Selection of Patients for Clinical Trials: An Interactive Web-Based System

Eugene Fink
eugene@csee.usf.edu

Princeton K. Kokku
kokku@csee.usf.edu

Savvas Nikiforou
savvasn@ucy.ac.cy

Lawrence O. Hall
hall@csee.usf.edu

Dmitry B. Goldgof
goldgof@csee.usf.edu

Jeffrey P. Krischer
jpkrischer@moffitt.usf.edu

Computer Science and Engineering
University of South Florida
Tampa, Florida 33620
phone (813) 974-4757, fax (813) 974-5456

Abstract

The purpose of a clinical trial is to evaluate a new treatment procedure. When medical researchers conduct a trial, they recruit participants with appropriate health problems and medical histories. To select participants, the researchers analyze medical records of the available patients, which has traditionally been a manual procedure.

We describe an expert system that helps to select patients for clinical trials. If the available data are insufficient for choosing patients, the system suggests additional medical tests and finds an ordering of the tests that reduces their total cost. Experiments show that the system can increase the number of selected patients. We also present an interface that enables a medical researcher to add clinical trials and selection criteria without the help of a programmer. The addition of a new trial takes ten to twenty minutes, and novice users learn the functionality of the interface in about an hour.

Keywords: Clinical trials, expert system, cost reduction, breast cancer.

1 Introduction

Cancer causes 550,000 deaths in the United States every year [17, 49], and the treatment of cancer is an active research area. Medical researchers explore new treatment methods, such as drugs, surgery techniques, and radiation therapies. An experiment with a new treatment procedure is called a *clinical trial.* When researchers conduct a trial, they recruit patients with appropriate cancer types and medical histories. The selection of patients has traditionally been a manual procedure, and studies have shown that clinicians can miss up to 60% of the eligible patients [11, 13, 19, 37, 44, 48].

If the available records do not provide enough data, clinicians perform medical tests as part of the selection process. The costs of most tests have declined over the last decade, but the number of tests has increased [40, 45], which is partially due to inappropriate ordering of tests [1, 32]. Clinicians can reduce the cost by first requiring inexpensive tests and then using their results to avoid some expensive tests; however, finding the right ordering may be a complex optimization problem.

The purpose of the described work is to automate the selection of patients for clinical trials and minimize the cost of related tests. We have developed an expert system that identifies appropriate trials for eligible cancer patients, designed a web-based interface that enables a clinician to enter new trials without the help of a programmer, and built a knowledge base for trials at the Moffitt Cancer Center, located at the University of South Florida. We have presented preliminary results in two conference papers [18, 26], and we now describe a new version of the system and more extensive experiments with the data from 261 patients at Moffitt.

The contributions of this work include three parts. First, we have built a rule-based system that helps to assign patients to clinical trials. Second, we have developed heuristics for determining an ordering of medical tests that reduces their expected cost. Third, we have run extensive experiments with the data from the breast-cancer patients admitted to Moffitt in the last three years.

We begin with a review of the previous work on medical expert systems (Section 2). We then explain the design of the developed system and present empirical confirmation of its effectiveness (Section 3). We also describe the interface for adding new knowledge (Section 4). In conclusion, we point out some limitations of the developed system and compare it with other trial-selection systems (Section 5).

2 Previous Work

Researchers have developed several expert systems that help to select clinical trials for cancer and AIDS patients. In particular, Musen *et al.* built a rule-based system, called EON, that matched AIDS patients to clinical trials [24]. Ohno-Machado *et al.* developed the AIDS^2 system, which also assigned AIDS patients to clinical trials [28]. They integrated logical rules with Bayesian networks, which helped to make decisions based on incomplete data and to quantify the decision certainty.

Bouaud *et al.* created a cancer expert system, called ONCODOC, that suggested alternative clinical trials for each patient and allowed a physician to choose among them [4, 5]. Séroussi *et al.* used ONCODOC to select participants for clinical trials at two hospitals, which helped to increase the number of selected patients by a factor of three [38, 39].

Hammond and Sergot created the OaSiS architecture [14], which had a graphical interface for entering patients' data and extending the knowledge base. Smith and her colleagues built a qualitative system that assisted a clinician in selecting medical tests, interpreting their results, and reducing the number and cost of tests [21, 40].

Theocharous developed a Bayesian system that selected clinical trials for cancer patients [31, 43]. It learned conditional probabilities of medical-test outcomes and evaluated the probability of a patient's eligibility for each trial. On the negative side, the available medical records were often insufficient for learning accurate probabilities. Furthermore, when adding a new trial, the user had to change the structure of the underlying Bayesian network. To address these problems, Bhanja *et al.* built a qualitative rule-based system for the same task [2].

Fallowfield *et al.* studied how physicians selected cancer patients for clinical trials, and compared manual and automatic selection [10]. They showed that expert systems could improve the selection accuracy; however, their study also revealed that physicians were reluctant to use these systems. Carlson *et al.* conducted similar studies with AIDS trials, and also concluded that expert systems could lead to a more accurate selection [6].

Researchers have also investigated various representations of medical knowledge. In particular, Ohno-Machado *et al.* proposed the GuideLine Interchange Format for medical knowledge [27]. Lindberg *et al.* considered an alternative format, called the Unified Medical Language System, and developed tools for converting various databases into this format [20]. Rubin *et al.* analyzed selection criteria for cancer clinical trials and proposed a format for these criteria [35, 36]. Wang *et al.* compared eight previously developed formats and identified main elements of medical knowledge, which included patient data, treatment decisions, and related actions [46].

Eriksson pointed out the need for general-purpose tools that would allow efficient knowledge acquisition, and described a system for building such tools [9]. Tallis and his colleagues developed a library of scripts for modifying knowledge bases, which helped to enforce the consistency of the modified knowledge [12, 41, 42]. Blythe *et al.* designed a general knowledge-acquisition interface based on previous techniques [3]. Musen developed the PROTÉGÉ environment for creating knowledge-acquisition tools [22]; later, researchers used it in the work on AIDS expert systems [33, 34], and an asthma treatment-selection system [15]. Musen *et al.* extended PROTÉGÉ and built a new version, called PROTÉGÉ-2000 [23].

3 Selection of Clinical Trials

Physicians at the Moffitt Cancer Center have about 150 clinical trials available for cancer patients. They have identified criteria that determine a patient's eligibility for each trial, and they use these criteria to select trials for eligible patients. Traditionally, physicians have selected trials by a manual analysis of patients' data. The review of resulting selections has shown that they usually do not check all clinical trials and occasionally miss an appropriate trial.

To address this problem, we have developed an expert system that helps to select trials for each patient. It prompts a clinician to enter the results of medical tests, and uses them to identify appropriate trials. If the available records do not provide enough data, the system suggests additional tests. We give an example of the selection process, describe the main elements of the knowledge base, and outline the system's web-based interface. We then give experimental results, which confirm that the system helps to find eligible patients and to reduce the related costs.

Example: In Figure 1(a), we give a simplified example of eligibility criteria for a clinical trial. This trial is for young and middle-aged women with a noninvasive cancer at stage II or III. When testing a patient's eligibility, a clinician has to order three medical tests (Figure 1b).

The system first prompts a clinician to enter the patient's sex and age. If the patient satisfies the corresponding conditions, the system asks for the mammogram results and verifies Conditions 3 and 4; then, it requests the biopsy and electrocardiogram data. The ordering of tests depends on their costs and on the amount of information provided by test results. The system begins with the mammogram because it is cheaper than the other tests and provides data for two eligibility criteria.

If the patient's records already include some test results, the clinician can answer the corresponding questions while entering the personal data, before the system selects tests. For example, if the records indicate that the cancer stage is IV, the clinician can enter the stage along with the sex and age, and then the system immediately determines that the patient is ineligible for this trial.

(a) Eligibility criteria.	(b) Tests and questions.
1. The patient is female. 2. She is at most forty-five years old. 3. Her cancer stage is II or III. 4. Her cancer is not invasive. 5. At most three lymph nodes have tumor cells. 6. Either • the patient has no cardiac arrhythmias, or • all tumors are smaller than 2.5 centimeters.	*General information* What is the patient's sex? What is the patient's age? *Mammogram,* Cost is \$150 What is the cancer stage? Does the patient have invasive cancer? *Biopsy,* Cost is \$400 What is the cancer stage? How many lymph nodes have tumor cells? What is the greatest tumor diameter? *Electrocardiogram,* Cost is \$200 Does the patient have cardiac arrhythmias?

Figure 1: Example of eligibility criteria, tests, and questions.

(a) Acceptance expression.	(b) Rejection expression.
sex = FEMALE and *age* $\leq$ 45 and *cancer-stage* $\in$ {II, III} and *invasive-cancer* = NO and *lymph-nodes* $\leq$ 3 and (*arrhythmias* = NO or *tumor-diameter* $\leq$ 2.5)	*sex* = MALE or *age* $>$ 45 or *cancer-stage* $\in$ {I, IV} or *invasive-cancer* = YES or *lymph-nodes* $>$ 3 or (*arrhythmias* = YES and *tumor-diameter* $>$ 2.5)

Figure 2: Logical expressions for the criteria in Figure 1(a). The acceptance expression represents the eligibility conditions, whereas the rejection expression describes ineligible patients.

Knowledge base: The knowledge base includes questions, medical tests, and logical expressions that represent eligibility criteria for each trial. Since clinicians specify eligibility criteria as hard constraints, without priorities or soft constraints, we allow only hard-constraint logical expressions. The system does not prioritize eligibility criteria, and it treats the results of medical tests in the same way as other data, such as sex, age, and medical history. We give a simplified example of tests and questions in Figure 1(b), and logical expressions in Figure 2.

The system supports three types of questions; the first type takes a yes/no response, the second is multiple choice, and the third requires a numeric answer. For example, the cancer stage is a multiple-choice question, and the tumor diameter is a numeric question. The description of a medical test includes the test name, dollar cost, and list of questions that can be answered based on the test results. For instance, the mammogram in Figure 1 has a cost of \$150, and it allows the answering of two questions. Different tests may answer the same question; for example, both mammogram and biopsy show the cancer stage.

We encode the eligibility for a trial by a logical expression that does not have negations, called the *acceptance expression.* It includes variables that represent the available medical data, as well as equalities, inequalities, "set-element" relations, conjunctions, and disjunctions. For example, we

$$\left(\begin{array}{l} \textit{sex} = \text{FEMALE and} \\ \textit{age} \leq 45 \text{ and} \\ \textit{cancer-stage} \in \{\text{II, III}\} \text{ and} \\ \textit{invasive-cancer} = \text{NO and} \\ \textit{lymph-nodes} \leq 3 \text{ and} \\ \textit{arrhythmias} = \text{NO} \end{array}\right) \text{ or } \left(\begin{array}{l} \textit{sex} = \text{FEMALE and} \\ \textit{age} \leq 45 \text{ and} \\ \textit{cancer-stage} \in \{\text{II, III}\} \text{ and} \\ \textit{invasive-cancer} = \text{NO and} \\ \textit{lymph-nodes} \leq 3 \text{ and} \\ \textit{tumor-diameter} \leq 2.5 \end{array}\right)$$

Figure 3: Disjunctive normal form of the acceptance expression in Figure 2.

encode the criteria in Figure 1(a) by the expression in Figure 2(a). In addition, the system uses the logical complement of the eligibility criteria, called the *rejection expression,* which also does not have negations (Figure 2b); it describes the conditions that make a patient ineligible for the trial.

The system collects data until it can determine which of the two expressions is TRUE. For instance, if a patient's sex is MALE, then the rejection expression in Figure 2(b) is TRUE, and the system immediately rejects this trial. If the sex is FEMALE, and the other values are unknown, then neither acceptance nor rejection expression is TRUE, and the system asks more questions.

If the knowledge base includes multiple clinical trials, the system checks a patient's eligibility for each of them. It first prompts the clinician to enter the personal data for a patient, then asks for the tests related to multiple trials, and finally requests additional tests for specific trials. After getting each new answer, the system re-evaluates the patient's eligibility for each trial. It displays the list of matching trials, rejected trials, and trials that require additional information.

Order of tests: If a patient's medical records do not include enough data, the system asks for additional tests; for example, if the records do not provide data for the eligibility criteria in Figure 1, the system asks for the mammogram, biopsy, and electrocardiogram. The total cost of tests may depend on their order; for instance, if we begin with the mammogram, and it shows that the cancer stage is IV, then we can immediately reject the trial in Figure 1 and avoid the more expensive tests.

We have explored heuristics for ordering the tests based on the test costs and the structure of acceptance and rejection expressions. The heuristics use a disjunctive normal form of these expressions; that is, each expression must be a disjunction of conjunctions. For example, the rejection expression in Figure 2(b) is in disjunctive normal form, whereas the acceptance expression in Figure 2(a) is not. If the system uses ordering heuristics, it has to convert this acceptance expression into the disjunctive normal form shown in Figure 3.

The system chooses the order of tests that reduces their expected cost. After getting the results of the first test, it re-evaluates the need for the other tests and revises their ordering. The choice of the first test is based on three criteria. The system scores all required tests according to these criteria, computes a linear combination of the three scores for every test, and chooses the test with the highest score.

1. *Cost of the test.* The system gives preference to cheaper tests. For instance, it may start with the mammogram, which is cheaper than the other two tests in Figure 1.

2. *Number of clinical trials that require the test.* When the system checks a patient's eligibility for several trials, it prefers tests that provide data for the largest number of trials. For

example, if the electrocardiogram gives data for two different trials, whereas the mammogram provides data for only one trial, the system may prefer the electrocardiogram despite its higher cost.

3. *Number of clauses that include the test results.* The system prefers the tests that provide data for the largest number of clauses in the acceptance and rejection expressions. For example, the mammogram data affect both clauses of the acceptance expression in Figure 3 and two clauses of the rejection expression in Figure 1(b). On the other hand, the electrocardiogram affects only one clause of the acceptance expression and one clause of the rejection expression; thus, the system should order it after the mammogram.

The system disregards the costs of tests performed in the normal course of treatment, and accounts only for the costs related to the selection of clinical trials. For example, if a patient needs the mammogram regardless of trial participation, the system views it as a zero-cost test. On the other hand, if the only purpose of the biopsy and electrocardiogram is to select trials, the system uses heuristics to order these tests.

Although the system suggests the single most effective test, it allows a clinician to order multiple tests at once. For instance, if it indicates that the mammogram is the best test, the clinician can determine that the electrocardiogram is also an effective test, and order both tests at the same time.

User interface: The system includes a web-based interface that allows clinicians to enter patients' data through remote computers; the interface consists of five screens (Figure 4).

The start screen is for adding new patients and retrieving old patients (Figure 5). After a user enters a patient's name, the system displays a list of the available trials (Figure 6). The user can choose a subset of these trials, and then the system checks eligibility only for the selected trials. The next screen is for basic personal and medical data, such as sex, age, and cancer stage (Figure 7).

After the system gets these basic data, it prompts the user for medical information related to specific trials (Figure 8). When the user enters medical data, the system continuously re-evaluates the patient's eligibility and shows the decision for each trial. If the patient is ineligible for some trials, the user can find out the reasons by clicking the *"Why"* button. The interface also includes a screen for the review and modification of the previous answers, similar to the screen in Figure 8.

Note that the system uses a "patient-centered" approach; that is, a user enters data for one patient at a time, and the system chooses trials for each new patient. A physician typically delegates the data entry to a nurse, who enters the related information after a patient's visit. She reports the suggested trials to the physician, who can discuss trial participation with the patient during the next visit. The system maintains a persistent database of patients' data, and it does not require entering all information at once. A nurse can enter or modify data at any time; furthermore, she can enter data in any order, regardless of the system's advice. Although the system suggests an ordering of tests and questions, it does not enforce this ordering.

Experiments: We have built a knowledge base for the breast-cancer clinical trials at the Moffitt Cancer Center, applied the system to the retrospective data from 187 past patients and 74 current patients, and compared the results with manual selection by Moffitt clinicians. The number of matching trials for a patient has ranged from zero to three. For most patients, the system rejects most trials during the initial entry of basic data, such as sex, age, and cancer stage. It usually

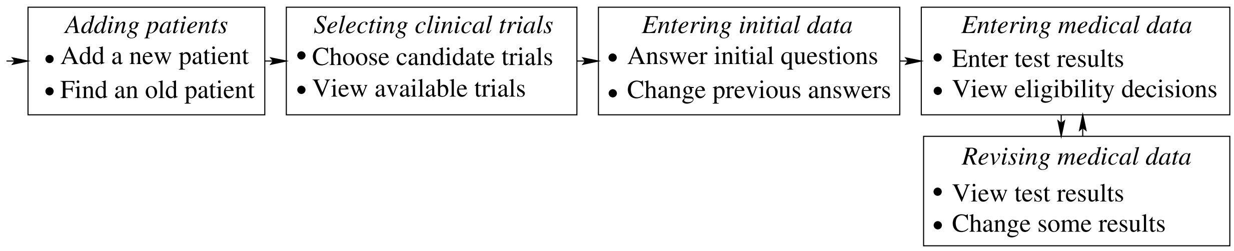


Figure 4: Entering a patient's data. The web-based interface for the data entry consists of five screens. We show these screens by rectangles and the transitions between them by arrows.

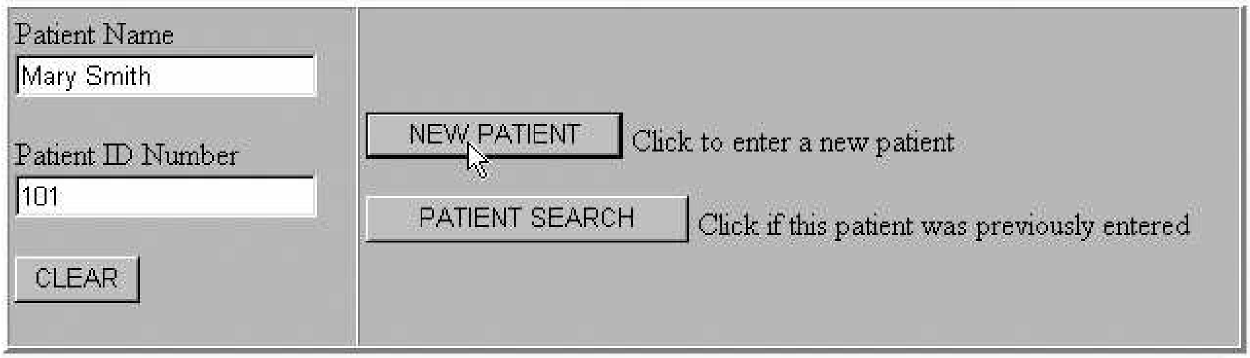


Figure 5: Adding new patients and retrieving existing patients.

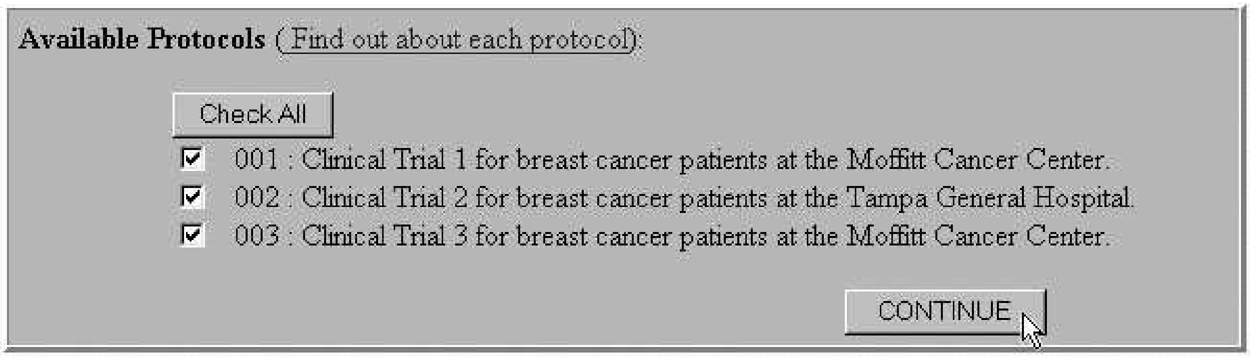


Figure 6: Selecting clinical trials.

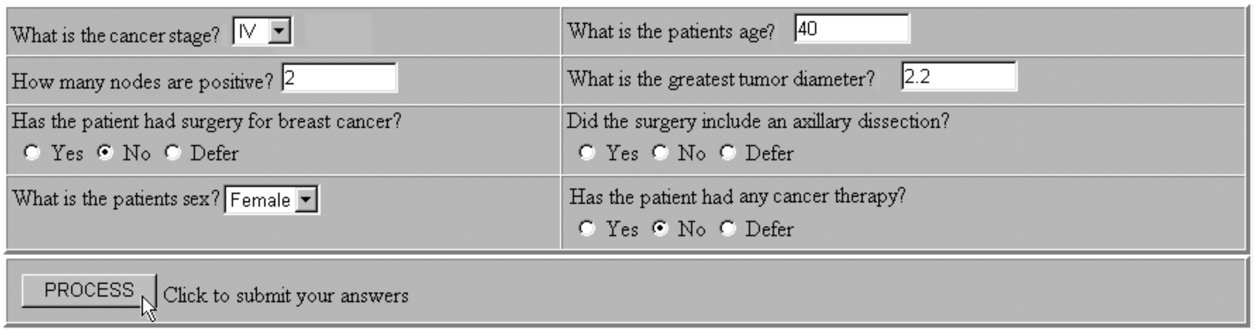


Figure 7: Entering basic information for a patient.

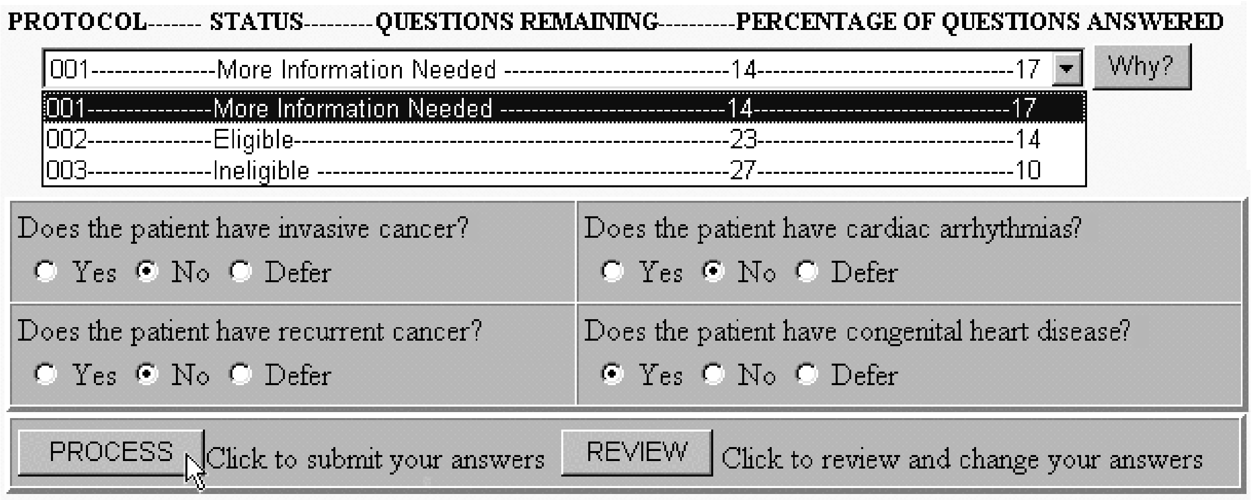


Figure 8: Entering medical data.

identifies two to five potential matches based on these basic data, and narrows the selection down to one or two trials based on the following trial-specific questions.

We summarize the results for the past patients in Table 1(a), and the results for the current patients in Table 1(b). The "same matches" column includes the number of patients who have been selected by both human clinicians and the expert system. The "new matches" column gives the number of patients who have been matched by the system but missed by human clinicians. Finally, the last column shows the number of matching patients whose available records are incomplete. Clinicians have found trials for these patients, but the system cannot identify these matches because of insufficient data. Since these patients are no longer at Moffitt, we cannot obtain the missing data; note that this problem is due to the use of retrospective data, and it does not arise when clinicians select trials for new patients.

The system has identified a number of situations when patients were eligible for clinical trials, but did not participate in these trials. We have checked these results with Moffitt clinicians, and they have confirmed that all matches are correct. In most cases, patients did not participate in the matching trials because clinicians missed these matches; however, for some of the past cases, we have been unable to verify that physicians actually missed the matches, rather than having undocumented reasons for omitting them.

We show the mean test costs with and without the ordering heuristics in Table 2, and give a graphical view of the cost savings in Figure 9. The results confirm that the heuristics reduce the cost of the selection process. Five clinical trials have incurred selection costs; the heuristics have significantly reduced the costs for four of these trials, and have not affected the cost for the fifth trial. The other trials have not incurred costs because all related tests were performed in the normal course of treatment before the trial selection.

Table 1: Results of matching 187 past patients and 74 current patients. We give the number of matches found by both the expert system and human clinicians, as well as the number of new matches identified by the system. We also show the number of matches missed by the system because of insufficient data.

(a) Results for the 187 past patients.

Clinical Trial	Same Matches	New Matches	Missing Data
10822	10	5	0
10840	0	19	3
11072	48	26	19
11378	4	19	3
11992	5	6	0
12100	8	20	13
12101	20	30	0

(b) Results for the 74 current patients.

Clinical Trial	Same Matches	New Matches	Missing Data
11132	4	1	1
11931	1	12	0
11971	3	0	0
12100	0	3	0
12101	6	26	1
12601	0	1	2
12775	4	5	1

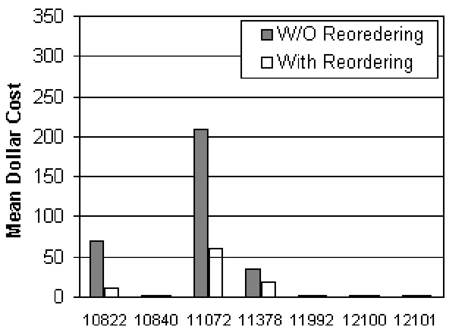
Table 2: Cost savings by test reordering.

(a) Results for the 187 past patients.

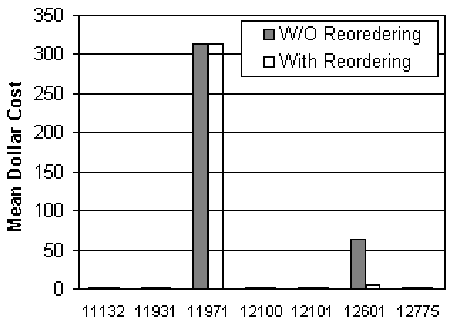
Clinical Trial	Average Dollar Cost: Without Test Reordering	Average Dollar Cost: With Test Reordering
10822	$70	$11
10840	$0	$0
11072	$209	$60
11378	$35	$19
11992	$0	$0
12100	$0	$0
12101	$0	$0

(b) Results for the 74 current patients.

Clinical Trial	Average Dollar Cost: Without Test Reordering	Average Dollar Cost: With Test Reordering
11132	$0	$0
11931	$0	$0
11971	$314	$314
12100	$0	$0
12101	$0	$0
12601	$64	$6
12775	$0	$0



(a) Results for the 187 past patients.



(b) Results for the 74 current patients.

Figure 9: Costs with and without test reordering.

1. The patient is female.
2. She is at most forty-five years old.
3. Either
   - her cancer is not invasive, or
   - her cancer is not recurrent.
4. Either
   - at most three lymph nodes have tumor cells, or
   - all tumors are smaller than 2.5 centimeters.
5. Either
   - the patient has no cardiac arrhythmias, or
   - the patient has no congenital heart disease.

(a) Eligibility criteria.

sex = FEMALE and *age* ≤ 45 and (*invasive* = NO or *recurrent* = NO) and (*lymph-nodes* ≤ 3 or *tumor-size* ≤ 2.5) and (*arrhythmias* = NO or *congenital* = NO)	*sex* = FEMALE and *age* ≤ 45 and *invasive-and-recurrent* = NO and (*lymph-nodes* ≤ 3 or *tumor-size* ≤ 2.5) and *arrhythmias-and-congenital* = NO
(b) Acceptance expression.	(c) Reduced expression.

Figure 10: Reducing the number of disjunctions. The conversion of the eligibility criteria (a) into a logical expression (b) leads to an explosion in the size of the corresponding disjunctive normal form. We may prevent the explosion by replacing some disjunctions with single questions (c).

Scalability: The time complexity of evaluating the acceptance and rejection expressions is linear in their total size. Experiments on a Sun Ultra 10 have shown that the evaluation takes about 0.02 seconds per question, and the time is linear in the number of questions. Typical eligibility conditions for a clinical trial include ten to thirty questions; thus, the evaluation time is 0.2 to 0.6 seconds per trial. We are working on improved data structures for eligibility criteria, and we aim to reduce the time to 0.1 seconds per trial.

The linear scalability is an advantage over Bayesian systems, which usually do not scale to a large number of clinical trials [8, 28, 30]. The authors of these systems have reported that the sizes of the underlying networks are superlinear in the number of trials [29, 47], and that the training time is superlinear in the network size [31, 43].

If the system uses the cost-reduction heuristics, it converts the acceptance and rejection expressions into disjunctive normal form, which can potentially lead to an explosion in their size. For example, if eligibility conditions are as shown in Figure 10(a), the system initially generates the expression in Figure 10(b). If the system converts it to disjunctive normal form, the resulting expression consists of eight clauses.

Although the conversion may result in impractically large expressions, experiments with cancer trials have shown that this problem does not arise in practice because the number of nested disjunctions is usually small. Furthermore, we can eliminate some disjunctions by combining their elements into longer questions. For instance, we can represent Condition 3 in Figure 10(a) by a single question: "Does the patient have both invasive and recurrent cancer?" If we apply this modification to Conditions 3 and 5, then we obtain the expression in Figure 10(c), and its conversion to disjunctive normal form results in an expression with two clauses.

4 Entering Eligibility Criteria

We describe a web-based interface for adding new clinical trials [25], which consists of two main parts; the first part is for information about medical tests (Figure 11), and the second is for eligibility criteria (Figure 12). The interface includes fifteen screens; three of them are "start screens," which can be reached from any other screen. We give an example of entering eligibility criteria, describe the two parts of the interface, and present experiments to illustrate its effectiveness.

Example: Suppose that a user needs to enter the criteria shown in Figure 1, and the system initially has no data about the related tests. The user has to describe the tests and questions, and specify the eligibility conditions.

First, the user utilizes the *"Adding tests"* screen to enter the new tests; we illustrate the entry of a test in Figure 13. Then, the user adds the related questions; to enter questions for a specific test, the user selects the test and clicks *"Modify"* (Figure 14), and the system displays the *"Modifying a test"* screen (Figure 15). To add a question, the user clicks the appropriate button at the bottom (Figure 15) and then types the question (Figure 16).

After adding the questions for all tests, the user goes to the *"Adding clinical trials"* screen and initializes a new trial (Figure 17). The user gets the *"Selecting tests"* screen and chooses the tests related to the current trial (Figure 18). Then, the user marks relevant questions and the answers that make a patient eligible (Figure 19). If the eligibility criteria include disjunctions, the user has to utilize the screen for composing logical expressions (Figure 20).

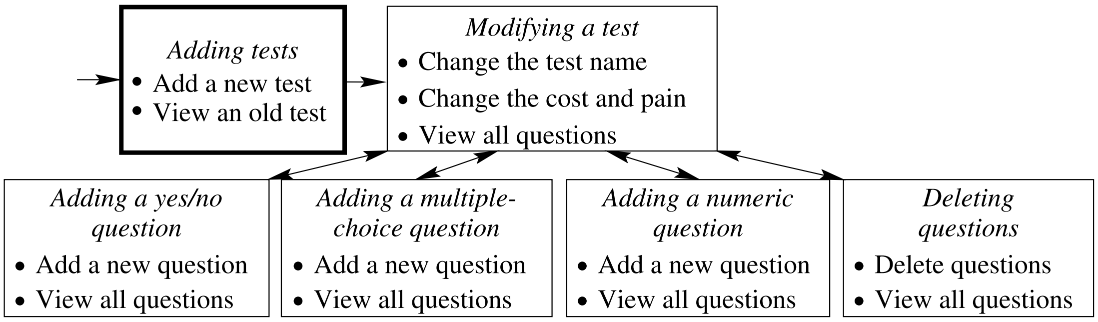
Tests and questions: We now describe the six-screen interface for adding tests and questions (Figure 11a). The start screen allows viewing the available tests and defining new ones, whereas the other screens are for modifying tests and adding questions.

We show the start screen in Figure 13; its left-hand side allows viewing questions and going to a modification screen. If the user selects a test and clicks *"View,"* the system shows the questions related to this test at the bottom of the same screen. If the user clicks *"Modify,"* it displays the *"Modifying a test"* screen (Figure 15). The right-hand side of the start screen allows adding a new test by specifying its name, cost, and pain level.

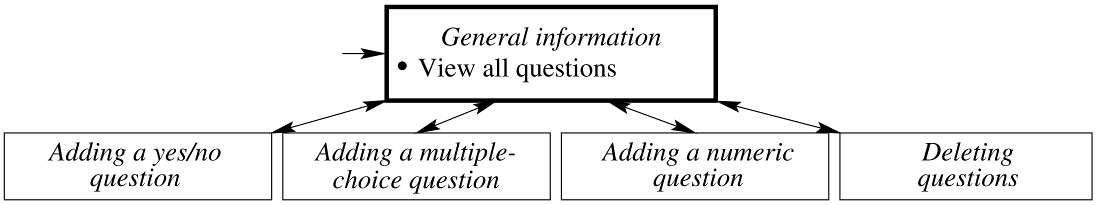
The *"Modifying a test"* screen shows the information about a specific test, which includes the test name, cost, pain level, and related questions. The user can change the test name, cost, and pain level; the four bottom buttons allow moving to the screens for adding and deleting questions.

We show the screen for adding yes/no questions in Figure 16(a) and multiple-choice questions in Figure 16(b); the screen for numeric questions is similar. The user can enter a new question for the current test, along with a set of allowed answers. If the question is also related to other tests, the user has to mark them in the lower box. The *"Deleting questions"* screen is for removing old questions, which allows modification of old eligibility criteria.

The mechanism for adding general questions, such as sex and age, consists of five screens (Figure 11b), and the user adds general questions in the same way as test-related questions.



(a) Tests and questions.



(b) General questions.

Figure 11: Entering tests and questions. We show the screens by rectangles and the transitions between them by arrows. The bold rectangles are the start screens.

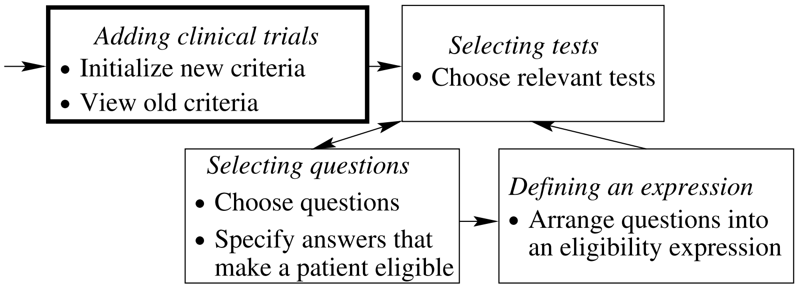


Figure 12: Entering eligibility criteria.

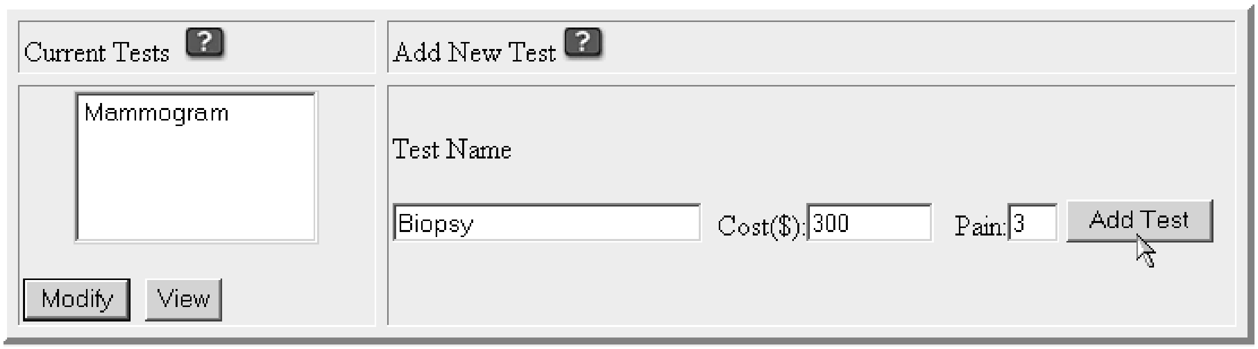


Figure 13: Adding a new test.

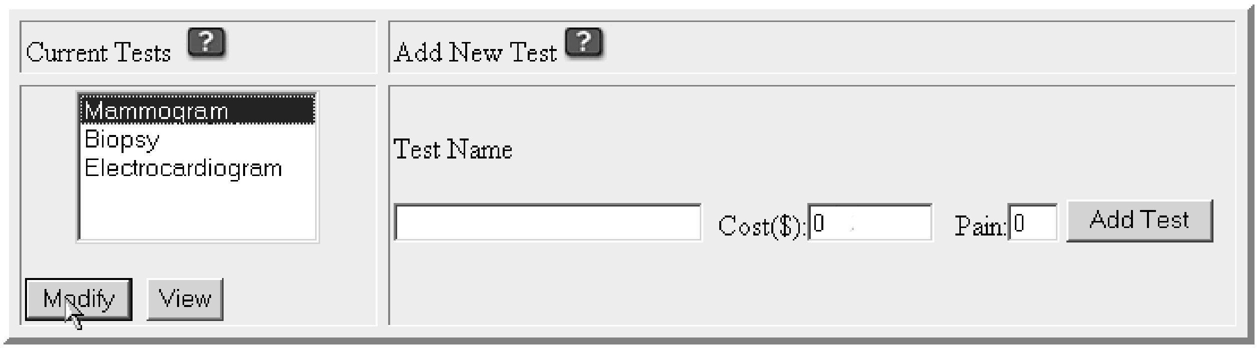


Figure 14: Selecting a test for entering the related questions.

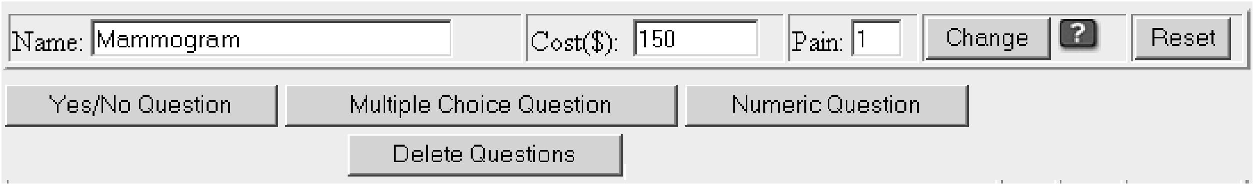
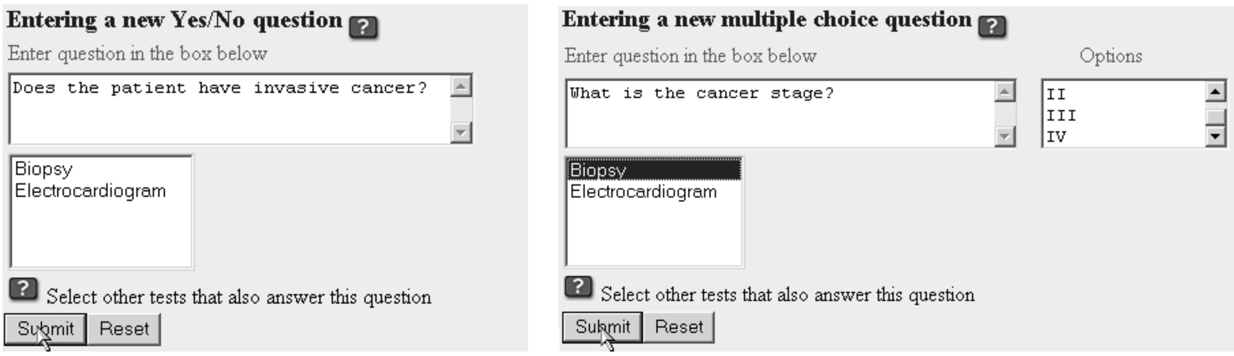


Figure 15: Modifying a test; the bottom buttons are for moving to question-entry screens.



(a) Yes/no question.

(b) Multiple-choice question.

Figure 16: Adding new questions; the user types a question and the answer options.

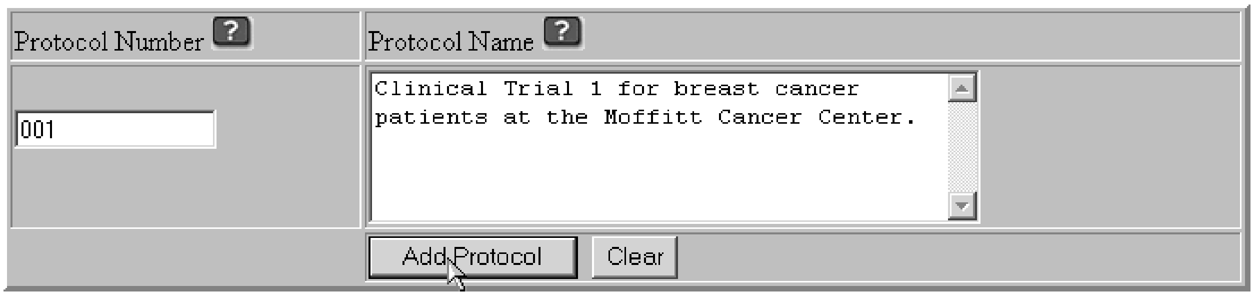


Figure 17: Adding a new clinical trial.

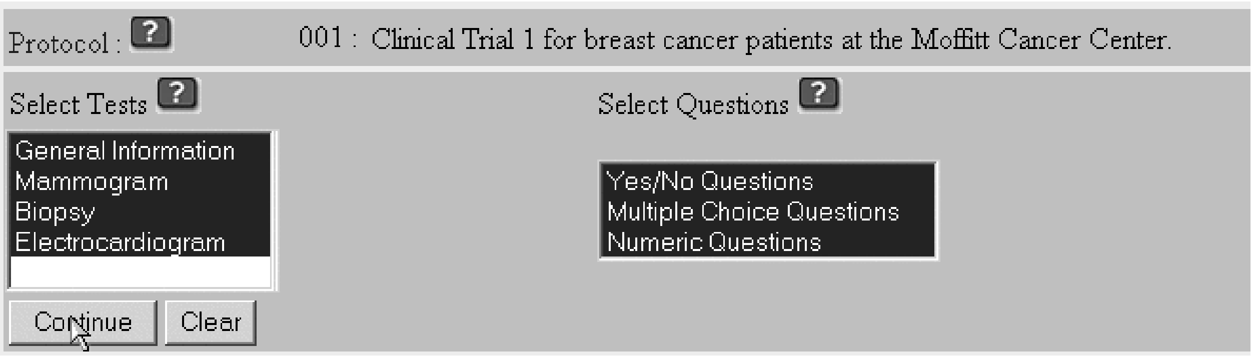


Figure 18: Choosing tests and question types.

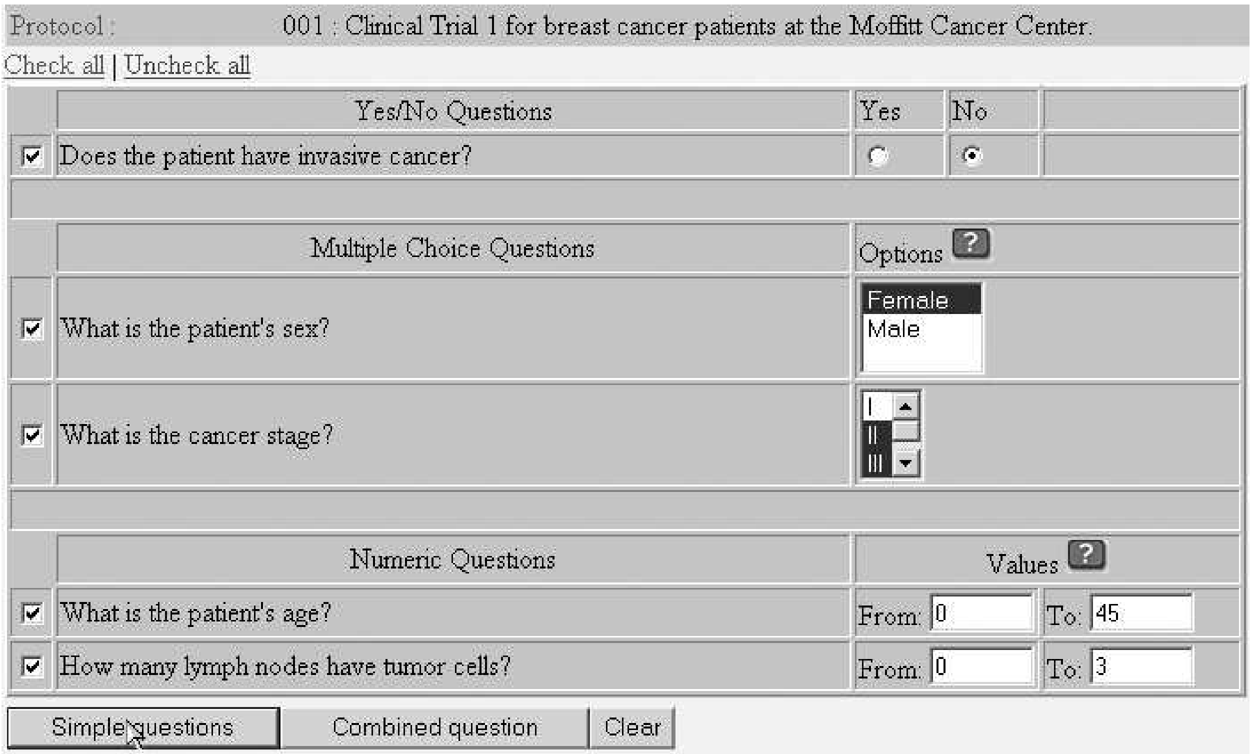


Figure 19: Selecting questions and answers. The user checks the questions for the current clinical trial and marks the answers that satisfy the eligibility criteria.

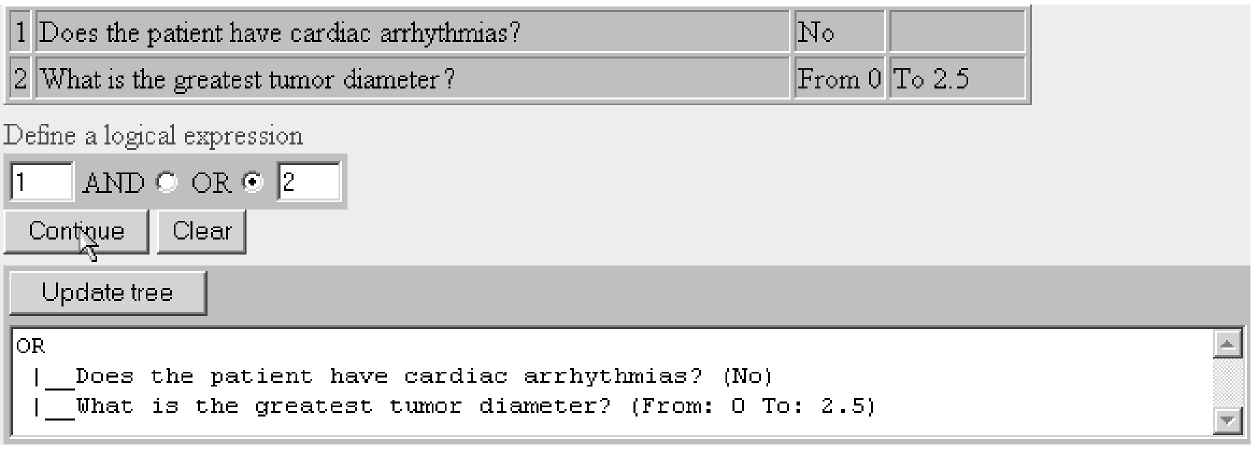


Figure 20: Combining questions into a logical expression.

Eligibility conditions: We next describe the mechanism for entering eligibility criteria, which consists of four screens (Figure 12). The start screen allows the user to initialize a new clinical trial and view the criteria for old trials. If the user needs to modify a clinical trial, the system first displays the test-selection screen (Figure 18). The user then chooses related tests and question types, and clicks *"Continue"* to get the question list.

The next screen (Figure 19) allows the user to select specific questions and mark answers that make a patient eligible. For a multiple-choice question, the user may specify several eligibility options; for example, a patient may be eligible if her cancer stage is II or III. For a numeric question, the user has to specify a range of values; for instance, a patient may be eligible if her age is between 0 and 45 years. If the user clicks *"Simple questions,"* the system generates a conjunction of the selected criteria. If the eligibility conditions involve a more complex expression, the user has to click *"Combined question"* and then use the screen for composing logical expressions (Figure 20).

The system combines the eligibility criteria into an acceptance expression, and then generates the corresponding rejection expression by recursive application of DeMorgan's laws. If the system uses the cost-reduction heuristics, it converts these expressions into disjunctive normal form using a standard conversion algorithm [7, 16].

The system allows a user to modify questions, answer options, and eligibility conditions. If clinicians need to change eligibility criteria, they can readily update them in the system; however, each change requires an approval of the Institutional Review Board, and clinicians rarely modify eligibility criteria in practice. If a new clinical trial is similar to some old trial, a user can enter it by copying the old trial and then modifying its criteria; however, the experience with trials at the Moffitt Cancer Center has shown that minor modifications are rare, and entering new criteria from scratch is usually faster than modifying old criteria.

Entry time: We have run experiments with sixteen novice users, who had no prior experience with the interface. First, every user has entered four sets of medical tests; each set has included three tests and ten questions. Then, each user has added eligibility expressions for ten clinical trials used at the Moffitt Cancer Center; the number of questions in an eligibility expression has varied from ten to thirty-five.

We have measured the entry time for each test set and each clinical trial. In Figure 21, we show the mean time for every test set and the time per question for the same sets. All users have entered

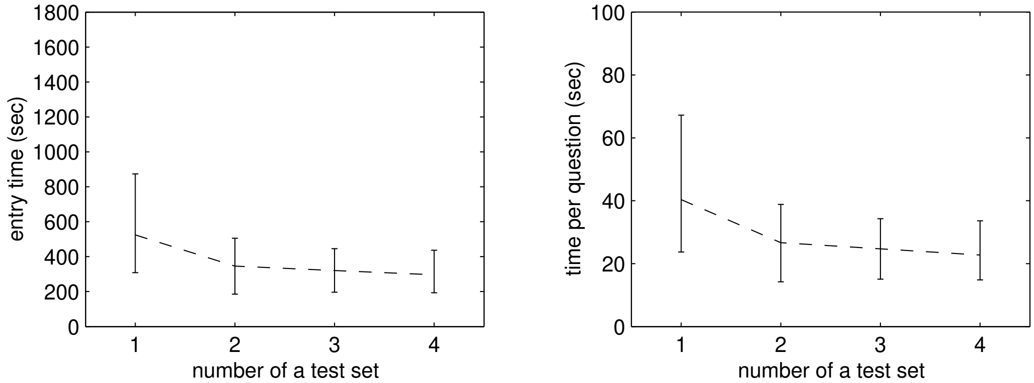


Figure 21: Entry time for test sets (left) and the mean time per question for each set (right). We plot the average time (dashed lines) and the time of the fastest and slowest users (vertical bars).

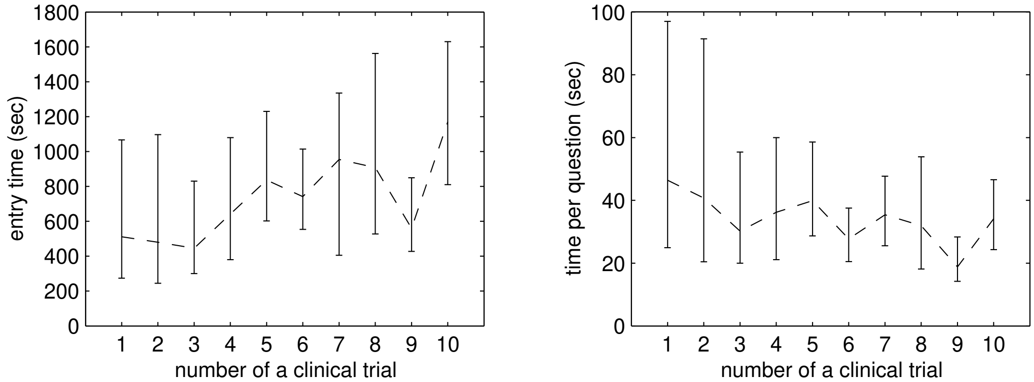


Figure 22: Entry time for eligibility criteria. We show the average time for each clinical trial and the time per question (dashed lines), along with the performance of the fastest and slowest users (vertical bars).

the test sets in the same order; since they had no prior experience, their performance has improved during the experiment. In Figure 22, we give similar graphs for the entry of trials.

The experiments have shown that novices can efficiently use the interface; they quickly learn its full functionality, and their learning curve reaches a plateau after about an hour. The average time per question is 31 seconds for the entry of medical tests and 37 seconds for eligibility criteria, which means that a user can enter all 150 cancer trials used at Moffitt in about two weeks.

5 Conclusions

We have developed an expert system that assigns cancer patients to clinical trials. We have described the representation of selection criteria, heuristics for ordering of tests, and a web-based interface for entering patients' data, which will enable physicians across the country to access a central repository of clinical trials. The system also includes an interface for extending its knowledge base, which allows a user to enter a new trial in ten to twenty minutes. Novices can use the interface without prior instructions, and they reach their full speed after about an hour. Although cancer research has provided the motivation for this work, the system is not limited to cancer, and we can use it for trials related to other diseases.

The system uses logical eligibility expressions, similar to those in EON [24] and ONCODOC [4, 5]; this approach is different from AIDS^2 [28] and Theocharous's system [31, 43], which are based on Bayesian networks. The use of logical expressions ensures scalability and ease of adding new trials, but it does not allow probabilistic decisions based on incomplete data.

We have applied the system to the data from 261 breast-cancer patients admitted to the Moffitt Cancer Center in the last three years. The experiments have confirmed that the system can improve the speed and accuracy of selecting trial participants. The results suggest that physicians miss about 60% of matching trials, which means that the system can increase the number of participants by a factor of 2.5. These results are consistent with the studies of the manual trial selection [11, 13, 19, 37, 44, 48], which confirm that clinicians miss up to 60% of matches. They are also consistent with the experiment on using ONCODOC at two French hospitals, which has increased the number of selected matches by a factor of three [38, 39]. We have been unable to compare the results with those of AIDS^2, EON, and Theocharous's system, because the authors of these systems have not reported large-scale clinical experiments.

The developed system includes heuristics for the ordering of medical tests, which is an advantage over the other trial-selection systems. The experiments have shown that the ordering of tests affects their overall cost, and the implemented heuristics reduce this cost. We have not yet evaluated the system in the real-time hospital environment, and have not analyzed satisfaction of physicians and nurses with the system's performance. As the next step, we plan to deploy the system at Moffitt and evaluate its effectiveness in helping clinicians.

We now point out some limitations of the developed system and related future challenges. First, the system does not access patient data in the Moffitt clinical database, and nurses have to enter all relevant information through the system's interface. The data in the clinical database are mostly in natural language, as dictated by physicians; we plan to develop a mechanism for transferring these data into the trial-selection system, which will require domain-specific tools for natural-language processing. Second, the system does not keep track of temporal changes in the data; for example, it does not update a patient's age, and does not flag the out-of-date test results. We have recently designed a mechanism for temporal reasoning, and we plan to integrate it with the system. Third, the test-ordering heuristics do not account for the probabilities of possible test results, and we are presently working on the integration of the current heuristics with probabilistic reasoning.

Acknowledgments: This work has been partially supported by the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command under contract DAMD17-00-1-0244, and by the Moffitt Cancer Center.

References

[1] D. Bareford and A. Hayling. Inappropriate use of laboratory services: Long term combined approach to modify request patterns. *British Medical Journal*, 301(6764):1305–1307, 1990.

[2] Sanjukta Bhanja, Lynn M. Fletcher, Lawrence O. Hall, Dmitry B. Goldgof, and Jeffrey P. Krischer. A qualitative expert system for clinical trial assignment. In *Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference*, pages 84–88, 1998.

[3] Jim Blythe, Jihie Kim, Surya Ramachandran, and Yolanda Gil. An integrated environment for knowledge acquisition. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 13–20, 2001.

[4] Jacques Bouaud, Briggite Séroussi, Éric-Charles Antoine, Mary Gozy, David Khayat, and Jean-Franşois Boisvieux. Hypertextual navigation operationalizing generic clinical practice guidelines for patient-specific therapeutic decisions. *Journal of the American Medical Informatics Association*, 5(suppl.):488–492, 1998.

[5] Jacques Bouaud, Briggite Séroussi, Éric-Charles Antoine, Laurent Zelek, and Marc Spielmann. Reusing ONCODOC, a guideline-based decision support system, across institutions: A successful experiment in sharing medical knowledge. In *Proceedings of the American Medical Informatics Association Annual Symposium*, volume 7, 2000.

[6] Robert W. Carlson, Samson W. Tu, Nancy M. Lane, Tze L. Lai, Carol A. Kemper, Mark A. Musen, and Edward H. Shortliffe. Computer-based screening of patients with HIV/AIDS for clinical trial eligibility. *Online Journal of Current Clinical Trials*, 4(179), 1995.

[7] Yves Crama and Peter L. Hammer. *Boolean Functions*. Springer-Verlag, Berlin, Germany, 2001.

[8] Francisco J. Díez, José Mira, E. Iturralde, and S. Zubillaga. DIAVAL, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, 10(1):59–73, 1997.

[9] Henrik Eriksson. Specification and generation of custom-tailored knowledge-acquisition tools. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, volume 1, pages 510–518, 1993.

[10] Lesley Fallowfield, D. Ratcliffe, and Robert Souhami. Clinicians' attitudes to clinical trials of cancer therapy. *European Journal of Cancer*, 33(13):2221–2229, 1997.

[11] John H. Gennari and Madhu Reddy. Participatory design and an eligibility screening tool. In *Proceedings of the American Medical Informatics Association Annual Fall Symposium*, pages 290–294, 2000.

[12] Yolanda Gil and Marcelo Tallis. A script-based approach to modifying knowledge bases. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, pages 377–383, 1997.

[13] Carolyn Cook Gotay. Accrual to cancer clinical trials: Directions from the research literature. *Social Science and Medicine*, 33(5):569–577, 1991.

[14] Peter Hammond and Marek J. Sergot. Computer support for protocol-based treatment of cancer. *Journal of Logic Programming*, 26(2):93–111, 1996.

[15] Peter D. Johnson and Mark A. Musen. Development of a guideline authoring tool with PROTÉGÉ-II based on the DILEMMA generic protocol and guideline model. Technical Report SMI-96-0620, Stanford Medical Informatics Group, Stanford School of Medicine, 1996.

[16] Rosen H. Kenneth. *Discrete Mathematics and Its Applications*. McGraw-Hill, New York, NY, third edition, 1988.

[17] Kenneth G. Keppel, Jeffrey N. Pearcy, and Diane K. Wagener. Trends in racial and ethnic-specific rates for the health status indicators: United States, 1990–98. *Healthy People 2000, Statistical Notes*, 23, 2002.

[18] Princeton K. Kokku, Lawrence O. Hall, Dmitry B. Goldgof, Eugene Fink, and Jeffrey P. Krischer. A cost-effective agent for clinical trial assignment. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2002.

[19] Cyrus Kotwall, Leo J. Mahoney, Robert E. Myers, and Linda Decoste. Reasons for non-entry in randomized clinical trials for breast cancer: A single institutional study. *Journal of Surgical Oncology*, 50:125–129, 1992.

[20] Donald A. B. Lindberg, Betsy L. Humphreys, and Alexa T. McCray. The unified medical language system. *Journal of Methods of Information in Medicine*, 4(32):281–291, 1993.

[21] Michael D. D. McNeely and Beverly J. Smith. An interactive expert system for the ordering and interpretation of laboratory tests to enhance diagnosis and control utilization. *Canadian Medical Informatics*, 2(3):16–19, 1995.

[22] Mark A. Musen. *Automated Generation of Model-Based Knowledge Acquisition Tools*. Morgan Kaufmann, San Mateo, CA, 1989.

[23] Mark A. Musen, Ray W. Fergerson, William E. Grosso, Natalya F. Noy, Monica Crubezy, and John H. Gennari. Component-based support for building knowledge-acquisition systems. In *Proceedings of the Conference on Intelligent Information Processing of the International Federation for Information Processing World Computer Congress*, 2000.

[24] Mark A. Musen, Samson W. Tu, Amar K. Das, and Yuval Shahar. EON: A component-based approach to automation of protocol-directed therapy. *Journal of the American Medical Informatics Association*, 3(6):367–388, 1996.

[25] Savvas Nikiforou. Selection of clinical trials: Knowledge representation and acquisition. Master's thesis, Department of Computer Science and Engineering, University of South Florida, 2002.

[26] Savvas Nikiforou, Eugene Fink, Lawrence O. Hall, Dmitry B. Goldgof, and Jeffrey P. Krischer. Knowledge acquisition for clinical-trial selection. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2002.

[27] Lucila Ohno-Machado, John H. Gennari, Shawn Murphy, Nilesh L. Jain, Samson W. Tu, Diane E. Oliver, and Edward Pattison-Gordon. The GuideLine Interchange Format: A model for representing guidelines. *Journal of American Medical Informatics Association*, 5(4):357–372, 1998.

[28] Lucila Ohno-Machado, Eduardo Parra, Suzanne B. Henry, Samson W. Tu, and Mark A. Musen. AIDS^2: A decision-support tool for decreasing physicians' uncertainty regarding patient eligibility for HIV treatment protocols. In *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*, pages 429–433, 1993.

[29] Agnieszka Oniśko, Marek J. Druzdzel, and Hanna Wasyluk. Learning Bayesian network parameters from small data sets: Application of noisy-OR gates. In *Proceedings of the Workshop on Bayesian and Causal Networks: From Inference to Data Mining*, 2000.

[30] Agnieszka Onisko, Mark J. Druzdzel, and Hanna Wasyluk. Application of Bayesian belief networks to diagnosis of liver disorders. In *Proceedings of the Third Conference on Neural Networks and Their Applications*, pages 730–736, 1997.

[31] Constantinos Papaconstantinou, Georgios Theocharous, and Sridhar Mahadevan. An expert system for assigning patients into clinical trials based on Bayesian networks. *Journal of Medical Systems*, 22(3):189–202, 1998.

[32] Franco Perraro, Paolo Rossi, Carlo Liva, Adolfo Bulfoni, G. Ganzini, and Adriano Giustinelli. Inappropriate emergency test ordering in a general hospital: Preliminary reports. *Quality Assurance Health Care*, 4:77–81, 1992.

[33] Angel R. Puerta, John W. Egar, Samson W. Tu, and Mark A. Musen. A multiple-method knowledge-acquisition shell for the automatic generation of knowledge-acquisition tools. *Knowledge Acquisition*, 4(2):171–196, 1992.

[34] Angel R. Puerta, Henrik Eriksson, John W. Egar, and Mark A. Musen. Generation of knowledge-acquisition tools from reusable domain ontologies. Technical Report KSL-92-81, Computer Science Department, Stanford University, 1992.

[35] Daniel L. Rubin, John H. Gennari, and Mark A. Musen. Knowledge representation and tool support for critiquing clinical trial protocols. In *Proceedings of the American Medical Informatics Association Fall Symposium*, pages 724–728, 2000.

[36] Daniel L. Rubin, John H. Gennari, Sandra Srinivas, Allen Yuen, Herbert Kaizer, Mark A. Musen, and John S. Silva. Tool support for authoring eligibility criteria for cancer trials. In *Proceedings of the American Medical Informatics Association Annual Symposium*, pages 369–373, 1999.

[37] Briggite Séroussi, Jacques Bouaud, and Éric-Charles Antoine. Enhancing clinical practice guideline compliance by involving physicians in the decision process. In Werner Horn, Yuval Shahar, Greger Lindberg, Steen Andreassen, and Jeremy C. Wyatt, editors, *Artificial Intelligence in Medicine*, pages 76–85. Springer-Verlag, Berlin, Germany, 1999.

[38] Briggite Séroussi, Jacques Bouaud, and Éric-Charles Antoine. Users' evaluation of ONCODOC, a breast cancer therapeutic guideline delivered at the point of care. *Journal of the American Medical Informatics Association*, 6(5):384–389, 1999.

[39] Briggite Séroussi, Jacques Bouaud, and Éric-Charles Antoine. ONCODOC: A successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. *Artificial Intelligence in Medicine*, 22(1):43–64, 2001.

[40] Beverly J. Smith and Michael D. D. McNeely. The influence of an expert system for test ordering and interpretation on laboratory investigations. *Clinical Chemistry*, 45(8):1168–1175, 1999.

[41] Marcelo Tallis. A script-based approach to modifying knowledge-based systems. In *Proceedings of the Tenth Innovative Applications of Artificial Intelligence Conference*, pages 1183–1195, 1998.

[42] Marcelo Tallis and Yolanda Gil. Designing scripts to guide users in modifying knowledge-based systems. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pages 242–249, 1999.

[43] Georgios Theocharous. An expert system for assigning patients into clinical trials based on Bayesian networks. Master's thesis, Computer Science and Engineering Department, University of South Florida, 1996.

[44] Samson W. Tu, Carol A. Kemper, Nancy M. Lane, Robert W. Carlson, and Mark A. Musen. A methodology for determining patients' eligibility for clinical trials. *Journal of Methods of Information in Medicine*, 32(4):317–325, 1993.

[45] Carl Van Walraven and C. David Naylor. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *Journal of the American Medical Association*, 280(6):550–558, 1998.

[46] Dongwen Wang, Mor Peleg, Samson W. Tu, Edward H. Shortliffe, and Robert A. Greenes. Representation of clinical practice guidelines for computer-based implementations. *Medical Informatics*, 10:285–289, 2001.

[47] Haiqin Wang and Marek J. Druzdzel. User interface tools for navigation in conditional probability tables and elicitation of probabilities in Bayesian networks. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 617–625, 2000.

[48] Salim Yusuf, Peter Held, K. K. Teo, and Elizabeth R. Toretsky. Selection of patients for randomized controlled trials: Implications of wide or narrow eligibility criteria. *Statistics in Medicine*, 9:73–86, 1990.

[49] Cancer facts and figures 2002. American Cancer Society, 2002.