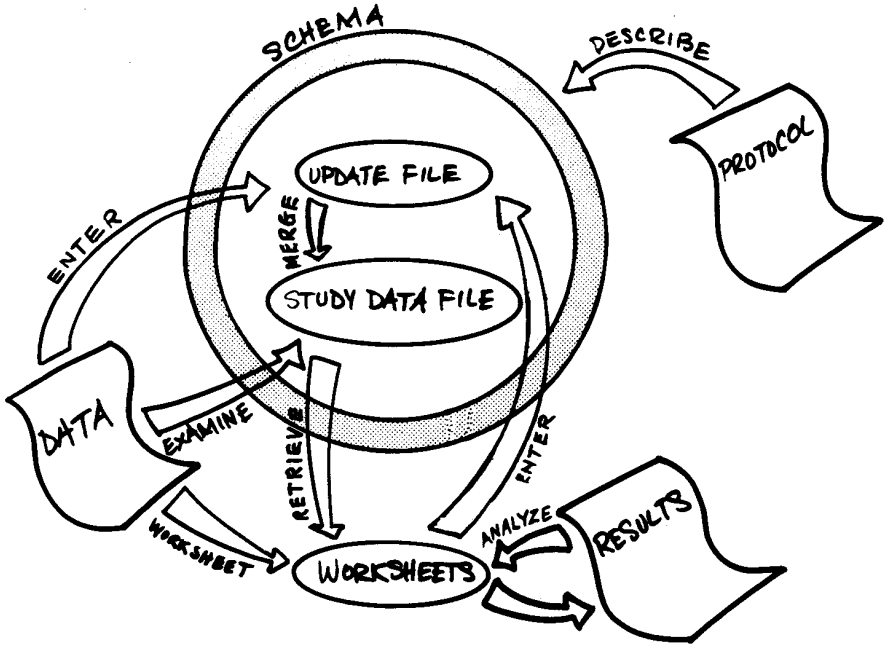

Fig. I-6—A more complete functional representation of the CLINFO system. The schema, derived from the experimental protocol, mediates all communication with the SDF. Data enter the SDF via the Update File.

Although there are many variations, the activities illustrated in Fig. I-6 are normally used in the following sequence:

- Describe and create the schema (*describe*).
- Enter patients into the study (*enter*).
- After trial data entry (*enter*) and retrieval (*retrieve*), modify the schema (*describe*) until it is satisfactory.
- Enter data into the Update File (it cannot be entered directly into the SDF) and review it (*enter*).
- Ask the system manager to merge (i.e., reorganize and store) the Update File into the SDF, then discard the Update File (*merge*).
- Examine and possibly modify data in the SDF (*examine*).
- Retrieve data from the SDF (possibly for a subset of patients—see the following discussion) into worksheets (*retrieve*).
- Create additional worksheets (*worksheets*), enter original data (*enter*) and the results of analyses (*analyze*) into worksheets, and enter worksheet data into the Update File (*enter*).
- Enlarge retrieved worksheets or reorganize data within them (*worksheet*), sort data in worksheets based on values of variables (*worksheet*), move data from one worksheet to another (*retrieve*—not illustrated), and perform calculations on worksheet data and store the results directly in the same worksheets (*calculate*—not illustrated).
- Analyze or plot worksheet data and possibly store the results in new worksheets (*analyze*).

Another important activity, *subset*, is illustrated in Fig. I-7. It comprises subactivities which allow the creation of sets of patients according to logical and arithmetic criteria specified for data in the SDF (options 4 and 5) or for data in worksheets (options 6 and 7). A set may also be created by naming its members explicitly (option 1) or by performing logical operations on existing sets (options 2 and 3). Once sets are created, retrieval from the SDF may be restricted to data belonging to members of a particular patient set. This is represented in Fig. I-7 by an arrow showing retrieval of data from the SDF to worksheets, going "through" the Subset File.

Several other activities are also shown in Fig. I-7. The *calculate* activity and a subactivity of *retrieve* are both represented by the same arrow. *Calculate* permits arithmetic, logical, and special operations to be performed on worksheet rows or columns, and stores the results in the same worksheet. Arrows going back and forth between worksheets and results represent the capability of some *analyze* subactivities to store their results in worksheets for purposes of plotting or additional analysis. *Retrieve*, option 4, permits the user to copy portions of worksheets (or entire worksheets) into other (or the same) worksheets. This is ilustrated in Fig. I-8 , where a new worksheet, C, has been built which contains sections of worksheets A and B.

Another group of activities which make the system easier and more convenient to use is shown in Fig. I-9. These activities allow the display of data from the SDF, lists of patient sets or worksheets (in the *files* activity), lists of members of a particular set (in the *subset* activity), and portions of worksheets or entire work-

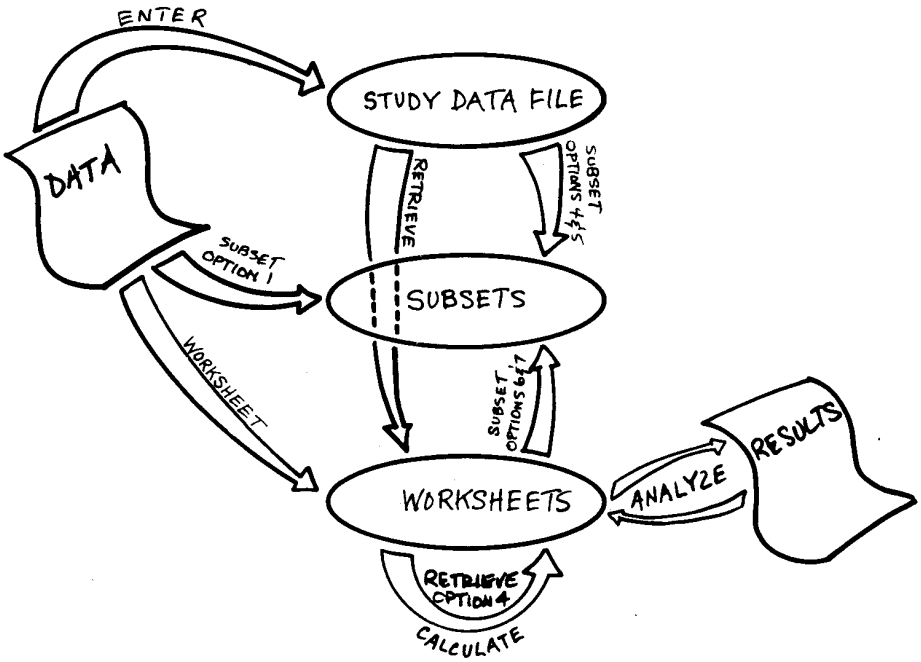

Fig. I-7—Relationships of the subset activity to other major CLINFO activities. Patient sets may be created by listing patient abbreviations, specifying conditions on data in the SDF, or specifying conditions on data in worksheets. Retrievals may then be limited to patients in specified sets.

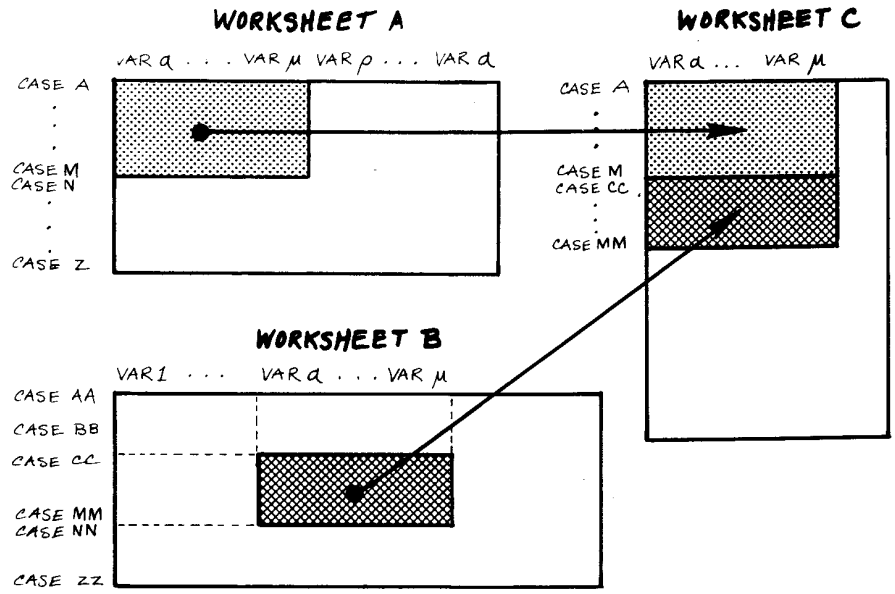

Fig. I-8—Copying of data from worksheets into other worksheets or different locations of the same worksheet using *retrieve*, option 4. Segments of worksheets, defined by row and column bounds, may be copied into other worksheets or copied into different locations in the same worksheet. In this example, values for variables α through μ for two groups of cases (A-M and CC-MM) are copied into a new worksheet C.

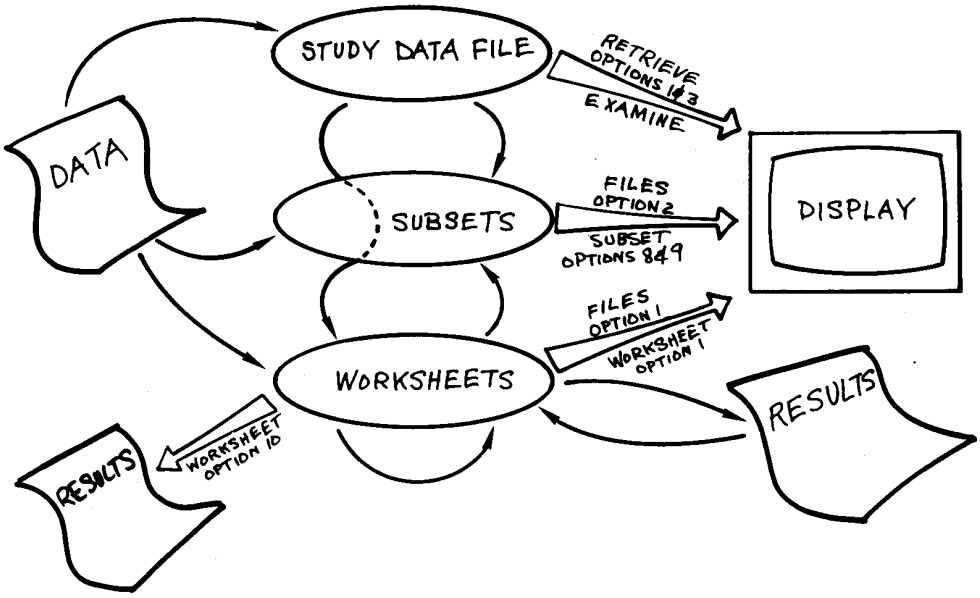

Fig. I-9—CLINFO options that permit display of data and other important study information. For example, *files*, option 2, presents a list of patient sets, while *subset*, option 8, lists the members of a particular set.

sheets (in the *worksheet* activity). If hard copy is desired, the contents of the CRT screen may be printed out on the small printer attached to the terminal; for more extensive output (e.g., an entire large worksheet), it is more convenient to use the system printer (using *worksheet*, option 10).

Figure I-10 introduces the concept of the *Communication File.* This file and the activities that operate upon it allow the exchange of data between CLINFO (the region within the dark frame), BASIC programs that run on the same computer as CLINFO, and external systems. For example, this exchange may be made via magnetic tape written by another computer or paper tape from an instrument such as a scintillation counter (arrow 1); or data values produced by using a BASIC program may be stored in a Communication File (arrow 2); or data in the SDF or a worksheet may be copied into a Communication File for transmittal to a BASIC program or external system *(retrieve,* options 3 and 5); or properly formatted data in a Communication File may be stored in a CLINFO worksheet *(retrieve,* option 6).

This capability allows the user to write (or have written) BASIC programs to carry out special analyses not presently provided by the CLINFO system. The results of such analyses may be displayed directly, may be passed back into CLINFO worksheets, or may be passed back (through a worksheet and the Update File) into the SDF.

A final activity, shown in Fig. I-11, is called *response files.* It enables a user to carry out a sequence of CLINFO activities repeatedly without reentering the required series of responses. In this figure, the user creates a *Response File* (using *response,* option 2), then carries out a sequence of activities (A, B, etc.) by entering

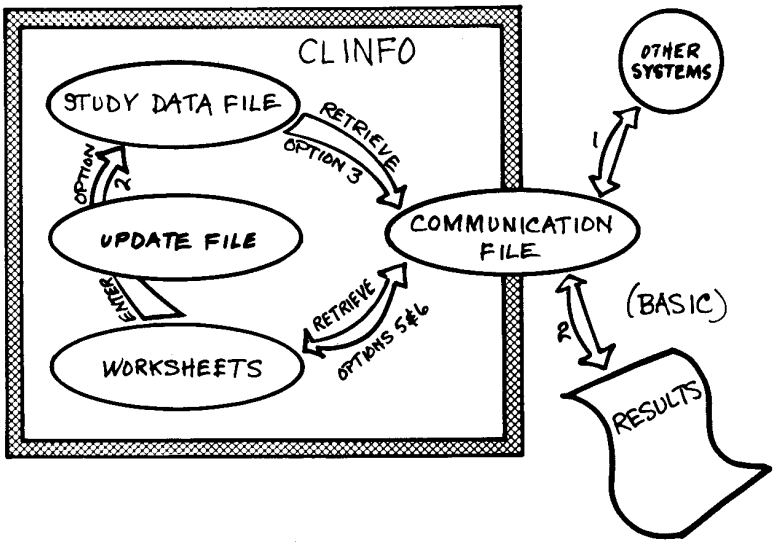

Fig. I-10—Using the Communication File, data may be moved back and forth from the CLINFO system to BASIC programs or, via computer tape, to other machines. Once data have been entered into worksheets, either directly or from the Communication File, they may be merged with the SDF.

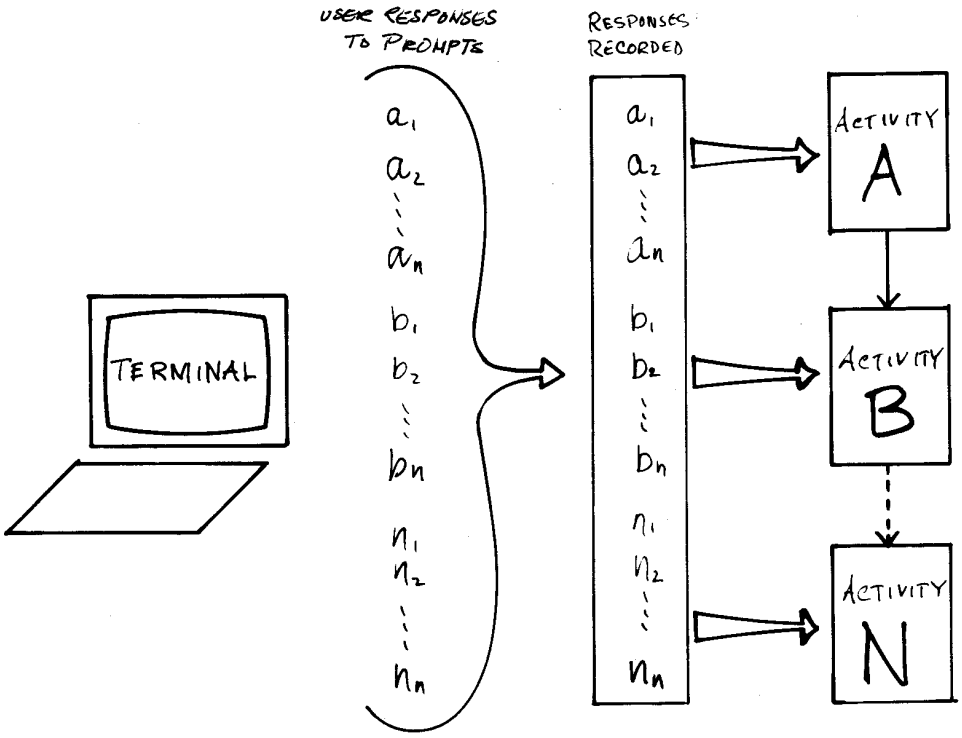

Fig. I-11—The creation of a Response File in the CLINFO system. Response Files are created by carrying out the desired sequence of activities; the user's responses are automatically recorded.

a series of responses (a_1, a_2, etc.) to the normal system prompts. The responses are recorded in a Response File which is named and saved.

Figure I-12 represents the editing of the Response File (using *response*, option 3). In this example, the response actually recorded, b_2, is replaced by a question mark which indicates that the *user* will supply a response each time the sequence is executed.

In Fig. I-13, the user executes the Response File (using *response files*, option 1). The system reads and acts upon the sequence of responses, a_1, a_2, . . ., a_n, b_1, from the Response File until a question mark is encountered. It then waits for the user to respond. After acting upon the user's response, the system continues reading responses from the file, b_3, n_1, n_n, etc., and acting upon them.

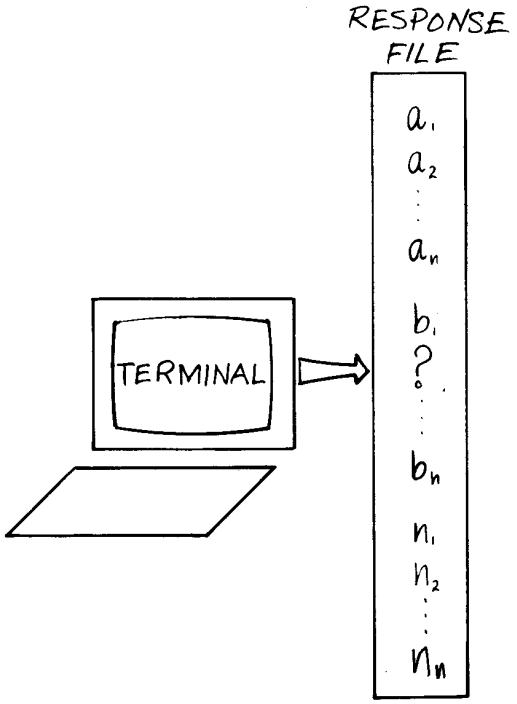

Fig. I-12—Editing the Response File. The user's responses may be edited, for example, to replace a specific response with a ?, which allows the user to enter a different response each time the Response File is run.

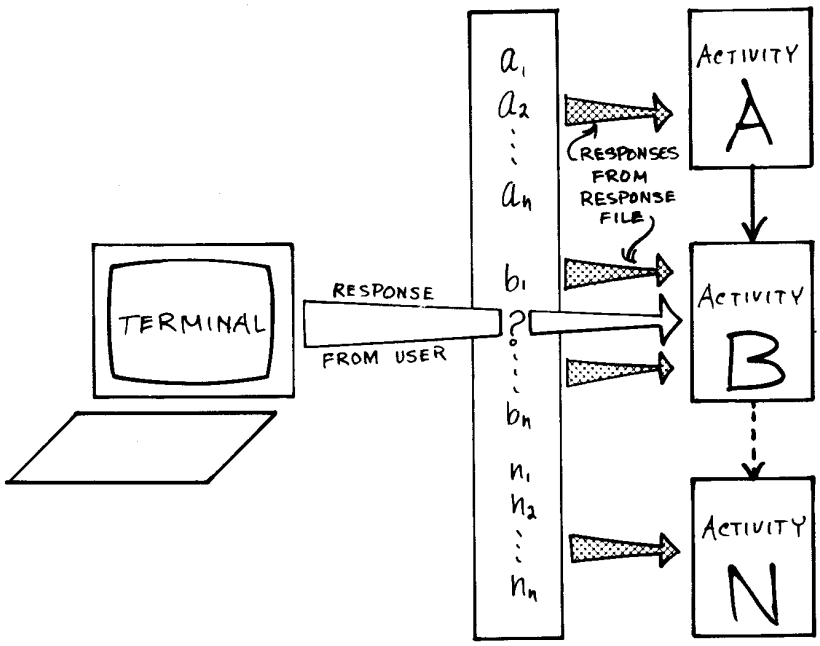

Fig. I-13—Execution of the Response File. The original sequence of activities is run with all prompts displayed and with most responses taken from the Response File; a question mark in the Response File causes it to wait for a user response from the terminal.

II. USING THE SYSTEM

THE ESSENTIALS

This section presents the essential information required to use the CLINFO prototype system.

Using the Terminal*

Turning It On. The Ann Arbor (AA) terminal and the attached printer are sketched in Fig. II-1. The power switch (a slide switch) is mounted on the side of the terminal. Place the switch in the forward position to turn the terminal on; the sound of a fan will indicate that the power is on. Press the Data Terminals and Communications (DTC) printer POWER switch; it should light up. Also press the LINE/LOCAL switch if the *line* portion is not lighted, and press the top of the DTC RESET switch.

Getting the Attention of the System. The ESC (escape) key, at the upper left-hand corner of the keyboard, is used to get the attention of the system both to log on (see below) and to interrupt ongoing operations. It is only active while you hold down the yellow key marked CLINFO, on the third row down, at the left-hand edge of the keyboard.

Where You Are. The small "dash" on the screen, called the *cursor*, is usually positioned following the last system prompt and moves one space as each character of your response is typed in.

Responding to System Prompts. After typing in your response (and seeing that it is correct on the screen), you must press the RETURN key (the large key at the right edge of the alphanumeric keyboard). Many prompts may be responded to simply with a RETURN.

Commands. A CLINFO activity is initiated by pressing the key labeled with that activity name (abbreviated, in yellow) and simultaneously pressing the yellow CLINFO key, then typing an option number (if an option is desired), then hitting the RETURN key.

Correcting Errors. After typing a response, and before hitting the RETURN key, you may correct your response by backing up over it (thereby erasing) character by character, using the BS (backspace) key directly above the RETURN key.

Precautions. It is important to keep in mind the following features of the system and terminal:

- The system distinguishes between upper and lower case. If the TTY LOCK (lower left) key is lighted, press it to deactivate it; when on, it acts as an upper-case lock for the alphabetic characters. The SHIFT LOCK key (just

*For more details, see pp. 38-41.

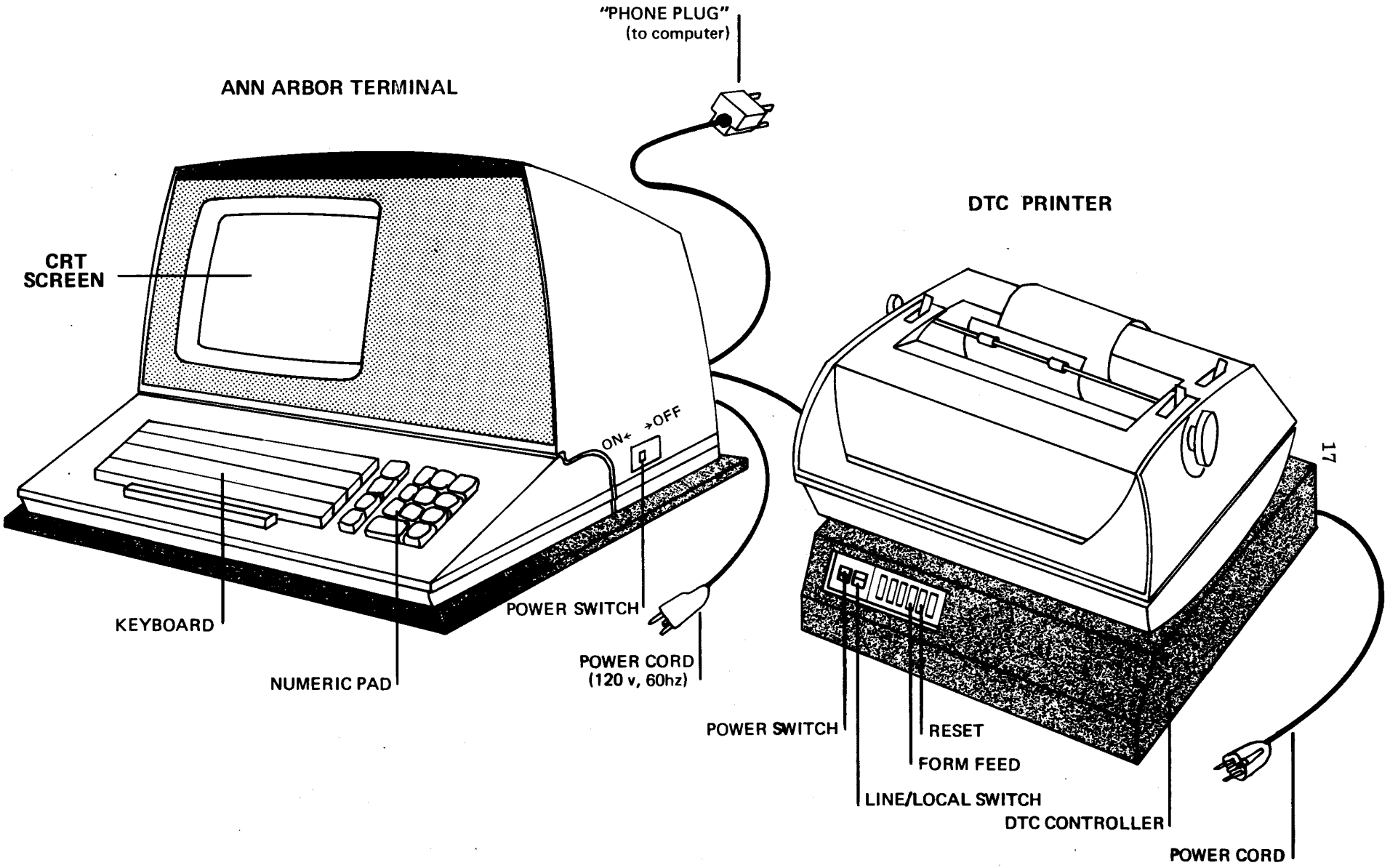

Fig. II-1—The Ann Arbor terminal and the attached DTC printer.

above the SHIFT key at the lower left of the keyboard) locks into place when pressed; to release it, hit the SHIFT key.

- Holding down keys (including RETURN) for more than a second causes them to repeat their function.

Logging On

Each time you use the CLINFO system, you must "log on," i.e., you must use the password issued to you by the system manager to identify the study with which you wish to work. If more than one person is to have access to the same study (e.g., the principal investigator, a fellow, and a technician), each will be issued his own password.

To log on:

1. Make sure that the terminal is on (see "Using the Terminal," above).
2. Hold down the yellow CLINFO key and hit the ESC key—the system should respond, "PASSWORD:".
3. Enter your password, e.g., 1234MYSTUDY, and hit the RETURN key. The password will not be displayed as you enter it, so be careful to enter it correctly. Common errors are:

   — Using the wrong case (e.g., the SHIFT LOCK or TTY LOCK may be on).
   — Confusing the letter "o" with the numeral zero.
   — Confusing the letter "*l*" with the numeral one.
   — Inserting a blank between the numbers and the letters, e.g., 1234 MYSTUDY.

4. The system should respond by clearing the screen, perhaps displaying a "message of the day," indicating the current worksheet and the size of any existing Update File, and prompting for the name of the first activity to be used. If this occurs, proceed as described in "Communication Between Activities," below. If, instead, the system responds

   UNKNOWN PASSWORD
   CLINFO RELEASE 3.0 AT YOUR SERVICE

   make sure you are using the correct password and repeat from step 2 (i.e., hit the CLINFO and ESC keys again).

   If the system responds

   ID IN USE
   CLINFO RELEASE 3.0 AT YOUR SERVICE

   someone else is currently logged on to your study. If no one else is authorized to access your study, check with the system manager; if other people are authorized to use it, wait until they are finished and try again.

   If the system does not respond to the CLINFO/ESC keys with "PASSWORD:," check with the system manager—he may be doing backup or some other system function, or the terminal may not be properly connected to the computer.

If you were not the first to use the terminal after turning it on, it might have a message of the form

CLINFO RELEASE 3.0 AT YOUR SERVICE

on the screen; proceed as above.

CLINFO Activities

Summary of Activity Functions. The CLINFO activities and their functions are listed below. The activity names, prefixed with !, are the commands that call up the activities themselves (see "Communication Between Activities," below). The activities are described more fully later in this section.

The options within each major activity are listed in the menus on the inside back cover and are briefly described later in a table entitled "Summary of Activity Options." They are described in detail in the *CLINFO User's Guide.*

Activity	Function
analyze	Plot and perform statistical analyses upon data in worksheets; store results in other worksheets.
calculate	Specify and carry out arbitrary arithmetic, logical, and other operations upon data in rows or columns of worksheets; store results in the same worksheet.
describe	Construct, edit, and print the schema and create and delete study files.
enter	Enter patients into a study; enter data into the Update File.
examine	Examine and edit SDF data panel by panel.
files	Display ordered lists of worksheets, subsets, Response Files, and utility files.
goodbye	Terminate the current CLINFO session.
message	Record messages for the system manager.
response files	Create, edit, and execute lists of CLINFO commands for frequently used sequences of activities.
retrieve	Extract data from the SDF, from worksheets, and from a Communication File; place them into worksheets or Communication Files, or display them.
subset	Create sets of patient abbreviations by listing, from other sets, or by specifying conditions upon data in the SDF or in worksheets.
worksheet	Create, label, enter data into, modify, and discard worksheets.

Communication Between Activities. In order to move from one activity to another (e.g., from *worksheet* to *analyze)* or from one option to another within an activity, (e.g., from Display this Worksheet (option 1) to Label Worksheet Rows (option 4)), the user generally types commands that start with an exclamation mark (!). The commands for these moves are listed below. Commands may be typed in full or may be entered by using the yellow CLINFO key in conjunction with one of the labeled activity keys. Commands may be entered whenever the system is waiting for a response.

To go from:	To:	Type:
Any activity	Main menu of activity choices	*!done*
Any activity	Menu of options for activity α*	*!*α (e.g., *!ana*)
Activity α	Option number n of activity β	*!*β*,n* (e.g., *!ana,7*)
Anywhere in activity α	Option number n of activity α	*!n* (e.g., *!7*)
Option number n of activity α	The same option of the same activity	*!* or *CLINFO/ESC*
Menu of options for activity α	Option number n of activity α	*n* (e.g., 7) (*!* is not necessary)
Any activity	Terminate the session	*!goodbye*

* α and β represent CLINFO activity names, which may be abbreviated to their first three letters.

It is important to note that

- Some activities will prompt for obligatory information before they will let you move to another option or activity.
- Any activity name may be abbreviated to its first three letters, e.g., !analyze may be entered as !ana.
- Pressing CLINFO/ESC will interrupt the current activity and return you to the beginning of that activity; you need not wait for a prompt, as with !.
- A complete command may always be used in place of an abbreviated command. For example, to go from anywhere in one activity to option n of the same activity, *!*α*,n* and *!n* are equivalent, and to go from option n of one activity to the start of the same option of the same activity, *!*α*,n* and *!n* and *!* are all equivalent.
- Simultaneously pressing the yellow CLINFO key and a key with a yellow label (e.g., *!ana,*) is equivalent to typing the label. Thus, simultaneously pressing the CLINFO key and the *!ana,* key, then typing 7 produces the same results as typing *!ana,7.*

How to Find Essential Information About Your Study

In your use of the CLINFO system, you will frequently need to refer to lists of patients, variables, worksheets, and other information peculiar to your study. The activities that will produce the various lists are shown below. Keeping paper copies of your lists will prove very helpful, since you will not need to leave your current activity in order to recall what patients have been entered, what worksheets have been created, how item names are spelled, and so forth.

To List	Enter
Activity menu	*!done*
Events	*!describe,* 1 or *!describe,* 2*
Items	*!describe,* 1 or *!describe,* 2*
Panels	*!describe,* 1 or *!describe,* 2*
Patient abbreviations	*!subset,* 8 (with set "all")
Response Files	*!files,* 3
Subsets	*!subset,* 9 or *!files,* 2
Worksheets	*!work,* 9 or *!files,* 1**

Helpful Hints

Below are some suggestions about how to use CLINFO activities to accomplish frequently performed tasks. Additional suggestions appear in the *CLINFO User's Guide.*

• Create a new worksheet.	Enter *!work,0* and the name of the new worksheet, its title (i.e., a short description), and size.
• Move data within a worksheet so that is is lined up for analysis.	Enter *!ret,4* and name the worksheet as both the source and destination, then proceed.
• Leave a message for the system manager to merge your data, lock your schema, or provide other assistance.	Enter *!mes,* leave a message, and then proceed. Also, a request to merge can be issued by responding *yes* if prompted about this when logging off.
• Make a Response File readable.	While creating it, use the complete *!act-name,n* notation instead of the shorter *!n* notation. When editing it, follow a question mark with a space and an explanatory comment.

* *Describe,* option 2, produces a nicely formatted listing of the schema on the system printer; option 1 produces a more compact listing either on the system printer or on the terminal.

** *Worksheet,* option 9, generates a list of complete worksheet descriptions on the terminal; *files,* option 1, produces a much more compact listing with less detail but permits sorting by name, size, etc.

• Add, delete, or move entire rows or columns within a worksheet.	Enter *!work,11* and proceed.
• Find out if your Update File data have been merged.	When signing on or off or after entering *!ent*, see if an Update File exists.
• Print the entire current worksheet.	Enter *!work,10* and proceed.
• Print an entire Response File.	Enter *!res,3*, followed by the name of the Response File and *print*.
• Make a worksheet with standard labels.	Execute a Response File that creates a worksheet and labels it.
• Copy a worksheet.	Use *retrieve,5* to store the worksheet in a Communication File, then use *retrieve,6* to retrieve the same Communication File and store it in a new worksheet.
• Create a patient set that lists patients in the same order in which they appear in a sorted worksheet.	Use *subset,7* to form a new set that contains all the listed patients.

Size Limitations

Limitations on various CLINFO parameters are listed below:

Parameter	Maximum Value
Patients per study	392
Variables (items)	392
Panels	56
Events	28
Items/panel	30
Repetitions of panels	No practical limit
Worksheet size	From 1 row by 799 columns (or vice versa) to 47 rows by 47 columns
Text item length	70 characters
Patient identifiers	8 characters
Worksheet names	
Item names	
Panel names	
Subset names	
Character-string item values	

System Manager Functions

The system manager

- Instructs and assists users.
- Assigns passwords for new studies.
- "Locks" and "unlocks" schema.
- Merges the Update File into the SDF.
- Writes special BASIC programs.
- Transfers externally created machine-readable data to CLINFO.
- Sees to the maintenance of the CLINFO computer.
- Documents and reports to the system implementers about system error messages.
- Copies all user data to tape each day (i.e., does "backup").
- Installs new CLINFO software and informs users of changes.

System Errors

Occasionally the CLINFO system will make an error. Because, ideally, this should *never* occur, and because it is important for us to eliminate all such "bugs," we would appreciate it if you would inform your system manager about system errors and the circumstances under which they occur. When a detectable system error occurs, a message will appear on the top line of your CRT screen, and you will be prevented from proceeding normally. The message describes the error and says *TYPE quit TO QUIT.* At this point, push the PRINT key to copy the screen contents on the printer, call the system manager, explain what happened, and give him the hard copy of the screen contents. Typing *quit* will log you off and will allow you to initiate a new session. We will fix the error as soon as possible.

CLINFO ACTIVITIES

The CLINFO system is organized into several more or less independent activities. The major processing activities are *analyze, calculate, describe, enter, examine, response files, retrieve, subset,* and *worksheet;* the *files, goodbye,* and *message* activities perform simple auxiliary tasks. To provide easy access to specific functions, the major activities usually comprise a collection of options (e.g., *analyze* options include ANOVA and T-Test). Each of the major activities is described briefly below.

Analyze

Analyze is the set of data analysis subactivities that allow the investigator to perform mathematical and statistical analyses upon, and produce reports, plots, and graphs from, the worksheets. The analytical functions, such as t-test, chi-square, etc., are accessed via a selection process. The alternatives are displayed and the user chooses the specific type of analysis that he requires.

The analysis procedures assume that the data upon which they are to operate are contained in user-specified rows or columns of a worksheet. Certain procedures that produce output in tabular form will place that output into a worksheet if requested. For example, the output worksheet of the Linear Regression option

includes the residuals and predicted values which may be plotted using Scatter Plot.

The data analysis subactivities are called Analysis of Variance, Chi-Square Test, Cross Tabs, Descriptive Statistics, Frequency Distribution, Histogram, Life Table Analysis, Linear Regression, Non-Parametric Paired Tests, Normality Test, Radioimmunoassay, Scatter Plot, and T-Test. (Data sorting, an analysis-like function, is included in the *worksheet* activity as option 8, Sort This Worksheet; *calculate*, a separate activity, performs user-defined analyses.)

In all the data analysis subactivities, missing values are identified as such and do not enter into calculations.

Calculate

Calculate allows the user to specify and compute functions of data in worksheet rows and columns; results are stored in specified rows or columns of the *same* worksheet. The *calculate* options give the user a modest amount of programming capability. They are useful when the user wants to compute a function that is not provided by the *analyze* activity or when he wants the results stored in the same worksheet as the data.

Specifically, *calculate* provides for

- Defining and evaluating arbitrary arithmetic expressions (e.g., row 9 = row 2 + 5×log (row 8)) that refer to rows or columns of a worksheet.
- Defining and evaluating arbitrary logical expressions (involving ANDs or ORs) that refer to rows or columns of a worksheet.
- Special calculations including replacing missing values with a specified constant and vice versa, cumulative sum, and cumulative product.
- Computing time differences, in specified units, between dates and/or times stored in worksheet rows or columns.
- Statistical calculations including counts, sums, means, standard deviations, and standard errors of data in worksheet rows or columns.

It is often useful to add rows or columns to a worksheet (by using *worksheet*, option 11) before using *calculate*.

Describe

Describe allows the investigator to specify, in a standardized manner, the characteristics and the structure of data that he wishes to collect. This specification is called the study *schema*. The schema furnishes the flexibility required for entry and retrieval of data:

- It provides a means of associating data items into groups; for example, vital signs may be a group of items (corresponding to a CLINFO *panel*) whose values are measured at essentially the same time by the same person.
- It provides a dictionary which enables the investigator to describe, name, and specify the range or list of acceptable values of data items to be entered into the SDF.

- It allows the user to specify acceptable character-string (i.e., short word) responses and the numeric codes associated with them.
- It provides a means for describing to the system the relationship between events and the values of their associated items. An *event* is an optional signal within the CLINFO system that provides a point of reference to which other item values may be related in time. For example, an event may represent the administration of a drug, the first occurrence of a measured item, a specific value, or the maximum or minimum of the values entered for an item.
- It allows the user to personally create study files (viz., schema, SDF, and list of patients) for the purpose of testing his schema design by trial data entry and retrieval. If the user finds his schema design unsatisfactory and he has entered only a small amount of data, he may delete the patients and data entered for them, then revise the schema. If a substantial amount of data has been entered, only the system manager can revise the schema; he can make certain kinds of changes while still preserving the user's data.

The optimal design of a schema is critical to both the execution of the protocol which it represents and the effective use of the CLINFO prototype. The system manager will assist inexperienced users in the initial design and entry of a schema and will provide examples of existing well-designed schemata.

Enter

Enter is the data entry activity which allows the user to

- Add patients to the study.
- Enter data into the Update File (which is then merged into the SDF).
- Review the Update File.
- Transfer data from a worksheet to the Update File.

As data values are entered, *enter* performs three functions:

- It relates the values to the item descriptions in the schema and does the indicated screening for acceptability and encoding.
- When the sort/merge procedure (see below) is used, *enter* automatically reevaluates the times associated with events.
- It automatically associates with the data values additional identifiers, i.e., date and time of data collection, date and time of data entry and the identity of the data enterer.

One function of the schema is to prevent erroneous data values from entering the SDF. However, what seems to be an error may sometimes be an exceptional value. CLINFO will question a value that is out of the previously specified range; if the data enterer insists, the system will accept such a value where practicable but will flag it as exceptional and will record the enterer's initials. This inclusion of additional data identifiers and flags for exceptional values provides the investigator with an avenue for auditing his data.

In order to make both interactive data entry and retrieval flexible and efficient, and to implement the CLINFO prototype system on a small, economical computer,

it is necessary to process and reorganize the data after they have been entered. This processing is performed by the CLINFO system manager, using the *sort/merge* procedure, either on request or at a specified time of day. Although data may be reviewed immediately after they are entered, they may not be modified in, or retrieved from, the SDF until after this procedure is executed. The user who wishes to analyze new data immediately may enter the data into a worksheet (using the *worksheet* activity) and analyze them in that form.

Examine

Examine permits the user to display, modify, and delete data values stored in the SDF. Data are displayed by specifying the patient, panel name, and data collection time (i.e., the context). Successive instances of panels may be displayed without respecifying the context. Items whose values are changed are flagged. *Examine* also provides for determining how many patients and panel instances have been entered and how full the SDF is.

Response Files

If a sequence of responses is entered frequently with little variation, it is sometimes convenient to save these in a Response File, edit the file to allow for the variations, and later execute the file as appropriate. When executing a Response File, the CLINFO prototype behaves as if the stored responses were entered by the user. The user may watch as execution progresses and enter only those responses that generally change each time the Response File is executed.

In order to maintain generality and flexibility, the data entry portion of the *enter* activity is quite passive, i.e., the user must indicate when information about a new patient or panel are to be entered. If the data enterer wishes, however, a Response File can be set up that prompts for new patients, panels, and data items at the right points, and the enterer needs only to enter the data values.

The *response file* activity provides for

- Creating a Response File by example, i.e., by responding to the normal prompts and saving the responses.
- Editing a Response File, e.g., to indicate where the response is to be provided by the user instead of the file or to indicate that a sequence of prompts is to be repeated.
- Executing a Response File.
- Discarding a Response File.
- Making a copy of a Response File to create a variation of it.
- Appending one Response File to another.

Retrieve

Retrieve retrieves previously entered data from various sources (the SDF, worksheets, and Communication Files) for purposes of review, display, and analysis; it also stores those values into worksheets and Communication Files or displays them on a terminal.

Retrieve allows the investigator to describe which data he wants to retrieve from the SDF and how he wants them organized, by

- Specifying the patient(s) or patient set of interest.
- Specifying the item(s) or panel(s) for which values are to be retrieved.
- Specifying the time unit that is used for alignment (minute, hour, day, week, month, year, or a multiple of one of these units).
- Specifying the time interval, relative to an event or absolute date and time, that is to be used for data selection.
- Specifying logical conditions to restrict the values retrieved.

The investigator can use an event to single out a particular occurrence of a data value so that its date and time serve as a reference point for the data retrieval process. He may also restrict retrieved data values to those that satisfy specified conditions; in this way, he can discover which patients have data values that satisfy certain conditions and at the same time retrieve their data.

The time associated with an event is recomputed whenever the SDF is updated by the sort/merge procedure. Thus, if the time of an event is defined to be the data collection time of the *last* value of a variable, it always refers to the most recently measured value.

When *retrieve* is used to store data into or retrieve data from a Communication File, the user may specify that the Communication File belongs to a study different from the one he is currently accessing. This provides a convenient mechanism for moving data from one study to another.

Subset

Subset groups patients into *sets*, according to specified criteria. When the criteria are specified and the system creates a set by selecting the patients whose characteristics match the criteria, the user names the set and enters comments that describe it. A patient set is a list of patients; it does not include any data. Data for the patients in the set are retrieved by using the *retrieve* activity and referencing the set name. Also, patients may be selected from a set to create a new set. All patients on the study belong to the set called *all.*

A set of patients may be created from

- A list of patient abbreviations.
- Those belonging to all of a list of sets.
- Those belonging to at least one of a list of sets.
- Those in a named set whose data in the SDF satisfy all conditions in a list or at least one condition in a list.
- Those named in a worksheet and whose worksheet data satisfy all conditions in a list or at least one condition in a list.

A condition is satisfied when a specified item has a numerical value that lies within a specified range or has a character-string value that matches a specified string. When a set is created by examining data in the SDF, a single condition is satisfied if any value within a specified time interval meets the condition. Depending on the user's specifications, conditions must be satisfied either simultaneously or at any time within the same time interval. As with the *retrieve* activity, the time interval may be specified as the time between one time-marker (i.e., event or absolute date and time) and another, a list of time units relative to a time-marker, or a range of time units relative to a time-marker.

Worksheet

Worksheet is the activity used for creating, listing, discarding, labeling, displaying, printing, entering data into, editing, and sorting data in worksheets.

The Select subactivity is used to name, specify the size of, and create a worksheet. It is also used to select the *current worksheet* that is to be further processed by *worksheet*, *analyze*, *calculate*, or *retrieve* functions. (For convenience, the *analyze*, *calculate*, and *retrieve* activities also have a Select a Worksheet subactivity.)

Two labeling functions are used to label worksheet rows and columns with arbitrary 8-character strings. Labels may also be changed or deleted.

Worksheet data may be entered (or deleted) by row, by column, or by cell. When entering data by row, the user specifies the range of rows and columns for which he wants to enter data; then for each row he is prompted for values to be entered. Cell data entry is convenient for entering a few data values or for changing values; the user simply specifies a row, a column, and a value. Values may be entered as numbers, dates, or character strings. Dates may be entered in a format such as 5/22/75 or 7/4/1976.

The Display subactivity displays a specified 20-row by 6-column portion of the current worksheet on the terminal screen. Different portions are displayed by specifying different starting (i.e., upper left) rows and columns. Values representing dates are presented in a format such as 5/22/1975. Missing values are displayed as three dots (. . .). If so requested, the Display subactivity decodes encoded values and displays them as character strings.

The entire worksheet may be printed on the system printer by using the Print subactivity. This is particularly useful for larger worksheets.

The Sort subactivity sorts all the rows (or columns) of worksheet data by a single column (or row). The sort may be performed in either ascending or descending order.

The Edit subactivity is used to rearrange the contents of an existing worksheet; to move, copy, or delete rows (or columns); and to insert additional rows (or columns). It may also be used to change a worksheet's name and title.

Summary of Activity Options

The options within each major activity are briefly described in the following tables. The *examine*, *goodbye*, and *message* activities are excluded as they do not have options.

Activity Option	Name and Function
ANALYZE	
0	**Select a Worksheet**: Provides for specifying which existing worksheet is to be the *current* one, i.e., which contains the data to be analyzed; also provides for creating a new worksheet.
1	**Descriptive Statistics**: Computes counts, means, medians, ranges, standard deviations, and other statistics for data in specified portions of selected rows or columns of the current worksheet; optionally stores results in a new worksheet.

ANALYZE (continued)

2 **T-Test**: Tests hypotheses concerning the *means* of one or two populations by use of the Student t-statistic.

3 **Chi-Square Test**: Tests the hypothesis that two categorizing variables are independent, i.e., that there is no relationship between the distribution of values of one and the values of the other. The test is based on data representing *frequencies* of cases categorized on two variables.

4 **Linear Regression**: Estimates the linear relationship between one (dependent) variable (for which data are stored in a portion of a worksheet row or column) and up to three (independent) variables; provides for transforming the variables prior to the fit:

y=a+bx	sqrt(y)=a+bx
y=a+b/x	2nd degree polynomial
ln(y)=a+b ln(x)	3rd degree polynomial
ln(y)=a+bx	

Original and predicated values and residuals may be stored in a worksheet and may be plotted using *analyze*, option 7.

5 **ANOVA—Analysis of Variance**: One-way analysis of variance tests the hypothesis that from two to ten groups of samples are derived from the same population, i.e., have the same mean. Group numbers and values in the output worksheet may be plotted (with *analyze*, option 7) to create a scatter-plot type of histogram.

6 **Cross Tab**: Provides frequency counts of cases categorized on two variables, calculates the total frequency for each category, and calculates the means of each variable as categorized by the other. The user specifies the categories by entering cutpoints. The output worksheet may be used as input to *analyze*, option 3, Chi-Square Test.

7 **Scatter Plot**: Plots bar charts, scatter plots, and connected line drawings; uses data for one dependent variable and up to five independent variables stored in portions of worksheet rows or columns.

8 **Histogram**: Creates and displays a horizontal frequency histogram from data in a portion of a worksheet row or column; computes counts, frequencies, and cumulative frequencies for continuous data (with user-specified cutpoints) and for categorical data.

9 **Frequency Distribution**: Counts the number of occurrences of each distinct value in a portion of a specified worksheet row or column; displays the values, sorted in ascending value, and their associated frequencies. The output worksheet may be used to plot vertical histograms (with *analyze*, option 7) or to compute cumulative frequency distribution (with *calculate*, option 4).

ANALYZE (continued)

10 **Normality Test**: Uses the Kolmogorov-Smirnov statistic to test the hypothesis that a specified sample was drawn from a normally distributed population; produces an automatically scaled horizontal frequency histogram for the sample data (in a user-selected worksheet row or column).

11 **Non-parametric Tests**: Includes non-parametric tests (i.e., tests that do not assume normality) on variables which represent *pairs* of measurements made on the same cases:

- Spearman Rank Correlation Test—tests the hypothesis that the two variables are independent.
- Sign Test—tests the hypothesis that the *median* of the difference between pairs of measurements is zero. Useful when both observations of a pair are made under similar conditions but different pairs are observed under different conditions.
- Wilcoxon Signed Ranks Test—tests the hypothesis that the *median* of a group of observations is equal to some specified value or that the median of the differences between two groups is equal to zero.

Includes the non-parametric *non-paired* test:

- Wilcoxon Two-Sample Rank Sum Test—tests the hypothesis that the two samples are drawn from the same population.

12 **Life Table Analysis**: The current worksheet contains survival data (time on study and whether or not the endpoint of interest has occurred) for groups or individuals in one or two samples; computes for each time on study (1) the probability of survival to that time, (2) the expected lifetime at that time, and (3) the probability of arriving at the endpoint within the time interval; stores the results in a worksheet for possible plotting (with *analyze*, option 7) or other manipulation; compares the two life tables (if there are two samples) with a Wilcoxon test.

13 **Radioimmunoassay**: Estimates the concentration of ligand (i.e., the substance whose concentration is being determined by a laboratory radioimmunoassay) for a number of unknowns. Suboptions provide for the following:

1. Builds a tube layout and saves it in a worksheet.
2. Edits an existing tube layout.
3. Using an existing tube layout, fits a standard curve to the data, plots the fitted curve, estimates the unknowns, and saves the results in worksheet.
4. Prints an existing tube layout.

CALCULATE

0 **Select a Worksheet**: Provides for specifying which existing worksheet is to be the *current* one, i.e., which contains the data to be processed; also provides for creating a new worksheet.

CALCULATE (continued)

1 **Arithmetic Expression**: Provides for defining and evaluating arbitrary arithmetic expressions (e.g., Row3=(Row4+5.6)X Row 7/15) that refer to data in worksheet rows or columns; stores the results in specified rows or columns of the same worksheet.

2 **Logical Expressions (AND)**: Provides for defining and evaluating arbitrary logical expressions involving the AND operator (e.g., Set Row7=1 if Row3 is greater than 5.6 AND Row4 is less than 13.6) that refer to data in worksheet rows or columns; stores the results in specified rows or columns of the same worksheet.

3 **Logical Expressions (OR)**: Provides for defining and evaluating arbitrary logical expressions involving the OR operator (e.g., Set Row7=1 if Row3 is greater than 5.6 OR Row4 is less than 13.6) that refer to data in worksheet rows or columns; stores the results in specified rows or columns of the same worksheet.

4 **Special Calculations**: Suboptions provide for the following:

1. Replaces all missing values in a (selected portion of a) row or column with a specified constant.
2. Replaces each occurrence of a selected constant in a (specified portion of a) row or column with a missing value.
3. For each cell in a (specified portion of a) row (or column), computes the cumulative sum of row (or column) cell values up to that one; stores the results in corresponding cells of a specified row (or column). That is, for each k, computes $SUM(x_i)$ where i=j to k.
4. For each cell in a (specified portion of a) row (or column), computes the cumulative product of row (or column) cell values up to that one; stores the results in corresponding cells of a specified row (or column). That is, for each k, computes $PRODUCT(x_i)$ where i=j to k.

5 **Time Difference**: Calculates differences between two dates, two times, or two date-time pairs in rows or columns of a worksheet. Results, expressed in time units ranging from minutes to years, are stored in a row or column of the same worksheet.

6 **Statistical Calculations**: Calculates statistics for data in specified rows (or columns) of a worksheet and stores the result in the corresponding rows (or columns) of a specified column (or row) of the same worksheet. For example, if means are computed for columns 5 through 10, the results might be stored in columns 5 through 10 of row 20. The available statistics are:

1. Counts of non-missing values
2. Sums
3. Means
4. Standard Deviations
5. Standard Errors

DESCRIBE

1 **Edit the Schema**: Provides for constructing, editing, and reviewing the schema for a particular study.

2 **Print a Formatted Schema**: Produces a specially formatted schema listing on the system printer. This listing is easier to read than those produced using option 1.

3 **Create/Delete Study Files**: Creates study files (SDF and patient file) for trial data entry and retrieval after the user constructs a schema; after the user enters a modest amount of data, provides for deleting the study files in order to enable the user to revise the schema.

ENTER

0 **Add New Patient Abbreviations into the Study**: Provides for entering the patient abbreviations that will be used throughout the study. A patient can be referenced only after his abbreviation has been entered in this manner.

1 **Enter Data into the Update File**: In response to prompts, the user types patient abbreviations, data collection dates and times, and item values for direct entry into the Update File. Values are screened for validity as they are typed. The system manager must run *merge* to store the values in the SDF.

2 **Copy and Screen Data from a Worksheet into the Update File**: Provides for specifying the data in an existing worksheet that are to be copied into the Update File. Values are screened for validity as the system stores them in the Update File. The system manager must run *merge* to store the values in the SDF.

3 **Review Data in the Update File**: Provides for reviewing data values prior to running *merge*. Values may be reviewed by patient or in the entered sequence. Review may start with the data entered first (and go forward) or with the most recently entered data (and go backward). Entire panels may be deleted.

FILES

1 **List Worksheets**: Lists worksheet names and other characteristics ordered by

1. Filename, i.e., worksheet name
2. Size
3. Date and time last modified
4. Date of last access
5. Unordered

2 **List Subsets**: Lists Subsets in the orders shown above.

3 **List Response Files**: Lists Response Files in the orders shown above.

FILES (continued)

4 **List Utility Files**: Lists utility files, i.e., Communication Files, Patient File, Schema File, SDF, and Update File, in the orders shown above.

RESPONSE FILES

1 **Execute a Response File**: The system reads responses from the specified file, executes some as if entered by the user, and (where a question mark appears in the file) prompts the user for others.

2 **Create a Response File**: After entering this subactivity, the user proceeds to use the system in a normal manner. The system records his responses and stores them in a Response File.

3 **Edit a Response File**: Provides for modifying a Response File created with option 2 to add, delete, or change responses, to list the file, to indicate cycling (i.e., looping) within the Response File, or to change some responses to question mark (thereby indicating that the user should be prompted for them during execution).

4 **Discard a Response File**: Discards a Response File that is no longer useful.

5 **Copy a Response File**: Copies a Response File and names the copy. The copy can then be edited to create a variation of the original.

6 **Append a Response File**: Appends one Response File to another to create a longer one. This avoids the necessity of creating a large Response File (with option 2) in a single step.

RETRIEVE

0 **Select a Worksheet**: Provides for specifying which existing worksheet is to be the *current* one, i.e., which is to be used in options 4 and 5; also provides for creating a new worksheet.

1 **Event Display**: Displays information about events for individual patients. This information is the name, type, date, and time associated with the event and the name and value of the triggering item.

2 **Study Date File to Worksheet**: Provides for retrieving numeric data about a specific patient, item, or time from the SDF and for storing them in a new worksheet. A time interval (relative to an event or an absolute date and time) and a time unit (which determines the resolution of the retrieved data) may be specified. Retrieved values may be restricted to those that satisfy specified conditions.

3 **Study Data File to Terminal/Communication File**: Numeric and textual data for specified patients, panels, and times are displayed and/or are stored in a Communication File, a panel at

RETRIEVE (continued)

a time. All values are identified by patient, panel, time, and item name. The Communication File may be processed by a BASIC program, e.g., to produce a report or to transfer the data to another computer.

4 **A Worksheet to a Worksheet**: Copies a portion of a worksheet and stores the copied data and labels in specified locations of either the same worksheet or a different worksheet. This is useful for properly positioning data to be analyzed and for building a summary worksheet from several others.

5 **Worksheet to Communication File**: Copies the contents of a worksheet into a Communication File and associates that file with the same study, a different study, or an account in which BASIC programs can be written. Generally, a Communication File is processed by a specially written BASIC program; it may be transferred to another computer via magnetic tape. This option may be used together with option 6 to copy entire worksheets or to move worksheets from one study to another.

6 **Communication File to Worksheet**: Copies the contents of a properly formatted Communication File (in the same study, a different study, or an account in which BASIC programs can be written) into a new worksheet. The Communication File may be created by option 5 or by a BASIC program. It may contain data transferred from another computer.

SUBSET

1 **A List of Patient Abbreviations**: Provides for creating a patient set by naming its constituent patients; useful when specific inclusion criteria cannot be stated.

2 **Patients Belonging to All of a List of Sets**: Creates the set of patients who are members of *all* of one list of selected existing sets *and* who are not members of any of another list.

3 **Patients Belonging to at Least One of a List of Sets**: Creates a set of patients who are members of *any* one of a list of selected existing sets *or* who are not members of at least one set in another list.

4 **Patients Whose SDF Data Satisfy All Conditions in a List**: A time interval is specified and a set of conditions is placed on item values; SDF data within the time interval are examined and a patient becomes a member of the new set if these data satisfy *all* the conditions.

5 **Patients Whose SDF Data Satisfy Any of the Conditions in a List**: A time interval is specified and a set of conditions is placed on item values; SDF data within the time interval are examined and a patient becomes a member of the new set if *any* of the conditions is satisfied by the data.

SUBSET (continued)

6 **Patients Whose Worksheet Data Satisfy All Conditions in a List**: Creates a set of patients whose data in a selected worksheet satisfy *all* the stated conditions.

7 **Patients Whose Worksheet Data Satisfy at Least One Condition in a List**: Creates a set of patients whose data in a selected worksheet satisfy *any* of the stated conditions.

8 **List Patient Abbreviations from a Set**: Lists all the members of a selected set; all patients on the study are listed if the set called *all* is specified.

9 **List All of Your Set Names**: Lists the name, creation date, number of members, and description of every patient set associated with the study.

10 **Discard a Particular Set**: Discards a specified set that is no longer useful.

WORKSHEET

0 **Select a Worksheet**: Creates a new worksheet having the user-specified name, title (i.e., descriptive commentary), number of rows, and number of columns; also provides for specifying the *current* worksheet, i.e., the one that is to be subsequently processed.

1 **Display This Worksheet**: Displays up to 20 rows and 6 columns of the current worksheet, starting with a specified row and column. Numeric codes for categorical items described in the schema may be displayed as character strings.

2 **Label Worksheet Rows**: Provides for labeling rows with words up to 8 characters long. Existing labels may be changed. Labels may be blanked out by entering three dots (. . .).

3 **Label Worksheet Columns**: Provides for labeling columns in the same manner as labeling rows with option 2.

4 **Enter Data by Rows**: Provides for entering data, a row at a time, into a specified range of worksheet rows. There is no distinction between entering new values and replacing existing ones. Character strings may be entered for categorical items described in this schema. Dates are entered in the form 10/27/72. Missing values are entered (or existing values are deleted) by typing three dots (. . .).

5 **Enter Data by Columns**: Provides for entering columns of data into a worksheet. This option is otherwise similar to option 4.

6 **Enter Data by Cells**: Provides for entering data into specified cells of the current worksheet; useful when only a few values are to be entered or changed. This option is otherwise similar to option 4.

WORKSHEET (continued)

7	**Discard This Worksheet**: Discards a worksheet that is no longer useful.
8	**Sort This Worksheet**: Reorders worksheet rows by sorting the values in a column or reorders columns by sorting the values in a row. This is a very useful function because it can be used to group patients and their data according to values of one or more variables. *Analyze* and *calculate* subactivities can then be applied to the resultant individual groups of data.
9	**List Your Worksheets**: Lists the names, titles, created and last modified dates, and sizes of all worksheets.
10	**Print This Worksheet**: Prints the entire current worksheet on the system line printer.
11	**Edit This Worksheet**: Provides for changing the current worksheet's name and title and for adding, deleting, moving, and copying rows and columns within the current worksheet. This function is particularly useful for enlarging the worksheet so that more data can be added.

DATA ORGANIZATION AND FILES

Data stored by the CLINFO system are organized to facilitate the management and analysis of the clinical research study. The *study* is a single experiment, based on an approved protocol and carried out by one or more investigators. An investigator may have more than one study active on the CLINFO system, but each is completely separate and distinct. Even if the same subject (i.e., patient or normal volunteer) happens to be on two different studies, data relating to him must be entered separately for each study, and access to each study is obtained by a different study name. Thus any single interactive session, from log on to log off, is carried out in the context of a single study. Communication Files may be used to transfer data from one CLINFO study to another.

The CLINFO system stores data in two types of data file. Each study has a single, carefully protected, highly structured collection of data called the *Study Data File* (SDF). Data are entered into the SDF under control of the schema, and data to be examined or analyzed are retrieved from the SDF. CLINFO users may also create, modify, and discard files called *worksheets*, as required. Data may be entered directly into a worksheet or may be retrieved from the SDF or another worksheet. Data must be stored in a worksheet in order to be analyzed.

Study Data File

The SDF of a study is immediately accessible as long as the study is active; it is also copied onto a permanent medium each day to ensure against loss. The SDF may be thought of as having three dimensions: patient, item (or panel), and time of data collection. Measurements for any item or set of items may be repeated many times or only a few times, in which case the SDF may be large but sparse. Storage of data in an SDF permits great flexibility in data entry and retrieval.

Data are entered into the SDF by means of the *enter* activity, which utilizes the schema to prompt for item values and to check the data against user-specified ranges of values or codes and which stores the time of data collection along with each entry. Data are retrieved from the SDF by means of the *retrieve* activity.

Worksheets

Data may be extracted from the SDF into CLINFO worksheets in order to carry out analyses, review data, or produce reports, graphs, and tables. Data in worksheets may be copied from them and stored in the SDF. The combination of extracting data from the SDF into worksheets and copying worksheet data into the SDF provides for minimizing the number of worksheets that must be saved. It also provides for restructuring worksheets by entering worksheet data into the SDF, then retrieving these data from the SDF into new worksheets. For example, data from worksheets that apply to individual patients can be moved (through the SDF) into a worksheet that stores time-varying values of an item for a group of patients.

Worksheets are rectangular, tabular representations of study data, which are useful for the review of data and are required for data analysis. As worksheets are created by the user, they are given *names* by which they may be recalled, long *titles* which describe them, dimensions (numbers of rows and columns), and, optionally, row and column *labels*.

Although the data stored in worksheets are numeric, a decoding facility, which works in conjunction with the schema, permits the display of character-string values rather than the stored numeric codes.

Data may be retrieved from the SDF into a worksheet, or they may be entered in five other ways:

- By entering data directly from the terminal keyboard.
- By copying portions of one worksheet into portions of another.
- By moving and copying rows or columns within a worksheet.
- By storing results computed by *analyze* or *calculate* activities.
- By retrieving data that have been processed by a BASIC program and stored in a Communication File.

Worksheets are automatically saved by the system when they are created or modified, but they are not protected in the same way as the SDF; i.e., worksheets may be easily modified or deleted by the user. All of the data analysis programs expect the data to be in rows or columns of an existing worksheet. If a worksheet happens to be arranged so that each *row* represents a series of values of different variables for a single patient, then an analysis would probably be performed on individual *columns* because each would contain values of a single variable. In any case, the user informs the system whether rows or columns are to be analyzed and then designates, by number, the row or column to be analyzed.

In addition to the SDF and a collection of worksheets, a number of other files are used by CLINFO both for internal housekeeping and for direct use.

Schema File

A Schema File contains descriptions of the data to be collected for a particular study. This file is initially entered by the user and is referenced by the system for

all data entry, retrieval, and subsetting functions that involve the SDF, as well as for decoding numeric values for display.

Update File

This file acts as a temporary, protective storage area for data to be entered into the SDF. This temporary file allows the updating of the SDF, which may be time-consuming, to be carried out when it will not interfere with normal access to the system.

Patient File

The Patient File is a list of patients, in order of their entry into the study, containing the *patient abbreviation*, i.e., the identifier by which the patient is known to the system, and information about the location of the patient's data in the SDF.

Subset File

The Subset File contains lists of abbreviations of patients who belong to the patient sets that the user has defined (i.e., groups of patients who have particular characteristics or values of measured variables in common, where the variables and their values are specified by the user); this file also contains information that permits the system to locate the SDF data referring to patients in the sets.

Communication Files

These files act as buffers between the CLINFO prototype and non-CLINFO programs. They provide a link to the BASIC programming language while protecting the CLINFO processor and files from BASIC procedures. Communication Files permit data values to be transmitted from the SDF or from worksheets to BASIC programs; they can also be used to transmit data from non-CLINFO processes (e.g., data generated by BASIC programs) to CLINFO worksheets and from one study to another. In addition, a Communication File can receive data from and transmit data to a system device (e.g., a magnetic tape) that can be taken to another computer for processing.

Response Files

These files are executable lists of responses to CLINFO prompts. When a Response File is executed, the prototype behaves as if the responses were entered by the user. This is a convenient vehicle for storing and later executing a sequence of responses that is repeated often, perhaps with variations. The user creates Response Files when they are needed and discards them when they are no longer needed.

THE CLINFO USER'S TERMINAL

The user interacts with the CLINFO prototype through a CRT terminal with

an attached printer. The terminal can display up to 40 lines of 80 characters each on its TV-like screen. Information and commands are transmitted to the system by use of a typewriter-like keyboard, which is described in greater detail below. The small, attached DTC printer is used for copying the contents of the screen onto paper.

The Ann Arbor Terminal

Switch Settings. The power switch is located on the right-hand side of the AA terminal and is "on" in the forward position. The two switches on the rear panel of the terminal labeled S1 and S3 should be in the right-hand position when viewed from the rear, corresponding to *line* and *full duplex*, respectively. Switch S1 may be switched to LOCAL to modify the screen's contents before copying them onto the printer, but it must be switched back to LINE to use CLINFO. Other switches on the rear panel are inoperable. The system manager may adjust an internal switch to set the terminal's operating speed to 1200 baud or 4800 baud (i.e., to 120 or 480 characters per second).

The Keyboard. Throughout this report, the keys on the terminal are referred to by enclosing the appropriate symbol in braces; e.g., {a} refers to the lower-case "a" key.

In addition to the standard typewriter-like keys and controls, the AA keyboard has some keys and functions that are specific to the CLINFO prototype and some that are not used in this application:

- The keys labeled in yellow with abbreviated activity names are used in conjunction with the yellow CLINFO key to initiate the activities.
- The control characters above most of the alphabetic keys (such as DC4 over the letter T) are disregarded by the CLINFO system.
- The 10 numeric keys and the {.} on the right-hand side of the keyboard may be used instead of the upper row of numeric keys. They are not affected by the SHIFT key.
- The key used for exponentiation in BASIC is { ∧ }. This is displayed as a "hat" on the screen.
- The four keys containing arrows, and the HOME, TAB, and LINE FEED keys are disabled.
- The TTY LOCK key (which lights when activated) has the same function as the SHIFT LOCK, but only for alphabetic keys. All other keys transmit their lower-case character, e.g., {,} and {1}, when {TTY LOCK} is on.
- The yellow CLINFO key is used to enable the escape function; that is, to log onto the CLINFO system or to terminate an ongoing activity, the operator holds down the CLINFO key while pressing the ESC key. This arrangement prevents the operator from accidentally hitting {ESC} and aborting an ongoing process. When the CLINFO key and a key labeled in yellow with *!* or an abbreviated activity name are simultaneously pressed, this is equivalent to typing the label.
- The PRINT key above the 10-key numeric pad is used to command the attached printer to copy the contents of the CRT screen. Printing is also activated by simultaneously pressing the CLINFO key and the key labeled PRINT in yellow.

- The BRK key halts the print function and returns control to the user.
- The RETURN key is used to complete all user responses; depressing the RETURN key informs the system that the response is to be read and acted upon. In many of the system dialogues, the choice of responses includes simply striking RETURN, usually to indicate the most common or expected choice.
- The BS key backspaces and deletes successive characters from the screen. Unintentional or incorrect responses may be corrected by backspacing and thereby deleting characters with the BS key, then retyping the correct response, prior to pressing the RETURN key.

Repeat Function. The alphabetic, numeric, and BS, RETURN, and ESC keys produce repeated characters if held down more than about a second. This can cause unexpected behavior, because a series of RETURN, or other, characters may be stored by the computer if the keys are held down; the system will then respond to the stored characters as if they had just been entered by the user. If the system does not respond promptly to a RETURN, be patient; a second RETURN will not speed the response and will probably produce an undesirable effect on the following prompt when the system does respond. Instead, use the space bar or BS key to see if the system is alive.

Getting Ahead of the System. As mentioned above, the characters issued by the keyboard are stored and acted upon in sequence by the system. This feature can be used to your advantage, since familiar sequences of commands and responses may be entered more quickly than the system can act on them; thus you do not have to wait for a familiar prompt but may enter your response and wait for the system to catch up. On the other hand, this can be dangerous because the system may behave in an unexpected manner if an inappropriate or incorrect response has been entered. If you enter more characters than the system can handle, it will "beep" to indicate that it has missed a character; wait until the next prompt to see where you are. As mentioned above, sequences of commands may also be stored in a Response File.

The DTC Printer

A DTC printer, resembling a typewriter without a keyboard, is attached to the terminal. The information displayed on the CRT screen is typed on this printer when the PRINT key on the terminal is pressed. Thus, hard copies of worksheets, analytic results, plots, and lists of files are immediately available. Also, connected lines (which cannot be displayed on the CRT) can be printed on the printer by the Scatter Plot function. While the printer is in operation, the keyboard is disabled; however, pressing the BRK key will halt the printer and reactivate the keyboard.

The DTC printer is turned on by pressing its leftmost switch, i.e., the one labeled POWER. The LINE/LOCAL switch should be in LINE mode for normal use, and the PITCH rocker switch should be set at 10. The three switches labeled CR (carriage return), LF (line feed), and FF (form feed) will operate only with the LINE/LOCAL switch in LOCAL.

After paper has been loaded into the printer and aligned, the horizontal perforation should be brought about an inch above the top of the platen, then the RESET

switch should be pressed. This will establish the top-of-page position. To advance the paper to a new page, before pressing the PRINT key on the terminal, press the LINE/LOCAL key on the printer, then FF, then LINE/LOCAL again to return the printer to line mode. Occasionally, particularly when the printer is first turned on or after the BRK key has been pressed to stop printing, the printer will fail to operate properly. This is usually corrected by pressing the top of the printer's RESET switch. If that does not work, turn the terminal power off and on and try again. If that does not solve the problem, inform the system manager.

III. EXAMPLE OF A CLINFO SESSION

This section presents a simple but realistic example of the use of the CLINFO prototype. In this example, we reference the CLINFO activities employed at each step in the procedure but do not describe them in detail. Generally, CLINFO displays information on the CRT screen in upper- and lower-case or all upper-case letters; prompts (i.e., requests for information from the user) are in upper-case letters, followed by a colon (:). The user responds with lower-case letters.

In this example, the role of glucagon in the pathogenesis of diabetic ketoacidosis in man will be evaluated by investigating the effect of suppression of glucagon secretion by somatostatin. We shall examine changes in plasma glucose, glucagon, alanine, and lactate concentrations after acute withdrawal of insulin from 37 patients with juvenile-type diabetes. The example is derived from a study reported by Gerich et al. in "Prevention of Human Diabetic Ketoacidosis by Somatostatin—Evidence for an Essential Role of Glucagon," *New England Journal of Medicine,* Vol. 292, No. 19, May 8, 1975, pp. 985-989.

We assume that the investigator has defined the protocol for the study. It involves about 40 patients, with a small amount of identifying, clinical, and historical information recorded for each. In addition, each patient is subjected to a treatment situation and a control situation in which a series of measurements are made on hourly blood samples following acute withdrawal from insulin. The treatment series is measured while somatostatin is infused at a rate of 500 micrograms/hour; the control series is measured while a 0.9 percent saline solution is infused.

The investigator's data collection form might look like Fig. III-1. Each patient would have two such forms: one with history data and the laboratory data for the control situation, and the other with only the laboratory data from the treatment samples.

The investigator, who already has a personal password, obtains an identifier for this new study from the system manager and authorizes other members of his study team—fellows, technicians, and/or secretary—to access it. He then enters a schema into the system, using the *describe* activity. The schema is shown in Fig. III-2, as it is listed by the system's line printer. Names have been chosen for the panels and items, and limits and acceptable values have been specified for each item. It is assumed that each patient has several instances of panels called *control* and *treatment* because each panel instance corresponds to the measurements on a single blood sample; the instances are distinguished by their blood sample collection times. The notation is such that *glucosec* refers to glucose in the control situation and *glucoset* refers to glucose in the treatment situation. The schema has been "compiled" and the system has assigned panel and item numbers.

The schema is next "locked," with the assistance of the system manager, and at this point, data entry can begin.

After the patient abbreviations (i.e., *case 1* through *case 37)* have been entered for the 37 subjects, using option 0 of the *enter* activity, option 1 of *enter* is used to enter *History* panel data for each patient, as shown in Fig. III-3. The question marks and the items preceding colons are displayed by the system; the specifications or data following question marks or system-generated colons are entered by the user.

Somatostatin Study

Patient Name_____

Case No.:_____

Patient History Date:_____

Age:_____ Sex: M F

Percent Ideal Weight: _____

Duration of Diabetes: _____ (yrs.)

Insulin Requirements: _____ (U)

Lab Data Date _____ Control ☐ Treatment ☐

Sample Time											
Insulin Given											
Glucose											
Glucagon											
Alanine											
Lactate											

Fig. III-1— The investigator's data collection form.

If the experimental data are available, they are entered by specifying the correct context (i.e., patient, date, and time) and the appropriate panel. Data entry continues as shown in Fig. III-4. The data, from samples collected at approximately 6 A.M. and 7 A.M. on May 4, 1975, are entered into the panel called *control* for the patient called *case 1*.

After entering all the data for *case 1* and some of the data for *case 2*, the data enterer realizes that some of the interaction is unnecessary. That is, because data are already organized by patient in a lab book with history data followed by control and treatment data, and because both the control and treatment experiments result in a fixed number of blood samples that are all drawn on the same day, the flexibility that is normally provided by CLINFO's prompting is not required. Prompting can be bypassed by storing responses (such as "History," "date:?," "control," and " ") in a Response File that controls data entry. The Response File is produced by using option 2 of *response files* and then going through the steps of entering data for a patient (but hitting the RETURN key to enter " " instead of entering actual data values) to provide CLINFO with an example. The Response File that has thus been created by example is then modified somewhat, using option 3 of *response files*, and is executed by using option 1.

PANEL 1 History Patient's History

I#	Name	Description	Units	Type	Range/ # Codes/Text	Len	C
1	case no.	Case No.	*	num	1,20		
2	age	age	yr	num	18,35		
3	sex	Sex	*	char		code (2)	
		0 m					
		1 f					
4	%idl wt	% ideal weight	%	num	90,105		
5	dur diab	duration of diabetes	yr	num	5,25		
6	ins req	daily insulin reqmnt	U	num	25,60		

PANEL 2 control control data

I#	Name	Description	Units	Type	Range/ # Codes/Text	Len	C
7	insulinc	control insulin	U	num	25,60		
8	glucosec	control glucose	mg/100 ml	num	25,325		
9	glucgonc	control glucagon	pg/ml	num	0,200		
10	alaninec	control alanine	mM	num	.1,.8		
11	lactatec	control lactate	mM	num	1,2		

PANEL 3 treat treatment data

I#	Name	Description	Units	Type	Range/ # Codes/Text	Len	C
12	insulint	treatment insulin	U	num	25,60		
13	glucoset	treatment glucose	mg/100 ml	num	25,325		
14	glucgont	treatment glucagon	pg/ml	num	0,200		
15	alaninet	treatment alanine	mM	num	.1,.8		
16	lactatet	treatment lactate	mM	num	1,2		

EVENTS

E#	Name	Description	Item	Modifier	Trigger
1	instopc	ctrl insulin stopped	insulinc	last	
2	insstopt	treat insulin stoped	insulint	last	

Fig. III-2— The investigator's schema as listed by CLINFO after compilation. The panel, item, and event numbers have been assigned by the schema compiler.

```
PLEASE TYPE YOUR INITIALS: gfg
ADDING NEW DATA FOR

PATIENT ABBREV: case 1__

DATE OF SAMPLE: 5/3/75__

TIME OF SAMPLE: 800_


ENTER panel_name OR item_name:item_value OR context OR
      patient:patient_abbrev OR date:date_value OR time:time_value.

? History

case no.: 1

age     : 27

sex     : male
male     IS NOT AN ACCEPTABLE VALUE FOR sex

sex     : m

%idl wt : 100

dur diab: 12

ins req : 35

? date:5/4/75
ADDING NEW DATA FOR

PATIENT ABBREV({RETURN} FOR case 1  ): ________

DATE OF SAMPLE({RETURN} FOR 5/4/1975): ________

TIME OF SAMPLE({RETURN} FOR 0800): 615_

?
PLEASE SPECIFY panel_name OR item_name:item_value OR context OR
patient:patient_abbrev OR date:date_value OR time:time_value.

?
```

Fig. III-3—The dialog between the system and the user as new data are entered for the patient identified as *case 1*. After establishing the context, the user chooses the *History* panel and enters the 6 items that it contains. At the next prompt of ?, the user changes the context and is prompted to verify the entire context. He maintains the same patient and date but changes the time to 615, i.e., to 6:15 A.M.. The prompt at the bottom reminds the user of the acceptable responses to ?.

```
? control

insulinc: 35

glucosec: 103.5

glucgonc: 89.3

alaninec: 3.22
3.22 IS OUT OF RANGE.

IF OK, TYPE YOUR INITIALS, ELSE HIT {RETURN}:
VALUE NOT ENTERED.

alaninec: .322

lactatec: 1.42

? time:700
ADDING NEW DATA FOR

PATIENT ABBREV({RETURN} FOR case 1  ): ________

DATE OF SAMPLE({RETURN} FOR 5/4/1975): ________

TIME OF SAMPLE({RETURN} FOR 0700): ____

? control

insulinc: 35

glucosec: 98.7

glucgonc: [glucosec:97.8]

glucgonc: 90.2

alaninec:
```

Fig. III-4— A continuation of the dialog as the data for the items in the control panel are entered. An out-of-range value (3.22) is caught by the system and reentered correctly. The context is then changed by *time:700*, and another set of data is entered. An error is corrected by entering the item name and the correct value in brackets; the system then reprompts for the next item.

Data entry goes more quickly when the Response File is used because the data enterer does not need to enter the panel name or hit the RETURN key to verify the current patient, date, or time. After some or all of the data have been entered, the user, when logging off, indicates to the system manager that he would like his data merged. The system manager then runs the sort/merge procedure to store the entered data in the SDF.

Using option 2 of the *retrieve* activity, the user creates a worksheet called CLINICAL, shown in Fig. III-5, in which the rows represent patients and the columns represent variables. To define the retrieval, the investigator indicates the following:

- He wants to create a patients-by-items worksheet.
- The data are not time-dependent (because, in this case, he wishes to retrieve only data in the *History* panel).

```
WORKSHEET CLINICAL
TITLE Clinical Characteristics of Diabetic Patients
CREATED 12/17/75     MODIFIED 12/17/75
# OF ROWS  37
# OF COLS  5
```

ROWS/COLS	1	2	3	4	5
LABELS	age	sex	%idl wt	dur diab	ins req
1 case 1	27	m	100	12	35
2 case 2	24	m	98	11	45
3 case 3	32	f	98	15	45
4 case 4	34	m	94	20	50
5 case 5	21	f	96	9	40
6 case 6	23	f	98	10	30
7 case 7	28	f	99	15	45
8 case 8	19	m	92	5	27
9 case 9	35	f	103	20	57
10 case 10	32	f	101	7	43
11 case 11	20	f	95	5	37
12 case 12	33	m	105	18	55
13 case 13	25	m	98	7	40
14 case 14	23	f	95	5	37
15 case 15	27	f	97	9	40
16 case 16	27	m	92	13	50
17 case 17	33	m	91	18	48
18 case 18	30	m	96	7	42
19 case 19	18	f	100	5	26
20 case 20	35	f	102	23	35

Fig. III-5— The first 20 rows of the CLINICAL worksheet. This worksheet contains the data entered into the *History* panel of each patient. Note that the column labels are the names of the items in the *History* panel, the row labels are patient identifiers, and the "decode" option was used to display the characters *m* and *f* in column 2 in place of the stored codes *0* and *1*.

- Each row of the worksheet is to correspond to a member of a patient set.
- The particular set of interest is the one called *all*.
- Each column is to correspond to an item.
- The items of interest are *age, sex, %idl wt, dur diab,* and *ins req*.

The column labels of CLINICAL were automatically copied from the schema by the *retrieve* activity. Values of the item called *sex* are displayed as the characters *m* and *f*. The worksheet may be checked against the original data, and corrections entered by using the *examine* activity.

At this point, the user might further check the accuracy and reasonableness of his data by obtaining descriptive statistics, using option 1 of *analyze* (Fig. III-6), a frequency histogram of continuous data using option 8 of *analyze* (Fig. III-7), and perhaps a scatter plot of variables that he would expect to be consistently related, such as age and duration of diabetes (Fig. III-8). This latter method is excellent for detecting outliers which may represent erroneous data.

If a questionable data point appears, the user sorts the worksheet on the variable in question (a good technique for large worksheets). He may then determine which case is associated with the largest (or smallest) value of the variable, and perhaps check back to the original data.

```
COL DATA FROM WORKSHEET CLINICAL CREATED 12/17/75  MODIFIED 12/17/75

TITLE Clinical Characteristics of Diabetic Patients

COL# Label     N     Minimum      Maximum      Range        Mean        Median
=== ======== === =========== =========== =========== =========== ===========
  1 age       37      18           35          17         26.6486       27
  2 sex       37       0            1           1           .54054       1
  3 %idl wt   37      90          105          15         97.3243       98
  4 dur diab  37       5           25          20         12.4865       12
  5 ins req   37      25           60          35         39.5946       40

COL# Label      Std Dev       Sum       Sum Squares   Variance     Std Error
=== ========  ===========  ==========  ===========  ===========  ===========
  1 age          4.82026      986         27112        23.2349       .79245
  2 sex           .50523       20            20          .25526      .08306
  3 %idl wt      3.80872     3601       3.50987E+05     14.5064       .62615
  4 dur diab     5.41922      462          6826         29.3679       .89091
  5 ins req      9.2690      1465         61099         85.9149      1.52382
```

Fig. III-6— The output of option 1 of *analyze*, displaying descriptive statistics for the five variables in the worksheet called CLINICAL. Observing the maximum and minimum verifies the expected limits for each variable.

```
      Cut Pt.   Freq    %   Cum %        10        20        30        40
   ------------ ---- ----- ----- ----+----+----+----+----+----+----+----+
                   3   8.1 100.0 ***  |         |         |         |
             33  -                    |         |         |         |
                   6  16.2  91.9 ******|         |         |         |
   a         30  -                    |         |         |         |
   g               5  13.5  75.7 *****|         |         |         |
   e         27  -                    |         |         |         |
                  10  27.0  62.2 **********    |         |         |
             24  -                    |         |         |         |
                   7  18.9  35.1 *******|      |         |         |
             21  -                    |         |         |         |
                   5  13.5  16.2 *****|         |         |         |
             18  -                    |         |         |         |
                   1   2.7   2.7 *    |         |         |         |
                 --- -----      ----+----+----+----+----+----+----+----+
          Total:  37 100.0

    Max =     35          Min =     18      Mean =  26.649     S. D. =  4.8203
```

Fig. III-7— Histogram, option 8 of *analyze*. Knowledge of the sample distribution might avoid data analysis problems.

Successive use of option 2 of the *retrieve* activity also creates a series of patient "flowsheets" (i.e., one worksheet for each patient) such as that shown in Fig. III-9. To specify the retrieval of such a worksheet, the investigator states that

- He wants to create a times-by-item worksheet for a patient.
- The worksheet is for a particular patient (in this example, for *case 1).*
- The rows correspond to different times.
- The appropriate time unit is *hour.*
- The times are specified as a list (of hours) relative to an event.
- The relative hour numbers are -2, -1, *0, 1, 2, 4, 6, 8,* and *10.*
- The event that establishes the base time is *instopc.*
- Columns correspond to dates, times, and items in panels.
- The panel of interest is the one called *control.*

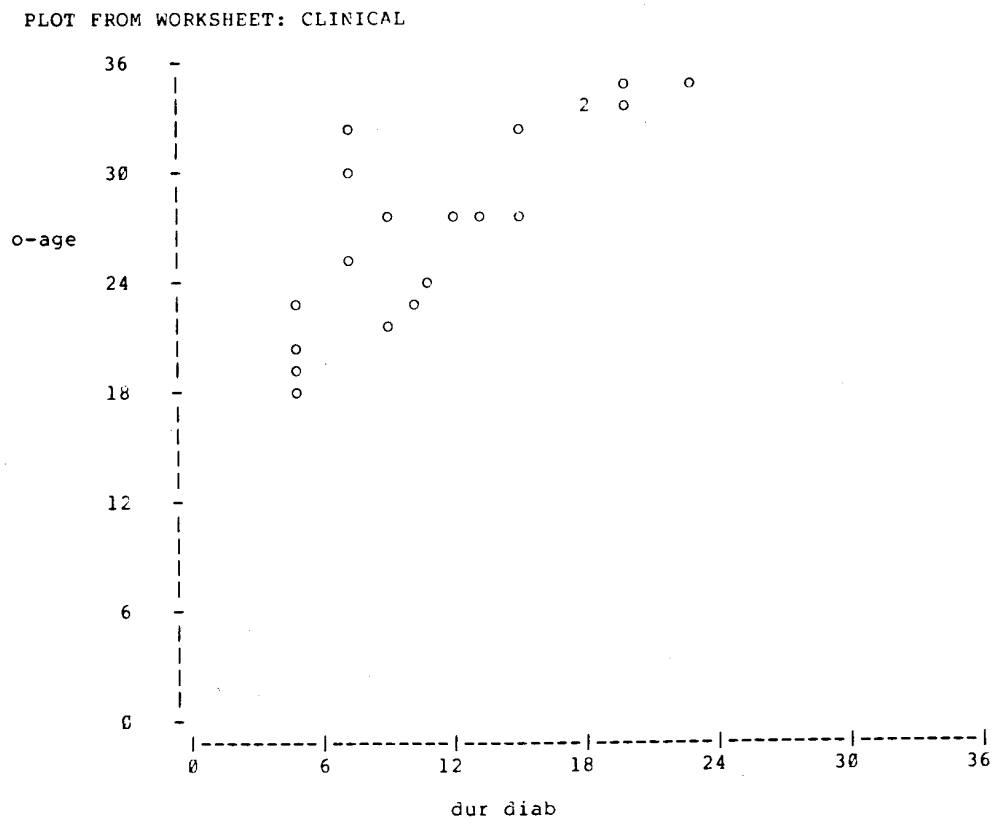

Fig. III-8— A scatter plot of age vs. duration of diabetes created using option 7 of *analyze*. Here the user would expect no points to lie below the diagonal, since age must be greater than duration of diabetes in each case.

```
WORKSHEET CASE1C
TITLE Flowsheet for case1, control data
CREATED 12/17/75     MODIFIED 12/17/75
# OF ROWS  9
# OF COLS  7
```

ROWS/COLS	1	2	3	4	5	6
LABELS	DATE	TIME	insulinc	glucosec	glucgonc	alaninec
1 HR -2	5/ 4/1975	615	35	103.500	89.300	.322
2 HR -1	5/ 4/1975	700	35	97.800	90.200	.324
3 HR 0	5/ 4/1975	810	35	96	90.500	.318
4 HR 1	5/ 4/1975	917	...	150.100	108.300	.298
5 HR 2	5/ 4/1975	1020	...	202.300	119.700	.291
6 HR 4	5/ 4/1975	1205	...	257.800	127.300	.307
7 HR 6	5/ 4/1975	1400	...	264.600	131.400	.335
8 HR 8	5/ 4/1975	1612	...	283.700	143.800	.378
9 HR 10	5/ 4/1975	1823	...	285.300	148.700	.418

(NOTE: ... indicates missing values)

Fig. III-9— The first six columns of worksheet CASE1C, created by retrieving *control* panel data for *case 1*, by *hour*, and aligning the times by use of the event *instopc*. A similar worksheet is prepared for each patient. The rows in the different worksheets then represent measurements taken at comparable relative times.

In the first resulting worksheet (or "flowsheet"), which the investigator titles CASE1C, the row labels denote times relative to the specified event, the first two columns list the dates and times that insulin was given and samples were collected for the patient, and the remaining five columns (the last of which does not appear on the display) itemize the data values. Note that *HR 0* corresponds to the last value of *insulinc* (insulin-control case); this is because the event *instopc*, as defined in the schema (see Fig. III-2), occurs at the same time as the last value of *instopc*. When the investigator similarly retrieves data for the other patients, he observes that the dates and times of sample collection are different for each one. However, because each worksheet has values for the same relative time stored in the same row (e.g., in every patient worksheet, row 5 corresponds to the second hour after insulin withdrawal), data values for all the patients can be easily time-aligned. Although the patient worksheets are convenient for examining and verifying collected data by patient, the investigator realizes that there is a simpler way to organize his data for purposes of analysis.

Using the *retrieve* activity again, the investigator constructs a worksheet for each *variable*, such as that shown for *glucosec* in Fig. III-10. Each row in the

```
WORKSHEET GLUCOSEC
TITLE Control Glucose by Patient and by Hour
CREATED 12/17/75     MODIFIED 12/17/75
# OF ROWS  37
# OF COLS  9
```

ROWS/COLS	1	2	3	4	5	6
LABELS	HR -2	HR -1	HR 0	HR 1	HR 2	HR 4
1 case 1	103.500	97.800	96	150.100	202.300	257.800
2 case 2	90.800	...	83.800	...	185.400	144.200
3 case 3	101.300	95.400	92.800	147.900	199.200	254.600
4 case 4	95.400	84.500	87.400	...	191.300	249.500
5 case 5	111.700	109.800	110.100	158.700	211.800	265.700
6 case 6	106.500	94.800	96.200	146.400	202.300	254.800
7 case 7	90.800	87.500	83.800	148.200	185.800	147.200
8 case 8	114.800	110.300	105.700	163.400	214.800	268.400
9 case 9	103.300	100.700	104.300	153.800	205.400	260.500
10 case 10	101.700	97.400	92.800	147.900	198.800	254.200
11 case 11	103.500	95.800	96	...	202.300	257.800
12 case 12	95.800	84.500	83.800	152	181.800	144.600
13 case 13	111.300	109.800	110.700	158.700	215.600	265.700
14 case 14	104.800	110.300	108.100	163.400	214.800	268.400
15 case 15	111.300	95.800	92.800	147.900	198.800	254.600
16 case 16	103.500	97.400	96.400	150.100	202.300	250.800
17 case 17	90.400	...	87.600	140.400	195.300	249.500
18 case 18	106.300	105.700	101.300	153.800	201.400	267.500
19 case 19	90.800	84.500	83.400	152.400	185.800	144.200
20 case 20	101.300	90.400	92.800	147.900	198.800	254.600

(NOTE: ... indicates missing values)

Fig. III-10— A portion of the GLUCOSEC worksheet. The first 20 rows contain all of the hourly *glucosec* (control glucose) values for each of the first 20 patients. Data for the other 17 patients are in the worksheet but are not shown. Also, columns 7, 8, and 9, which contain data for *HR6*, *HR8*, and *HR10*, are not shown. A similar worksheet is constructed for each of the experimental variables, using the *retrieve* activity. Note that the data are now aligned by (relative) time and that missing data are easily spotted.

worksheet GLUCOSEC contains data for a particular patient, and each column contains glucose data for a time relative to the event *instopc*. The data for all patients are now aligned by time, and instances of missing data may be easily observed (and the data entered if available).

Statistical computations may now be performed on the worksheets containing the time-aligned data. In our example, the investigator wishes to plot the mean value of each variable (across patients) against time, so he creates a worksheet that contains the mean values. Starting with a worksheet like GLUCOSEC, he first uses Edit This Worksheet (option 11 of *worksheet)* to enlarge it by one row (row 38). Next, using Statistical Calculations (option 6 of *calculate),* he computes the means of all columns and stores the results in row 38. Sums, standard deviations, and standard errors can be similarly computed and stored, whereas other statistics such as median and sum of squares can be computed using Descriptive Statistics (option 1 of *analyze)* but cannot be stored directly in the same worksheet.

Next, option 4 of *retrieve* is used to copy row 38, containing the mean values of *glucosec,* to row 1 of a new worksheet called MEANS (see Fig. III-11). This

```
retrieve,4  WORKSHEET TO WORKSHEET COPY

THE CURRENT SOURCE WORKSHEET IS:

WORKSHEET GLUCOSEC
TITLE Control Glucose by Patient and by Hour
CREATED 12/17/75    MODIFIED 12/17/75
# OF ROWS  38
# OF COLS  9

TO SELECT ANOTHER SOURCE WORKSHEET, TYPE ITS NAME, ELSE {RETURN}:

THERE IS NO CURRENT DESTINATION WORKSHEET, PLEASE SELECT ONE: means

THE CURRENT DESTINATION WORKSHEET IS:

WORKSHEET MEANS
TITLE Means of Control and Treatment Data by Hour and Variable
CREATED 12/17/75     MODIFIED 12/17/75
# OF ROWS  8
# OF COLS  13

TYPE c TO COPY COLUMNS OR r TO COPY ROWS: r

SOURCE ROW RANGE..a,b..: 38,38

SOURCE COL RANGE..a,b..: 1,9

TYPE yes TO TRANSPOSE ROWS INTO COLS, ELSE {RETURN}:

ROWS WILL BE COPIED INTO ROWS
COLS WILL BE COPIED INTO COLS

DESTINATION STARTING ROW: 1

DESTINATION STARTING COL: 1

   TYPE yes TO COPY ROW LABELS, ELSE {RETURN}:

   TYPE yes TO COPY COLUMN LABELS, ELSE {RETURN}: yes
```

Fig. III-11— The dialog for using *retrieve,* option 4, to copy a row (38) from the GLUCOSEC worksheet, which contains means of *glucosec* for each hour, into the first row of the MEANS worksheet. Row 38 from each of seven successive source worksheets is copied into an appropriate row (2 through 8) of the same destination worksheet.

process is repeated for each of the variables (that is, row 38 computed in each of the worksheets retrieved for the individual variables is copied into successive rows of the MEANS worksheet).

Response Files can be used to automate this repetitive procedure. Starting with Create a Response File (option 2 of *response files),* the user goes through the above steps—retrieving a worksheet for a variable, enlarging the worksheet by one row, computing and storing the means of all columns, and copying the new row into the MEANS worksheet. Next, using Edit a Response File (option 3 of *response files),* he changes each of the stored responses that changes from variable to variable (namely the item name, the name of the retrieved worksheet, and the destination row in MEANS) to "?" to indicate which responses are to be entered from the terminal (instead of being read from the Response File). Figure III-12 shows the

```
RESPONSE FILE: figlØ        THIS FILE DOES NOT CYCLE

 LINE #     RESPONSE (THIS FILE CONTAINS 41 RESPONSES)

   1.ØØ     !ret,2
   2.ØØ     2
   3.ØØ     glucosec
   4.ØØ     2
   5.ØØ     all
   6.ØØ     2
   7.ØØ     3
   8.ØØ     -2
   9.ØØ     1Ø
  1Ø.ØØ     instopc
  11.ØØ     yes
  12.ØØ     4
  13.ØØ
  14.ØØ     glucosec
  15.ØØ     Control Glucose by Patient and by Hour
  16.ØØ     yes
  17.ØØ
  18.ØØ
  19.ØØ     !wor,11
  2Ø.ØØ     r
  21.ØØ     add 38
  22.ØØ     !calc,6
  23.ØØ     r
  24.ØØ     3
  25.ØØ
  26.ØØ     #38
  27.ØØ     means
  28.ØØ
  29.ØØ
  3Ø.ØØ     !ret,4
  31.ØØ
  32.ØØ     means
  33.ØØ     r
  34.ØØ     38,38
  35.ØØ     1,9
  36.ØØ
  37.ØØ     1
  38.ØØ     1
  39.ØØ
  4Ø.ØØ     yes
  41.ØØ     !res,3
```

Fig. III-12— A Response File used in retrieving values of particular variables, computing the means across patients, and storing the results in the MEANS worksheet. To apply to more than one variable, the changes indicated in the text must be made.

Response File for our example. Lines 31 through 40 contain the responses in Fig. III-11. To generalize, Response File lines 3, 14, 15, 27, and 37 should be changed to "?"; line 41 should be deleted, and the Response File should cycle back to line 1 to repeat the procedure for the next variable.

The first 6 columns of the MEANS worksheet are shown in Fig. III-13. Scatter plots of variable versus variable and variable versus time may be constructed with this worksheet.

To examine the time variation of mean variable values, the investigator uses Scatter Plot (option 7 of *analyze)* to create four plots, one of which is shown in Fig. III-14. Row 1 (control values of glucose) and row 5 (treatment values of glucose) of the MEANS worksheet are plotted against the column labels (time in hours from insulin withdrawal).

```
WORKSHEET MEANS
TITLE Means of Control and Treatment Data by Hour and Variable
CREATED 12/17/75    MODIFIED 12/17/75
# OF ROWS  8
# OF COLS  13
```

ROWS/COLS	1	2	3	4	5	6
LABELS	HR -2	HR -1	HR 0	HR 1	HR 2	HR 4
1 glucosec	103.400	99.750	97.186	152.383	201.457	242.957
2 glucgonc	89.104	90.235	90.613	108.425	119.680	127.324
3 alaninec	.322	.325	.320	.298	.294	.310
4 lactatec	1.418	1.382	1.503	1.455	1.186	1.271
5 glucoset	112.327	107.800	106.905	75.245	69.400	77.820
6 glucgont	98.704	109.753	110.105	68.300	45.810	46.302
7 alaninet	.373	.361	.337	...	.354	.406
8 lactatet	1.318	1.434	1.369	...	1.260	1.184

(NOTE: ... indicates missing values)

Fig. III-13— A portion of the destination worksheet, MEANS, showing the first 6 columns after data have been copied into it row by row. This worksheet is now used as input to the Scatter Plot subactivity to create plots of each variable versus time, using the column labels as values of the time variable.

To make his plots easier to interpret, the investigator indicates that he would like the printer attached to his CRT terminal to draw lines between the data points; the terminal itself cannot do this—Fig. III-15 was produced in this way by a printer. Apparently the somatostatin treatment changes plasma glucose level.

The investigator is also interested in the relationships between variables in both the treatment and control experiments following withdrawal of insulin. To explore these, he uses the Scatter Plot subactivity to plot one variable in columns 4 through 13 of MEANS against another. Figure III-16 illustrates the relation between mean plasma glucose and alanine levels after withdrawal of insulin in the control experiment; Fig. III-17 shows a similar relation in the treatment experiment. These figures demonstrate that at any given alanine level there was greater hyperglycemia when glucagon secretion was unrestrained by somatostatin.

Finally, to arrive at estimated mathematical descriptions of the relations illustrated in Figs. III-16 and III-17, the investigator uses Linear Regression (option 4 of *analyze).* The results for the regression of mean glucose (from the MEANS

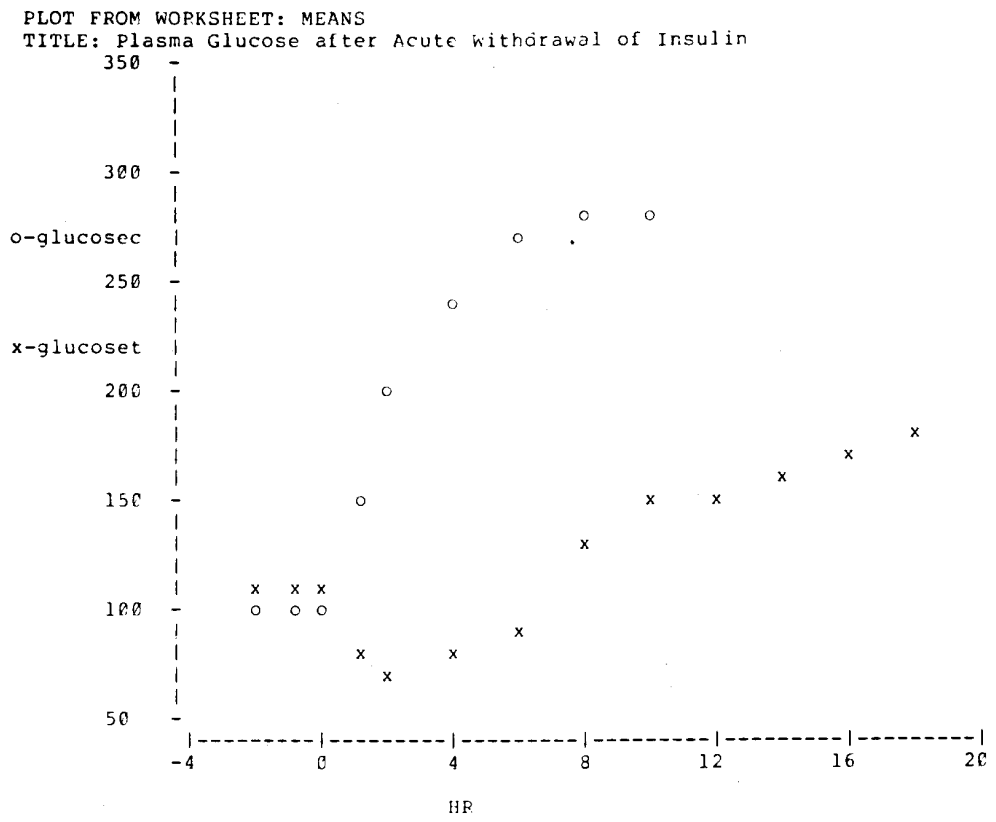

Fig. III-14— The scatter plot using column labels of the MEANS worksheet to define the x-axis and rows 1 (glucosec) and 5 (glucoset) as the y-axis variables.

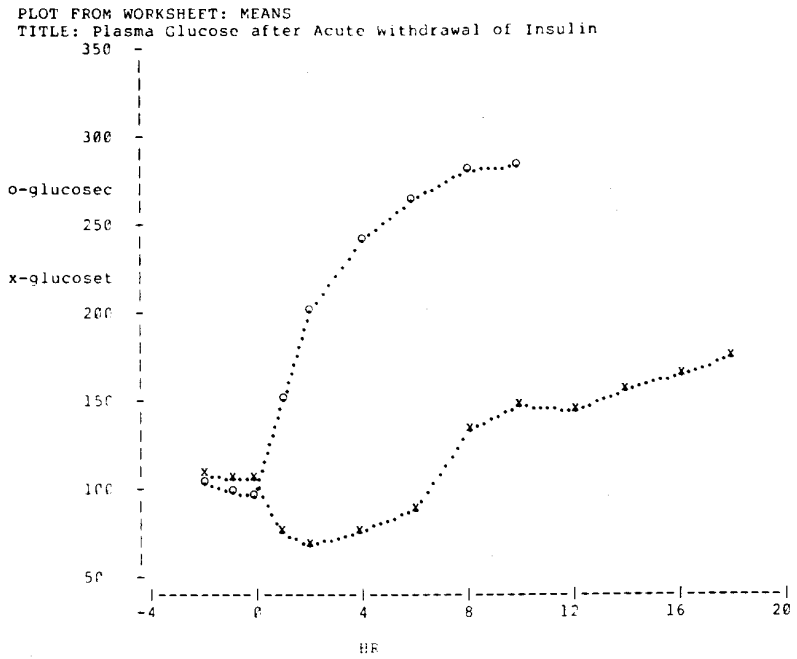

Fig. III-15— The scatter plot (similar to Fig. III-14, but with dotted lines drawn by the printer).

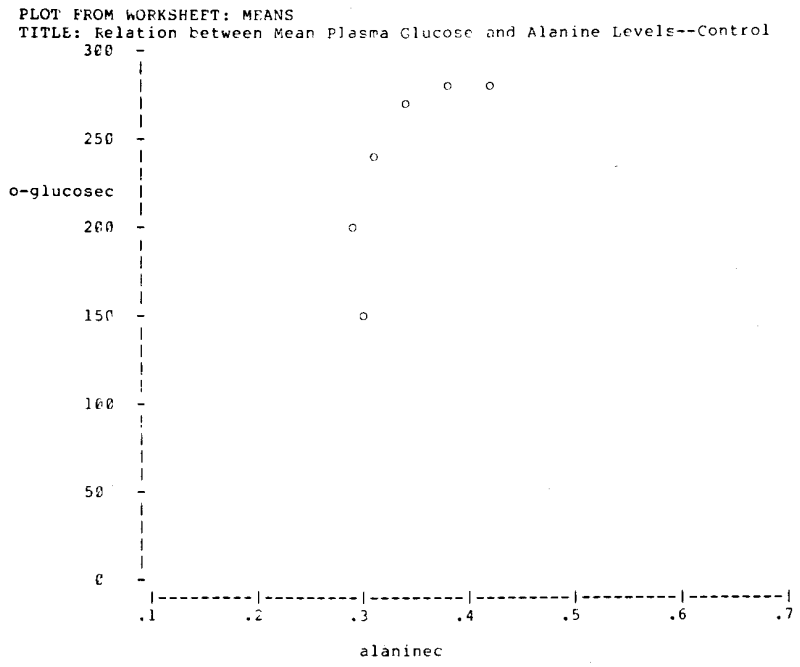

Fig. III-16—The scatter plot using row 3 (alaninec) as the x-axis variable and row 1 (glucosec) as the y-axis variable.

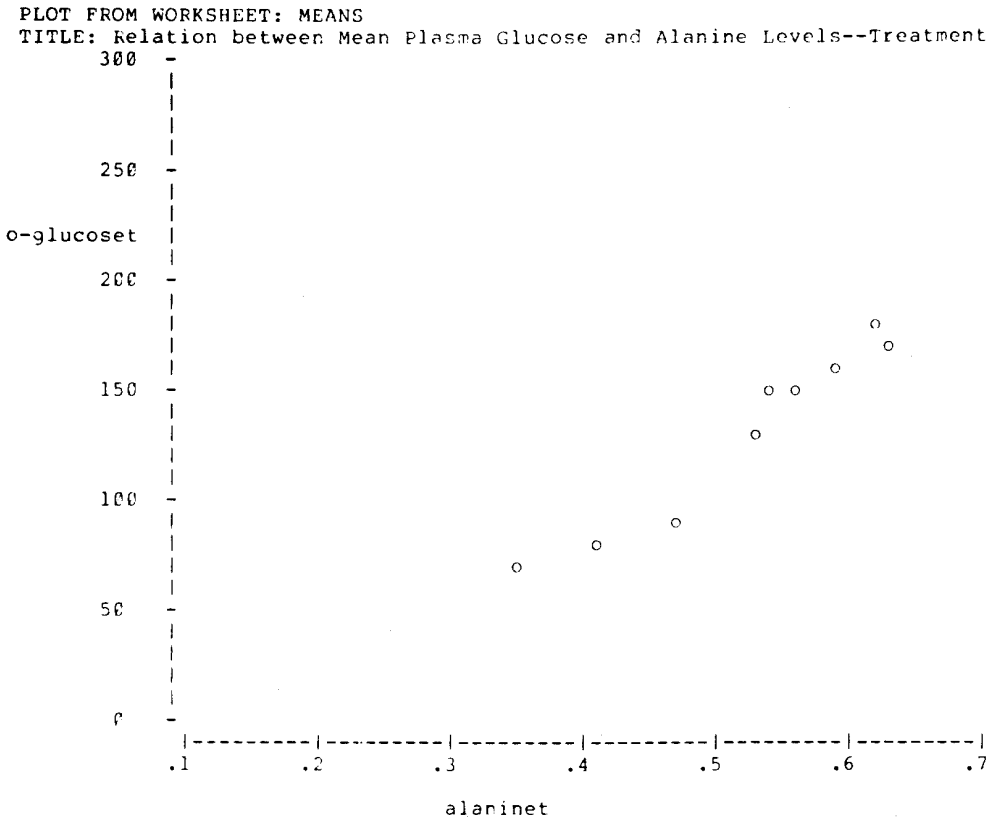

Fig. III-17—The scatter plot using row 7 (alaninet) as the x-axis variable and row 5 (glucoset) as the y-axis variable.

worksheet) upon mean alanine in the treatment experiment are shown in Fig. III-18. The estimated slope and intercept are 408 and −84, respectively; the coefficient of correlation is 0.97, with a significance level of .005. For the control experiment, the estimated slope and intercept are 803 and −35, respectively; the coefficient of correlation is 0.79, with a significance level of .06. Finally, the regression line computed in Fig. III-18 is plotted in Fig. III-19.

The CLINFO prototype has several important features that are not illustrated by this example. In addition to entering data as shown in Figs. III-3 and III-4, a user may enter data directly into a worksheet and may transfer (using option 2 of *enter* and sort/merge) worksheet data (that have been entered or computed) into the SDF. Data in an Update File (where they are stored until the system manager runs sort/merge) may be reviewed using option 3 of *enter;* data in the SDF (where they are stored permanently) may be examined and changed using the *examine* activity. Sets of patients with common characteristics (e.g., all females older than 25, or all patients in a particular set who have *glucosec* greater than 200 between 1 and 4 hours after insulin is stopped) may be created by using one of the *subset* options. Finally, there are a number of statistical analyses available (as *analyze* options) that are not illustrated here, and special analyses and reports may be produced by writing computer programs in the BASIC language and transferring data in Communication Files (using options 3, 5, and 6 of *retrieve).*

```
 analyze,4  LINEAR REGRESSION

ROW DATA FROM WORKSHEET MEANS       **** COLUMN RANGE IS FROM 4 TO 13 ****

TITLE Means of Control and Treatment Data by Hour and Variable
CREATED 12/17/75     MODIFIED 12/17/75


Regression of ROW 5 (glucoset)  upon ROW 7 (alaninet) ;  Sample size =  9

Regression Equation ( Y=A+B*X ) :glucoset= -84.186 +  408.09 * alaninet

Standard Error of Estimate (Sy.x):  9.7378  ( = sqrt( 663.775 / 7 ) )

               Std. Error      95% Conf. Interval
               ----------      --------------------
Intercept (A):   3.2459        -91.863  -76.510
    Slope (B):   36.132          322.64   493.55


               Mean       Std. Dev.(Population Est.)
             --------     --------------------------
glucoset     129.47          39.937

alaninet     .52356          .09528

       Coefficient of Correlation:

    ------------------------------------
     R       R-SQ     F-RATIO   Sig. Level
    ---      ----     -------   ----------
   0.974     0.948    127.60        .005

95% Confidence interval for R:       0.881     0.994
                           R-Sq:     0.776     0.989
 ENTER A VALUE FOR alaninet TO ESTIMATE glucoset; ELSE HIT {RETURN}:
```

Fig. III-18—The results of using the Linear Regression subactivity to compute the regression of columns 4 to 13 of row 5 (glucoset) in the MEANS worksheet upon row 7 (alaninet).

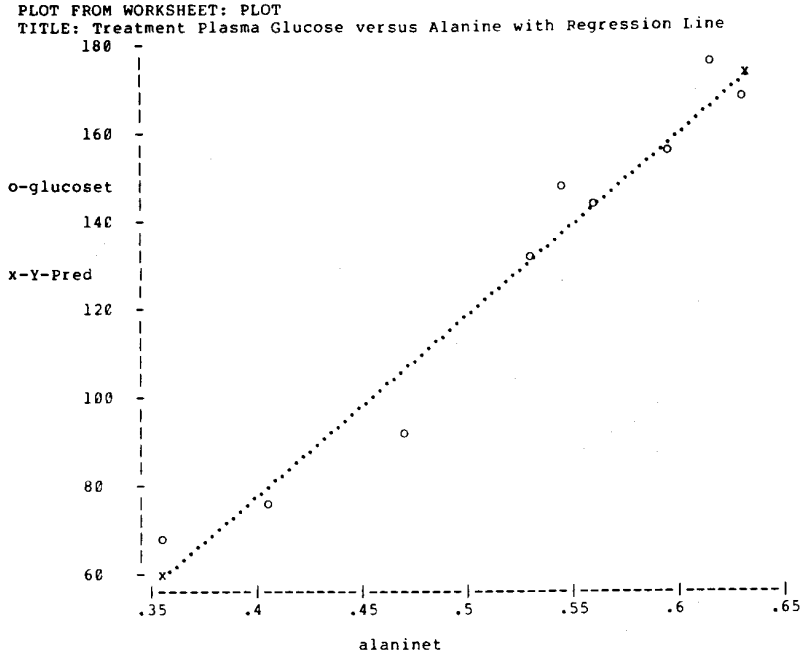

Fig. III-19—The regression line, together with the data plotted in Fig. III-17.

NEXT ACTIVITY CHOICE

!analyze. . . . . Analyze data
!calculate . . . . . Do row/col calculations
!describe . . . . . Edit the schema
!enter . . . . . Enter study data
!examine . . . . . Examine/edit study data
!files. . . . . Examine file utilization
!goodbye . . . . . Terminate this session
!message . . . . . Record messages
!response . . . . . Create/use a Response File
!subset . . . . . Subset patients
!worksheet . . . . . Manipulate worksheets

ANALYZE

0 . . . Select a worksheet
1 . . . Descriptive statistics
2 . . . T-test
3 . . . Chi-square test
4 . . . Linear regression
5 . . . Anova
6 . . . Cross tabs
7 . . . Scatter plot
8 . . . Histogram
9 . . . Frequency distribution
10 . . . Normality test
11 . . . Non-parametric tests
12 . . . Life table analysis
13 . . . Radioimmunoassay

CALCULATE

0 . . . Select a worksheet
1 . . . Arithmetic expression
2 . . . Logical expressions (AND)
3 . . . Logical expressions (OR)
4 . . . Special calculations
5 . . . Time difference
6 . . . Statistical calculations

DESCRIBE

1 . . . Edit the schema
2 . . . Print a formatted schema
3 . . . Create/delete study files

ENTER

0 . . . Add new patient abbreviations into the study
1 . . . Enter data into the update file
2 . . . Copy and screen data from a worksheet into the update file
3 . . . Review data in the update file

FILES

Display names of

1 . . . Worksheets
2 . . . Subsets
3 . . . Response Files
4 . . . Utility Files

RESPONSE FILE

1 . . . Execute a Response File
2 . . . Create a Response File
3 . . . Edit a Response File
4 . . . Discard a Response File
5 . . . Copy a Response File
6 . . . Append a Response File

RETRIEVE

0 . . . Select a worksheet

Copy data from one file to another

1 . . .	Study Data File	to	Event Display
2 . . .	Study Data File	to	Worksheet
3 . . .	Study Data File	to	Terminal/Communication File
4 . . .	Worksheet	to	Worksheet
5 . . .	Worksheet	to	Communication File
6 . . .	Communication File	to	Worksheet

SUBSET

Create a set of patients

From:

Those belonging to

1 . . . a list of patient abbreviations
2 . . . all of a list of sets
3 . . . at least one of a list of sets

Those whose study data satisfy

4 . . . all conditions in a list
5 . . . at least one condition in a list

Those whose worksheet data satisfy

6 . . . all conditions in a list
7 . . . at least one condition in a list

List or discard

8 . . . List patient abbreviations from a set
9 . . . List all of your set names
10 . . . Discard a particular set

WORKSHEET

0 . . . Select a worksheet
1 . . . Display this worksheet

Labeling

2 . . . by rows
3 . . . by columns

Data entry

4 . . . by rows
5 . . . by columns
6 . . . by cell

7 . . . Deletion of worksheet
8 . . . Sorting worksheet
9 . . . Listing worksheets
10 . . . Printing worksheet
11 . . . Editing worksheet

RAND/R-1541-NIH