AD_____

Award Number: DAMD17-98-1-8039

TITLE: FACTS (Find the Appropriate Clinical Trials) for You: A Computer-Based Decision Support System for Breast Cancer Patients

PRINCIPAL INVESTIGATOR: Lucila Ohno-Machado, M.D., Ph.D.

CONTRACTING ORGANIZATION: Brigham and Women's Hospital
Boston, Massachusetts 02115

REPORT DATE: May 2001

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2001	Annual (20 Apr 00 - 19 Apr 01)

4. TITLE AND SUBTITLE
FACTS (Find the Appropriate Clinical Trials) for You: A Computer-Based Decision Support System for Breast Cancer Patients

5. FUNDING NUMBERS
DAMD17-98-1-8039

6. AUTHOR(S)
Lucila Ohno-Machado, M.D., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Brigham and Women's Hospital
Boston, Massachusetts 02115

E-Mail: machado@dsg.harvard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The *Find Appropriate Clinical TrialS* (FACTS) system was redesigned to make it more accurate and compliant with existing standards. An explicit data model of patient eligibility for breast cancer clinical trials was developed, and served as the basis for encoding eligibility criteria. Standard vocabularies were utilized to represent concepts used in the system, and to retrieve their hierarchical relationships. The system now uses Bayesian networks to handle missing patient information. Protocols are presented to the user ranked by the likelihood that the patient is eligible for each one of them. As a result of a detailed data model, most of the eligibility criteria in 10 clinical trial protocols taken from the National Cancer Institute database were encoded and the performance of the system was compared to that of two oncologists.

In a preliminary evaluation, there was a good agreement between the system's selection of clinical trials and those of two oncologists (kappa 0.86, 0.76). All cases in which the system's selection of a protocol did not agree with any of the physicians were analyzed, and the system's limitations were identified. The disagreement on ranking the protocols (kappa 0.24, 0.14) is discussed.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Artifical intelligence, Bayesian networks, Decision Support System	41
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Frontcover 1
Standard Form (SF) 298 2
Table of Contents 3
Introduction 4
Body 4
  Section 1. Overview of Tasks 4
  Section 2. New FACTS 8
    2.1. System requirements 8
    2.2. Clinical trial protocols 9
    2.3. Implementation 9
    2.4. Evaluation 28
  Section 3. Results 30
    3.1. Encoding process 30
    3.2. Preliminary system evaluation 31
    3.3. Analyzing disagreement 34
Key Research Accomplishments, Year 3 37
Reportable Outcomes 38
  Manuscripts 38
  Abstracts 38
  Presentations 38
  Informatics such as databases 38
Conclusions 39
References 40

Introduction

This report refers to the third term of a three-year award. The main accomplishments of this year, relative to the last progress reports are outlined in the body of this document.

Although participation in clinical trials has been shown to improve health outcomes, accrual of patients is difficult and is estimated to be below 5% of the eligible population [1]. Lack of information and automated tools to search clinical trials appropriate for each particular patient are some of the main reasons for low accrual. The purpose of this project is to build and evaluate a computer-based decision support system to help patients and primary care providers seek appropriate trials for their specific situation, even in conditions of uncertainty (missing data). We have proposed to make available, via the WWW, a search engine for clinical trial eligibility that searchers trials listed in the PDQ database of the NCI. On-line description of the project and working prototype can be found in http://dsg.harvard.edu/public/dsg/projects/facts.html

Body

Section 1. Overview of Tasks

Briefly, we have proposed to build our computer-based eligibility determination engine in two stages: (1) build an ad-hoc deterministic (i.e., non-probabilistic engine not able to deal with uncertainty or consider associations among eligibility criteria and patient data values), and (2) build a probabilistic engine, based on belief networks, that is able to statistically infer values for missing data, given the information it can gather from the patient or health care provider, and can take into account associations among variables and patient data values.

Previously accomplished goals were updated to reflect the improvements of the overall system. A description of the research accomplishments associated with each Task outlined in the Approved Statement of Work (restated in **bold** face) follows:

Task 1. **Analyze, structure, and construct data entry forms for eligibility criteria derived from clinical trials for breast cancer treatment available in PDQ, Months 1-6 (UPDATED):**

a. **PDQ clinical trial summaries for health care professionals will be dissected**

We have created an explicit data model for the representation of criteria. This model is scalable and is based on standardized vocabularies. A more detailed description of the model is given later in section 2.3.2.

b. A structured format for storing eligibility criteria in a relational database will be defined

As documented in previous reports, a relational database was not necessary to store the eligibility criteria, as the XML files were deemed more general and could be parsed in real time with no performance degradation.

c. WWW-based data entry forms will be constructed an linked to database

New forms to address the needs of primary care physicians were added.

d. Database for interim storage of patient data will be constructed

XML files continue to be used for this purpose.

Task 2. **Construct simple models that do not model uncertainty to assess the need for belief network models, Months 7-9 (UPDATED):**

a. **Simple rule-based system construction using knowledge from domain expert**

The outcomes of the rule-based system were updated to include probabilities of a criterion being met by a particular patient. Formerly, a deterministic system was used.

b. Preliminary evaluation of simple rule-based system

A comparison of system's performance with and without the probabilistic feature was made. Details are described in Section 3.2.

Task 3. **If results from Task 2 show that belief networks are needed, construct belief network to model uncertainty in most common eligibility criteria and perform inference on entered data, else refinement of simple models and interface construction will take place, Months 9-12 (UPDATED):**

a. Belief network model will be constructed using knowledge from domain expert

Dr. Nachman Ash, an internist and current postdoctoral fellow in medical informatics, reconstructed simple Belief networks featuring relations among laboratory values that are frequently encountered in eligibility criteria. A previous version constructed by Dr. Huan Le was deemed inappropriate. The small belief networks deal with few demographic data and laboratory values related to liver, renal, and hematologic function.

b. Belief network model will be integrated with WWW and database environments to create application

The belief network engine used in a previous version of the system was built with Netica. The current one is based on JavaBayes and is more flexible and robust.

c. Algorithm for ranking possible trials for a patient will be implemented

A new ranking algorithm was developed and implemented by Dr. Ash. Details are given in section 2.3.8.

d. GUI for displaying results and linking to specific summaries in PDQ will be built

The graphical user interface has been redesigned.

Task 4. **Redesign of evaluation methods and interim analysis and system refinement, Months 12-24 (UPDATED):**

a. Evaluation methodology will be redesigned

The evaluation strategy was redesigned to conform to the realities of the clinical services at Brigham and Women's Hospital (BWH) and Dana Farber Cancer Institute (DFCI). The need for unbiased oncologists to properly implement the proposed clinical trial was the critical point for its implementation in year 3. These oncologists were identified and participated in the evaluation of the system. Retrospective data from Brigham and Women's Hospital was obtained for preliminary testing of the model, with filing and approval from the Institutional Review Board. The IRB approval at the Dana Farber Cancer Institute was delayed due to administrative issues.

b. Interim analysis of the system using abstracted cases will be conducted

These cases were constructed based on actual retrospective data collected from the Brigham and Women's Hospital. Data from 20 patients admitted to Brigham and Women's Hospital with a diagnosis of breast cancer stage IV was used for thorough evaluation of the system and comparison of performance to that of oncologists. The items collected correspond to those on the WWW forms and were collected from the electronic medical record.

c. System will be refined in terms of belief network model and GUI given interim analysis results and internal user feedback.

The initial implementation was completely substituted given problems with its performance and connectivity to the other components of the system.

Task 5. **Subject recruitment, abstraction of medical records, and creation of survey instruments for final analysis, Months 16-24 (UPDATED or partially ACCOMPLISHED):**

a. Lay people ("patients") will be recruited (recruitment has started at BWH, and is pending IRB approval for DFCI).

We have contacted a number of organizations to help with the lay user interface, through contacts at the Harvard Medical School and the Massachusetts Department of Public Health Breast Cancer Program. Small focus groups for discussion of interface issues are currently being scheduled.

b. Medical records will be abstracted and randomized (ACCOMPLISHED)

Medical records were collected and abstracted by an internist. The data originated from the electronic medical record at BWH.

c. On-line forms for recording selection of clinical trials for "patients" and providers will be built (UNNECESSARY)

The construction of these forms was deemed unnecessary. Log files from the system could be analyzed for this purpose.

d. Surveys for assessing "patient" and provider satisfaction with the system will be built (ACCOMPLISHED)

The questionnaires are currently under review by the Internal Review Board of BWH and DFCI.

e. Primary care providers and oncologists will be scheduled for final experiments (partially ACCOMPLISHED)

We have used two oncologists and one internist for secondary and primary evaluation, respectively.

Task 6. **Evaluation experiments, Months 25-33 (partially ACCOMPLISHED – see particular items below:**

a. Oncologists will assess system's performance (ACCOMPLISHED)

b. "Patients" will use the system and fill on-line forms and surveys

We are still working on subject recruitment.

c. Primary care providers will use the system and fill on-line forms and surveys (ACCOMPLISHED).

The following Task is expected to be completed during the one-year extension that was recently approved:

Task 7. **Final analysis and report writing, Months 34-36:**

a. Final analyses of data from oncologists, "patients," and providers will be performed

c. A final report and manuscripts will be prepared

In the next sections, we describe the new version of FACTs, and illustrate with some screen samples from the existing system.

Section 2. New FACTS

2.1. System requirements

System requirements were outlined based on the goals of the new FACTS project and previous experience with similar systems.

The system should:

- Collect patient data and return a list of clinical trials for which the patient may be eligible. Trials in which at least one of the entry criteria is not met should be filtered out.
- Rank the trials by the likelihood of patient's eligibility.
- Reason with any amount and content of patient data, inferring values for missing data.
- Adhere to and make use of standards in medical informatics (e.g., controlled terminologies).
- Be generalizable: use common clinical trial protocols, and be expandable to different medical domains (not only the one that serves for prototype development).
- Be able to represent most of the eligibility criteria (at least 90%).
- Create a sharable encoded clinical trial protocols database.
- Be available to both patients and health professionals.
- Be accessible from anywhere (e.g., patient's home, clinician's office, inpatient ward).
- Have an intelligent user interface:
  - Ask for data and present results differently by the type of user: health professional or patient.
  - Ask for data items in an iterative way: ask first for the most common data items in the encoded protocols, generate results, and then let the user decide whether to enter more data, and thus narrow the list of appropriate protocols, or browse the results as they are. If the patient elects to enter more data, ask her for the most important data items.
  - Avoid redundancy (e.g., the system should not repeat questions about previously answered data items, it should not ask for stage of disease if it is known that the patient has metastasis).

- Generate explanations: show why a criterion was evaluated to true or false, and why a protocol was ranked the way it did.

2.2. Clinical trial protocols

Clinical trial protocols were taken from the NCI's PDQ database [2].

This source of protocols was selected since it is the most comprehensive resource on cancer clinical trials, which includes information about clinical trials sponsored by the NCI and others. Since one of the goals of this project is to create a general system, it makes sense to use a comprehensive source of protocols, rather than local institution-specific protocol database.

Another advantage of using PDQ's protocols is their availability on the Web through CancerNet in a single format that facilitates automatic retrieval of eligibility criteria by parsing the HTML protocol document.

As a start, analysis and testing were restricted to a subset of protocols: Phase II and Phase III trials for the treatment of metastatic or recurrent women's breast cancer. Working with this subset is initially warranted since it simplifies development, but the goal of creating a scalable system that could be applied to other domains needed to be considered as design decisions were made.

The selected domain is specific, but extensive:

- Breast cancer is the oncology domain that contains the largest number of clinical trials (201 listed in the NCI database as of April 2001).
- Patients with advanced disease would be more interested in seeking participation in clinical trials after exhausting traditional treatment venues.
- Phase II and Phase III trials are further developed than trials in other phases, and typically involve more patients.

Seventy-nine phase II and phase III protocol trials for the treatment of metastatic or recurrent women's breast cancer were found in the NCI's database as of February 2001 (82 on April 2001).

2.3. Implementation

The system was redesigned to follow several principles:

- Medical knowledge was encapsulated in an object-oriented data model.
- Concepts were represented using standard vocabularies.
- Eligibility criteria were encoded in a logical expression language derived from Arden syntax.
- Bayesian networks were incorporated into the system's evaluation process for inferring missing patient data.
- Evaluated protocols were ranked by the likelihood that the patient might be eligible for each of them.
- The system had a platform-independent implementation based on Java.

The following sections describe the implementation in detail.

2.3.1. High level design

The system is designed as a thin client, server-based application (thus, computing power and storage are centralized on the server, not the client). The user accesses the application via the Web. The design is based on a viewer-controller-model paradigm. The viewer is composed of several Java Server Pages (JSP), which constitute the user interface. The controller is responsible for coordinating the flow of data between the user interface and the model, and is implemented as a Java servlet. The model is the heart of the application where the eligibility criteria are evaluated.

Figure 1 illustrates the architecture of the system. The data collected from the user interface are stored and processed in the data model object. The belief network infers additional values. The processed variables and their values are sent to the evaluator manager, which coordinates the evaluation of the eligibility criteria. It takes criteria from the coded protocol database, and sends them with the appropriate data to be evaluated by the logical expression evaluator. The result of the evaluation of all protocols is the basis of a protocol's selection and ranking, which is presented to the user.

The "medical knowledge of the system" is embedded within the data model and in the medical vocabularies used by the system.

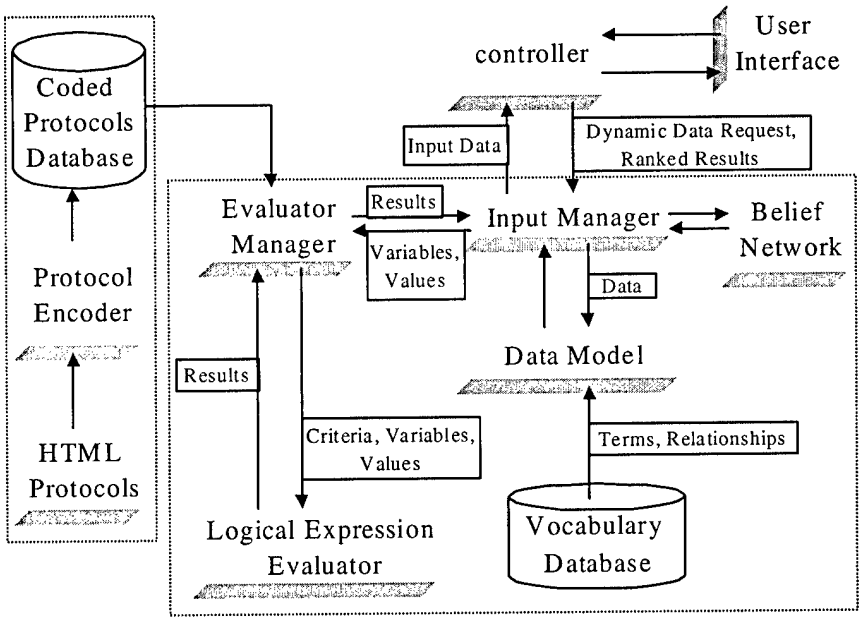


Figure 1: High level design of the new FACTS system.

2.3.2. Data model

In order to achieve the goals of the project, mainly encoding most of the entry criteria, the data model of the system had to be extended. The approach used in the previous implementation of the FACTS project was, unfortunately, difficult to extend as the data model was built as a data dictionary defined in an XML document. Extending this model would require entering all the data-types and terms that need to be used by the system (which would hinder extensibility and flexibility). Moreover, this data model was domain specific. Applying the system to a different medical domain would require creating a new data model, or extensively modifying the old one. Therefore, a different approach was chosen by creating a domain-independent object-oriented data model.

The use of an object-oriented approach has the following advantages:

- Modeling a complex domain such as eligibility for clinical trials requires compound classes (or data-types). Although an object-oriented approach is not the only alternative (frames could be used as well) it is well suited for this purpose.

- The compound data-types of the old model could easily be transformed to objects with attributes.

- Inheritance plays a key role in creating a model that is easily expandable. For example, in the FACTS system data model BREAST CANCER is a subclass of CANCER. In order to extend the model to clinical trials in the domain of prostate cancer, all that is needed is to add a couple of new objects, PROSTATE CANCER PATIENT that extends PATIENT and PROSTATE CANCER that extends CANCER. These new objects will probably contain few attributes, since most of the needed attributes are inherited.

  - Inheritance makes it easy to construct the model (the same common attributes do not need to be rewritten).

The data were modeled based on analysis of the breast cancer protocols and the Common Data Elements (CDE) of breast cancer clinical trials developed by NCI [3]. The data items in the model are those required for determining patient eligibility for a clinical trial. The model was designed (using the Unified Modeling Language design tool by TogetherSoft [4]) based on common medical knowledge. Figure 2 illustrates the breast cancer model.

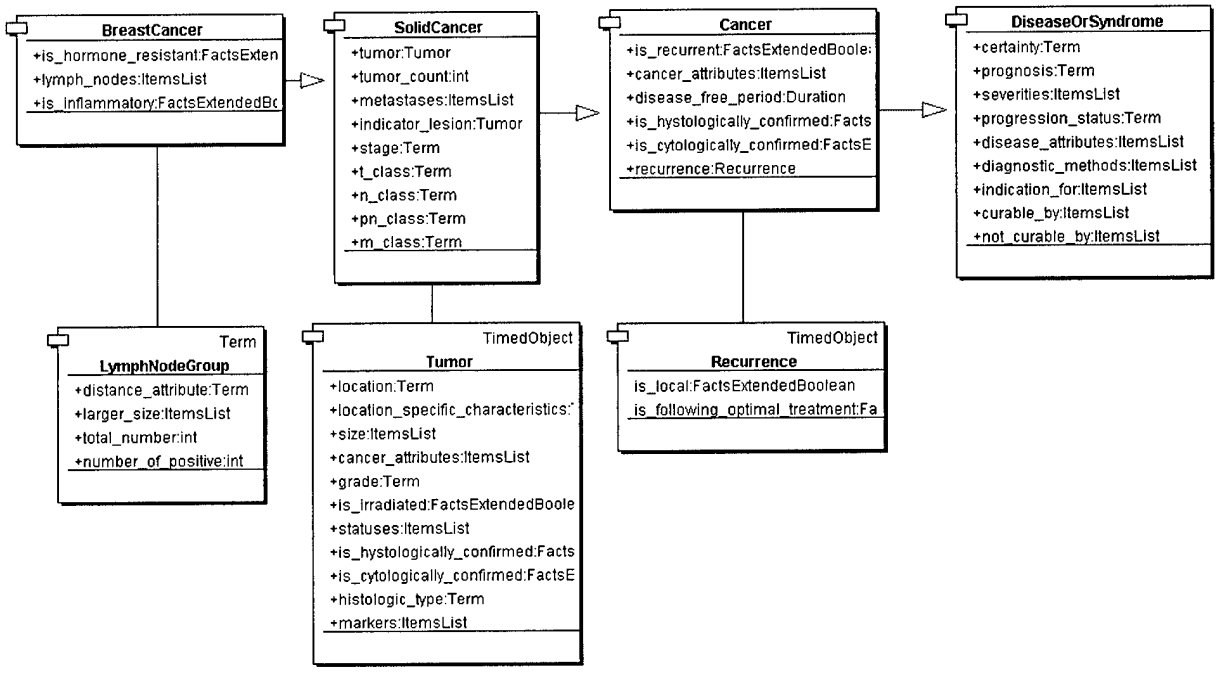


Figure 2: Part of the data model of breast cancer clinical trials.

The design of the model and the attribute names used in its classes impact the language created for encoding eligibility criteria (the variable names in this language are created by automatic transformation of attribute names – see discussion below). Therefore, it was important to use a design and names that resulted in "easily understandable" variable names. For example, the name of the histology type of the breast cancer tumor is represented by the variable name "breast_cancer.tumor.histologic_type.name".

Time plays an important role in evaluating eligibility for clinical trials. A frequent requirement, for example, is that certain treatment modalities had not been undertaken in a given time period ("more than 6 months since prior adjuvant chemotherapy"). Time was modeled by adding time stamps to data items (start_time, end_time and observation_time), and creating functions that use these time stamps to select the appropriate instance (latest, earliest, etc.).

It was also mandatory to model "not existing" in order to be able to say, for example: "the patient does not have congestive heart failure". That was done by adding an "is_present" attribute that is inherited by all objects in the model.

Patient data are stored in a model object ("BreastCancerPatient" in our case).

2.3.3 Use of standard medical terminologies

As opposed to the previous implementation of the system, the new system makes use of standard medical terminologies to represent terms and capture relationships between them. The advantages of using existing controlled terminologies are enormous:

- Time savings of not "reinventing the wheel": most of the needed terms and relationships already exist in standard vocabularies.
- A system that makes use of standard components is more acceptable.
- Terms in standard terminologies are mapped to the UMLS [5] and thus enable:
  - Linking of the system to other systems (like Electronic Medical Record systems).
  - Using various terms and strings that represent the same concept (e.g. "CHF" and "Congestive heart failure" can be used interchangeably).
  - Free text input is mapped to UMLS concepts, and thus gains a meaning.

Each term entered by the patient or used in the protocol eligibility criteria is looked up in the vocabulary database. The term's concept unique identifier (CUI) and its ancestors (terms which are more general in the thesaurus hierarchy than the patient's term) are retrieved, saved, and used while evaluating the encoded eligibility criteria (see Frame 1 for example).

Frame 1: An example of using CUI and relationships while evaluating

Text criterion: No history of diabetes mellitus
Encoded criterion: not have ("any name isa *diabetes mellitus* in diseases")

While the encoded criterion is evaluated the function "isa" checks if the value of the variable "diseases.name" isa "diabetes mellitus". That means that if the CUI of the value or one of its ancestors is equal to the CUI of "diabetes mellitus" the statement is evaluated to true.

Using relationships from standard terminologies has some pitfalls. The main one is that a terminology may contain hierarchic relationships that are inappropriate for the needs of the FACTS system. While generalization is suitable (e.g., "heart diseases" is a parent of "congestive heart failure), many other kind of hierarchic relationship are not. For example, in the COSTART vocabulary (one of the UMLS vocabularies), "diabetes mellitus" has a parent "Islets of Langerhans". While this relationship may be appropriate for the original intended use of this terminology, in the FACTS system the "isa" function may be inaccurately evaluated because of it. This problem was solved by restricting the use of relationships to two databases: MeSH (Medical Subject Headings) and Physician Data Query, giving priority to MeSH. These two were chosen because they contain most of the terms used in the clinical trial protocols, and appropriate terms' ancestors. For each term, the ancestors are taken from the MeSH database first. When there are no ancestors in MeSH, they are taken from the Physician Data Query database.

Some of the terms used by eligibility criteria in clinical trial protocols may not be found in the UMLS, and in some cases the necessary relationships may be missing from both MeSH and Physician Data Query databases. In that case, the user who encodes the criterion is able to add terms and relationships to the database.

2.3.4 Encoding language

Eligibility criteria are encoded using a variation of the Guideline Expression Language (GEL) [6], which is based on Arden syntax's logic grammar. Arden syntax was developed in order to facilitate sharing of medical logic among different health care institutions [7]. As the FACTS project is about using medical logic to evaluate eligibility for clinical trials, and since it is aimed at being sharable among institutions, the selection of the Arden syntax's logic grammar as the core of the encoding language was a natural choice. Moreover, Arden syntax was accepted as a standard of the American Society for Testing and Materials (ASTM) in 1992.

GEL was developed by the InterMed collaboratory (collaboration among medical informatics groups at Harvard, Stanford, and Columbia Universities [8]) for the GuideLine Interchange Format (GLIF) project [9.10] as a preliminary language that will capture the knowledge and logic of clinical practice guidelines. GEL differs from Arden syntax by letting the user define his or her own functions. This is a powerful property that enables extension of the language as shown below.

The encoding language is composed of 3 main components:

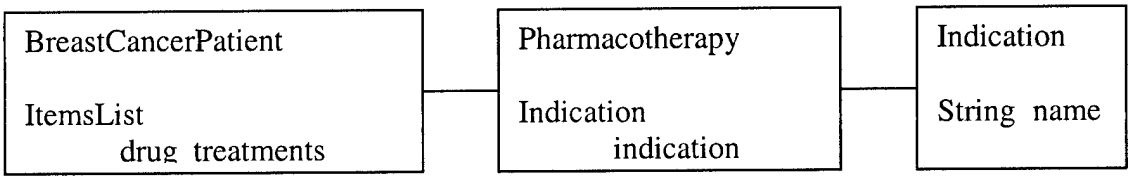
- GEL syntax
- Variable names
- Functions added to the syntax

The GEL syntax is a simple, yet powerful, logical expression syntax. It supports temporal functions and lists. However, it can deal with simple data types only (it supports neither complex data types nor objects). Therefore, the objects' fields in the data model need to be transformed into simple data type variables. This is done automatically by creating variables, the names of which are composed of the path of attributes from the root object to the leaf attribute (see Frame 2). The conversion function uses a depth-first search to create a total of 776 variables in the system.

Three functions were added to GEL for this project. Two of them (GET, HAVE) are used to retrieve values of variables from lists. These lists (of diseases, drug treatments etc.) contain complex data type (all attributes of disease or pharmacotherapy, for example). Since GEL does not support lists with complex data types, a function that retrieves the appropriate variable and sends it

for evaluation is needed. The GET function gets the value of the variable, while the HAVE function checks if the requested item exists and returns an extended boolean (*true*, *false* or *unknown*).

Frame 2: Transformation of attributes in objects to variables with simple data types



The root object is "BreastCancerPatient". It has an attribute "drug_treatments" which is a list of "Pharmacotherapy" objects. This object has an attribute "indication" which is an "Indication" object, which, in turn, contains the leaf attribute "name". The leaf attribute must have a simple type, in this case "String".
The variable name that holds the value of the indication to this drug treatment is "drug_treatments.indication.name". Similarly, the start time of this indication is named "drug_treatments.indication.start_time"

The third function is ISA, mentioned above. It takes a variable name and a string, checks the variable value, and returns an extended boolean (for example, it returns *unknown* if the value of the variable is a parent of the string, such as, when the patient is known to have "heart disease", but the criterion is "not congestive heart failure" – it is unknown whether the patient's disease is congestive heart failure). The behavior of the function is complex, since it must take into account "no existing" values (the patient says that she doesn't have congestive heart failure), and components in a list (the patient says that she doesn't have any disease).

One of the goals of this work was to create a language that might be comprehensible to medical professionals who may encode their own trial's eligibility criteria. Limited by the syntax of GEL, functions were designed to take one long string argument that might be more comprehensible for reading than composite strings would be. This long string is parsed by specific functions. It contains keywords that are used in various ways. Some of them indicate which item in a list should be retrieved (any, first, earliest, all, etc.), and others put constraints on the requested items (WHERE clause, CONTAINS clause). ISA can serve as a key word as well. NOTISA is another keyword, which is evaluated to not ISA.

As can be seen in the few examples given in Frame 3, the encoding language can be divided into two parts. The first one is retrieval of values from variables (GET and HAVE functions). The second one is a logical expression statement that is evaluated to *true*, *false* or *unknown*, and is the result of the criterion's evaluation.

Frame 3: Examples of encoded criteria.

Text criterion: Age 18 and over
Encoded criterion: age >= 18

Text criterion: Absolute neutrophil count at least 1,500/mm3
Encoded criterion: abs_neutrophil_count := get ("latest numerical_value from test_results where name
isa *NEUTROPHIL COUNT* and unit.name isa *cells/uL*");
abs_neutrophil_count >= 1500

Text criterion: At least 4 weeks since prior chemotherapy
Encoded criterion: had_chemotherapy := have ("any in chemotherapies");
chemo_end_date := get("ended_latest end_date from chemotherapies");
if had_chemotherapy then conclude not (chemo_end_date is within past 4 weeks); else
conclude not had_chemotherapy;endif;

2.3.5 Encoding process

The protocols selected for encoding were chosen by order of appearance in the search results of the PDQ database.

Encoding of the eligibility criteria is usually a manual process: each text criterion is examined and "translated" using an encoding language as described above. A special editor, created specifically for this project, retrieves the HTML page from the CancerNet™ Web site, delimits the eligibility criteria of that protocol, and presents them to the user, who needs to type in the GEL-based encoding (Figure 3). If a criterion is already encoded, its GEL-based encoding is retrieved from the database.

Most of the criteria encodings are simple, but some are more difficult, and the result does not completely reflect the original text. Reasons include:

- Use of vague terms in the text criterion ("Adequate cardiac function" -- what is adequate? "Newly diagnosed disease" -- what is newly? Not treated? Time-related?)
- Deficiency of the data model for capturing some of the concepts ("No evidence of disease improvement by radiography" -- the model currently does not capture the method used to collect evidence).
- Avoidance of long and cumbersome encoded criteria ("...unless tumor involvement in treated or incompletely treated patients" -- although this expression could be encoded, it would make the criterion very long and confusing. In certain cases, keeping the criteria simple was preferred).

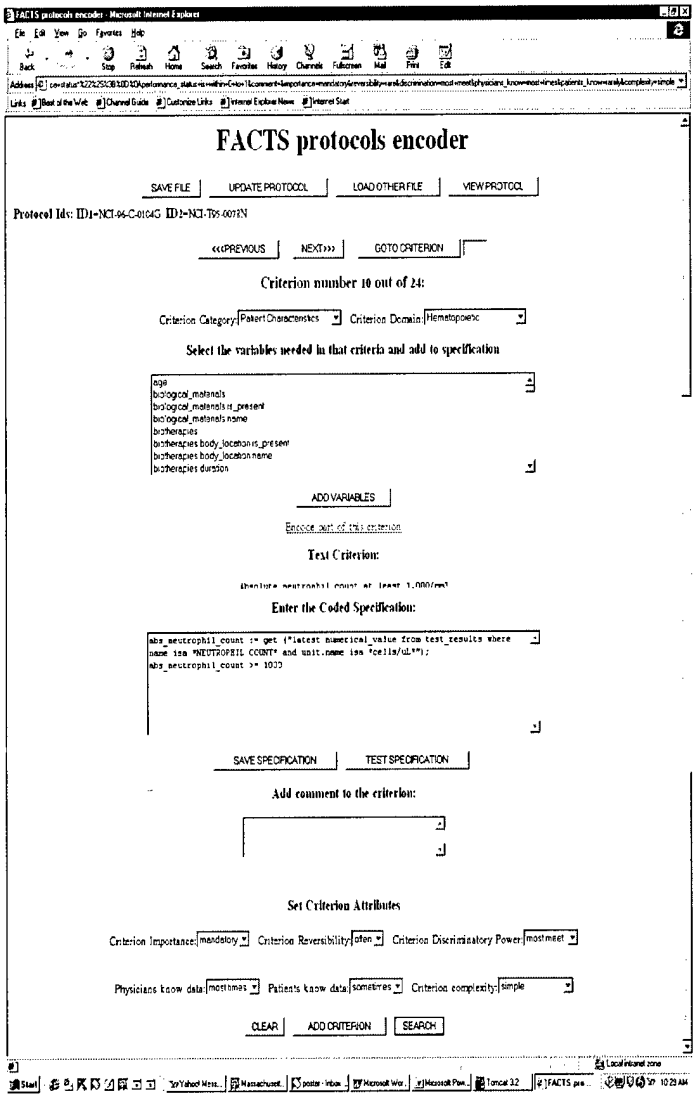


Figure 3: The FACTS protocols encoder. Text criterion is presented to the user who needs to type the GEL-based encoding in the middle window.

These difficulties were solved by different strategies:

- Transformation to a computable expression, even if not covering the whole semantics of the criterion (e.g., "Adequate cardiac function" is encoded by an expression that checks for normal ejection fraction).
- Use of vague terms in the encoded criterion ("uncontrollable hypertension") -- the user has to enter this information.
- Disregard of some information when it is considered not important (e.g., the method of measuring the ejection fraction is ignored with the assumption that most measurements are done by valid, interchangeable techniques).
- Addition of comments. The encoder can add comments that will be presented to the user of the system. The comment can clarify some aspects of the criterion, or just state that this encoding is not completely accurate.

The editor lets the user check the syntax of an expression for correctness, verify the legitimacy of variables' names used in the expression, and assess whether the terms used in the expression map to concepts in the UMLS.

For each criterion, the user needs to add the following information:

- The importance of the criterion (can it be ignored in some cases, or is it mandatory?).
- The reversibility of the criterion (if it is evaluated to false, can it change to true in the future?).
- Estimation of the discriminatory power of the criterion (do most patients who access the system meet this criterion? Or some of them? Or few of them?).
- Estimation of whether patients and physicians would know the values needed to evaluate this criterion (on a 1 to 5 rank scale).

This information is used by the system to rank the protocols and ask for more data (see below).

The encoded protocol is saved in both a Java object format (to be used by the system for eligibility determination) and an XML format (to view and share). Encoded criteria and information about the encoded protocols are saved in a relational database.

The time spent on encoding of each criterion is measured automatically and saved for analysis.

2.3.6. Missing data

The process of evaluating eligibility of a patient for clinical trials is data-intensive, as exemplified by the 776 variables defined in the system. Most users will probably enter only a small portion of the necessary values, both because they will not know the values of others, and because they will not be willing to spend sufficient time to enter all the required data. Therefore, it is expected that the system will have to deal with several missing values.

The new FACTS system infers missing values using two strategies. The first is deterministic: a missing value may be able to be deduced from a known value of a related parameter. The second is probabilistic and uses simple Bayesian networks.

2.3.6.1 Deterministic inference of missing values

There are two types of deterministic inference:

- Updates of linked data items using domain knowledge. For example: if a patient is known to have metastases, we know the stage of her disease (stage 4), or if a patient is known to be postmenopausal, she is also not pregnant, not fertile and not breast-feeding.
- Transformation of measurement units: different criteria may use different measurement units of the same test. For example, **ECOG 0-1** and **Karnofsky 70-100%** are two equivalent criteria regarding the performance status of a patient. When the system knows the value of the patient's performance status (in either measurement scale) it adds the value in all other possible scales. Thus any criterion using related measurement scales gets evaluated properly. This is used extensively for laboratory results that may be expressed in different units.

This kind of inference of missing values is important for several reasons:

- As the evaluation engine gets more information, its performance becomes more accurate, since more eligibility criteria are evaluated to a value other than *unknown*.
- It reduces the input burden: the system avoids asking the user to enter information on related items.
- Inconsistencies in input data are avoided.

2.3.6.2 Probabilistic inference of missing values

The protocol ranking may be more accurate by inferring missing values, since the ranking algorithm weighs results differently if they are based on inferred values (see below for more details). The system makes use of simple Bayesian networks to infer missing values.

A Bayesian (belief) network is a directed acyclic graph in which nodes represent variables, and arcs between nodes represent probabilistic relationships [11]. The network is created by selecting the desired variables needed to model the domain, adding appropriate causal arcs between them, and assigning prior and conditional probabilities. If some values of the variables are observed, the values of others can be inferred using Bayesian inference.

As discussed earlier, Bayesian networks have been proposed for eligibility evaluation systems by modeling the entire set of eligibility criteria of a protocol (or more than one) in a complex collection of networks [12,13,14]. This approach is not feasible for determining eligibility for multiple clinical trials. Therefore, creating several small independent networks that infer missing values of specific patient data items was preferred. These are general-purpose networks, modeling common medical knowledge related to frequently appearing data items in clinical trial protocols.

Currently, the system uses four separate directed acyclic graphs, representing age-related items (Figure 4), liver function tests, white blood cell counts, and pulmonary function tests. There are a total of 31 nodes in these graphs. The Bayesian networks were implemented using JavaBayes [15] as the Bayesian inference software.

Prior and conditional probabilities that populate these networks were taken in part from the medical literature (e.g., [16]). The remaining probabilities were estimated by the author based on medical knowledge. In the future, these probabilities could be updated by using relevant patient data, as they become available, in a manner suggested by Neapolitan [17]. Possible sources of such information may be clinical databases, and the database that will be created by data collected by the system.

The known patient data (data entered by the user) are inserted into the Bayesian networks as the observed evidence. The posterior probabilities are then calculated for all unknown variables in the network. If the posterior probability of a specific value is above a certain threshold (currently set to 5% above the chance probability), it is selected as the inferred value for the variable.

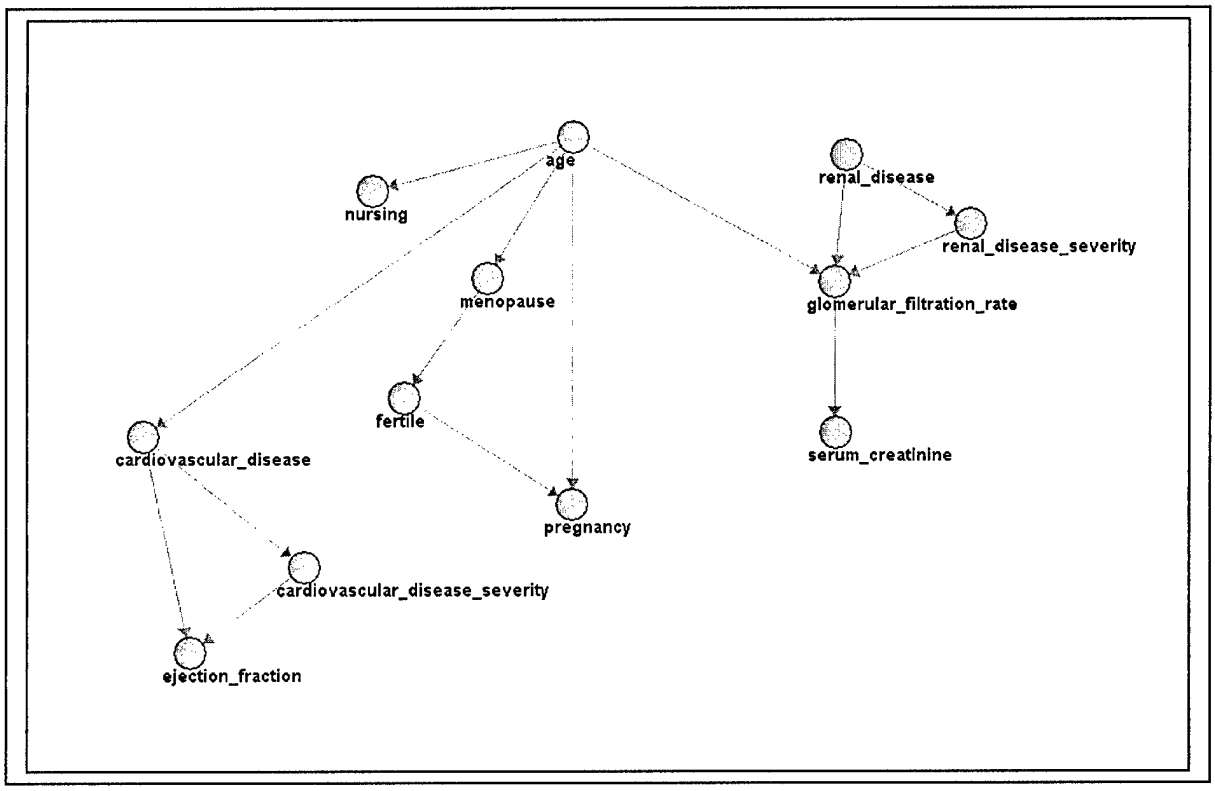


Figure 4: Age-related items organized in a typical Bayesian network used by the new FACTS

The posterior probabilities are not considered in the ranking of the protocols. Thus, a value inferred with a probability of 90%, and a value with a posterior probability of 30% (provided that it is above the threshold) are given the same weight during the ranking process. This limitation will be discussed later.

2.3.7. Evaluation of encoded criteria

A GEL parser / evaluator , built for use in the GLIF project (developed by Omolola Ogunyemi, Decision Systems Group, Boston, MA), evaluates encoded criteria. Variable names are replaced with values (if existing), and each expression in the criterion is evaluated. The evaluation result of the criterion is an extended boolean value (*true, false* or *unknown*). If the criterion can not be evaluated because of missing data, the result is *unknown*.

Each criterion is evaluated twice: once with data entered by the patient including deterministically-inferred data (definite data), and afterwards with probabilistically-inferred data. In the second round some of the criteria previously evaluated to unknown are evaluated to true or false.

The final result of a criterion evaluation is given as a letter symbol:

- **T** - criterion that evaluated to *true* based on entered and deterministically-inferred data only.
- **t** - criterion that evaluated to *unknown* based on entered and deterministically-inferred data, but evaluated to *true* when probabilistically-inferred data were added.
- **U** - criterion that evaluated to *unknown* based on entered, deterministically- and probabilistically-inferred data
- **f** - criterion that evaluated to *unknown* based on entered and deterministically-inferred data, but evaluated to *false* when probabilistically inferred data was added.
- **F** - criterion that evaluated to *false* based on entered and deterministically-inferred data only.

Thus, we get a rough qualitative measure of the likelihood that a patient meets the criterion: *T* and *F* represent the two extremes (100% and 0% respectively), and *t*, *U* and *f* represent ordinary intermediate values.

The result of a protocol evaluation is a list of these symbols, one for each criterion in the protocol.

2.3.8. Ranking of protocols

As stated above, the protocols should be ranked for a patient by the likelihood of that patient's eligibility. This is accomplished by examining and aggregating the evaluation results of the individual criteria in the protocol.

The patient is considered eligible for protocols for which all of the criteria evaluate to *T*. These are ranked highest and presented by the number of criteria that they contain.

Protocols for which one or more criteria evaluate to *F* are considered as inappropriate for the patient, and are therefore filtered out. Nevertheless, it is important to present these protocols to the user, and let him or her investigate why they were rejected. They are ranked separately, as discussed below.

The rest of the protocols contain any combination of criteria that were evaluated to *T*, *t*, *U*, or *f*. These are ranked by a weighted score that is dependent on the number of criteria that were evaluated to *t*, *U* and *f*. The weights represent the notion that the patient has a higher likelihood of eligibility for trials in which the criteria evaluated to *t*, than for those in which the criteria evaluated

to U . Similarly, a higher likelihood of eligibility for trials in which criteria evaluate to U is expected than for those in which criteria evaluate to f. Criteria that evaluate to U are weighted by their discriminatory power, using a scale predetermined by the encoder (see in "encoding process", above). Thus, a criterion with higher discriminatory power (i.e., one that is believed a priori to be true for only a small portion of breast cancer patients) gets a lower weight, and one that is believed to be true for most of the patients gets a higher weight.

It is important to notice that criteria that evaluate to f are not filtered out, but they have an increased probability of being ranked lower, determined by the weight of the criterion.

The algorithm described above was used to give each protocol a bottom line measure of appropriateness for a given patient on a scale of 1 to 5. Protocols for which all criteria evaluate to T get the maximal score of 5. Protocols for which at least one criterion evaluated to F get the minimal score, 1. Other protocols may get a score of 4 (the patient is probably eligible for the protocol), 3 (possibly eligible) or 2 (possibly ineligible), depending on the weighted score of the criteria, as described above.

As mentioned above, protocols that contain criteria that evaluate to F are filtered out, but are presented to the user for inspection. These protocols are ranked by the likelihood of the patient's eligibility despite this result (i.e., the protocol can be useful in the future if, for example, the patient's status changes, or if the clinical trial researcher believes that the criterion that evaluated to F is not too important). This ranking is achieved by evaluating the importance and reversibility scores that were given to the criteria during encoding (see above). If the criterion that evaluated to F is deemed not very important and is reversible, the patient may become eligible for the protocol. On the other hand, if the criterion is important or irreversible, then the patient is definitely ineligible for the protocol, and it will be ranked lowest.

Frame 4 contains a simple example of a ranked protocol list.

Frame 4: Example of ranked protocol list. The first one contains 1-*t*, 8-*U*, 1-*f*. The second one contains 2-*t*, 9-*U*, 1-*f*. Therefore, there is a higher likelihood that the patient is eligible to the first protocol that contains fewer unknown and probabilistically-inferred criteria. The two bottom protocols are filtered out, since they contain at least one criterion that evaluated to *F*. Notice that protocols containing criteria that evaluated to *f* are not filtered out.

```
protocol: NCI-G00-1878 ranked 1
20 criteria in this protocol were evaluated as follows:
  U  U  T  T  T  U  T  T  T  f  T  T  U  T  U  U  t  U  U  T

protocol: NCI-96-C-0104G ranked 2
24 criteria in this protocol were evaluated as follows:
  U  U  T  U  U  T  T  U  U  T  T  T  U  U  t  T  f  T  T  T
T  t  U  T

The following protocols are NOT appropriate for the patient:

protocol: NCI-G00-1834
24 criteria in this protocol were evaluated as follows:
  U  U  T  U  U  U  T  U  T  U  T  U  U  T  T  T  F  T  T  T
T  t  U  U

protocol: NCI-V97-1341
20 criteria in this protocol were evaluated as follows:
  T  U  U  T  T  U  T  T  U  T  f  T  F  U  f  t  T  U  T  T
```

3.9. User interface

The user interface was implemented as several JSP files that are controlled by a Java servlet. All pages, except the first introductory one, are generated dynamically, depending on which protocols are encoded, what input from the user is available, what the evaluation result of the protocols is, and what the user wants to see or do.

There are two user interfaces: one for use by patients and their representatives (herein called the "patient" interface), and another for use by health professionals. They differ in several aspects:

- The data items requested of the user (e.g., the patient is not asked to estimate her life expectancy, or to describe the histology type of her tumor).

- The way the request for data is presented to the user (e.g., when asked to enter the daily performance status, the patient gets a detailed description of the choices, while the health professional is asked to enter the value of the ECOG performance status).
- The way that the user enter the data (e.g., the patient is requested to enter diseases by using a simple menu, while the physician enters them as free text).
- The way the results are presented to the user (e.g., the patient gets a list of protocols for which she may be eligible, while the health professional gets also the evaluation results of the criteria, and the list of protocols that were filtered out).

The first input form refers to values of most frequent data items in the encoded protocols (Fig. 5). The encoded criteria are analyzed automatically to find those that appear most frequently. For each data item, the program checks if there is no limitation on presentation to the user. Some items are not presented to patients either because they probably would not know the value, or for other reasons (e.g., life expectancy is too sensitive a topic for the patient interface).

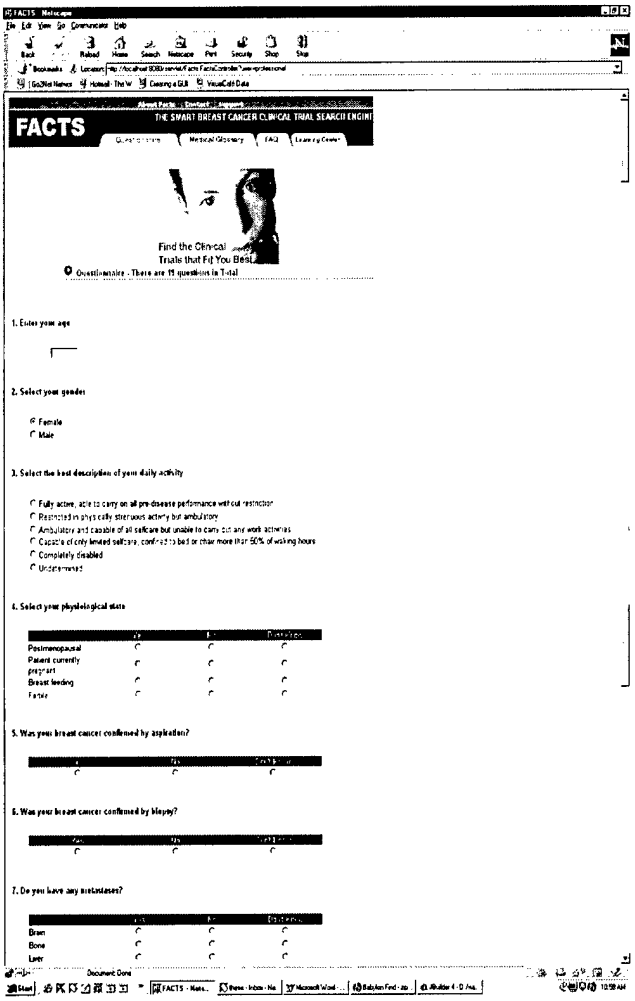


Figure 5: First input form generated dynamically based on the encoded criteria.

When the user submits her first set of answers, the system checks the data for allowed values, and evaluates the encoded criteria with the patient data. The user is presented with the number of appropriate protocols found, and can choose either to see the results or to enter more data in order to further narrow the protocol list.

Other input forms are created dynamically for data in criteria that evaluated to *unknown*. Once again, if the criterion is considered a priori as probably not known by the patient (as determined by the encoder of the criterion), it will not be asked. The system does not repeat questions for items that were already answered (even if they are still unknown).

The user may answer any item she wishes, and skip others. The system can reason with any number and content of data items.

The full results are presented to the user as a ranked list of protocol names. The clinical trial names are linked to the corresponding protocol summaries at CancerNet according to the type of the user (e.g., results for patients are linked to patient summaries).

Health professionals are exposed to a more detailed result (Fig. 6), including the evaluation results for the criteria (the numbers of those that evaluated to each of the categories *T,t,U,f,F*), and protocols that were filtered out.

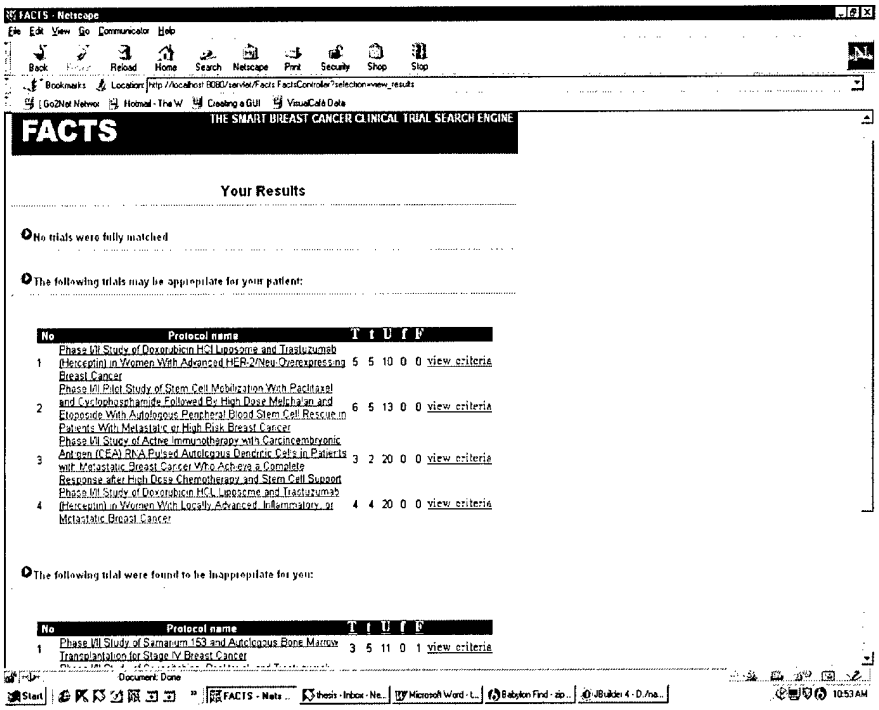


Figure 6: Presenting results to health professional: the names of the protocols presented with the number of criteria evaluated to T,

2.4. Evaluation

A preliminary evaluation of the system's selection and ranking algorithms was conducted, in order to get a preliminary measure of its agreement with selection and ranking by expert physicians.

Patient data were abstracted from medical records of 20 patients with active metastatic or recurrent breast cancer, who were consecutively hospitalized during 1995 at the Brigham and Women's Hospital, Boston, Massachusetts. Forty-three data items were examined for each patient (items related to patient characteristics, disease characteristics, past treatment, other diseases and test results). Researchers not familiar with the encoding process and the particular encoded protocols collected the data. They decided which data items to collect by general familiarity with PDQ's protocols.

Two independent oncologists evaluated the appropriateness of the protocols for each of the patients, and ranked them. The physicians were given a short narrative description of the patients' data, and the

full abstracts of 10 protocols as downloaded from NCI's CancerNet Web site. When evaluating the appropriateness of the protocols for each patient, they were requested to give a score for each protocol (from 1 to 5, similar to the system's score, as described above), and then to rank the protocols that they found appropriate for the patient.

The system used the same patient data to evaluate the eligibility of the patients for each of the clinical trials.

The agreements on selection and ranking of protocols between the system and each physician and among the physicians were calculated using the kappa and weighted kappa statistics [18,19]. Statistical analysis was conducted using Microsoft Excel and Analyze-it [20].

Section 3. Results

3.1. Encoding process

The first 10 protocols listed on the search results from NCI's database were encoded. Each protocol contains between 20 and 41 eligibility criteria (mean 27.2). Out of 272 criteria, 228 (83.8%) criteria were unique. Criteria were considered unique if they were written in the protocols in a unique manner. If, for example, two criteria express the same idea, but are written differently, they represent two unique criteria (e.g., "No other concurrent antineoplastic agents" and "No other concurrent antineoplastic therapies").

It was feasible to encode 269 (98.9%) criteria. Thus, between 96.4% and 100% of the criteria in each protocol were encoded. The encoding process resulted in 141 (61.4% of the unique criteria) distinct encodings (in our example above, the two unique criteria had the same identical encoding).

Three criteria were not encoded. Two of them ("no prisoners" and a criterion related to a specific geographic location) lacked representation in the model. The third ("No other concurrent medical or psychological condition that would preclude study compliance") is difficult to encode because it involves complex human judgment. A total of 39 other criteria (27.6%) did not represent their text version with 100% accuracy (e.g., "No medical or psychiatric condition that would increase risk" was encoded as "No severe medical or psychiatric condition" -- since assessment of risk is subjective, it is difficult to encode for computation purposes).

A moderate number (30.3%) of the encoded criteria were lengthy (> 255 characters), which is indicative of their being among the more complex criteria.

Table 1 presents the encoding time for 77 criteria from the last three protocols. Approximately 20% of the criteria were labeled as difficult or complex. Retrieval of the code from the database was possible in 23.3% of the criteria, as these criteria were already encoded in other protocols. Most of the criteria were encoded in less that 4 minutes, but in some cases nearly one hour was necessary (this includes the time taken to make some changes in the data model in order to enable encoding of these criteria). The average encoding time was 5.88 minutes (median 2.1). Therefore, encoding an average-sized protocol may take about 3 hours.

Table 1: Average encoding time of 77 criteria stratified by difficulty.

Criterion Difficulty	Number of Criteria	Average Encoding Time (Min)
Automatic Coding	18	≈ 0
Trivial	8	1.47
Easy	35	3.52
Difficult	9	11.12
Complex	5	28.12
Very Complex	2	36.80

3.2. Preliminary system evaluation

Data from 20 patients with metastatic, locally invasive, and recurrent breast cancer were collected from medical records of the Brigham and Women's Hospital, Boston. About 25% of the 43 data items requested for each patient had missing values. Age distribution was 25-71 years (mean 44.4). Other patient characteristics are shown in Table 2.

Table 2: Patient characteristics.

Data Item	No. of patients (percent)	Data Item	No. of patients (percent)
Disease Stage:		**Known Metastases**	11 (55%)
Stage IV	5 (25%)	Liver	7 (35%)
Stage IIIb	5 (25%)	Lung	4 (20%)
Unknown	10(50%)	Bone	5 (25%)
Tumor Histology:		**Recurrent Disease**	3 (15%)
Invasive Ductal Ca.	1 (5%)		
Unknown	19 (95%)		
Confirmed Histology/Cytology	17 (85%)	**Locally Advanced Disease**	8 (40%)
Measurable/Evaluable Disease	14 (70%)	**Known Lymph Node Involvement**	9 (45%)
Menopausal Status		**Other Diseases:**	
Postmenopausal	5 (25%)	Hypertension	3 (15%)
Premenopausal	8 (40%)	NIDDM*	1 (5%)
Unknown	7 (35%)	Asthma	1 (5 %)
Past Treatment			
Chemotherapy	16 (80%)		
Radiotherapy	6 (30%)		
Biotherapy	8 (40%)		
Hormonal therapy	7 (35%)		
Surgery	7 (35%)		

*Non Insulin Dependent Diabetes Mellitus

Table 3: Distribution of criteria evaluation results.

Criteria Evaluation	Criteria Number (percent)
TRUE	2283 (41.96%)
FALSE	210 (3.86%)
UNKNOWN	2947 (54.18%)
true (inferred)	515 (9.47%)
false (inferred)	39 (0.72%)

The process of protocol selection for these 20 patients involved 5,440 evaluations of 272 criteria (each criterion was evaluated 20 times, each time with different patient data). As can be seen in table 3, about 54% of the evaluations resulted in *unknown* because of missing patient data. After inference by the Bayesian networks, 18.8% of these evaluated to either *true* or *false*.

The system selected from 1 to 9 protocols per patient (Figure 7). On average 3.05 protocols were selected per patient. None of the selected protocols received an appropriateness score of 5 (*definitely eligible*) or 4 (*probably eligible*), 25 were graded 3 (*possibly eligible*), and 36 were graded 2 (*possibly ineligible*).

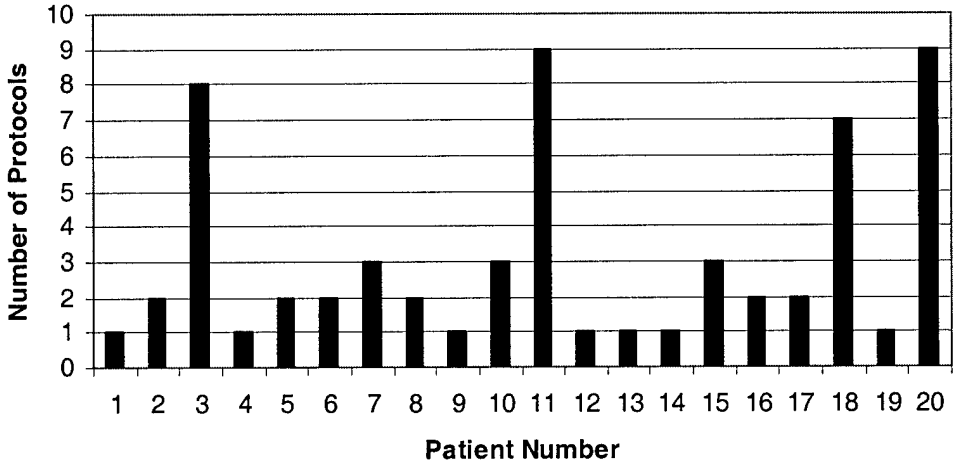


Figure 7: Number of protocols selected per

In order to see the impact of inferring missing values by the Bayesian Network, the system was tested with and without Bayesian network inferred values. As expected, fewer protocols received grade 3 without the Bayesian network inference (19 without versus 25 with the probabilistic inference). The protocol ranking was affected for 4 patients. In two of them, the protocols ranked first and second were swapped as a result of adding inferred values.

The system's results were compared to physicians' selection of protocols with respect to two aspects: (1) the agreement on whether the patient would be eligible for each protocol, and (2) the agreement on protocol ranking for each patient. The kappa statistic for patient eligibility was 0.86 (95% CI 0.72 - 1.00) for one physician and 0.76 (95% CI 0.62 - 0.9) for the other. The agreement between the two physicians was 0.72 (95% CI 0.58 - 0.86).

The agreement on ranking the protocols was low: weighted kappa of 0.24 and 0.14 between the system and the two physicians respectively, and 0.31 between the two physicians.

3.3. Analyzing disagreement

There are two possible kinds of disagreement on selection of protocols: (1) the physician might select a protocol that the system found to be inappropriate for the patient (**extending disagreement**), and (2) the physician might not select a protocol that the system found to be appropriate (**narrowing disagreement**). There were 2 narrowing disagreements and 10 extending disagreements with one physician, and 14 and 6, respectively, with the other. Thus there were 16 disagreements of each kind altogether. The physicians shared only 4 of the disagreements (2 of extending type and 2 of narrowing type).

Table 4: Classification of disagreements between the system and the physicians.

Type of disagreement	Number of disagreements
Lack of model representation	1
Encoding mistake	1
Simple inference of missing value by physician	1
Complex inference by physician	12
Physician mistake	6
Interpretation of a borderline pathologic test result	3
Use of information other than eligibility criteria	1
Misinterpretation of patient data	3

In each case, the physicians were asked to explain their decisions. Based on the explanations, several common reasons for disagreement were found (table 4):

- **Insufficient model representation** causing inaccurate criterion encoding. For example consider the following inclusion criterion: "Previously treated with paclitaxel and an anthracycline (if medically appropriate) as adjuvant therapy or for metastatic disease". The encoding of this criterion checks if the patient got treatment with these drugs, but does not check if this treatment is "medically appropriate" for the patient (this was added

as a comment for the user). In one case, it was known that the patient did not get these therapies (and therefore the system evaluated the criterion to *false*), but one of the physicians considered these therapies inappropriate for the patient, and therefore decided that the patient met the criterion (extending disagreement).

- **Encoding mistake -** wrong code for a criterion.
- **Simple deterministic inference of missing value** – a physician deduced a missing value from another known value, while the system failed to do the same.
  For example, both physicians concluded that a patient with chest wall involvement is eligible for a trial that required locally invasive disease, while the system failed to infer that chest wall involvement implied locally invasive disease.
- **Complex inference of missing value** – a physician made some assumptions and inferred new information about the patient.
  For example, the physician inferred that a patient with metastatic, non recurrent and non progressive disease who received chemotherapy in the past, received it for treatment of the metastatic disease (and therefore was not eligible for a protocol that excluded patients with previous chemotherapy for metastatic disease).
- **Physician mistake**, usually as a result of ignoring some known information about the patient, or failure to notice a criterion in the protocol.
- **Interpretation of a borderline pathologic test result** as not clinically justifying exclusion from the trial.
  The system has a deterministic approach to test results: any value outside a limit specified by the criterion will result in evaluating the criterion to *false*. Sometimes physicians may disregard a result that is only slightly beyond appropriate limits. For example, one of the physicians decided that ejection fraction of 47% is appropriate even if the criterion required a normal ejection fraction (above 50%).
- **Use of information other than eligibility criteria** -- Physicians considered information given in the clinical trial protocol outside of the eligibility criteria section.
  For example, in one protocol, the title of the trial restricted the trial to patients with metastatic disease, but no corresponding eligibility criterion was stated.

- **Misinterpretation of patient data** resulting from unclear presentation of the case. For example, a patient with recurrent disease and skin involvement was considered by one of the physicians to have skin metastasis.

Key Research Accomplishments, Year 3

- Created data model
- Incorporated standard vocabulary
- Redesigned and reimplemented Bayesian networks
- Redesigned graphical user interface
- Created new algorithm for selection and ranking
- Conducted pilot evaluation with two oncologists
- Collected and abstracted real cases from Brigham and Women's Hospital
- Started recruitment of subjects

Reportable Outcomes

Manuscripts

Ash, N. New FACTS (*Find Appropriate Clinical TrialS*): A Computer Based Decision Support System for Breast Cancer Patients. Master of Science in Medical Informatics Thesis. Harvard-MIT Division of Health, Sciences and Technology, May 2001 [Appendix 1]

Abstracts

Ohno-Machado L., Ogunyemi O, Greenberg S, Boxwala A, Greenes RA. Finding Appropriate Clinical Trials. The Internet and the Public's Health Meeting, 2000. Boston, MA.

Ohno-Machado L, Wang S, Greenberg S, Boxwala A. Using the Internet to Find Appropriate Clinical Trials for a Patient: The FACTs project. Proceedings of the Era of Hope, Department of Defense Breast Cancer Research Program Meeting, Atlanta, 2000; 803.

Presentations

Poster presentations at

The Internet and the Public's Health Meeting, 2000. Boston, MA.

Era of Hope, Department of Defense Breast Cancer Research Program Meeting, Atlanta, GA.

Informatics such as databases

Database of Encoded Protocols available at http://dsg.harvard.edu/FACTs/NewFacts/source

Conclusions

We have accomplished the overall tasks of year 3 towards the construction of an automated system to automate patient eligibility match to suggest appropriate protocols for a specific patients. We have refined the prototype built in year 2. We have re-implemented an engine that deals with uncertain items and infers appropriate values. We have evaluated the system and compared its performance with that of two oncologists using data from the electronic medical record at Brigham and Women's Hospital. We have started to recruit subjects to our evaluation trial.

Our next steps are to (1) evaluate the interfaces, (2) recruit other physicians for evaluating agreement, (3) encode more trials.

References

1. Ohno-Machado L, Boxwala AA, Wang SJ, Mar P. Decision Support for Clinical Trial Eligibility Determination in Breast Cancer. *Journal of the American Medical Informatics Association* 1999; Suppl 6: 340-4.
2. CancerNet - search for clinical trials (PDQ - Physician Data Query) - A National Cancer Institute service, in http://cancernet.nci.nih.gov/trialsrch.shtml.
3. Silva J and Wittes R, Role of clinical trials informatics in the NCI's cancer informatics infrastructure. Proc AMIA Symp, 1999:950-4.
4. Together - the control center for accelerating e-solutions, v. 3.2, TogetherSoft, Raleigh, North Carolina.
5. Lindberg D, The Unified Medical Language System (UMLS) of the National Library of Medicine. J Am Med Rec Assoc, 1990;61:40-2.
6. Ogunyemi O, The Guideline Expression Language (GEL) User's guide, Technical Report, DSG-TR-2000-001, 2000, Decision Systems Group, Boston, MA.
7. Hripcsak G, Writing Arden Syntax Medical Logic Modules. Comput Biol Med, 1994;24:331-63.
8. Shortliffe EH, Barnett GO, Cimino JJ, et al., Collaborative medical informatics research using the Internet and the World Wide Web. Proc AMIA Annu Fall Symp, 1996:125-9.
9. Ohno-Machado L, Gennari JH, Murphy SN, et al., The guideline interchange format: a model for representing guidelines. J Am Med Inform Assoc, 1998;5:357-72.
10. Peleg M, Boxwala AA, Ogunyemi O, et al., GLIF3: the evolution of a guideline representation format. Proc AMIA Symp, 2000:645-9.
11. Pearl J, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. 1988, San Mateo,CA: Morgan Kaufmann.
12. Tu SW, Kemper CA, Lane NM, et al., A methodology for determining patients' eligibility for clinical trials. Methods Inf Med, 1993;32:317-25.
13. Ohno-Machado L, Parra E, Henry SB, et al., AIDS2: a decision-support tool for decreasing physicians' uncertainty regarding patient eligibility for HIV treatment protocols. Proc Annu Symp Comput Appl Med Care, 1993:429-33.
14. Papaconstantinou C, Theocharous G, and Mahadevan S, An expert system for assigning patients into clinical trials based on Bayesian networks. J Med Syst, 1998;22:189-202.
15. Cozman FG, JavaBayes - a system that handles Bayesian networks implemented in Java, v. 0.346.
16. Ries LAG WP, Miller DS, Howe HL, Weir HK, Rosenberg HM, Vernon SW, Cronin K, Edwards BK, The Annual Report to the Nation on the Status of Cancer, 1973-1997, With a Special Section on Colorectal Cancer. Cancer, 2000;88:2398-2424.

17. Neapolitan R, Probabilistic reasoning in expert systems: Theory and algorithms. 1990, New York: Wiley, John & Sons, Inc.

18. Daniel Ww, Biostatistics: a foundation for analysis in the health sciences. Wiley series in probability and mathematical statistics - applied. 1995, New York: Wiley, John & Sons, Inc.

19. Altman D, Practical Statistics for Medical Research. 1991: CRC press, LLC. 403-409.

20. Analyse-it! Statistical analysis add-in software for Microsoft Excel, v. 1.61, Analyze-it Software LTD, Leeds, UK.