[116] H. R. Warner, "Experiences with computer-based patient monitoring," *Anes. Anaglesia Current Res.*, vol. 47, pp. 453-461, 1968.
[117] H. R. Warner, C. M. Olmsted, and B. D. Rutherford, "HELP—a program for medical decision-making," *Comput. Biomed. Res.*, vol. 5, pp. 65-74, 1972.
[118] H. R. Warner, B. D. Rutherford, and B. Houtchens, "A sequential approach to history taking and diagnosis," *Comput. Biomed. Res.*, vol. 5, pp. 256-262, 1972.
[119] H. R. Warner, J. D. Morgan, T. A. Pryor *et al.*, "HELP—A self-improving system for medical decision making," in *MEDINFO 74*. Amsterdam, The Netherlands: North-Holland, 1974.
[120] H. R. Warner, "Knowledge sectors for logical processing of patient data in the HELP system," in *Proc. 2nd Annu. Symp. Comput. Appl. Med Care* (IEEE, Washington, DC), 1978, pp. 401-404.
[121] R. J. Watson, "Medical staff response to a medical information system with direct physician-computer interface," in *MEDINFO 74*. Amsterdam, The Netherlands: North-Holland, 1974, pp. 299-302.
[122] H. Wechsler, "A fuzzy approach to medical diagnosis," *Int. J. Biomed. Comput.*, vol. 7, pp. 191-203, 1976.
[123] L. L. Weed, "Medical records that guide and teach," *New Eng. J. Med.*, vol. 278, pp. 593-599, pp. 652-657, 1968.
[124] ——, "Problem-oriented medical records," in *Problem-Directed and Medical Information Systems*, M. F. Driggs, Ed. New York: Intercontinental Medical Book Corp., 1973.
[125] S. M. Weiss, C. A. Kulikowski, S. Amarel, and A. Safir, "A model-based method for computer-aided medical decision-making," *Artific. Intell.*, vol. 11, pp. 145-172, 1978.
[126] S. Weyl, J. Fries, G. Wiederhold, and F. Germano, "A modular self-describing clinical databank system," *Comput. Biomed. Res.*, vol. 8, pp. 279-293, 1975.
[127] G. Wiederhold, J. F. Fries, and S. Weyl, "Structured organization of clinical data bases," in *Proc. 1975 NCC*, AFIPS Press, vol. 44, 1975, pp. 479-485.
[128] P. H. Winston, *Artificial Intelligence*. Reading, MA: Addison-Wesley, 1977.
[129] D. Wirschafter, J. T. Carpenter, and E. Mesel, "A consultant-extender system for breast cancer adjuvant chemotherapy," *Ann. Intern. Med.*, vol. 90, pp. 396-401, 1979.
[130] P. M. Wortman, "Medical diagnosis: An information processing approach," *Comput. Biomed. Res.*, vol. 5, pp. 315-328, 1972.
[131] V. L. Yu, L. M. Fagan, S. M. Wraith, *et al.*, "Antimicrobial selection by a computerized consultant: A blinded evaluation by infectious disease experts," *J. Amer. Med. Ass.*, vol. 241, 1979 (in press).
[132] V. L. Yu, B. G. Buchanan, E. H. Shortliffe, *et al.*, "An evaluation of the performance of a computer-based consultant," *Comput. Prog. Biomed.*, vol. 9, pp. 95-102, 1979.
[133] L. A. Zadeh, "Fuzzy sets," *Inform. Cont.*, vol. 8, pp. 338-353, 1965.
[134] N. Zoltie, J. C. Horrocks, and F. T. deDombal, "Computer-assisted diagnosis of dyspepsia—report on transferability of a system, with emphasis on early diagnosis of gastric cancer," *Meth. Inform. Med.*, vol. 16, pp. 89-92, 1977.

Computers and Clinical Decision Making: Whether, How, and For Whom?

PETER SZOLOVITS AND STEPHEN G. PAUKER

Abstract—**Optimal design and implementation of computer-based medical decision aids must give explicit attention to the nature of the intended audience of such programs. The need for sophisticated capabilities and interfaces argues for the incorporation of artificial intelligence techniques. To guard against the premature clinical use of inadequately tested decision aids and to insure both safety and efficacy, we propose a hierarchical scheme for the evaluation of such programs.**

OVER the past two decades, computers increasingly have been a focus for research on the clinical decision-making process; over the next two decades computers undoubtedly will play an important role in the practice of medicine. In the present issue of the PROCEEDINGS, Shortliffe *et al.* provide a coherent summary of computer-aided decision making in medicine [9]. In this paper, we consider how the intended audience of a program should affect its design.

Manuscript received February 15, 1979; revised March 29, 1979. This research was supported in part by National Institutes of Health under Grant 1 P41 RR 01096-02 from the Division of Research Resources. Dr. Pauker is the recipient of Research Carreer Development Award No. 1 K04 GM 00349-02 from the National Institute of General Medical Sciences.

P. Szolovits is with the Clinical Decision Making Group, Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

S. G. Pauker is with the Clinical Decision Making Group, Department of Medicine, New England Medical Center Hospital, Tufts University School of Medicine, Boston, MA.

Who Will Use the Program? Potential users of medical decision-making aids range from subspecialty trained physicians to patients themselves. Paradoxically, the most straightforward audience may be the highly trained specialist who requires well-defined, specific advice—advice which is narrowly focused—recalling specific medical information, interpreting clinical data according to detailed models, keeping track of the course of treatment being given to a patient, etc. In fact, the expert physician's sophistication makes the program design simpler; it assures that he will recognize anomalous results and avoid inappropriate use of the program.

The computer program could also aid the general practitioner who now turns to a human consultant. However, a major obstacle facing such decision aids will be earning acceptance on two separate levels: first, acceptance as a means of improving care or lightening the physicians' load and second, acceptance of the advice provided in an individual consultation, especially if that advice runs counter to the physician's own intuition. The first level of acceptance might be encouraged by making the program convenient and straightforward to use, for example, by providing natural language interaction or some other efficient form of information transfer. Acceptance at the second level will be far more difficult to achieve. Programs will have to *convince* the physician that their suggestions are appropriate. The only mechanism through which

0018-9219/79/0900-1224$00.75

such a dialogue might occur would be interactive explanation—the clinician must be able to pose specific queries and the program must then justify its conclusions [2], [10]. To reply to such queries properly, the program should include models of relevant physiologic processes (at least at the level of detail known by the physician), models of its own behavior, and models of physicians' expectations and preconceived frames of reference. Both the need for easily usable interfaces and for convincing justification of advice suggest the need for programs employing the techniques of Artificial Intelligence (AI).

An alternative tack could be taken—If the program has been tested and validated by the community of medical experts, then its recommendations might require no further justification. Although technically simpler, such an approach is less desirable. The physician's ability to sort out reasonable from unreasonable suggestions makes the task of building a decision aid tractable. The program need not be a compleat physician: it can concentrate on a relatively narrow problem domain because the physician remains responsible for integrating its advice with the total clinical situation. If a program completely "misses the boat" or makes irrational or dangerous suggestions, the physician will (and should) disregard it completely.

With decision aids for nurses, paramedical personnel, or patients themselves, other issues arise. Here, the natural protection provided by the critical clinician is missing, and outlandish suggestions, made because of either inappropriate application of or frank errors in the program, may go undetected. Thus great care must be taken to provide a supervisory environment which could detect errors. Certainly such a supervisory system would need a model of the problem domain, a model of the decision aid and a model of the kind of errors which might occur—"common sense" about medicine. Again, we would argue that programs addressed to these audiences must rely on the tools of AI to accomplish these goals.

Perhaps the simplest, and, therefore, the most likely successful computer aids to clinical decision making will be in conjunction with clinical data bases. Here the program could act as a background monitor, seeking either to prevent serious errors or to identify significant combinations of findings or occurrences which may have gone unnoticed [1], [12], [13]. Neither explanation nor modeling of physician behavior is of primary importance. The program can search for simple associations and conflicts and flag them for attention. Direct interaction of the program with physicians is not essential, and, therefore, rather simple techniques may suffice.

Which Methodologies Should a Program Use? Although simple problems might be solved by simple methods, the capable handling of many medical problems soon requires the addition of complex techniques [11]. For example, although the accurate pharmacokinetic model of digitalis allowed a straightforward program to be used for therapeutic recommendations [5], refinements of that model to account for individual patients' therapeutic goals and special sensitivities, and to interpret clinical manifestations of therapeutic benefit and toxicity have led to the adoption of AI techniques [4]. Two other important methodological arguments for the AI approach can be made.

First, compared with clinical algorithms, AI programs incorporate a much more explicit model of what they "know," thereby facilitating consistent debugging and augmentation. When an error appears to have been made, programs capable of symbolic reasoning permit the identification of specific inadequacies. Weiss has argued [14] that such programs encourage the development of models with sufficient detail to account for the important knowledge needed for solving the problem. Acting on this same insight, Davis has implemented a knowledge acquisition program which uses a specific disagreement between the program's performance and that of an expert to identify possible errors or omissions in the program [2]. If the program's domain is logically consistent, then, in principle, the expert can correct the system's knowledge to achieve complete agreement on every case so far considered. The AI methodology emphasizes the refinement of the underlying model to account for all observed phenomena, whereas the statistical methods tend to acquiesce to simpler models and accept errors as consistent with expected variability.

Second, the typical algorithm never identifies a model of how it uses and interprets clinical information. Inconsistencies may easily arise if the use of clinical data is changed in some places but not others. Therefore, it is both difficult to maintain a large program based on the flowchart technique and nearly impossible for either the program or its writer to give a cogent account of how all clinical data are utilized. In contrast, AI programs are often organized by frames or contexts which provide a framework both for maintenance and explanation. With such an organization the program might be designed to recognize from its own lack of adequate knowledge that it is inappropriate for the case at hand.

Is the Program Worthwhile and Should It Be Released? What problems stand in the way of the widespread adoption of computer aids in medicine? First, most existing programs do not manifest performance which is clearly superior to that of physicians. Second, these programs address sufficiently narrow areas that any particular program is likely to be of little real importance to its user. Finally, no generally accepted criteria exist for establishing that a program is safe and effective, thus leaving ethical and legal doubts in the minds of both program designers and potential users.

Why must a program outperform an expert clinician before it is likely to be accepted? After all, there are major variations in physicians' abilities at expert problem solving [3], [6]. When a doctor turns to a consultant for advice, his expectation is that the consultant will offer truly superior insight and advice. Yet if a program performs better at some specific task than its intended user, then, even if it cannot improve the performance of the best physicians, it can nevertheless raise the performance of the average physician, thus helping to distribute expertise.

Most AI aids for clinical decision making concentrate on a single decision [4], [8], [14], although most potential users of such a system normally deal with a diversity of decision making tasks. Except in a few special circumstances (e.g., [7]) coherent families of diagnostic and therapeutic aids must be developed before we can anticipate widespread acceptance.

A computer decision aid, as any new tool employed in medicine, must be shown to be safe and effective before it can be ethically and legally sanctioned for general use. We propose a hierarchical scheme for program evaluation. This scheme will both protect patients from the premature release of ineffective or potentially dangerous programs and will develop a data base or *panel* of cases which can be used to test any modifications in the program. The basic principle is that the program, as it evolves, should be able to handle correctly *any* clinical case which it handled correctly in an earlier ver-

sion. Initial testing should be performed on a series of prototypical cases. Next, another computer program [2] should be built to search systematically for inconsistencies in the clinical program. Then, a retrospective review should be undertaken, comparing the program's performance to that of unaided clinicians. Next, a prospective review should be mounted in which the program's suggestions are "overread" by experienced clinicians to be sure that no gross errors occur. Finally, a prospective controlled trial should be performed. In both the retrospective and the prospective trials, the computer's performance should be compared to the performance of unaided clinicians, preferably by a panel of experts blinded to which decision maker they are evaluating. In the final phase of evaluation, the impact of the computer program on health outcomes should be assessed. This final phase can only be allowed once the earlier phases of evaluation have certified the program as "safe" for the patient. The early phases of this evaluation sequence might be viewed as analogous to animal trials in the evaluation of a new drug. In the final phase, the prospective controlled trial, great care must be exercised to avoid the "Hawthorne effect," that is, an improvement in physician performance because his behavior is under scrutiny.

References

[1] S. N. Cohen, M. F. Armstrong, L. Crouse, and G. Hunn, "A computer-based system for prospective detection and prevention of drug interactions," *Drug Inform. J.*, vol 72, pp. 81-86, 1972.

[2] R. Davis, "Applications of meta level knowledge to the construction, maintenance and use of large knowledge bases," Ph.D. dissertation, Stanford Artificial Intelligence Laboratory Memo 283, 1976.

[3] A. S. Elstein, L. A. Shulman, and S. A. Sprafka, *Medical Problem Solving*. Cambridge, MA: Harvard University Press, 1978.

[4] G. A. Gorry, H. Silverman, and S. G. Pauker, "Capturing clinical expertise: A computer program that considers clinical responses to digitalis," *Amer. J. Med.*, vol. 64, pp. 452-460, Mar. 1978.

[5] R. W. Jelliffe. J. Buell, R. Kalaba, R. Sridhar, and R. Rockwell, "A computer program for digitalis dosage regimens," *Mathema. Biosci.*, vol. 9, pp. 179-193, 1970.

[6] J. P. Kassirer, and G. A. Gorry, "Clinical problem solving: A behavioral analysis," *Ann. Intern. Med.*, vol. 89, p. 245, 1978.

[7] E. Mesel, D. Wirtschafter, *et al.*, "Clinical algorithms for cancer chemotherapy—systems for community based consultant-extenders and oncology centers," *Meth. Inform. Med.*, vol. 15, pp. 168-173, 1976.

[8] E. H. Shortliffe, *Computer Based Medical Consultations: MYCIN*. Amsterdam, The Netherlands, Elsevier North-Holland Inc., 1976.

[9] E. H. Shortliffe, B. G. Buchanan, and E. A. Feigenbaum, "Knowledge engineering for medical decision making: A review of computer-based clinical decision aids, *Proc. IEEE*, this issue.

[10] W. R. Swartout, "A digitals therapy advisor with explanations," Laboratory for Computer Science, Tech. Rep. TR-176, Massachusetts Institute of Technology, Cambridge, MA, Feb. 1977.

[11] P. Szolovits and S. G. Pauker, "Categorical and probabilistic reasoning in medical diagnosis," *Artific. Intell.*, vol. 11, pp. 115-144, 1978.

[12] H. R. Warner, C. M. Olmsted, and B. D. Rutherford, "HELP—a program for medical decision-making," *Comput. Biomed. Res.*, vol. 5, pp. 256-262, 1972.

[13] L. L. Weed, "Medical records that guide and teach," *New Eng. J. Med.*, vol. 278, pp. 593, 1968.

[14] S. M. Weiss, "A system for model-based computer-aided diagnosis and therapy," Ph.D. dissertation, Computers in Biomedicine, Department of Computer Science, Rutgers University, New Brunswick, NJ, CBM-TR-27-dissertation, June 1974.

COSTAR—A Computer-Based Medical Information System for Ambulatory Care

G. OCTO BARNETT, MEMBER, IEEE, NORMA S. JUSTICE, MICHAEL E. SOMAND, J. BARCLAY ADAMS, BRUCE D. WAXMAN, PETER D. BEAMAN, MONICA S. PARENT, FREDRIC R. VAN DEUSEN, AND JACQUELINE K. GREENLIE

Abstract—**The storage, retrieval, and communication of information are key features of both the practice of medicine and the administration of health care. This paper describes a COmputer-STored Ambulatory Record (COSTAR) which replaces the traditional document-based patient medical record with a comprehensive, centralized, and integrated information system. COSTAR meets both the medical care and the financial/administrative needs of a variety of different medical practices (whether fee-for-service or prepaid) and can be implemented and operated without on-site programming support. COSTAR has a modular design to facilitate phased implementation, and uses a comprehensive dictionary of terms to standardize and store data. The physician records medical, administrative, and financial information on a single source document (the encounter form); data are input by clerical personnel; information is retrieved via different computer-generated displays and printouts which automatically select and organize the data. The system provides a high-level language which allows the user to access the database from a logical point of view and perform searches or prepare reports without programming support. COSTAR is available on minicomputers using commercially supported software and will be marketed by commercial organizations.**

Manuscript received December 7, 1978; revised March 20, 1979. This work was supported in part by the National Center for Health Services Research, OASH under Grant HS 00240.

G. O. Barnett, N. S. Justice, M. E. Somand, J. B. Adams, P. D. Beaman, M. S. Parent, and F. R. Van Deusen are with the Laboratory of Computer Science, Massachusetts General Hospital, Fruit Street, Boston, MA 02114.

B. D. Waxman is with the National Center for Health Services Research, Hyattsville, MD 20782.

J. K. Greenlie is with MDP Software Development, Digital Equipment Corporation, Marlboro, MA 01752.

I. Introduction

THE communication of medical information is a critical element in the practice of high-quality medical care. Traditional recording practices rely almost completely on a manual record folder where physician notes are handwritten or dictated and merged with laboratory data and other

0018-9219/79/0900-1226$00.75