Information processing needs and practices of clinical investigators—Survey results

by NORMAN A. PALLEY and GABRIEL F. GRONER
The Rand Corporation
Santa Monica, California

INTRODUCTION

Although medical researchers were among the first to pursue the promise of computerization, clinical investigators, namely physicians who conduct research into the biology of human subjects and who evaluate the efficacy of new therapeutic measures, have not shared even modestly in the benefits so far available. With the expectation that improved information processing techniques and facilities can improve the quality and efficiency of clinical research and ultimately patient care, The Division of Research Resources (DRR) of the National Institutes of Health (NIH) has been sponsoring a scientific inquiry (called CLINFO) aimed at first developing a detailed understanding of the information processing needs and current practices of clinical investigators and then developing methods for generally and economically alleviating some of the most important needs. The CLINFO investigation team consists of clinical investigators (T. Graham Christopher, M.D. at the University of Washington, Arthur W. Nunnery, M.D. at the University of Oklahoma and Howard K. Thompson, Jr., M.D. at the Baylor College of Medicine), the authors and other information scientists at The Rand Corporation, and William R. Baker, Jr., the CLINFO Project Officer, and other staff members of the DRR.

We have taken the approach that before we can effectively recommend or develop computer-based aids for clinical investigators, our entire team must develop a common deep and realistic understanding of clinical investigators’ backgrounds, working environments, practices, problems and expectations as well as a similar understanding of the costs, capabilities and reliabilities of present and future computer hardware and software. To do this, we have informally discussed clinical research processes and a broad set of problem areas with many investigators and have conducted a formal survey to collect additional objective data. Over the next three years we will develop a prototype minicomputer-based clinical-research data management and analysis system which we will evaluate in a few research centers. Only after we have tested a set of hardware, software, personnel and operating-procedure specifications, would we be prepared to recommend the widespread utilization of such a system.

We began our investigation by observing and informally interviewing more than 100 personnel at over a dozen clinical research sites. As a result, we identified the following as areas that should be considered further: research-data storage, retrieval and analysis; research center administration; research protocol generation; report writing; and locating potential subjects and tracking outpatients.

After carefully examining tentative approaches to these problems, we hypothesized that present computer technology could have substantial impact on the clinical research community if addressed to the data management and analysis areas for the following reasons:

- Good data management is required to ensure that complete, accurate sets of data are collected, that data are available for timely decision-making, and that information is accessible in a form suitable for analysis both during and following an experiment.
- An experiment is of little value if the data collected are not thoroughly analyzed and interpreted.
- Investigators often do not have the time, background, facilities, and staff for proper data analysis.

Furthermore, although we did not find an existing computer system that we felt adequately handled the investigator’s data management and analysis problems while also being sufficiently flexible, economical and accessible, we were pleased to find that several systems (e.g., PROPHET, TOD, MUMPS, GEMISCH and commercial time sharing systems) had demonstrated, by dealing with important aspects of these problems, that our preliminary goals were attainable.

We wanted to ensure that we were not being misled by our own biases, that we were not overlooking important problem areas and considerations, and that our recommendations would be generally applicable to a broad set of investigators and clinical research centers who participate in NIH’s General Clinical Research Centers (GCRC) Program. We also wanted to learn more about investigators’ present utilization of computers. We decided that we could best accomplish these goals by conducting a formal, in-person sample survey involving a substantial number of clinical investigators and clinical research centers. This would also provide an opportunity to gather a data base for the initial design specifications of a computer system.
that would assist clinical investigators in their data manipulation tasks if such a system appeared to be warranted.

The detailed results and a discussion of the survey instrument development, sample selection and survey execution are presented in Reference 1. It will suffice to point out here that the particular design and execution of a survey of user needs must be sensitive to the sociology and traditions which are inherent in the discipline under investigation.

In choosing our sample GCRCs in which to interview investigators and to collect environmental information, we selected GCRCs in larger institutions known for their productivity. Analysis of our data shows that while the 23 GCRCs surveyed represent 28 percent of the centers in the GCRC program, they represent higher percentages of the beds, annual publications, number of investigators using the GCRC, number of active projects, and annual number of research-patient days.

A total of 89 clinical investigators participated in the survey. Although this is not a substantial number, those who were interviewed appear to the authors to be representative of the more active investigators at the more active research centers.

CHARACTERISTICS OF THE INTERVIEWED CLINICAL INVESTIGATORS

Below we have characterized the typical interviewed investigator, based on median or modal responses to a series of questions.

Background and training

A typical interviewed investigator received his MD in 1959.

He started using this GCRC in 1968.

He is an endocrinologist.

He is an associate professor.

He does not have extensive training in math or statistics.

He has little or no formal or informal training in the use of computers.

Modus operandi

A typical interviewed investigator frequently develops new laboratory-analytic methodologies.

He has 5 research studies under way. This does not mean that the single investigator is working on 5 projects simultaneously, but rather that a number of studies are in different states of completion.

He has studied 28 inpatients and 13 outpatients in the past year.

He works with two senior staff, two fellows and three technicians.

He informally shares his unpublished data frequently within his institution and occasionally outside his institution.

He frequently stores his data in laboratory notebooks and on flow sheets. (Flow sheets are tabular data collection forms, generally with time as the horizontal axis and with variable or measurement names on the vertical axis. The flow sheets may be very formal and study specifics or they may be simple grids with dates and variable names entered by hand.)

Use of calculators

A typical interviewed investigator uses desk calculators and manual calculating methods. Seventy-four percent of the investigators stated that they or their staff frequently or almost always used desk calculators in the analysis of their research data, while 50 percent frequently or almost always used manual methods, including slide rules, nomograms and lookup tables in the reduction and analysis of research data.

He has good access to a programmable calculator and uses it extensively. Of the investigators interviewed, 40 percent reported that they had their own programmable calculator. Most of the investigators who had programmable calculators expressed great enthusiasm for them, were pleased with the ease with which they were used, and had grown quite dependent on them.

Use of computers

Eighty-five percent of the investigators reported that they have access to computer centers for batch processing of their research data. But of these 76 investigators, only 27 indicated that they ever used that facility. Most of the computer centers that were utilized to any extent (15 out of 25) were administratively within the medical center. Regarding interactive facilities, only 51 percent of the investigators stated that they did have access to them, and of these 44, only 19 reported that they or their staff ever used the facilities. In some cases, investigators did not know that there were facilities available to them.

We tried to determine the major reasons that investigators did not use the facilities that were, at least theoretically, available. Responses fell into the following categories: Computer Not Needed In My Research—26%, Lack Of Understanding About Computers And Their Utility—20%, Lack Of Required Assistance—14%, Have My Own Computer—10%, Administrative Problems—10%, Physical Location Of Computer Is Inconvenient—10%, Miscellaneous—5%, Cost Of Computer Time—4%. Some typical comments recorded in response to this question included the following:

- Difficulties with the logistics of getting data into the computer. The problems include specifying the data items, creating disk and tape files, preparing coding forms, transcribing data, and detecting and correcting errors.
The investigator's methods are constantly changing so that he felt it would not be worthwhile to invest the required time in setting up an automated data handling system. This perception partially results from the inflexibility of available provisions for handling variable numbers of repeated measures and specification of limits. Also, given the investigator's background, financial support and available assistance, software development is formidable.

Incompatible media, e.g., the computer center can't handle the paper tape that is the output medium from the investigator's scintillation counter.

Investigator has only a small volume of data to analyze and feels the computer center represents an overkill.

The investigator is not yet ready to use the computer.

Funds and/or people are lacking.

The programs that the computer center makes available are not adequate to the investigator's problem.

Aside from hardware and software reliability problems, investigators particularly mentioned the lack of stability of computer centers, i.e., there had been changes in operating systems as well as administrative changes that had obsoleted large programming investments.

The computer center doesn't understand the clinical investigator's needs.

The center is several miles away.

Long turnaround on programming; this is significant in conjunction with the need for special programs to set up data files, enter data, extract data and perform analyses. By the time the programming is completed, the investigator might have lost his train of thought, found an alternative solution, or gone on to another problem.

THE CLINICAL-RESEARCH DATA BASE

In determining the magnitude of the data base, we asked the respondent to focus on a single study that he felt was representative of his research. In most cases, the study was one in which the investigator was currently engaged.

Some of the results of these questions are shown in Figure 1. Forty-seven percent of the studies examined were inpatient, six percent outpatient, and 47 percent mixed. Many of the values in the tails of the “Number of Subjects” and “Total Time . . .” distributions belong to the latter two classes. However, these values have little effect on the median and 80th percentiles. To handle 80 percent of the reported studies, a computer system would have to provide (for each current study) storage for about 100,000 items of numeric information (400K bytes), per study and 5,000 to 10,000 nonnumeric items per study (300K bytes).

IMPEDEMENTS TO CLINICAL RESEARCH

We and our colleagues had previously generated a model comprising 14 identifiable research processes. We
asked the respondents to indicate the degree to which each of these process or steps unnecessarily impeded their research. Carry Out Experimental Procedures And Care For Patients, Collect Specimens, and Carry Out Laboratory Analyses were rated low as impediments to research. This can be attributed to the fact that the studies were undertaken in GCRCs which were specifically designed, in terms of personnel and facilities, to minimize such difficulties. In our sample, we could find no convincing difference in responses to the “impediment” questions between computer users and nonusers. Retrieve, Reduce, And Analyze Data received the highest median “impediment” score. Many investigators indicated that Develop Computer Programs represents a “great” or “very great” difficulty, but many others said it was not a problem because they never did it.

Another series of questions asked how much of a problem it was for the investigator or his staff to perform each of a list of data manipulation tasks and operations. Figure 4 shows the percentage of investigators who responded “great” or “very great” (problem) to the indicated item. Only those items receiving scores of 25 percent or more are shown. Many investigators indicated that Develop Computer Programs represents a “great” or “very great” difficulty, but many others said it was not a problem because they never did it.

We asked those investigators who used computers about the extent to which presently automated tasks would be more difficult without the use of computers. A typical response of those investigators who used computers for tasks such as complex statistical analysis and modeling and simulation was that without computers it would not be possible for them to pursue their current research programs. Some of the tasks that computer and programmable calculator users said would be “very much more difficult” without their tools were: Complex Statistics (38 of 44 users who performed this task), Simple Descriptive Statistics (26 of 46), Subsetting (24 of 35), Sorting Research Records (20 of 28), Finding All Patients With A

---

### Figure 3—Percentage of GCRC investigators responding Great or Very Great (impediment) to a list of possible steps in the clinical research process

<table>
<thead>
<tr>
<th>Step</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtain support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>Obtain literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Develop res. plan</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtain approvals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Instruct assistants</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepare equip.</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Select &amp; admit pts.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Carry out exp. procedures</td>
<td></td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collect specimens</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory anal.</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organize &amp; store data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>Analyze data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Report to sponsor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Publish</td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

### Figure 4—Percentage of GCRC investigators who rated various data manipulation tasks as being Great or Very Great problems

<table>
<thead>
<tr>
<th>Task</th>
<th>0</th>
<th>25</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop computer programs</td>
<td></td>
<td></td>
<td>49%</td>
</tr>
<tr>
<td>Find all patients with particular characteristics</td>
<td></td>
<td>47%</td>
<td></td>
</tr>
<tr>
<td>Complex statistical analysis</td>
<td></td>
<td></td>
<td>40%</td>
</tr>
<tr>
<td>Find all values of a single variable</td>
<td></td>
<td></td>
<td>39%</td>
</tr>
<tr>
<td>Add new measure to all research records</td>
<td></td>
<td></td>
<td>36%</td>
</tr>
<tr>
<td>Gather all data for a single patient</td>
<td></td>
<td></td>
<td>32%</td>
</tr>
<tr>
<td>Update research records</td>
<td></td>
<td></td>
<td>29%</td>
</tr>
<tr>
<td>Explore records to test &amp; generate hypoth.</td>
<td></td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>

---

THE NEED FOR A DATA-ORIENTED COMPUTER SYSTEM

Several series of questions were aimed at determining the relative importance and the relative need for various features of a computer system for clinical research. The percentages of investigators who responded “frequently” or “almost always” to the question, “When you or your staff are analyzing your research data, how commonly do you perform the following procedure?” were: Graphing And Plotting—97%, Manual Transcription—96%, Descriptive Statistics—89%, Subsetting (i.e., selecting research subjects with common characteristics)—57%, Complex Statistics (e.g., regression analysis)—46%, Arithmetic Preprocessing Of Data—37%, and Modeling And Simulation—12%. It is interesting to note that the most frequently performed tasks are also very amenable to computerization.

---

From the collection of the Computer History Museum (www.computerhistory.org)
BENEFITS OF COMPUTER-BASED DATA MANIPULATION

Another series of questions required that the investigator indicate the extent of his agreement with a series of statements which asserted that certain improvements (in data manipulation) could critically influence the performance of his research. A majority of investigators indicated that they agreed, or strongly agreed, with nine of the 14 assertions. The improvements which apparently can critically influence the performance of clinical research are: More Insightful Ways of Examining Data (79% agreement), Easier Retrieval Of Research Data (73%), Easier Combination Of Old And New Data (71%), Easier Subsetting Of Data (70%), Detection Of Missing Or Inaccurate Lab Determinations (65%), More Convenient Arithmetic Processing (60%), Reduced Need For Transcribing Data (57%), Easier Statistical Analysis (56%) and Quicker Notification Of Missing Samples (54%).

Three of our assertions that did not meet with majority agreements are: More Accurate Control Of Sample Collection Times, Less Delay In Receiving Results Of Laboratory Determinations and More Accuracy In The Recording Of Laboratory Data. Again, this appears to reflect the investigator's satisfaction with the sample collection and laboratory facilities which they have at their disposal. Also, the areas in which it was felt that improvement could have a critical effect on clinical research are those (named above) to which computer technology can make some substantial contributions.

Other questions sought to determine to what extent the investigators agreed that improvements in these data collection and manipulation areas could have certain beneficial results. The potential benefits agreed to by a majority of the respondents were: Reduced Nonproductive Time That An Investigator Must Spend On A Study (85%), Greater Insight Into The Investigator's Research Data (84%), Reduced Elapsed Time Needed To Complete A Study (75%) and Increased Publishable Results (73%). The majority of respondents disagreed with assertions that such improvements would reduce the required number of experiments, subjects or data samples. Thus, the investigators felt that the improvements would be in the form of better and more efficient use of the available clinical material rather than a reduction in the need for that material.

REQUIRED OPERATIONAL CHARACTERISTICS

Next, a series of questions asked specifically the extent to which the investigators agreed with the need for certain operational characteristics of a proposed computer system. Those features for which 50 percent or more of the respondents answered “agree” or “strongly agree” are shown in Figure 5.

Some of the most interesting results are among the features that did not generate much enthusiasm. In particular, only 31 percent “agreed” or “strongly agreed” that visual displays of graphs and plots on a TV-like screen must be available. This is particularly significant in light of the degree to which graphing and plotting are performed as well as the expressed criticality of having more insightful ways of examining data. Our interpretation of this result is that a majority of investigators have had little experience with online graphic systems and were unfamiliar with their capabilities and advantages. This would also be consistent with the heavy stress on the need for printed reports and printed graphs and plots. Another is that investigators prefer hardcopy plots, which have traditionally been prepared by technicians or data clerks, so that they can examine and compare them at their convenience.

CONCLUSIONS

Our conclusions are primarily drawn from our analysis of the objective responses, colored by our prior informal interviews, visits to clinical research institutions, and discussions with the CLINFO clinical contractors; and by our conceptual view of the significant role that information processing plays in clinical research.

The status quo

Many investigators have and use small accessible computers and programmable calculators for arithmetic
processing of their research data and for the computation of descriptive statistics. Those devices do not provide for the flexible storage or retrieval of research data files; their accessibility and their ease of use partially counterbalance these deficiencies, however.

The investigator has at least nominal access to large computer centers based administratively either at an affiliated university or at the medical school. These centers are little used by the researchers we surveyed. There appear to be three primary explanations for this. First, most investigators have had little or no training in the mathematical sciences or in the use of computers.

Second, the computer centers can rarely provide the kind of interested and knowledgeable assistance that the clinical investigator requires. Most computer center staffs either are not aware of the needs of their local clinical research communities or view them as unlikely prospective customers. Information concerning the use of the center, available programs, methods of data entry, etc., usually is not available at a level useful to the uninitiated.

Third, partially as a result of the conditions mentioned above, the investigator has little motivation to make what he views as a major investment of his career time to achieve the questionable advantages of machine-aided information processing. The barriers that he often faces include inadequate software, uncertainty about an unstable computer center environment, inconvenient geographic distance, slow turnaround (in both computer runs and special programming), and ad hoc consultative assistance lacking continuity and compatibility. In addition, he views his information processing needs as rapidly changing and requiring great flexibility. As awkward and unresponsive as his manual methods may be, he can keep personal control over his hard-won research data, can peruse it without the need for an intermediary, and does not fear its loss due to a programming error or disk crash.

As a result, the clinical investigator forgoes the potential of easy retrieval and manipulation of his study data. He also recognizes that the waste of his time and talents unnecessarily prolongs the duration of his studies and probably reduces the quantity and quality of publishable results.

Immediate need for people

The survey results show that clinical investigators could make much better use of existing computational facilities if they had better access to well trained people who are capable of educating them about the use and capabilities of computers; who are able (on a long term basis if necessary) to assist them with data organization, data reduction, retrieval, and analysis; and who can promote their use of an existing facility while providing a realistic appraisal of its capabilities and shortcomings. Also, these people must be responsive to the investigator’s needs and sensitive to problems unique to clinical research.

Since few such well trained, dedicated people exist, additional support should be provided for:

- Training of laboratory and other personnel already participating in clinical research.
- Utilization of existing institutions such as biostatistics or biomathematics departments and computer centers.
- The establishment of national resources devoted to satisfying these needs.

Long-term need for computer systems

Based on our observations, substantial benefits would result from the development of a small, economical, accessible, widely-used computer system especially designed for the clinical research environment. Such a system would facilitate difficult data-oriented tasks that currently are not perceived as practical. A flexible system would encourage the sharing of data through the implementation of agreed on variable definitions, normal ranges, and data organizations while allowing the individual investigator to specify his own data entry, analytic and other procedures.

Widely used systems, based on the same hardware and underlying software would encourage the development of sharable computer programs, as appropriate, while still satisfying the specific needs of particular individuals and centers. It would be extensively utilized because of its accessibility and if it were accompanied by people to assist and educate the prospective users. Also, if its cost were compatible with the research center’s budgetary constraints and if it were under the center’s administrative control, it would not be as subject to fiscal, administrative, and technical disruptions and uncertainties as are large computer systems operated by traditional computer centers.

While a preliminary system design may be based on the information we have so far accumulated, only experimentation in the clinical research environment will reveal the acceptability and value of a specific system design to clinical investigators. We will carry out such experimentation during the next phase of the CLINFO project.

ACKNOWLEDGMENTS

The research reported on in this paper was supported by the National Institutes of Health under Contract No. NO1-RR-2-2106. This survey would not have been possible without our earlier extensive discussions and information gathering with Drs. T. G. Christopher (U. of Washington), A. W. Nunnery (U. of Oklahoma), H. K. Thompson, Jr. (Baylor), and N. Z. Shapiro and W. L. Sibley (Rand); the guidance and participation of Drs. W. G. Walker (Johns Hopkins) and R. S. Goldsmith (formerly Mayo Clinic); the support and encouragement of Drs. W. R. Baker, Jr., W. DeCesare, W. F. Raub (NIH) and B. Schachtel (formerly NIH); the other interviewers E. Fairbrother and L. Heiser (Rand) and J. Farquhar and A. MacInnes (for-
merly Rand); and the numerous clinical investigators and
other clinical research personnel who spent their precious
time with us.

REFERENCES

and their Information Processing Activities, R-1539-NIH, The Rand
Corporation, July 1974.

2. Sibley, N. L., M. D. Hopwood, G. F. Groner and N. A. Palley, A Pro-
totype Data Management and Analysis System for Clinical Investi-
gators: An Initial Functional Description, R-1621-NIH, The Rand
Corporation, August 1974.


4. L. Sibley, A Plan for the Development and Evaluation of a Data
Management and Analysis System for Clinical Investigators, R-1542-

5. Ransil, B. J., "Applications of the PROPHET System in Human
Clinical Investigation," AFIPS Conference Proceedings, Vol. 43,


7. Fries, J. F., "Time-Oriented Patient Records and a Computer
Databank," JAMA, 12, 222, December 18, 1972, pp. 1536-1554.

"A System for Clinical Data Management," AFIPS Conference

Generalized Medical Information System (GEMISCH) for Practic-
ing Physicians," National Conference of the Association for Comput-