INFORMATION PROCESSING OF CANCER CHEMOTHERAPY DATA

Alice R. Holmes and Robert K. Ausman
Health Research, Incorporated
Roswell Park Memorial Institute
Buffalo, New York

INTRODUCTION

Roswell Park Memorial Institute, one of the largest cancer research and treatment hospitals in the world, conducts an extensive program of cancer chemotherapy, the use of chemical agents in the treatment of cancer. Roswell Park, acting as the statistical center for participating hospitals throughout the United States, is responsible for implementation, control, and follow-up of several chemotherapy studies. To facilitate the handling of the diverse and extensive amount of data made available by the participants, the architects of these chemotherapy studies found it desirable and necessary to automate as many facets as was feasible. This paper describes the procedures employed in a system which extensively treats medical data.

GENERAL PRINCIPLES

For each study the participating hospitals throughout the United States follow the same procedures, thus insuring essentially identical treatment of all patients. This procedure guarantees uniformity which facilitates statistical analyses. Prior to the initiation of any given study, guidelines are determined and a protocol to be followed is established, stipulating qualifications a patient must meet prior to entry in the study such as age, medical history, blood picture, and previous therapy. Also specified is the conduct of the study, the method of drug administration, dosage regimen, frequency of blood counts, and follow-up policies. Investigators are expected to conform as closely as possible to the protocol.

In conjunction with the writing of the protocol is the design of the forms that enables the physician to collect the information requested. The proper design of these forms is extremely important, for it determines the quantity and order of the material to be stored in the computer for later use in several types of statistical analyses. To facilitate the handling of the data collected, most questions are worded for an objective answer. An important aspect of the system is the return of these forms to the statistical unit as soon as the pertinent patient information is available to the physician. In this manner there is continuing assurance that the protocol is being followed correctly. In addition, the immediate response allows for current reports on the progress of the study.

A patient is entered in the study by a referring physician who telephones the statistical unit. At
this point, he receives information as to the specific drug, dosage, and frequency of administration. Subsequently, the physician receives written confirmation of the entry and procedures of therapy.

During the exchange of information between Roswell Park and the participants there are several checks on the completeness and accuracy of the protocol procedures. If the forms are received as incomplete, inaccurate, or if they contain discrepancies, physicians from the statistical center staff note these errors. The corrections are indicated in memos which are prepared by a separate semiautomated system and recorded on the IBM Magnetic Tape Selectric Typewriter and sent to the participants. If corrections have not been received at the close of a 30-day period, names of delinquents are run through the Magnetic Tape Selectric Typewriter, which prepares a reminder form.

Results of the 3-month follow-up examinations, perhaps the most important phase of the studies, are submitted to determine the effect, if any, of the chemotherapy. In this manner it is possible to tell if there has been a recurrence of cancer, the time lapse between therapy and recurrence, or most significantly, if there has been no recurrence at all.

The methods of handling the data received, previous to automation, were time-consuming, clumsy, and inaccurate. The information on each patient chart was transferred manually onto three separate code cards. Because of the limited available space on the cards, only a summary of the data could be coded. For example, the total amount of drug given was recorded, but the individual doses and dates of administration were eliminated. In addition, only the lowest blood count was punched in the card, eliminating the daily record. This system made it impossible to analyze the effectiveness of the drugs accurately and completely.

As the number of studies grew, the staff became aware of the absolute necessity for a method which would allow more intricate and sophisticated recording and analysis of patient information. A means was devised by which a greater amount of data could be recorded with less manual labor. With the introduction of an automated system, initiated in 1963 and utilizing an IBM 1401, the chemotherapy studies have produced more reliable information, stricter adherence to the protocol, and less misinterpretation of data.

METHODS AND PROCEDURES

The individual patient charts were chosen as the initial input documents since they afforded the most logical and practical method of handling raw data. Because of the format of the charts, it was determined that most of the data could be punched directly from the charts with an absolute minimum amount of coding.

The first stage of automation included the design of the punched card layout. Since the cards are punched directly from the patient forms, the card layout follows these forms as closely as possible. The number of columns to be allowed for an item provides for the highest possible value of the particular data. Two major controls incorporated on each punched card are the card type number and patient study number. The card type number indicates a particular card format and the information recorded on that card. As an example, the patient's former medical history may be recorded on one card, while daily observations may be found on another. The unique number assigned to each card reveals at a glance which category of information is contained on the card.

The patient study number, different for each patient, is punched on all cards for that patient. In the event that one card is separated from the others belonging to a patient, the information is not incorporated incorrectly with the data recorded for another patient.

The assigning of codes to that data which have not been written previously in numerical form must be established in conjunction with the design of the card layout. Examples are the indication of sex, to be coded with a 1 or 2, or any type of yes or no question which can be coded with a 1 for yes and 2 for no. This translation can be accomplished easily by keypunchers.

Magnetic tape is employed as the storage device for the records. The tape format can be designed in several different ways. Each tape record may be made identical to the initial input card, resulting in a card image on magnetic tape. In this case, card type and study number are retained on each tape record, insuring proper identity. The second method of formatting the tape is to combine several input cards into one tape record. Utilizing this approach, type and study number appear only one time. A third way is to combine all cards for a patient into one record; unless there is a fixed number of rec-
ords for every patient, this method is not advisable.

The present application creates a tape record for every input card. The number of records per patient may vary from 5 to as many as 100, depending on the number of daily observations recorded for a patient. If there were one record per patient, the record size would range from 400 characters to 8000 which would be cumbersome to program and analyze.

SYSTEM ORGANIZATION

The first program in the system writes the data cards on tape. A standard card-to-tape routine is used providing there are no changes or modifications in the data at the time it is written on tape.

When the tape is not formatted in the same manner as the cards, it is converted to the proper layout. This conversion may include combining two card records into one tape record, or possibly rearranging the data so that it can be manipulated more effectively. What may be efficient for keypunching may not be efficient when working with the massive file of patient records on tape.

Once the tape has been created and sorted, a complete edit is performed by the computer in order to check all the data for accuracy, completeness, and logic or validity. If the individual who reviewed the patient form overlooked a discrepancy, or the data was keypunched erroneously, the program detects and prints the error.

There are several different types of checks. Most data are screened for validity. For example, months must be numbered 1-12, days 1-31. Some data can be coded only with certain codes (male = 1, female = 2). A code of 3 for sex would indicate an error.

A portion of the data is recorded on several different records as they occur; the dates are cross-checked in each location to insure compatibility. A case in point is the date of surgery which appears on three types of records. These dates must agree since there is only one possible date of surgery in the study. Calculating the patient's age from date of birth and comparing it to the given age is another important check. For these studies age is a principal factor in determining the type of therapy the patient is to receive because an error may exclude the patient from the study.

In some instances the given information determines the nature of other information that should be present. If the patient died during the course of the study, there must be a date of death, or in the event that a patient had a disease recurrence, the site and system of the recurrence must be recorded.

The amount of drug given is reviewed to insure that the dosage is in accord with the protocol. Extremely low or high blood counts are listed and checked later. Checks on all dates that must be later than the date of the surgery are made to detect any errors. All parameters specified in the protocol are examined to insure that the patient has met the qualifications for the study.

When the tape has been edited, the list of errors is sent back to the statistical unit for correction. This function may involve writing a letter to the investigator for clarification, or it may represent re-coding and repunching. After editing, the tape is merged with the master tape, thus creating a new master.

The update program demands a variety of routines. First, it allows for the inclusion of new patients on the master tape in the proper sequence. Second, it permits the insertion of additional records for a patient and also the deletion of extraneous records. Third, this program has the flexibility necessary if data on a particular record demand alteration.

After the master tape has been updated, it is ready for statistical analysis. A duplicate of the master tape, denoted as a "frozen tape," is created for the statistical work. This tape is not updated as frequently because statisticians run correlations, frequency distributions, and analyses which do not allow for the daily change of the data. On the other hand, administrators require up-to-date information to prepare reports, or to provide the investigators with current data relative to their patients. Requests for information concerning certain facets of the study are made at frequent and irregular intervals.

As soon as a patient is entered in the study, the initial entry form is sent to be keypunched. This information consists of the patient's study number, name, type of therapy, and date of surgery. The card is written on tape. The master tape is updated and contains at least one record for every patient entered to date. Reports must be prepared monthly, giving a distribution of the number of patients entered in each different drug category by hospital. Having an automated record of each patient entered allows for the production of reports that are current. Prompt transit of the forms to the statistical
unit is of utmost importance. If forms are overdue, a letter is sent to each investigator. As a form is received, it is recorded and the patient's tape record is updated daily. Each month a delinquency program is run against the master tape to determine which patients have outstanding forms. A personal letter with a list of the missing forms is written to the investigator on the computer printer. These letters have been extremely effective in reminding the study participants that they must send information promptly. They continue to receive a letter each month until all delinquent forms for a patient have come into the office.

Another control is the determination of which patients are overdue on the follow-up section of the study. Since the patients must be examined every three months, forms summarizing the results of the examination are submitted. A list of overdue patients by hospital, as well as those patients who must be seen during the current month, is prepared by the computer. This list gives the patient's study number, name and date that the patient should have been seen.

These two systems are the most effective way of reminding the study participants. They have provided more effective control and obtain requested information with more accuracy and less time.

Specific and individual patient or drug information which may be of extreme importance to a particular physician is easily available. He can receive a printout for each patient which might include daily drug dose, any toxicity, and other data. In another case, a doctor may phone to inquire about response and/or toxicity of a particular drug to determine if he wishes to administer this drug to a patient. If results show that patients having a tumor and receiving a certain drug experience extreme toxic reactions, a doctor may not wish to administer the agent or may wish to give a smaller dose.

Frequently, requests are made to submit reports or punched cards to the National Institutes or Health, coordinators for these chemotherapy studies. Because all conceivable patient information is on tape, it is possible to provide any requested data in whatever format is desired. With the former method or present methods at statistical centers other than Roswell Park, much of this data would have to be obtained from the patients' charts and either coded, if punched cards were requested, or typed, if a report were requested. This process is long and tedious, especially if each patient must be recorded separately.

The ability to punch certain information from the master tape and make graphs with a plotter that is connected to the IBM1620 is important, as it eliminates the expenditure of many man hours. Graphs are plotted for individual patients to compare the response of their tumors (e.g., did one tumor decrease, while another increased during study, or did the tumors decrease while the patient was receiving drug, and then increase with suspension of the drug). Applications of this kind are extremely useful when developing reports or presentations concerning a particular study.

SUMMARY

As a result of this system, more patient information is recorded with less summarizing and coding of the data prior to punching and writing on tape. The data, which always is current, facilitates the control of delinquent forms and produces reliable and representative reports.

The ability to perform more intricate and sophisticated statistical analyses has been enhanced by the system because of the improved condition and increased amount of data recorded.

The accurate and close scrutiny on the progress of these chemotherapy studies offers physicians information which previously was available to them only through the "trial and error" method.
Figure 1. Chemotherapy studies — system flow chart.
Figure 2. Chemotherapy studies — system flow chart.