White Paper ■

Designing Medical Informatics Research and Library-Resource Projects to Increase What Is Learned

WILLIAM W. STEAD, MD, R. BRIAN HAYNES, MD, SHERRILYNNE FULLER, PHD, CHARLES P. FRIEDMAN, PHD, LARRY E. TRAVIS, PHD, J. ROBERT BECK, MD, CAROL H. FENICHEL, PHD, B. CHANDRASEKARAN, PHD, BRUCE G. BUCHANAN, PHD, ENRIQUE E. ABOLA, PHD, MARYELLEN C. SIEVERT, PHD, REED M. GARDNER, PHD, JUDITH MESSERLE, CONRADE C. JAFFE, MD, WILLIAM R. PEARSON, PHD, ROBERT M. ABARBANEL, MD, PHD

Abstract Careful study of medical informatics research and library-resource projects is necessary to increase the productivity of the research and development enterprise. Medical informatics research projects can present unique problems with respect to evaluation. It is not always possible to adapt directly the evaluation methods that are commonly employed in the natural and social sciences. Problems in evaluating medical informatics projects may be overcome by formulating system development work in terms of a testable hypothesis; subdividing complex projects into modules, each of which can be developed, tested and evaluated rigorously; and utilizing qualitative studies in situations where more definitive quantitative studies are impractical.

■ **J Am Med Informatics Assoc.** 1994; 1:28–33.

The members of the Biomedical Library Review Committee (BLRC) of the National Library of Medicine devoted a portion of each of their meetings during the 1990–1991 year to discussions of how to increase what is learned from information access, information systems, and medical informatics research projects.*

This paper presents the consensus that emerged from those discussions. It is intended to provide guidance to investigators in designing and describing how they will evaluate their projects. The reader is referred to a previous paper[1] for suggestions regarding preparation of other aspects of a medical informatics research project proposal.

This paper begins by placing applied medical informatics within the spectrum of medical informatics research and development activities. It then proposes a framework for design of applied medical informatics research projects and the evaluative studies that are part of them. The framework is intended to represent a conceptual approach, and it should not be taken as a direct prescription. The fundamental concept is that innovation in medical informatics results from a number of disparate, but logically sequential, activities. Research and development should be staged, and rigorous evaluation should take place at each stage. The type of evaluation that is appropriate will vary according to the stage of the work, but all evaluations will involve thoughtful collection of information and subsequent analysis and interpretation of this information.

Separate approaches for applied medical informatics research and library-resource projects are not rec-

Affiliations of the authors: WWS, Vanderbilt University, Nashville, TN; RBH, McMaster University, Hamilton, Ontario; SF, University of Washington, Seattle, WA; CPF, University of North Carolina, Chapel Hill, NC; LET, University of Wisconsin–Madison, Madison, WI; JRB, Baylor College of Medicine, Houston, TX; CHF, Hahnemann University, Philadelphia, PA; BC, Ohio State University, Columbus, OH; BGB, University of Pittsburgh, Pittsburgh, PA; EEA, Brookhaven National Laboratory, Upton, NY; MECS, University of Missouri–Columbia, Columbia, MO; RMG, University of Utah, LDS Hospital, Salt Lake City, UT; JM, Harvard Medical School, Boston, MA; CCJ, Yale University, New Haven, CT; WRP, University of Virginia, Charlottesville, VA; RMA, The Boeing Company, Seattle, WA.

Correspondence and reprints: William W. Stead, MD, Vanderbilt University Medical Center, Ste. 2000, Village at Vanderbilt, 1500 21st Avenue South, Nashville, TN 37232-8143.

Received for publication: 1/04/93; revision accepted for publication: 5/17/93.

*The views expressed in this article are the personal opinions of the authors. The article represents neither the viewpoints nor the policies of the National Library of Medicine or of the National Institutes of Health.

ommended, because these two categories of work are thought to represent different points on a staging continuum, rather than totally different types of projects. Library-resource projects that seek to increase information access should include an assessment of the impact of use of the stable end products of informatics research. Those library-resource projects that involve information systems may resemble information-access projects or they may involve earlier stages in the continuum of medical informatics research and development.

The Spectrum of Medical Informatics Research and Development

Medical informatics is "the study, invention and implementation of structures and algorithms to improve communication, understanding and management of medical information."[2] Innovations in medical informatics result from a cyclical spectrum of research, development, and implementation activities. Innovations begin with basic research and theory development, continue through engineering and applied research, and conclude with capitalization into robust production systems. Analysis of the results of building and deploying systems should then feed back to basic research and lead to new hypotheses regarding why things work or how they could work better.

Each of these types of activities has distinct objectives, techniques, and endpoints. Basic research tests hypotheses with regard to areas such as knowledge representation, linguistics, or information assimilation through exploratory experiments that are designed to accept or reject proposed hypotheses. The products are scholarly presentations and publications.

Applied research can be subdivided into three categories of work. First, model testing seeks to evaluate an innovative approach. Knowledge about the effectiveness and the inadequacies of an approach constitute the result of the work; and any systems that are generated in the process are considered incidental tools. Second, engineering and development activities involve building a tool and demonstrating its ability to perform as specified. Endpoints include not only the tool but also lessons learned during its construction and documentation of its performance. Third, demonstration or implementation activities involve assessing impact of a stable tool. Information documenting the demand for, and benefit of, the tool in an operational setting results from the work.

Capitalization efforts take proven techniques and tools that result from basic and applied research and convert them into debugged, tuned products that can survive in the marketplace. This time- and labor-consuming process should not begin until capability, suitability, and marketability are demonstrated. The result should be a product that will be taken over and supported on some income-generating basis.

Individual projects may cross over the boundaries between these types of activity. Those aspects that involve basic medical informatics research projects often can be designed according to methodologies developed for the natural or social sciences. Capitalization efforts can be patterned after commercial product planning and development, and the major question surrounding them relates to the merit of their receiving research grant funding.

Applied medical informatics research, on the other hand, presents unique problems because such research often takes place under conditions that make direct application of methods from the natural sciences difficult. First, the products of applied medical informatics research constantly evolve. Freezing them for formal study is contradictory to the developmental aims of many projects. Second, an applied informatics product must be judged from multiple vantage points. For example, an informatics tool must be acceptable to both the expert who determines whether it does what it is supposed to and the client or user who must decide if it is usable and can provide benefit in an operational environment. Third, precise outcome measures are not always available. "Gold standards" can be used to test diagnostic accuracy, but no clear standards exist for assessing the validity of management suggestions. It may be difficult to factor out the impact of a particular information resource on a decision-making process that has multiple inputs.

Design of Applied Medical Informatics Projects

Design problems inherent in applied medical informatics research may be overcome. First, system development work may be formulated in terms of a testable hypothesis. For example, the investigator might hypothesize that certain elements are necessary to make up a system with specified performance characteristics and then propose prototyping a system that has those elements to test the hypothesis. Either proof or disproof of the hypothesis will contribute to society's general knowledge whether or not a working system is developed. Second, the work may be subdivided into modules, each of which can be developed, tested, and evaluated rigorously. Third, rigorous qualitative studies, while not as definitive as quantitative studies, can provide insight into the reasons for success or failure of a project.

Table 1 ■

Relationship of System Development Stage to Level of Evaluation*

System Development	Evaluation I Definition	II Laboratory Bench	III Laboratory Field	IV Remote Validity	V Remote Efficacy
A Specification	→	↓			
B Component development		↓			
C Combination of components into a system		↱	↱	↓	
D Integration of system into environment		→	↱	↱	↓
E Routine use			→	→	→

*The rows indicate stages of system development and the columns represent the different levels of evaluation. The presence of an arrow in a column of a row indicates that the level of evaluation indicated by the column is appropriate for the stage of development represented by the row. A horizontal arrow indicates that it is appropriate to proceed to the next level of evaluation while the system development stage is unchanged. A vertical arrow indicates that it is appropriate to proceed to the next stage of development without changing evaluation level. A double arrow indicates that it is appropriate to proceed to either the next stage of development or the next level of evaluation.

Staging System Development/Evaluation Projects

Optimal development and testing of complex information systems require subdivision of the project into a series of sequential stages. Each stage builds upon successful completion of its predecessors. This approach makes it easier to determine the cause of failure if it occurs. Similarly, careful evaluation of informatics innovations requires sequential trials with increasing levels of complexity. Many errors in judgement or technique can be identified through early and inexpensive levels of evaluation, reserving more extensive and expensive studies for hypotheses and technologies that pass early trials.

System development can be thought of as taking place in five stages. The first stage involves specification of the work to be performed. A needs analysis is performed; functional specifications are developed; and technical specifications are outlined regarding the hardware and software necessary to provide the function. The second stage consists of developing and testing components or modules. A component should be a small, isolatable subset of a system with a defined input and output. The third stage is the combination of the various components into a system and then testing the resultant complex structure. The expected combinations of inputs and outputs must be tested systematically, either exhaustively or by using sampling techniques. The fourth stage consists of integrating the system into the environment in which it will be used. The system must function as a component of the overall hardware and software infrastructure within the institution. The system must also play a role in the overall social, cultural, and functional environment of its intended users. Testing in this stage emphasizes issues that cross system boundaries, for example, consistency in patient identification or data serialization across distributed databases. The fifth stage begins when the system is used routinely in an operational setting.

Five levels of evaluation can also be identified. The first level consists of problem definition. This level of evaluation should identify formally the need that is to be met. A clear statement should relate proposed efforts to prior investigative research. Literature review, observation, medical audit, and surveys are among the techniques that can be used. Bench testing in the laboratory is the second level of evaluation. A computer scientist may carry out bench testing using his or her own programming skills and a single subject's ideas and experiences regarding medical care. Rapid prototyping is a form of bench testing. Paper cases or scenarios may serve as the basis for initial testing, although use of actual case materials is preferred. The third level of evaluation consists of early field trials under the direct control of the investigator. The goal of these trials is to determine whether the system performs as designed in a realistic environment. Systematic data collection protocols should be utilized. Studies can range in complexity from asking users what they think to utilizing nonrandomized controls. The fourth and fifth levels of evaluation involve field testing in an environment in which the developers are not closely involved. The fourth level of evaluation should ensure that the assessment of validity was not influenced by direct input from the developer and does not reflect conditions unique to the development site. The techniques applicable to this level are similar to those in the third level. The fifth level of evaluation involves study of system efficiency during routine operational use. It includes determination of the effectiveness of a product and the reason for its effect. Randomized trials, inception cohorts, impact studies, and critical incident techniques are among the methodologies that can be utilized.

As depicted in Table 1, there is a relationship between the stage of system development of a project and the level of evaluation that is appropriate. For example, field trials are inappropriate prior to com-

pletion of the specifications, component development, and integration of the components into a system. On the other hand, field trials can be performed at this developmental stage without waiting for the system to be integrated into the environment. As each stage of system development or level of evaluation is completed, the tested and/or evaluated components should be frozen for the duration of subsequent stages of development and/or evaluation. If changes are made, the previously completed steps may well be invalidated and should be repeated.

A "stage 0" may be appropriate for curiosity-based research or exploratory system development. Such a stage would permit people to ask questions such as, "What if we had a new type of machine?" Early, fast, incomplete prototyping should be used in "stage 0" to firm up ideas before specification. It is not necessary that each project progress through each stage. Early stages may be skipped if the investigator is building on work completed by others. Late stages may be omitted if the system changes too rapidly to freeze for evaluation, the necessary evaluative techniques do not exist, or the cost of a study would be prohibitive. Nonetheless, investigators should identify explicitly the current stage of project development and provide references to previous work that forms a basis for their efforts. The evaluative goals and methodology should be appropriate for the stage that is selected.

Additional axes are necessary to define important

Table 2 ■

Questions for Judging the Adequacy of Advanced Studies of Clinical Informatics Innovations*

Purpose of Study				
Therapy	Diagnosis	Screening	Prognosis and Prediction	Quality Assurance
Is the assignment of patients to the intervention and control groups really randomized?	Is the diagnostic aid to be compared blindly with a gold standard?	Is the study a randomized trial? If YES, see Therapy If NO:	Will an inception cohort be assembled?	Is the assignment of patients to the intervention and control groups really randomized?
Are clinically important outcomes assessed objectively?	Is there an adequate spectrum of disease among patients to be tested?	Are there efficacious treatments for the disorder?	Will baseline features be measured reproducibly?	Have the clinical acts under study been shown to do more good than harm? If not, does the study compare process with outcome?
Is the innovation feasible to implement in usual clinical practice?	Is the referral pattern described?	Does the current burden of suffering warrant screening?	Are outcome criteria clinically important and reproducibly measured?	Are the clinical processes or acts measured in a clinically sensible and valid way?
Are follow-up procedures adequate to ensure at least 80% follow-up of participants?	Is the description of the use of the diagnostic aid clear enough to reproduce it?	Does the screening procedure have high sensitivity and specificity?	Are follow-up procedures adequate to ensure at least 80% follow-up of participants?	Are follow-up procedures adequate to ensure at least 80% follow-up of participants?
Is the sample size based on a clinically important but realistic benefit?	Is the diagnostic aid reproducible or is assessment of reproducibility included?	Can the health system cope with the screening program?	Are there separate "derivation" and "validation" samples?	Are both clinical and statistical significance considered?
Is there a search for adverse effects and costs?	Is the contribution of the diagnostic aid to the overall diagnosis assessed?	Will positive screenees comply with intervention?	Will a multivariable statistical model for prediction be used?	Are contamination, co-intervention, and compliance dealt with adequately?

*Adapted from Sacket DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. Second edition, Boston: Little Brown & Co., 1991, p. 367.

aspects of any particular study. Specific evaluation aspects depend on the intended purpose of the innovation. As an example, Table 2 outlines some of the questions for judging the adequacy of various types of advanced clinical medical informatics innovations. Analogous sets of criteria can be used to judge projects involving topics such as library science, learning, human factors, or product factors.

The Evaluation Section of the Grant Application

Many applicants tend to focus upon the methods section of the application because it outlines the work in which they are most interested. Evaluation sections are often appended as an afterthought rather than being developed as an inherent portion of the proposed research. Applicants should begin by identifying the problem that they are addressing, and then include evaluation methods suitable for the specific aims of their proposal. The form of evaluation should be identified prior to writing the methods section of the application. As a result, the writer knows what data will be needed for evaluation, and its collection can be incorporated directly into the methods. The evaluation section then represents a logical extension and validation of the processes described in the methods section.

An applicant must clearly define the criteria used to judge the effectiveness or the outcome of a project. For example, if an applicant plans to employ utilization statistics to measure the effect of a system, the application must indicate how baseline data will be collected, how each new user will be identified, and the threshold level of system use that will be considered important. Reliance upon users to sign a log often produces incomplete data. A better approach would involve use of automatic monitoring programs. A statement such as "utilization statistics will be evaluated to determine if interest is sustained" should be avoided. Instead, a clearer statement would be that the outcome would be considered successful if 50% of the users introduced to the new system continued to use it six months after initial training, with "use" defined according to equally specific criteria.

Virtually all projects are designed with the assumption that they will generate new knowledge, produce a useful tool, or change existing ways of doing work. Projects that allow others to learn from the experience can be valuable even if the outcome is not as intended. The investigator must assume a pose of equipoise with regard to whether the evaluation will indicate that an innovation is helpful or not. Since it is possible that the outcome will be positive or negative, the study design needs to rest upon making sure that the reasons for success or failure are clear. The study design should also be broad enough to detect both intended and unintended effects.

The investigator should select the most rigorous measure suitable to the situation. A quantitative measure is preferred if it is meaningful to the situation and if it can be assessed with reliability and validity. On the other hand, qualitative methods may be more useful in clarifying organizational or cultural issues or in trying to understand why a system component did or did not work. The emphasis should be upon the rigor of the evaluation process and the appropriateness of the technique to answer the question at hand, rather than upon whether the technique is qualitative or quantitative.

Dealing with Uncertainties

Applicants for grant funding are expected to propose a tightly defined piece of work to fit under the research grant criteria. Nonetheless, the horizon they sketch in the proposal for testing their work frequently assumes that the work will produce a completed system. Reasonable detail is often not provided with regard to how all the pieces of the system will be put together or how the work will be evaluated if it proves impossible to put them together.

A different problem occurs when an applicant has in mind solving a clinical information problem and has already put together a system to do so. Some components of the system may work well while others do not. The applicant asks for funding to evaluate the system as a whole, even though only parts of it are really functional. Interpretation of the outcome of the evaluation of such a system will be difficult.

Applicants who are preparing proposals with a three- to five-year time horizon should not feel that they have to predict accurately how a project will progress through the various development stages or evaluations. They should concentrate on demonstrating that they understand the current stage of development of their work. They should propose to employ evaluation methods appropriate for that stage. They should then indicate their assumptions regarding the way the project should progress and the types of evaluation that would be appropriate if those assumptions held true. Proposal reviewers should recognize that unexpected things happen. However, if the applicant makes credible assumptions, and presents a credible plan based upon those assumptions, the reviewers should be willing to assume that all involved will be capable of adjusting the plan to deal with changes that become necessary as the proposed research advances through its evolutionary cycle.

Discussion

The purpose of research is to advance a field by providing lessons, insights, or new knowledge that can be of use to others. The scientific value of a project is related to the degree to which its results can be generalized into defensible principles.

Some applicants who propose information-access projects point out that they are trying to provide a resource and that they are not proposing to do research. On this basis, they may question the need for evaluation. However, a resource project is meant to demonstrate a new way of doing something or whether an important need is met by putting something new in place. Evaluation is necessary to identify the impact of the project and to justify its perpetuation at the home location under non–grant funding or its extension to other locations.

Individuals who are primarily interested in developing or deploying systems often postulate how a system ought to behave, build it, and if it works say, "It works." If the proposed system does not work, its complexity is such that the developers may not learn why it does not work, and the effort may be wasted. Even in situations where the system does work, it may be difficult to separate the system from its physical and organizational operating environment. For an evaluation to provide useful insights to the community at large, the evaluation must isolate systems effects from local environmental effects, and discuss each separately.

Careful study of applied medical informatics research and library-resource projects is necessary to increase the productivity of the research and development enterprise. First, the cost of disproving bad ideas must be limited. Dollars should not be spent implementing information resources or systems until their elements have been adequately bench tested. Similarly, large-impact studies should not be funded until a resource has been proven to perform as specified.

Second, projects should be carried out in ways that produce deliverables, either in the form of shareable tools or as system-independent, transferable knowledge about techniques that lead to success and those that do not. The reasons for success or failure and the contexts in which lessons can be applied should be clear. Ideally, new knowledge should allow generalization to permit prediction of human or system behavior in novel situations.

Finally, ethics require that innovations do more good than harm. It can be argued that the products of medical informatics research support people who are performing tasks or making decisions, and that there is, therefore, minimal chance of direct harm. However, poor systems consume resources that could otherwise be used beneficially, and their mere presence should be considered harmful. Particularly in the case of information-access projects, user-needs analysis should be adequate to ensure that the proposed resource will, if successful, meet a documented need. The lure of technology should not override a careful assessment of how individuals actually use information resources.

The framework proposed in this paper is intended to give investigators ideas about how to improve their research projects by subdividing them into steps and tailoring the evaluation for each step. The key idea is that a relationship exists between a developmental stage of a project and the level of evaluation that is appropriate. As with any model, no project will fit the framework exactly.

Nonetheless, the proposed framework for staged development and evaluation suggests implications for funding agencies. First, it is generally inappropriate to fund an expensive, rigorous study of an innovation that has not had preliminary testing. Second, early prototyping and testing should be supported by funding agencies, preferably by providing rapid review and limited funds. Third, funding agencies that provide money for developmental work on innovations should insist on protocols for bench testing as part of the developmental work. Fourth, funding agencies should discourage the submission for evaluation of protocols that are too many steps ahead of the developmental process and preliminary testing. Such protocols are so dependent on contingencies in development and timing that they may be a waste of time to prepare and review. Fifth, at the level of operational testing of a complete system (e.g., clinical trials), the project team should be relatively independent of the system developers and include collaboration with an investigator who has formal training in applied research methods.

The authors thank Dr. Roger Dahlen for organizing the discussion sessions, Charleta Brown for transcribing the discussions, and Dr. Richard DeMillo for presenting the NSF view of design and evaluation of large software projects.

References ■

1. Miller RA, Patil R, Mitchell JA, et al. Preparing a medical informatics research grant proposal: general principles. Comput Biomed Res. 1989;22:92–101.
2. Warner HR. Medical informatics: a real discipline? First American College of Medical Informatics Lecture. AAMSI Spring Congress, 1988.

CALENDAR OF AMIA RELATED MEETINGS, 1994–1995

APRIL 19–24, 1994: **American College of Physicians** Annual Session, Miami, FL. Contact ACP at 215-351-2800. Meeting will include several AMIA/ACP jointly sponsored workshops.

MAY 4–7, 1994: **AMIA** Spring Congress, Parc Lane's Fifty Five Hotel, San Francisco, CA. Contact AMIA at 301-657-1291. Medical Information and Record Systems—Integration at the Enterprise and Individual Levels. Program Chair: Thomas C. Rindfleisch, Stanford University.

JUNE 19–22, 1994: Fifth International Conference on Nursing Use of Computers and Information Science, **Nursing Informatics '94**, San Antonio, TX. Contact 512-471-7311. IMIA/AMIA jointly sponsored conference.

NOVEMBER 6–9, 1994: 18th Annual Symposium on Computer Applications in Medical Care (**SCAMC**), Sheraton Washington Hotel, Washington, DC. Contact AMIA at 301-657-1291. Transforming Information, Changing Health Care. Program Chair: Judy G. Ozbolt, PhD, RN, FAAN, University of Virginia.

JULY 23–28, 1995: **Medinfo '95**, Vancouver, Canada. Contact Marion Ball at 410-706-2004.