Report 85-21
Stanford -- KSL

Scientific DataLink

A Study of the Treatment Advice of a Computer-Based Cancer Chemotherapy Protocol Advisor. David H. Hickam, Edward H. Shortliffe, et al., 1985

A STUDY OF THE TREATMENT ADVICE OF A COMPUTER-BASED CANCER CHEMOTHERAPY PROTOCOL ADVISOR

David H. Hickam, M.D., M.P.H.
Edward H. Shortliffe, M.D., Ph.D., F.A.C.P.
Miriam B. Bischoff, M.S.
A. Carlisle Scott, M.S.
Charlotte D. Jacobs, M.D.

Medical Computer Science Group
Knowledge Systems Laboratory
Departments of Medicine and Computer Science
Stanford University School of Medicine

Running title: Cancer Chemotherapy Protocol Advisor

This research was supported by NIH Grants LM-03395, LM-00048, RR-00785, and RR-01631. Dr. Hickam, previously a Henry J. Kaiser Family Foundation General Internal Medicine Fellow, is now with the Division of General Medicine, Veterans Administration Medical Center, Oregon Health Sciences University, Portland, Oregon 97201. Dr. Shortliffe is a Henry J. Kaiser Family Foundation Faculty Scholar in General Internal Medicine. Dr. Jacobs is an American Cancer Society Junior Faculty Scholar. This study was presented in part at the Fifth Annual Meeting of the Society for Medical Decision Making, October 1983. Reprint requests should be directed to Dr. Shortliffe, Room TC-135, Stanford University School of Medicine, Stanford, California 94305.

A STUDY OF THE TREATMENT ADVICE OF A COMPUTER-BASED CANCER CHEMOTHERAPY PROTOCOL ADVISOR

David H. Hickam, Edward H. Shortliffe, Miriam B. Bischoff,
A. Carlisle Scott, and Charlotte D. Jacobs

ABSTRACT

A computer-based cancer chemotherapy protocol advisor, termed ONCOCIN, has been implemented for experimental use in a university oncology clinic. It uses artificial intelligence techniques to guide the treatment of patients enrolled in chemotherapy protocols. The program combines formal protocol guidelines with judgments of cancer experts who have experience adapting such protocols for aberrant clinical situations. The quality of the program's advice is one of several important evaluation questions for a system of this kind.

We compared the chemotherapy administered by clinic physicians with the treatment plan that would have been recommended by ONCOCIN in 415 clinic visits for 39 lymphoma patients seen prior to the program's introduction. In 189 visits the computer agreed with the therapy actually administered. A random sampling of the clinic visits was reviewed by four lymphoma therapy experts who were asked to assess the adequacy of a proposed treatment without knowing whether the plan was that of a physician or of ONCOCIN. There was no significant difference between the experts' assessments of ONCOCIN and of the physicians. Subanalyses showed that ONCOCIN tended to attenuate drug doses more than the experts would have done, whereas the physicians were less likely to attenuate doses to the extent suggested by the expert evaluators. ONCOCIN was also more likely to delay treatment when the experts judged that therapy could have been given. Our results show that ONCOCIN provides lymphoma treatment advice comparable to the treatment provided in a university oncology clinic.

BLANK PAGE

INTRODUCTION

It is a major challenge for clinicians to maintain proficiency in those areas of medicine in which diagnostic and treatment modalities are rapidly changing. Physicians have traditionally used consultants to help maintain this proficiency, and consultation networks are a central component of the modern health care system. The recent rapid evolution of computing technology has led to serious consideration of the use of computer-assisted medical decision making systems to provide medical consultation when expertise is not otherwise readily available [1, 2]. However, such systems have not yet gained widespread use, largely because optimal design criteria and performance characteristics have only recently begun to be defined [3]. Before such systems are introduced for routine clinical use, it is imperative that their decision making performance be formally validated. We report here an initial evaluation of the advice given by the ONCOCIN system, a computer-based cancer chemotherapy protocol consultant.

The clinical problem ONCOCIN addresses is the management of cancer chemotherapy for patients participating in formal protocol-based clinical trials. A major achievement of the current medical era is the improved prognosis of cancer patients resulting from innovations in oncologic therapy. While paying clear dividends in reduced mortality from neoplasms, these innovations usually have come at the price of the increased complexity and potential morbidity of multi-drug or multi-modality therapy. Such trade-offs have resulted in a unique set of challenges for the clinical oncologist, who must become expert in managing the specific protocols applicable to the treatment of each of many different tumors. Failure to apply treatment protocols properly may jeopardize a patient's chances for successful treatment. Recent reports of experience in the therapy of breast cancer [4] and Hodgkin's disease [5] have shown direct correlations between protocol drug dosages and treatment success rates. These results imply that techniques for improving protocol management may directly improve patient care.

There are several possible avenues for optimizing protocol management, among them

the use of computers as tools in the practice of oncology. Until recently, most oncology-related computer applications have involved the maintenance of patient databanks and have not directly affected care of individual patients [6]. However, a few computer systems have been developed to assist providers through the use of protocol-derived algorithms [7, 8, 9, 10]. While these systems successfully capture the algorithmic logic, they require that protocol managers enter data into the databanks and manage program operation. Such systems may also require considerable time to prepare the algorithm for each clinic visit [7, 8]. While such time commitments are possible in a research setting, they seriously limit the feasibility of implementing such prototype systems in the general community.

ONCOCIN was developed to meet the need for a comprehensive computer system which could provide protocol guidance while enhancing the management of clinical trial data and facilitating enrollment of oncology study patients under care by community physicians. Because ONCOCIN is a comprehensive data collection and management advice system, several components of its performance need evaluation. This report describes an evaluation of ONCOCIN's treatment recommendations. We have chosen to focus on this aspect of ONCOCIN's performance because a crucial component of the program's acceptability to clinicians is the demonstration that it suggests management decisions that are acceptable to physicians.

METHODS

The ONCOCIN System:

ONCOCIN is a medical consultation system that has been influenced by a predecessor infectious disease advice program known as MYCIN [11]. ONCOCIN uses knowledge about chemotherapy protocols for Hodgkin's disease and non-Hodgkin's lymphoma to provide advice on drug selection, dosing, and test selection. Technical aspects of the system's operation have been described elsewhere [12, 13] and are summarized in the Appendix to this report. Most knowledge is recorded in decision rules, some of which reflect judgmental expertise acquired from experts in addition to the "core" knowledge available from the protocol descriptions

themselves. Because it depends upon artificial intelligence techniques and judgmental knowledge acquired from expert oncologists, ONCOCIN is appropriately described as an "expert system" [14]. The program is integrated into the physicians' clinical activities so that it provides online data management as well as advice. All consultations involve the physicians themselves, with ancillary personnel using the system only for chart review and to enter laboratory data that become available between clinic visits.

ONCOCIN's role can be understood by considering how a physician or other provider typically participates in a clinical trial (Fig. 1). There are generally rules that define the treatment options and that assure a well-controlled experiment with adequate data for analysis. The provider records clinical data such as patient symptoms, physical findings, laboratory data, and treatment onto a flowsheet (datasheet), which is later transcribed into a patient databank for statistical analysis. ONCOCIN replaces the paper flowsheet with a consultation system designed to acquire and record the same data that were previously written by the physician on a datasheet (Fig. 2). Because protocol-guided management decisions are generally based on flowsheet data, ONCOCIN can combine a patient's clinical information with encoded protocol guidelines, plus additional knowledge provided by expert oncologists, to provide management advice and to remind the user about additional data that need to be collected.

All interaction with ONCOCIN takes place at a conventional high speed video display terminal equipped with an easily learned 21-key keypad that allows the physician to avoid typing [15]. Fig. 3 shows the appearance of the screen during data entry. The mid-portion of the screen mimics the appearance of a flowsheet, with the current date's data in the rightmost column, and there is an explanation region at the top which indicates acceptable responses for the entry about to be made. The physician directly enters clinical data at the terminal and ONCOCIN verifies their accuracy and completeness. After the physician has entered the patient's data, ONCOCIN "fills in" the chemotherapy portion of the flowsheet, providing an explanation for its recommendation and allowing the physician to select a different therapeutic plan if desired (Fig. 4).

Clinical setting:

ONCOCIN has been used experimentally since May of 1981 to assist with the management of lymphoma protocol patients in the Oncology Day Care Center at Stanford University Medical Center. Newly diagnosed patients with Hodgkin's disease or non-Hodgkin's lymphoma are invited to participate in randomized treatment protocols if they meet eligibility criteria. After furnishing informed consent, protocol participants are treated and followed in a lymphoma chemotherapy clinic. This clinic is staffed by post-doctoral oncology fellows, under the supervision of faculty in the oncology division of the Department of Medicine. Periodically the patients are also seen in a special clinic staffed by the principal investigators of the lymphoma protocol studies.

In order to compare the quality of ONCOCIN's lymphoma therapy recommendations with the treatment decisions made by Stanford oncologists prior to the program's introduction, all pre-ONCOCIN lymphoma chemotherapy clinic visits between July 1, 1980, and May 15, 1981 were included in our study. If a study subject had lymphoma clinic visits for treatment under the same protocol prior to July 1, 1980, then these earlier clinic visits were also included. If a patient was changed to an alternate therapy plan because of toxicity limitations or progressive disease, subsequent visits were deleted from consideration. In addition, no visits that occurred after ONCOCIN was introduced were included. None of the physicians who provided treatment had an association with the ONCOCIN project and none had knowledge of the developing computer program.

Tissue specimens for all subjects were reviewed by our pathology department and were categorized according to the Rappaport classification [16]. Subjects with non-Hodgkin's lymphomas were identified as having either favorable histologies (nodular lymphocytic poorly differentiated lymphoma, diffuse lymphocytic well-differentiated lymphoma, nodular mixed histiocytic lymphocytic lymphoma) or unfavorable histologies (diffuse lymphocytic poorly differentiated lymphoma, diffuse mixed histiocytic lymphocytic lymphoma, and nodular or diffuse histiocytic lymphoma) [17]. Protocols were classified as complex if they involved

alternation between chemotherapy and radiation therapy or between two different chemotherapy regimens. Otherwise, the protocols were classified as simple.

Database used for expert evaluation:

The usual time-oriented paper flowsheet incorporating data on symptoms, physical examination, Karnofsky performance status, laboratory data, treatment given, and drug toxicity was maintained for each lymphoma clinic patient. These flowsheets comprised the database used for our evaluation. Each patient's hospital chart was reviewed by one of the study physicians to compare the flowsheets to the written progress notes, treatment records, and laboratory reports. The flowsheets were then updated to insure that the database was as complete as possible.

The data from the time-oriented flowsheet were entered into ONCOCIN's patient database using the interface program described above, and two individual evaluation sheets for each clinic visit were generated and printed by the computer. The two evaluation sheets were identical with regard to the baseline clinical data (patient's age, tissue diagnosis, clinical stage, and protocol assignment) and data from all previous clinic visits (symptoms, disease status, blood counts, signs of drug toxicity, and treatment administered). Both sheets also included the patient's symptoms, physical findings, and laboratory data from the current clinic visit. However, Evaluation Sheet 1 (ES-1) included the treatment actually delivered by the physician for that visit (without advice from ONCOCIN), whereas Evaluation Sheet 2 (ES-2) included the treatment that ONCOCIN would have recommended for that visit if the program had been available. Note that the past data and drug doses included in ES-2 reflected the *actual* therapy given to the patient; only the final column included the therapy recommendation from the computer program.

A systematic comparison of the two evaluation sheets (ES-1 and ES-2) was conducted for each clinic visit to determine whether there was a significant difference between the treatment actually delivered by the physician and the treatment that would have been recommended by ONCOCIN. We had previously determined by a consensus of collaborating

expert oncologists how large a difference in dosage of the various drugs used in the protocols should be considered substantial. For example, a difference in prednisone dosing within 10 mg, reflecting a physician's decision to simplify the regimen for a patient, was not considered a serious disagreement with ONCOCIN's recommended dose of the drug. The visits in the study were divided into subgroups containing those in which ONCOCIN agreed with the physician's treatment and those in which there were substantial treatment differences. To limit the number of cases requiring review by expert oncologists, random samples for detailed study were selected from each of the two subgroups. The individual clinic visit, rather than the set of all clinic visits for a patient, comprised the unit of analysis in this study.

Expert evaluation of clinic visits:

The expert evaluation consisted of global ratings of the treatment regimens for each visit by one of four experienced oncologists. All four physicians were Stanford faculty members in the Departments of Medicine or Radiation Therapy with expertise in managing lymphoma patients at our institution. Each was given a packet containing data from a random sample of all clinic visits included in the evaluation. Each packet contained a balanced mixture of ES-1 and ES-2 sheets, but only one evaluation sheet from each visit was included. The experts were not told the patients' identities or whether a particular sheet reflected actual therapy administered by the physician (ES-1) or treatment recommended by ONCOCIN (ES-2). ONCOCIN rounds off all dosage recommendations to amounts that correspond to available pill sizes and injectable drug dilutions. Therefore, all ONCOCIN's recommendations could have been administered precisely as prescribed, and the experts were unable to infer the source of treatment plan from the drug dosages listed on the evaluation sheets. Several weeks after completing the initial evaluation sheets, each expert was given a second packet identical to the first, except that each ES-1 was replaced with the corresponding ES-2 for that clinic visit, and vice versa. Thus, the same expert evaluated both the physician- and ONCOCIN-recommended treatment for each visit.

Each evaluation sheet was accompanied by a rating form with two questions (Table 1).

The experts answered these questions for the treatment described on the flowsheet's current clinic visit only. Responses on the form allowed both an assessment of the clinical significance of a departure from ideal therapy, and the expert's reason for considering a treatment less than ideal. Previous pilot testing of the evaluation form with oncologists had shown that a four-category overall rating scale was sufficient to represent the experts' opinions. The oncologists had indicated that a scale with more than four categories was unrealistic for this rating task.

Data analysis:

Categorical variables were compared using the chi-square or Fisher's exact test. Scaled variables were compared using Student's t-test. Standard computer programs were used to perform multi-way analysis of variance [18]. We measured inter-observer reliability by having each rater review a small set of visits previously rated by a different expert. These repeat ratings were used only for inter-observer comparisons and were not used for the direct comparison of physician- and ONCOCIN-recommended treatments. The weighted kappa statistic was used for inter-observer comparisons [19].

RESULTS

Patient characteristics:

Thirty nine patients having 415 clinic visits were entered into the study. The patients had a mean (+1 S.D.) age of 41.2 (+12.8) years, and 55% were male. Of the 415 clinic visits, 57% involved patients with Hodgkin's disease, 24% involved patients with favorable histologies of non-Hodgkin's lymphoma, and 19% involved patients with unfavorable histologies of non-Hodgkin's lymphoma. Fifty six percent of the visits were for treatment on simple protocols, and 44% were for treatment on complex protocols. Patients were seen by first year oncology fellows in 52% of visits, by second year oncology fellows in 40%, by hematology or radiation therapy fellows in 4% and by faculty physicians alone in 4% of visits.

Comparison of physician- and ONCOCIN-recommended treatment:

Of the 415 total visits, ONCOCIN's treatment recommendations agreed with the treatment provided by physicians in 189. A 25% random sample (N=47) of these agreement visits was selected for expert evaluation. Systematic review of the patient flowsheets indicated that, in 29 of the remaining 226 visits, the source of disagreement between ONCOCIN and the physician was identical to the source of disagreement in an earlier visit for that patient. Thus, there were unique disagreements between ONCOCIN and the physicians in 197 of the 415 study visits. A 70% random sample of these disagreement visits (N=137) was selected for expert evaluation.

As anticipated, the experts gave higher ratings to the treatment delivered in visits in which ONCOCIN and the physicians agreed than in visits showing disagreement between the two sources of treatment (Table 2). Using a numerical scale for the rating categories shown in Table 1 (1 = unacceptable, 4 = ideal), the analysis showed a mean rating of 3.57 for the visits in which ONCOCIN agreed with the physicians' treatment. For the visits in which ONCOCIN and the physicians disagreed, the mean (+S.E.) rating was 3.10 (+0.09) for the physicians' treatment and 3.06 (+0.09) for the ONCOCIN-recommended treatment (p=0.8). For the disagreement visits, we also calculated mean ratings by source of treatment for each individual expert rater (Table 3). A two-way analysis of variance (with source of treatment and identity of rater as factors) revealed no significant difference in ratings between treatment sources or among raters.

As described previously, random samples of the clinic visits were submitted for expert evaluation. Proportional extrapolation of the ratings in Table 2 to all 415 clinic visits included in this study reveals that the physicians' treatment would be ideal or acceptable in 78% of visits. ONCOCIN's recommendation would be ideal or acceptable in 79% of visits. The sample size used in this study provides a statistical power of 80% to conclude that there is no more than a 10% difference in the rates of acceptable and ideal treatment between ONCOCIN and the physicians [20].

Although we found no overall ratings difference between the physicians and ONCOCIN, it is important to determine whether there may have been subtle differences in the treatments which were not reflected in the global ratings. In no subgroup of clinic visits was there a significant difference between the ratings given ONCOCIN and the physicians. We stratified the visits according to four clinical criteria: histologic diagnosis, protocol complexity, protocol cycle, and experience of physician. The ONCOCIN and physician ratings were equal in each stratum of these criteria.

A multivariate statistical model confirmed the equality of the ratings for ONCOCIN and the physicians. We performed a five-way analysis of variance of the ratings, using the source of treatment (ONCOCIN vs. physician) and the four clinical criteria listed above as factors. Neither treatment source as a main effect, nor any interaction terms, had a significant effect upon the ratings. We also performed separate four-way analyses of the ratings given ONCOCIN and the physician treatments. None of the clinical criteria had a significant effect upon the ratings given either source of treatment. We did not identify subsets of visits in which either ONCOCIN or the physicians performed particularly poorly.

The expert raters indicated the reasons for their ratings in all cases in which their ratings were less than ideal. These reasons are tabulated in Table 4 for the 137 evaluated visits in which the physician treatment disagreed with the ONCOCIN-recommended treatment. The physicians tended more often to order drug dosages that were higher than those recommended by the experts, and ONCOCIN more often suggested dosages that were lower. Because the data of Table 4 suggested a systematic difference in drug dosing between ONCOCIN and the physicians, we analyzed the visits in which the experts identified drug dosages as being too high or too low (Table 5). This analysis revealed significant differences between ONCOCIN and the physicians. ONCOCIN had a greater tendency to attenuate drug dosages excessively, while the physicians less frequently attenuated or escalated doses to the extent recommended by the expert evaluators.

ONCOCIN had a slightly higher rate of unacceptable ratings than the physicians

(Table 2). In order to determine whether this represented a systematic flaw in ONCOCIN's treatment advice, each visit given an unacceptable rating was classified by the treatment decision (Table 6). The major difference between ONCOCIN and the physicians was the tendency to excessive treatment delays by ONCOCIN. Of the seven visits in which ONCOCIN was judged to have made an unacceptable treatment delay, the patient's white blood cell count was below protocol definitions of normal in six. In five of these six visits, the experts indicated that the patient should have been treated with usual drugs anyway. In the sixth visit, the expert indicated that the patient should have been treated with a substitute drug. In the seventh unacceptable delay, ONCOCIN chose to delay treatment following involved field radiation therapy, as directed by the protocol, while the expert preferred to go ahead and give treatment with the usual drugs.

We also measured the agreement among the experts' ratings. An approximation of inter-rater agreement is provided by the data in Table 3, in which there was no significant difference in the mean ratings of the four raters. However, this is not a true measure of inter-observer reliability because each rater reviewed a different subset of the clinic visits. We directly measured interobserver reliability in a small random sample of clinic visits. Four inter-rater comparisons were made, with between ten and twenty pairs of ratings in each comparison. Inter-observer agreement was substantial (kappa .79) for one pair, fair (.38 and .34) for two pairs, and slight (.11) for one pair.

DISCUSSION

Our results demonstrate that the ONCOCIN system provides lymphoma protocol advice similar to the treatment that was delivered in the Stanford lymphoma clinic before ONCOCIN's introduction. In 46% of the clinic visits, ONCOCIN's recommendations were identical to the treatment delivered by physicians. Of the remaining visits, expert oncologists found that ONCOCIN's treatment recommendations were equally acceptable to the treatment actually delivered (Table 2). These findings are particularly notable in that we compared ONCOCIN to the treatment delivered in the clinic supervised by the group of faculty

oncologists who developed the lymphoma protocols being studied. Stanford is a national referral center for patients with lymphoma, and ONCOCIN accordingly appears to provide treatment advice that closely emulates expert-level patient care. While ONCOCIN does not improve upon the care provided by oncologists at a tertiary care center, it clearly has the potential to facilitate the dissemination of expert-level knowledge to other medical-care settings.

While ONCOCIN's treatment advice is of high quality, there are subtle differences between ONCOCIN's advice and the treatment delivered in the clinic. Compared to the physicians, ONCOCIN was more likely to attenuate or escalate drug dosages appropriately but also more often attenuated dosages excessively (Table 5). This finding suggests that ONCOCIN makes larger changes in drug dosages than the physicians by being more sensitive to protocol guidelines for dosage changes. While this attribute allowed ONCOCIN to make appropriate dosage changes that were overlooked or disregarded by the physicians, it also caused ONCOCIN sometimes to make dosage attenuations that were considered excessive by the experts. An important benefit of our evaluation of ONCOCIN is that it has defined those areas of the clinical micro-environment in which ONCOCIN's rules (and the chemotherapy protocols themselves) need further development and refinement.

Our evaluation has discovered a specific clinical situation in which ONCOCIN's knowledge requires further adjustment: the management of protocol delays. Although the experts occasionally criticized both the physicians' and ONCOCIN's management of treatment delays (Table 4), ONCOCIN's treatment was more often judged unacceptable in this setting (Table 6). ONCOCIN's behavior was consistent regarding treatment delays and represented a strategy of waiting for blood counts to return to normal levels before beginning the next chemotherapy cycle. The experts' ratings seemed to indicate that they use a lower definition of normal as the threshold for reinitiating treatment after a delay (i.e., a threshold lower than that described in the written protocols). The physicians working in the Stanford oncology clinic were aware of this change in strategy and avoided such treatment delays. If the physicians had

followed closely the written protocol guidelines, they would have made the same unacceptable treatment delays as made by ONCOCIN. These results underscore the importance of incorporating supplementary knowledge beyond that included in written protocols into an expert computer system.

The overall quality of ONCOCIN's treatment advice compares favorably to the performance of previously developed medical expert systems. ONCOCIN's framework for representing medical knowledge is similar to that used by MYCIN, an expert system that provides infectious disease therapy advice [21]. The advice provided by MYCIN has been formally evaluated in two studies [22, 23]. In each study, a panel of infectious disease experts was asked to rate the adequacy of MYCIN's treatment recommendations. In the first study [22], 73% of MYCIN's treatment recommendations for patients with positive blood cultures were judged acceptable. In the second study [23], 65% of MYCIN's treatment recommendations for patients with meningitis were judged acceptable. In our study, ONCOCIN's treatment advice was judged ideal or acceptable in 79% of the clinic visits.

Our study has relied upon the treatment decisions of physicians as a standard against which to compare ONCOCIN's performance. It is accordingly important to verify that the physicians' treatment is an adequate standard for comparison. It is difficult to define "gold standards" for patient treatment decisions. Expert decisions are based upon the synthesis of a network of clinical data and past experience which usually can be defined only incompletely [24]. This often leads to disagreement among experts on the adequacy of individual treatment decisions. In the second MYCIN evaluation study [23], for example, outside infectious disease experts rated the treatment decisions of five Stanford infectious disease faculty members to be acceptable in only 43% to 63% of cases. These data reflect the opportunity for disagreement, even among experts, in areas of medical practice where decisions are based partly upon experience and intuition. It accordingly may not be reasonable to expect a higher level of agreement with expert opinion than the 79% documented for ONCOCIN in this clinical domain.

There has been limited previous experience with computer-based oncology treatment advisers. Three other systems with which we are familiar have been designed to operate on smaller computers than the one required by ONCOCIN and to provide treatment advice using protocol-derived algorithms. An important feature of our evaluation is that, unlike studies of the other computer-based protocol systems [7, 8, 10], ours used the judgment of lymphoma experts rather than adherence to the protocol as its standard of adequate treatment. By supplementing protocol guidelines with expert knowledge, ONCOCIN provides an important extra dimension not possessed by the algorithm-based protocol systems.

Of the other systems, the most extensively studied was implemented in a multi-center setting in Alabama and provided treatment advice for patients enrolled in Hodgkin's disease and breast cancer adjuvant chemotherapy protocols [7, 8]. While the investigators detected a very low rate of protocol violation, their criterion for protocol compliance was adherence to the computer-generated algorithm. This approach does not provide an external expert evaluation of the algorithm-creation system. The Alabama system also requires a consulting oncologist to manage some situations which cannot be handled by the computer, and neither of its published reports provides data on how often this was necessary. ONCOCIN was able to provide treatment recommendations in all clinic visits in our study.

Neither of the two other algorithm-based protocol advisers has been studied as extensively as the Alabama system. Friedman [10] has reported on initial experience with a system used in Wisconsin. He found increased physician compliance with protocol guidelines after the system was introduced, but once again the criterion for protocol compliance was agreement with the computer-generated algorithm. The study did not use an external measure of the adequacy of patient care. A third algorithmic system, the CDMS program at Boston University [9], has not yet been the subject of a formal evaluation.

Other computer-based medical advice systems have been developed and used in patient care settings. Among them are at least three that have received some published evaluation, although none has been designed to provide advice for oncologic treatment. Perhaps the most

extensively studied of these systems is the Regenstrief Medical Record (CARE) [25, 26, 27]. CARE gives providers feedback on a variety of patient treatment issues and has been shown to alter physician behavior significantly. However, there has not yet been a published blinded evaluation of the quality of its advice, as in our study of ONCOCIN. Two other well known programs are the COSTAR [28] and HELP [29] systems. Both use internal logic to provide advice, but neither has been evaluated using an external standard of care.

ONCOCIN has has several features which provide benefits not available with other computer-based medical advice systems. ONCOCIN directly interacts with the physician and does not require that data managers transcribe data recorded on paper by the provider. Although HELP is used directly by physicians, both COSTAR and CARE have relied on paper reports for communication to and from providers. We have found that direct interaction with physicians has a beneficial impact on the capture of those data required by a protocol [30]. ONCOCIN also has an explanation feature that provides the physician with the reasons for its recommendations. Therefore, it more closely emulates a medical consultant and allows the physician to consider carefully whether ONCOCIN's recommendations should be followed. Because ONCOCIN uses a rule-based knowledge structure (see Appendix), it can be modified with relative ease to adapt to modifications in treatment protocols or to the introduction of new protocols. Since the completion of the present study, ONCOCIN's knowledge base has been revised to reflect recent changes in the Stanford lymphoma protocols. The Stanford breast cancer and small cell lung cancer protocols have also been implemented, and those for breast cancer are currently available in the clinic. ONCOCIN therefore shows considerable promise as a system capable of providing treatment advice for a wide range of cancer protocols; this will greatly increase its utility as an oncology treatment advisor.

This report does not address three additional important elements in the evaluation of a medical information management system: system acceptance, impact on data management, and effect on patient outcomes. Our preliminary experience with ONCOCIN is that it is well accepted by clinicians, and a formal attitude and acceptance study is underway. An assessment

of ONCOCIN's impact on data collection has shown that ONCOCIN significantly improves the completeness and consistency of clinical trial data [30]. Patients managed using ONCOCIN have not been followed long enough to report a meaningful outcome study at this time. However, confounding variables related to other changes in cancer treatment over time make it unlikely that a valid outcome study of this type can realistically be undertaken.

Recent advances in computer hardware are permitting ONCOCIN to be transferred from a large mainframe computer (as used in this study) to desk-top personal workstations. While the version of ONCOCIN for use on such machines currently is undergoing testing prior to its clinical introduction, there appear to be few obstacles to the program's eventual effective use on office computer systems of moderate cost. This development will allow ONCOCIN to evolve from its current use as a research tool to a future role as a treatment advisor for community oncology trials. Our evaluation of ONCOCIN's costs and benefits awaits completion of the small computer version and its implementation in a community setting. Physicians have shown a growing acceptance of computer-based therapeutic advice systems [31]. We expect that ONCOCIN will ultimately allow community physicians and their patients to participate in chemotherapy protocol trials that have previously been difficult to perform outside of university centers.

Acknowledgments

The authors gratefully acknowledge the contributions to ONCOCIN's development and evaluation by B. Campbell, R. Carlson, J. Cho, J. Differding, B. Elspas, L. Fagan, J. Ferguson, P. Gerring, J. Gordon, R. Hoppe, S. Horning, D. Kent, M. Koretz, P. Kushlan, C. Lane, C. Langlotz, C. Lawton, J. Lee, M. Musen, M. Safran, R. Teach, C. Tovey, S. Tsuji, W. van Melle, and D. Vian.

Appendix

We describe here in greater detail some of the technical aspects of ONCOCIN's performance and logistical issues regarding the system's use in the Stanford clinic. It should be emphasized, however, that the current prototype system uses computing resources that are inappropriate for dissemination to other sites. It is only after ONCOCIN's reimplementation and testing on desk-top personal workstations suitable for integration into private offices and clinics that we expect more widespread availability of the system.

ONCOCIN has been designed to run on a customized video display terminal that has been installed in Stanford's oncology clinic. The terminal uses a routine character-oriented display screen, but the keyboard has been modified so that a physician can enter flowsheet information using only a 21-key pad to the side of the conventional keys. Almost no typing is required. The terminal is driven at 9600 baud, a line speed that gives the appearance of nearly instantaneous filling of the screen.

After examining a patient, the physician interacts with ONCOCIN's data entry program (the *Interviewer*) and reviews time-oriented data from the patient's previous visits to the clinic. The screen display is designed to look like the flowsheets that have traditionally been used for recording patient data (Fig. 3). The physician uses the terminal keypad to enter information regarding the current visit; this interaction replaces the usual process of filling out the flowsheet by hand. As the physician enters data, the *Interviewer* passes the information to a second program known as the *Reasoner*. This second process is a rule-based program, written in a computer language called Interlisp, that attempts to determine optimal therapy for the patient based upon its knowledge of the protocol and the information that the physician is entering on the flowsheet. By the time the physician reaches the part of the flowsheet for the current therapy plan, the *Reasoner* has generated a therapy recommendation which it has passed back to the *Interviewer* for inclusion on the flowsheet (Fig. 4). Thus the therapy required by the protocol is already filled in for the physician. Both the *Reasoner* and the *Interviewer* also use a shared database of prior patient data. The two programs run in parallel on a single

machine (a Digital Equipment Corporation DEC-20); this separation of the interface program from the reasoning program allows the user to receive fast responses when entering flowsheet values; the slower Interlisp program is not allowed to degrade the speed of the interaction[1]. Before terminating interactions, physicians can examine the explanation provided with each recommendation. They may approve or modify ONCOCIN's recommendation; any changes are noted by the system and kept available for future review.

Knowledge about the oncology domain is represented using four main types of data structure: Contexts, Parameters, Rules, and Control Blocks [12]. As in the MYCIN system [11], rules are represented in a stylized format so that they may be translated from Interlisp into English for explanation purposes. Below is the English translation of one of ONCOCIN's 400 current rules:

RULE075

```
To determine the current attenuated dose for all drugs in MOPP
   or for all drugs in PAVe:

If:  1) This is the start of the first cycle after a cycle was
        aborted, and
     2) The blood counts do not warrant dose attenuation
Then: Conclude that the current attenuated dose is 75 percent of
        the previous dose.
```

ONCOCIN is currently available for use only with patients enrolled in protocols that have been encoded in the system, viz. those for Hodgkin's and non-Hodgkin's lymphomas and breast cancer. A physician preparing to see a clinic patient whose management data are in the ONCOCIN database finds a computer-generated summary sheet attached to the front of the patient's chart. These sheets provide an overview of the patient's course on previous visits, drug toxicity that has occurred, dose attenuation that may have been necessary, radiotherapy, and certain lab or physical abnormalities. Thus the sheet guides the physician in the assessment of the patient upon entering the examining room. Instead of filling out the flowsheet by hand, the physician uses the ONCOCIN terminal after seeing the patient.

---

[1]The new workstation version of ONCOCIN mentioned earlier also uses this parallel processing architecture.

ONCOCIN provides a therapy recommendation at the end of the interaction which the physician may chose to follow or reject. The physician is reminded about tests that are required by the protocol for the patient's next clinic visit, and approves or alters the followup plans proposed by the system. An encounter sheet is then printed at the clinic check-out desk. By the time the patient reaches the desk to check out, clinic staff already know what tests must be ordered and when the patient should be scheduled back to clinic.

BLANK PAGE

Legends to Figures

Figure 1: Diagram demonstrating the typical transfer of information in collecting data for clinical trials.

Figure 2: Proposed model of a computer-based consultant for clinical trials. The program provides advice to the physician while simultaneously capturing clinical information for the patient databank.

Figure 3: A typical data entry screen during the physician's interaction with ONCOCIN. The cursor location (black box) indicates that the user has just entered that there was no evidence of disease on the patient's most recent bone marrow biopsy. Region "a" is the *explanation field* which changes dynamically to indicate the interpretation of the entry (aligned with the cursor) being made by the clinician on the *flowsheet* (region "c"). Note that data from past visits are displayed on the flowsheet along with the current day's visit in the rightmost column. Region "b" is a *message field* which displays the patient's identity but is also used when the system needs to send a special message to the user (e.g., to indicate that an entered value is outside a legal range). The *special function designators* in region "d" correspond to labeled keys on the terminal keyboard. A special function key is used, for example, if the physician wishes to update flowsheet information from prior visits (see key VI). The location of the cursor on the screen, and thus the interpretation of the data being entered, is controlled by a special 21-key keypad. The flowsheet portion of the screen scrolls upward during data entry to permit access to the entire set of data items. Past data may be shifted from left to right on the screen to allow review of early visits that do not otherwise appear on the display.

Figure 4: A typical screen showing how ONCOCIN provides a therapy recommendation. Whereas the physician fills in the righthand column during data entry (Fig. 3), ONCOCIN fills in the therapy information in the righthand column when giving advice. The user may position the cursor beside a specific treatment recommendation and the *explanation field* changes to provide a justification for that piece of advice. The physician may override the program's recommendation by typing a new value for any entry. Notice that the interpretation of the special function keys changes when the system's recommendation is being displayed.

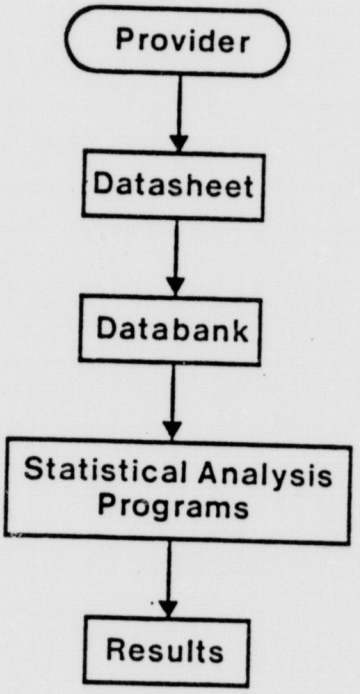
Provider
Datasheet
Databank
Statistical Analysis Programs
Results

Figure 1

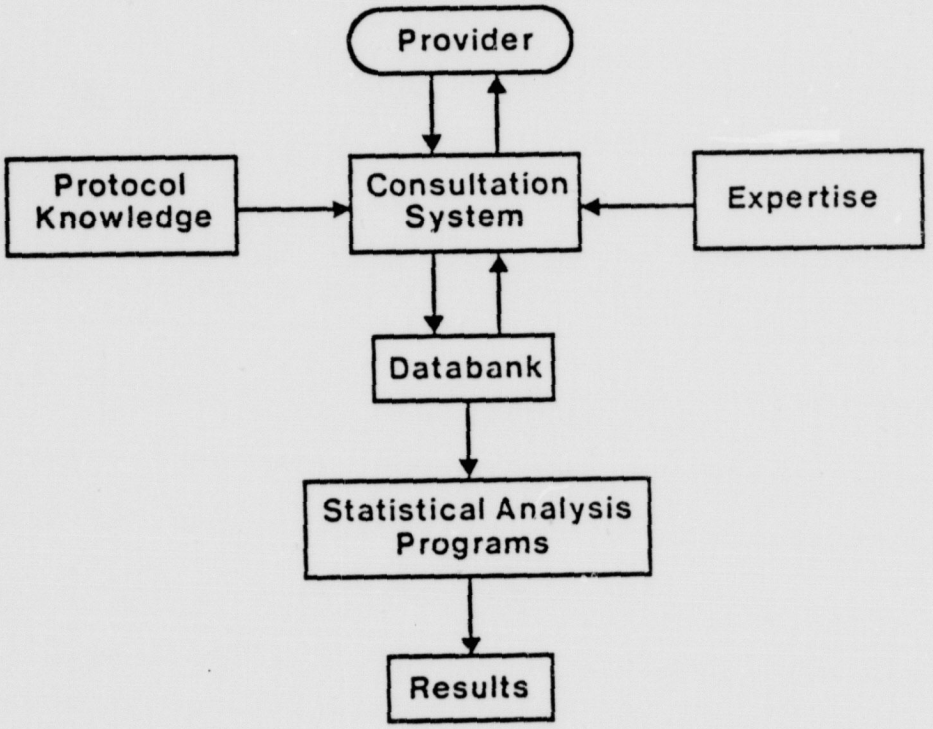
Provider
Protocol Knowledge
Consultation System
Expertise
Databank
Statistical Analysis Programs
Results

Figure 2

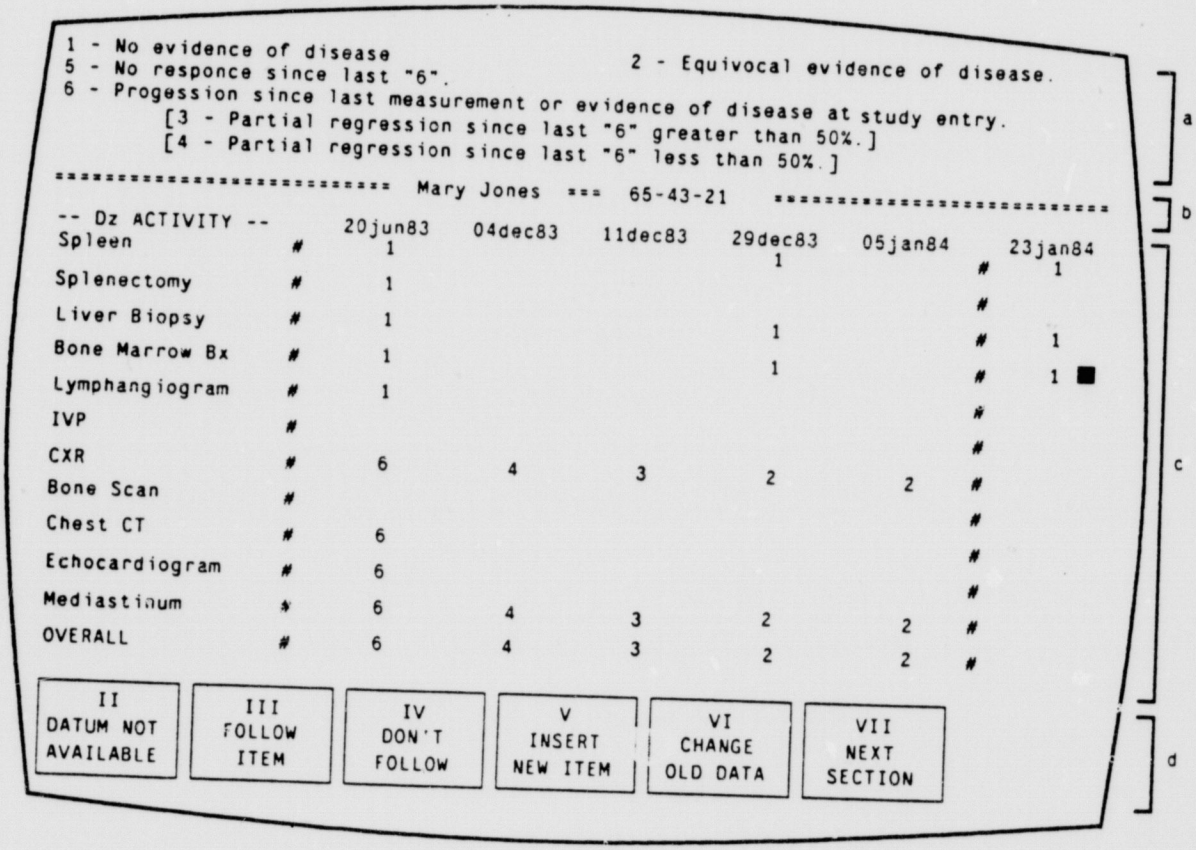
1 - No evidence of disease
2 - Equivocal evidence of disease.
5 - No responce since last "6".
6 - Progession since last measurement or evidence of disease at study entry.
[3 - Partial regression since last "6" greater than 50%.]
[4 - Partial regression since last "6" less than 50%.]
Mary Jones === 65-43-21
-- Dz ACTIVITY --
20jun83
04dec83
11dec83
29dec83
05jan84
23jan84
Spleen
Splenectomy
Liver Biopsy
Bone Marrow Bx
Lymphangiogram
IVP
CXR
Bone Scan
Chest CT
Echocardiogram
Mediastinum
OVERALL
II DATUM NOT AVAILABLE
III FOLLOW ITEM
IV DON'T FOLLOW
V INSERT NEW ITEM
VI CHANGE OLD DATA
VII NEXT SECTION
a
b
c
d

Figure 3

Give Procarbazine, 125.0 mg. PO (150 mg alternating with 100 mg) for 7 days.
[56.3 mg./m.sq. = attenuated to 75% following aborted cycle,
further attenuated to 75% due to low WBC]
[100% dose = 200.0 mg.]

========== John Doe ==== 12-34-56 ==========

The patient should receive chemotherapy PAVE-3A.

--CHEMOTHERAPY--	29dec83	05jan84	23jan84	30jan84	06feb84	13feb84	19feb84
BSA (m2)	2.1	2.1	2.1	2.1	2.1	2.1	# 2.1
OVERALL	2	2	2	2	2	2	# 2
Karnofsky (%)	100	100	100	100	100	100	# 100
PVC	39.1	40	39.4	40.4	39.9	40.8	# 38.8
WBC	6.9	4.5	2.1	6.6	2.8	2.4	# 3.8
Platelets	335	225	318	333	421	461	# 365
Combination Name	PAVE	PAVE	PAVE	PAVE	PAVE		# PAVE
Cycle #	2A	2B	DELAY	3A	ABORT		# 3A
Procarb., 100 mg/m2 POx7	200.0	200.0	0	200.0	0		# 125.0
Alkeran, 7.5 mg/m2 POx7	14.0	14.0	0	14.0	0		# 8.0
Velban, 6 mg/m2 IV	10.0	10.0	0	10.0	0		# 5.5

II	III	IV	V	VI	VII
DATUM NOT AVAILABLE	CHANGE OLD DATA	NEW DRUG	GIVE NO DRUGS	SELECT CHEMO	SEE SUMMARY

Figure 4

Table 1: Questions for Raters on Evaluation Sheet

A. Please indicate your opinion regarding the management decision indicated in the last column of the flowsheet:

[ ] Ideal = same treatment I would have administered.

[ ] Acceptable = I would not have chosen this treatment, but it is an acceptable alternative.

[ ] Sub-optimal = treatment represents a minor protocol violation,
or subjects the patient to an avoidable risk of moderate toxicity,
or subjects the patient to an avoidable risk of moderate tumor growth,
or is less than optimal for some other reason.

[ ] Unacceptable = treatment represents a major protocol violation,
or subjects the patient to an avoidable risk of major toxicity,
or subjects the patient to an avoidable risk of major tumor growth,
or is unacceptable for some other reason.

B. If you rated the management as less than ideal, in what way would your management have differed?

[ ] a. Would have delayed treatment.
[ ] b. Would have aborted cycle.
[ ] c. Would have given larger dose of one or more drugs.
[ ] d. Would have given smaller dose of one or more drugs.
[ ] e. Would have added, substituted, or omitted at least one drug.

Table 2: Expert Evaluations of Chemotherapy Treatments

Rating	Numerical Score	Number of Visits Given Rating: Agreement Visits (N=47)	Disagreement Visits (N=137) Physicians	ONCOCIN
Ideal	4	31	68	67
Acceptable	3	12	24	26
Sub-Optimal	2	4	36	29
Unacceptable	1	0	9	15

For Comparison of Physicians and ONCOCIN in Disagreement Visits:
Chi Square = 2.34 (3df, p=0.5)

Table 3: Mean Treatment Ratings of 137 Disagreements Between Physician and ONCOCIN-Recommended Chemotherapy Treatments

Expert Rater	Mean (S.E.) Rating		T-Test
	Physician Treatment	ONCOCIN-Recommended Treatment	p
A	3.20 (.18)	3.13 (.19)	.8
B	3.36 (.16)	3.17 (.16)	.4
C	2.92 (.15)	2.76 (.17)	.5
D	3.07 (.19)	3.38 (.17)	.2
All Raters	3.10 (.09)	3.06 (.09)	.8

Table 4: Reasons for Less than Ideal Ratings of Clinic Visits in which Physician Treatment did not Agree with ONCOCIN's Recommendation

Response	Number of Visits* Physician Treatment (N=68)	ONCOCIN-Recommended Treatment (N=66)
Should have Given Larger Drug Dose	34	39
Should have Given Smaller Drug Dose	19	8
Should have Changed Drugs	5	5
Should have Delayed or Aborted Treatment	3	3
Should not have Delayed Treatment	7	11

Chi Square = 5.68 (4df, p=0.2)

*The experts failed to give a reason for low ratings in one physician visit and four ONCOCIN visits.

Table 5: Comparison of Physician and ONCOCIN Behavior for Visits in which Experts Recommended Drug Dosage Change

Behavior	Number of Visits	
	Physicians (N=53)	ONCOCIN (N=47)
Failed to Escalate Dosage Adequately	22	17
Excessively Attenuated Dosage	12	22
Failed to Attenuate Dosage Adequately	17	8
Inappropriately Escalated Dosage	2	0

Chi Square = 8.49 (3df, p=0.04)

Table 6: Comparison of Physician and ONCOCIN Behavior for Visits in which Experts Judged Treatment to be Unacceptable

Behavior	Number of Visits	
	Physicians (N=9)	ONCOCIN (N=15)
Failed to Escalate Dosage Adequately	3	2
Failed to Attenuate Dosage Adequately	1	3
Inappropriately Attenuated Dosage	1	1
Failed to Delay or Abort Treatment	4	2
Inappropriately Delayed Treatment	0	7

Chi Square = 7.86 (4df, p=0.1)

References

1. Miller RA, Pople HE, Myers JD. INTERNIST-1, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 307;1982:468-76.
2. Pauker SG, Gorry A, Kassirer JP, Schwartz WB. Towards the simulation of clinical cognition: taking a present illness by computer. Amer J Med 60;1976:981-96.
3. Shortliffe EH. The computer and medical decision making: good advice is not enough. IEEE Engineering in Medicine and Biology Magazine I;1982:16-18.
4. Bonadonna G, Valagusso P. Dose-response effect of adjuvant chemotherapy in breast cancer. N Engl J Med 304;1981:10-15.
5. Armitage JO, Corder MP. Reasons for failure of MOPP to cure Hodgkin's disease: the importance of dose and schedule. Amer J Clin Onc 5;1982:315-319.
6. Hokanson JA, Costanzi JJ, Smith MS, et al. Design considerations for a medical school hospital cancer patient data system. Cancer 51;1983:1556-1561.
7. Mesel E, Wirtschafter DD, Carpenter JT, et al. Clinical algorithms for cancer chemotherapy: systems for community-based consultant-extenders and oncology centers. Meth Info Med 15;1976:168-173.
8. Wirtschafter D, Carpenter JT, Mesel E. A consultant-extender system for breast cancer adjuvant chemotherapy. Ann Int Med 90;1979:396-401.
9. Friedman RH, Frank AD. Use of conditional rule structure to automate clinical decision support: a comparison of artificial intelligence and deterministic programming techniques. Comput Biomed Res 16;1983:378-394.
10. Friedman RB, Entine SM, Carbone PP. Experience with an automated cancer protocol surveillance system. Amer J Clin Onc 6;1983:583-592.
11. Buchanan BG, Shortliffe EH. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Reading, Massachusetts: Addison-Wesley, 1984.
12. Shortliffe EH, Scott AC, Bischoff MB, Campbell AB, van Melle W, Jacobs CD. ONCOCIN: An expert system for oncology protocol management. Proceedings of the

International Joint Conference on Artificial Intelligence, American Association for Artificial Intelligence, Vancouver, British Columbia, August, 1981, pp. 876-881. Reprinted as Chapter 35 in (11)

13. Gerring PE, Shortliffe EH, van Melle W. The Interviewer/Reasoner model: an approach to improving system responsiveness in interactive AI systems. AI Magazine 3;1982:24-27.
14. Hayes-Roth F, Waterman DA, Lenat DB. Building Expert Systems. Reading, Massachusetts: Addison-Wesley, 1983.
15. Bischoff MB, Shortliffe EH, Scott AC, Carlson RW, Jacobs CD. Integration of a computer-based consultant into the clinical environment. Proceedings of the 7th Symposium on Computer Applications in Medical Care, IEEE Press, Baltimore, MD, October, 1983, pp. 149-152.
16. Rappaport H, Winter WJ, Hicks EB. Follicular lymphoma. A reevaluation of its position in the scheme of malignant lymphomas, based on a survey of 253 cases. Cancer 9;1956:792-821.
17. Portlock CS, Rosenberg SA. No initial therapy for stage III and IV non-Hodgkin's lymphomas of favorable histologic types. Ann Int Med 90;1979:10-13.
18. Nie NH, Hull CH, Jenkins JG, et al. Statistical Package for the Social Sciences. New York: McGraw-Hill, 1975. Second Edition.
19. Kramer MS, Feinstein AR. Clinical biostatistics LIV: The biostatistics of concordance. Clin Pharm Ther 29;1981:111-123.
20. Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons, 1981. Second Edition.
21. Shortliffe EH. Computer-Based Medical Consultations: MYCIN. New York: Elsevier/North Holland, 1976.
22. Yu VL, Buchanan BG, Shortliffe EH, et al. Evaluating the performance of a computer-based consultant. Comput Prog Biomed 9;1979:95-102.
23. Yu VL, Fagan LM, Wraith SM, et al. Antimicrobial selection by a computer: a blinded evaluation by infectious disease experts. JAMA 242;1979:1279-1282.

24. Elstein AS, Shulman LS, Sprafka SA. Medical Problem Solving. Cambridge: Harvard University Press, 1978.

25. McDonald CJ. Use of a computer to detect and respond to clinical events: its effect on clinician behavior. Ann Int Med 84;1976:162-7.

26. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. N Engl J Med 295;1976:1351-5.

27. McDonald CJ, Wilson GA, McCabe GP. Physician response to computer reminders. J Amer Med Assoc 244;1980:1579-81.

28. Barnett GO, Winickoff R, Dorsey JL, et al. Quality assurance through automated monitoring and concurrent feedback using a computer-based medical information system. Med Care 16;1978:962-70.

29. Cannon SR, Gardner RM. Experience with a computerized interactive protocol system using HELP. Comput Biom Res 13;1980:399-409.

30. Kent DL, Shortliffe EH, Carlson RW, Bischoff MB, Jacobs CD. Improvements in data collection through physician use of a computer-based cancer chemotherapy treatment consultant. J Clin Oncol (submitted for publication);1985.

31. Teach RL, Shortliffe EH. An analysis of physician attitudes regarding computer-based clinical consultation systems. Comput Biomed Res 14;1981:542-558.