COMPUTER SUPPORT FOR MUSCULAR SUBAORTIC STENOSIS RESEARCH

Neil Harding McAlister, H. Dominic Couvey, Charles Pollick, Anne M. Sheppard, and E. Douglas Wigle

Cardiovascular Unit
Toronto General Hospital
Toronto, Canada

Summary

Data collection and data management are time-consuming aspects of clinical research, often occupying much of the time that a physician spends in doing clinical investigation. For retrospective studies, backlogging data from old charts is a tedious process subject to errors. Even in prospective studies it can be difficult to maintain rigid adherence to data collection protocols. This paper describes data management procedures designed to overcome these problems in a large clinical research project on muscular subaortic stenosis. Data collection forms have been designed. A data-independent database management system on a minicomputer is being used to support the database. Information can be accessed as required using a query language and report generator.

Clinical Background

Hypertrophic cardiomyopathy is one type of primary heart muscle disease (cardiomyopathy) of unknown etiology. It is characterized pathologically, as the name suggests, by an abnormal hypertrophy of heart muscle fibers, particularly in the ventricular septum. In some patients with this disorder there is an obstruction (stenosis) just below the aortic valve (subaortic) caused partly by the hypertrophic ventricular septum. When such an obstruction is present the condition is called obstructive hypertrophic cardiomyopathy, or, hypertrophic subaortic stenosis, or by the term we have used, muscular subaortic stenosis (MSS). Such a stenosis may be present at rest (resting MSS) or it may only be present after stimulating the heart (latent MSS). Still other patients with hypertrophic cardiomyopathy may have no obstruction to ventricular outflow, in which case the term nonobstructive hypertrophic cardiomyopathy is used. Thus there are basically three forms of this type of heart muscle disease, and in this study we are primarily interested in MSS (resting or latent).

When symptomatic, MSS patients can present with symptoms including shortness of breath, chest pain, fatigue, palpitations, faintness, congestive heart failure, and sometimes arrhythmias.

Introduction

At one time it was normal for a clinical researcher to spend much of his research effort performing his own statistical analysis manually. Today there are expert biostatisticians, computer programs, and calculators to serve that purpose, freeing the researcher's time for more creative pursuits.

But today data management is still a time-consuming part of clinical research that, while largely repetitious -- involving things like tabulation, filing, and retrieval -- often requires a significant portion of the time that a physician spends in clinical investigation. The data collection process itself is also subject to several kinds of errors. Data can be misinterpreted, transcribed incorrectly, or simply not recorded. At the least, such errors can prejudice the results of a study, and in the worst case, they can undermine a study completely.

The purpose of this work is to establish data management procedures for our research on muscular subaortic stenosis (MSS) that can reduce the physician's burden of data collection and data manipulation, minimize the potential for errors or breakdown to occur in these processes, and serve as a model for the support of future clinical research projects in the Division of Cardiology at the Toronto General Hospital, or elsewhere.

530
MSS is of interest to cardiologists because it is an isolated primary pathology of heart muscle itself. Because MSS tends to "run in families", it has research interest as a genetic disorder. It is, furthermore, a potentially fatal disease. Sudden death can occur in MSS patients, sometimes without warning. Thus MSS has grave potential significance for the individuals who have it, and for their families.

Toronto General Hospital is a referral centre for MSS cases, where over 300 cases of this rare disorder have been seen in the past 15 years. A large backlog of potential research material, and a continuing influx of new cases therefore exist, providing impetus for further research.

**Computer Support**

A computer is being used to assist MSS research for several reasons. First, the machine has the ability to store large amounts of data and to retrieve all or a specified part of that data quickly. Second, the computer can be used to help enforce quality control and data collection protocols, thereby reducing the number of potential research cases lost to analysis because of missing data. Third, advances in database management systems have made it possible and economic to share data held in common by several specific systems have made it possible and economic to share data held in common by several specific applications, reducing the amount of redundant information gathering and management that would occur without such computer support. Fourth, a computer provides powerful tools for data analysis and display.

A commercial database management system called TOTAL was modified locally to provide logical and physical data independence in accordance with the ANSI/X3/SPARC recommendations. It is implemented on a Varian V76 minicomputer. The data independence module, called DIET (Data Independence Extension of TOTAL) interposes between the user and the TOTAL database management language. A data-independent database management system permits continuous modification and restructuring of various datasets used in different projects, without substantial disruptions in those other projects which may share data items in common with the datasets that have to be modified.

MSS research is a case in point. Its dataset incorporates many facets of clinical cardiology. For example, echocardiographic data required for documenting MSS cases is already being collected on computer-compatible optical mark recognition forms in order to generate the standard echocardiogram reports that are already used in the Division of Cardiology. The data sharing permitted by modern database technology makes sense in such an environment.

It was very difficult to define precisely the data that would be collected in support of MSS research. It was necessary to establish a dataset from which one can draw answers to general clinical questions that must be rather imprecise and broad in scope to begin with. At one extreme, there is the risk of creating a "write-only memory" -- an expensive repository for clinical data that is never used for anything. At the opposite extreme there is the danger of being too restrictive in scope, and spending money, time, and energy to collect too little data to be of any use in the long run. Only as we begin to use the database facility to analyze the data will we know whether or not the compromise is successful.

**Data Collection Methods**

**OMR Forms**

Optical mark recognition (OMR) forms, 8½ x 11 inches in size, are used as the principal vehicle for entering MSS data into the database system. On each of these forms, a description of the data required or the questions to be answered is printed along the left half of the page. On the right half of each page, in positions corresponding to the appropriate descriptions or questions, the user is required to fill in multiple-choice or numerical data by using pencil marks in reserved places. These reserved locations are lightly overprinted with numbers or alphabetical identifiers, so that the person filling out these OMR forms can identify the choices or numbers to be selected. Up to 64 questions, with up to 12 selections each are possible on this standard OMR format.

OMR forms have been designed for each kind of MSS data listed in Table 1. Figure 1 shows one such representative form, used for documenting the past history of MSS patients.

**Table 1**

<table>
<thead>
<tr>
<th>OMR Forms Designed for MSS Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Patient registration</td>
</tr>
<tr>
<td>2. Relevant past history</td>
</tr>
<tr>
<td>3. Functional inquiry and physical examination</td>
</tr>
<tr>
<td>4. Treatment (drugs and surgical)</td>
</tr>
<tr>
<td>5. Electrocardiogram</td>
</tr>
<tr>
<td>6. 1-Dimensional echocardiogram</td>
</tr>
<tr>
<td>7. 2-Dimensional echocardiogram</td>
</tr>
<tr>
<td>8. Chest X-ray</td>
</tr>
</tbody>
</table>

The forms have been designed to make them fairly easy to use. Insofar as possible, the forms are self-explanatory so that any physician or medical trainee can fill them out. In recording retrospective data from charts, the use of multiple-choice OMR forms forces the person extracting information from the charts to do a thorough and
**MSS -- PAST HISTORY**

**PATIENT'S LAST NAME**

**FIRST LETTER**

**SECOND LETTER**

**THIRD LETTER**

**PATIENT'S USUAL FIRST NAME**

**HOSPITAL CHART #**

**(OR BIRTH DATE, IF NO CHART EXISTS)**

**HOSPITAL (if chart # given)**

**INSTRUCTIONS:** Fill in age when each of the following symptoms first occurred. If negative, leave blank. If positive but age unknown, fill in age "99".

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Age at First Presentation (yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyspnea unspecified</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Exertional</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Orthopnea</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>P.N.D.</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Angina unspecified</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Exertional</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Orthopnea</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Presyncope unspecified</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Exertional</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Orthopnea</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Fatigue unspecified</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Exertional</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Orthopnea</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
</tbody>
</table>

**FIGURE 1**

Past History CMR Form (facsimile, first page)
consistent job. A medical student has done an exemplary job of backlogging MSS data using these forms.

Similarly, when filling out these forms prospectively when patients are seen, the physician is given a means of structuring clinical observations in a consistent manner that will not vary from visit to visit. On a computer form expressions such as "mild, moderate, severe" are revealed as meaningless in the absence of firm definitions. A data collection procedure such as the one described here compels us to define the terms we use according to clinically-reproducible criteria.

Keyboard Input

Some data is entered into the database directly from the keyboards of CRT terminals. This method is necessary both for large amounts of numerical data and when more than a few alphabetical characters must be entered, as OMR forms are best suited to "multiple-choice" responses and cannot handle text string input.

Registration Process

Preselection of Cases

It is not enough to specify data that must be collected, nor even to provide simple means of collecting that data. One must ensure that data defined as critical for a clinical research endeavour is in fact gathered.

A physician conducting clinical research on a relatively uncommon condition may not personally see a sufficient number of suitable patients to fulfill the requirements of a clinical research project. In these circumstances the researcher must rely on professional colleagues to furnish necessary cases. Periodically in our investigational Unit, for example, lists are circulated to the staff, summarizing the clinical research projects currently underway and requesting cooperation in notifying the interested clinicians about suitable case material.

However, goodwill alone often does not suffice to give researchers clinical material that they require. Busy colleagues easily forget about another researcher's earlier request for assistance.

In a busy clinical environment it would therefore be desirable to establish some procedure to ensure that every researcher would be aware of those cases being seen by colleagues that might be suitable for his or her own clinical research purposes. A registration procedure for all patients seen in our Unit is therefore being developed. The purpose of "universal" registration for all of our patients would actually be two-fold. Its first aim would be to support "preselection" of possible clinical research cases, in order to overcome the difficulty just outlined.

Unique Identification

The second purpose of the registration procedure would be to establish a method of unambiguous identification that the database management system can use to link and retrieve the various pieces of cardiologic data that pertain to each patient. Our unique identifier is made up of the first three letters of the patient's last name, followed by the first letter of his first name. Appended to these letters is the hospital chart number (in the case of in-patients) or the patient's birth data, represented numerically in the case of out-patients. This method was selected after an unsuccessful attempt to utilize various "universal" numbers such as Social Insurance Number, Ontario Health Insurance Plan Number, etc. These alternatives are simply not available for all patients, or they are not unique.

At each patient contact, the patient would have to be questioned to find out if he has been hospitalized, or if he has changed his name since the previous encounter. If so, all subsets of his data in the database would have their unique identifier changed.

Software Tools for Data Analysis and Display

Figure 2 indicates the software utilities that interact with the database. These software tools suit the requirements of MSS research, but they have been designed as general-purpose utilities that can serve other database users equally well.

Data Entry and Editing

A software utility for entering data items into the database has been developed. It supports input via either keyboard or optical mark recognition forms.

Data can be retrieved from the database, viewed on a video screen, and modified if desired using this software.

Query System

The query system that has been developed can test any field or combination of fields in the database. If the fields to be tested satisfy specified conditions (i.e. specific values, ranges, Boolean conditions) then any specified fields from the same records that contain the tested fields are "dumped" into a temporary disk file. The fields to be dumped need not necessarily be the same as those tested: for example, we may wish to dump the chart numbers of patients selected on the criterion of being male.

Report Generator

Once a query has been run and the specified data dumped onto the temporary disk file, the report generator can be used to perform sorts and
counts, and to produce printed reports. Printed documents are generated with titles, column headings, and with data in columns as specified by the user.

Statistics

Presently data from the temporary disk file can be copied onto magnetic tape in a format compatible with the commercially-available SPSS statistical package. Such a data tape produced by the Varian minicomputer can be taken to a central, larger computer that supports SPSS, for batch processing.

Work is presently underway to develop our own, limited, interactive statistics package. Average, mean, minimum and maximum, mode, standard deviation, t-test, life table analysis, etc. are nearly complete.

Graphics

The development of graphic display capabilities as a tool to assist manipulation of database data has been proposed. It would be desirable to have the ability to produce both output on video terminals, and hard-copy graphic output. The approach to be used if this proposal were to be acted upon would probably be to modify an existing graphics package for our own purposes.

Conclusions

The application of the database system to muscular subaortic stenosis research is now at the point where we soon look to reaping benefit from it. Presently work is underway extracting data from about 300 old charts of MSS patients. We anticipate being able to employ the database to correlate, both retrospectively and prospectively, this large amount of clinical data that was previously widely scattered in paper charts and inaccessible for detailed analysis. A query system and report generator for selecting and reporting subsets of data from the database have been developed. We are developing our own statistical and data display facilities, or adapting existing ones to the Varian's database system.

Working to adapt a database system for clinical cardiovascular research in the specific area of muscular subaortic stenosis has provided several insights into the use of hardware and software tools in clinical research endeavours:

1. As research needs become clearer during the course of a project, it will be perceived that some data items must be added to a dataset, and that other items being collected are useless. Data independence, therefore, is a desirable ingredient in medical research database, for it provides the optimal "tunability" required in the medical research environment. It is simply too expensive to modify or rewrite...
an entire applications package as would be necessary in a conventional database that lacked data independence.

2. The large effort (approximately 1 year) to design appropriate forms and to transcribe old, charted material onto computer-compatible OMR forms indicates that retrospective studies are treacherous not only from a methodological point of view, but also because the quality and completeness of data to be found in old patient charts is inconsistent. What individual doctors consider worthwhile recording in a dictated or written letter or note varies considerably. The effort to gather and quantize charted information for entry into a computer-based system forces some order out of chaos, and at least clearly demonstrates what data is available and what data is irretrievably absent. In addition, OMR forms can be read directly by the computer, thereby avoiding the error-prone and expensive step of hiring clerks to encode and transcribe clinical material onto computer-processable media.

3. An in-house data processing team is recommended for computer support of clinical research. The specific minicomputer software tools needed for this work -- a data-independent database management system with a query system, a flexible report generator, and statistical and data display packages -- is not commercially available. The effort to develop these tools is not trivial, but it is finite, especially when one builds on successful foundations rather than starting "from scratch". For example, the data independence extension for the commercial database management system, TOTAL, was developed by Dubien et al. over a 6 month period.7

4. It is desirable to build general-purpose software tools for medical research directed to a specific application. The muscular subaortic stenosis problem is specific enough to focus software development efforts, and to make obvious problems as they occur -- not after work has been supposedly completed and found inadequate. At the same time, the MSS project is broad enough to demand software tools that will serve other clinical research projects as well. Care has been taken to avoid tailoring the query system, the reporting system, etc. only to the needs of MSS work. The software that has been written has been general: MSS research has inspired the creation of the software, but MSS remains only one project sharing the general computing needs of clinical research in our Division.

Future

The significance of the computer support created for MSS research will extend beyond this one application. The success of this database system in helping us answer some questions about our MSS population over the next year will in large measure determine the steps that will be taken to adapt this same database to assist other clinical research endeavours.

References


Copies of the OMR forms may be obtained from:

Dr. E.D. Wigle
Chief of Cardiology
Toronto General Hospital
UC-1-104
101 College St.
Toronto, Ontario, Canada
M5G 1L7