MEDINFO at the MGH Oncology Data Research Center

Susan A. Stickler, Ethan A. Foster, Alfred M. Cohen, M.D.,
G. Octo Barnett, M.D., Marion O. Cameron, Jeannette Herrmann, Debra Midman

Abstract

MEDINFO is a database management system developed at MGH primarily to organize and analyze data from clinical investigations. It allows for flexible definition of data items contained in a research study and provides users with the ability to perform their own analyses interactively. MEDINFO has been used by a number of investigators both at MGH and at allied institutions for the past seven years on populations ranging from 100 to 8,000 patients.

Most recently, the MGH Oncology Data Research Center, containing over 43,000 records, has been converted to a MEDINFO system. This paper highlights the oncology data research activities and the MEDINFO system's role as a general-purpose database management system versus a highly specialized or uniquely written system. We compare the advantages and disadvantages in costs, implementation time, ease of usage, flexibility, ease of maintenance, and support.

MEDINFO

MEDINFO (MEDical INFORMATION) is a general-purpose interactive system for information collection, storage, retrieval, and analysis. Written in MGH MUMPS, a time-sharing language, MEDINFO now runs on a DEC VAX 11/780. The system was developed at the Laboratory of Computer Science at Massachusetts General Hospital (MGH) primarily to organize and analyze data from clinical investigations at a reasonable cost. Similar computer-based medical research systems include File Manager (2), WISAR (3), MEDUS/A (4), and CLINFO (5).

Researchers can explore their own data using a variety of interactive numerical and graphical analysis functions. Basic routines to extract, examine, and analyze subsets of the population incorporate graded levels of complexity and on-line help to enable users to work without extensive assistance from MEDINFO staff. The core statistical package, written in MUMPS, performs data searches, tabulations, descriptive statistics, and cross-tabulations. It also offers a variety of tests of significance. These include the Student's T, Chi-square, Fisher Exact, and Wilcoxon tests, and Kolmogorov-Smirnov statistic.

In our earlier version of MEDINFO, we also provide linear and logistic regression, discriminant analysis, and risk-ratios as part of the MUMPS package (6). In the VAX version, these functions are just a part of the menu offered as we link to the BMDP statistical package. We believe that it is cost-effective to tie into BMDP for complex computations, and to use MUMPS for direct calculations.

The sparse data files and non-fixed format fields of MUMPS make it possible for MEDINFO to provide flexible definition of the entire data base as well as of individual data items. Data items can be added to (or deleted from) a study at any time. Each item is associated with a syntax definition (e.g., multiple choice, numeric, patterned text, etc.) which is used to check data at the time of input. MEDINFO supports real-time entry of information into the computer with concurrent computer syntax checking and error detection. In addition, all data are stored in on-line files to facilitate updating and error correcting. MEDINFO handles a range of multi-dimensional storage problems, from the simple case of a data item collected several times to the more complicated case of a set of connected items linked by one or more fields that uniquely identify that set.

The programming style and string handling capabilities of MUMPS make custom programming for MEDINFO users simple. Although MEDINFO offers users many options within the standard entry, manipulation, and display routines there is provision to detour to even further enhancements. Such supplementary routines facilitate transcription, update rosters, calculate new data items, and generate specialized reports and displays.

MEDINFO has been used by a number of clinical investigators both at MGH and at allied institutions (i.e., Massachusetts Eye and Ear Infirmary and Harvard Community Health Plan). Their projects represent populations from 100 to 8000 patients with 25 to 1800 data items stored per patient. Clinical studies include the results of cardiac surgery, melanoma, hysterectomy, hip surgery, colonoscopy, urinary tract infection, and parathyroid disease. Evaluation of a cardiac intensive care unit, patient monitoring for anticoagulation therapy and quality assurance study of hypertension treatment...
represent the system's potential for improving the quality of care delivered. MEDINFO also provides security and coordinative functions for a collaborative randomized clinical trial of nephrotic syndrome.

**Oncology Data Research Center**

Recently, the MGH Oncology Data Research Center, with a registry of over 43,000 incidences of cancer, has converted to a MEDINFO system. The Accreditation Board of the American College of Surgeons requires the abstraction of a limited amount of initial and follow-up data on each patient with a diagnosed malignancy. This information forms the basis for a cancer registry.

In August 1972, a computer-based information system was established containing information about each patient, the cancer site, histology, associated treatment, and follow-up status.

Some of the major benefits of a computer-assisted registry are:

1. Easier handling of the voluminous and often redundant information regarding cancer patients (i.e., discharge diagnosis, pathology, therapy reports).
2. Improved quality and depth of information due to the ability to check the logical sense of interrelated data.
3. Easier and more complete initiation of requests for follow-up status, i.e., automatic identification of those patients who have not been seen for the past year, and generation of appropriate letters.
4. On-line inquiry of information on individual patients either by name or unit number.
5. Ability to produce accession rosters, annual statistics, and survival analyses, and to transfer data to the state cancer registry via magtape.
6. Easier examination of files, to satisfy research requests.

Increasing user sophistication, retrieval requirements, and programmer costs combined with decreasing storage costs made the original 1972 system design impractical in the 1980s. More importantly, it became apparent that custom programming for each research request was inefficient. At this point, there were two choices: to write a new system specifically for the registry, or to use an existing general-purpose system. In September 1982, the "specialized" registry system was abandoned and the Oncology Center's data base was transformed into a MEDINFO data base.

This new installation features user-friendly data entry via "intelligent" terminals, combining the branching power of a menu-driven selector with the methodical "fill-in-the-blank" approach that improves data completeness and accuracy. It also includes sophisticated data integrity checking for logical internal consistency, improved follow-up mechanisms, maintenance of potential case and annual accession rosters, plus all the analysis capabilities of the general-purpose MEDINFO statistical package. The center now collects data 3000 new cases per year and follow-up on another 20,000 patients per year. Below is an example from the seven-page "Type 1 Abstract".

**Type 1 Abstract**

1. Last, First Name
2. Date Abstracted:
3. Patient Street:
4. City:
5. State:
6. Zip Code:
7. Date of Birth:
8. Sex:
9. Date First Diagnosed:
10. Primary Code:
   1. First Primary
   2. Second primary
   3. Third primary
   4. Fourth primary
   5. Other

Press 'FORWARD' for next page. 'QUIT' to stop
Press 'QUIT' when done.

MEDINFO offered two important advantages over re-design of their ten year old computer-based system, **LOW COST** and **SHORT IMPLEMENTATION TIME**. The most important part of the implementation process is study design, this entails deciding what data is to be collected and, based on how it will be accessed and analyzed, determining the storage format for that data. From there, the MEDINFO team creates appropriate dictionaries, directories, and special purpose programs. The whole design process—from initial contact to implementation—can take from one week to several months depending on the complexity of the user's needs and volume of data to be collected. The cost is low since all data handling routines
already exist and a programmer's time is needed only to write special purpose programs and design non-standard schemata. The single-user maintenance cost is low since the small MEDINFO team can support one set of programs for all twenty (or more) applications. In addition, MEDINFO is conceived independently of any single application, so all users benefit from "free" promulgation of additional capabilities not anticipated in the particular study's design phase.

There are obvious disadvantages to a general-purpose database management system. In MEDINFO, the greatest disadvantage is the necessity of accessing the schema to determine the logical address of the data, thereby slowing the data entry process. However, this cost is far out-weighed by the flexibility to re-design the structure easily when the inevitable surprises and necessary changes arise after implementation. Compiled code provides a partial solution to this problem. By writing analyses which create their own MUMPS subroutines prior to execution, we save redundant references to the schema.

We have concluded from this experience that by using MEDINFO the MGH Oncology Data Research Center has reaped several benefits. First, since most of the implementation time was spent on the design phase the center's needs were extensively examined and documented. Second, by supplementing MEDINFO to meet the special needs of a registry rather than designing a special purpose system the savings in programmer costs were probably ten fold. In a short time and at a low cost, the center was converted to a new system, allowing them to improve the quality of their data base as well as providing users with direct access to many new analytic capabilities.

Additional functions are easy to implement and new features designed for one study group benefit all study groups. For example, when life table analysis was written for one study, it automatically became available to all studies.

The members of the Oncology Center increased their usage of the system from about 100 hours to 400 hours per month during the six months following its implementation in September 1982. They have expanded their activities from data collection and entry to analysis and evaluation of data integrity, taking advantage of the additional functions that MEDINFO offers. Meanwhile, the MEDINFO group has implemented systems for the anticoagulation therapy unit, an arthritis and joint disease clinic, a study of involuntary psychiatric admissions, and has automated part of the hospital's Staff Clinic. Currently in the design phase are studies on thromboemboli following hip replacement and on cardiac rehabilitation. Further development of MEDINFO is in such areas as front-end processing and increased graphics capabilities.

1. Johnson SC, Barnett GO. MEDINFO - A MEDical INFORMATION system. Laboratory of Computer Science, Massachusetts General Hospital, Boston, 1976.