

Complications in Using Automated Methods to Increase Clinical Trial Accrual

Sergiy Fefilatyev^{1,3}, M.S.,
Li Chen¹, M.S.,
Tim V. Ivanovskiy¹, M.S.,
Lawrence O. Hall¹, Ph.D.,
Dmitry B. Goldgof¹, Ph.D.,
Shibendra Pobi¹, M.S.,
Halina Greenstien²,
Amit P. Pathak², M.D.,
Christopher R. Garrett, M.D.²,

¹Computer Science and Engineering, University of South Florida, Tampa, FL 33620 USA

sfefilat@cse.usf.edu, lchen2@cse.usf.edu, ivanovsk@cse.usf.edu, hall@cse.usf.edu,
goldgof@cse.usf.edu, spobi@cse.usf.edu ,

²H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612 USA
Halina.Greenstien@moffitt.org, Amit.Pathak@moffitt.org, Chris.Garrett@moffitt.org

³Sergiy Fefilatyev, 4202 E.Fowler Ave., Tampa, FL 33620
Phone: (813) 974-3925

Keywords: expert system, clinical trial accrual, difficulties, cost

Abstract

Purpose:

This paper reports on the issues involved in improving accruals for cancer-related clinical trials. Clinical trials aim to determine the efficacy of novel therapies and must be conducted before new medical treatments become available to the public. In order for the clinical trial to be successful, a predetermined number of patients with an appropriate set of medical conditions must be accrued. Many of the cancer-related clinical trials in the US fail to accrue the required number of participants in the projected time frame or sometimes at all. An automated intelligent system is proposed to help physicians screen patients and increase accrual for clinical trials.

Methods:

We have implemented a web-based expert system at the H. Lee Moffitt Cancer Center & Research Institute in the Gastrointestinal Tumor Clinic to help physicians screen patients for phase II clinical trials. The participatory design model was used to make expert system user-friendly; optimization techniques such as data mining and statistical analysis were used to decrease the amount of data entry. Our system allows physicians to screen a patient for multiple trials simultaneously. Data available from clinical deployment of the expert system is shown and analyzed.

Results:

A slight increase in clinical trial accrual was observed. However, only 24.5 % of patients that went through the clinic were screened with the help of the described expert system. No patients other than those also found by the physicians were identified as eligible for clinical trials

but the conducted analysis suggests new patients may be found if the percentage of screened patients is higher.

The HIPAA authorization form was found as a major hurdle in our research and an additional burden for clinicians, although it is not possible to determine effect it had, since under current regulations in the US it is not possible to work without it.

It is likely possible to extract much of the needed information from the clinical records through automated methods. As shown by our analysis, nearly 80% of the patients could be automatically screened within two weeks of a visit.

Conclusions:

Adaptation of the system into a clinical environment and the success of the system are highly dependant on the usability, which can be heavily affected by side effects of research regulations, and are also related to the amount of time physicians are willing to spend working with that tool. A commercial system that did not require consent forms might well increase accruals. Further, a system which extracts data from patient records would significantly decrease the need for clinician input leaving only a small percentage of patients in the category of potentially eligible with more information.

I Introduction

A clinical trial is an experimental research study that is carefully designed to answer specific questions concerning the effect of a treatment, diagnostic method, or therapy. The participants of clinical trials are human volunteers with a set of medical conditions for which the trial is designed. Selection for a clinical trial is guided by a study plan (protocol), which consists of a set of inclusion and exclusion criteria that will either make a research subject eligible or ineligible for a trial.

Determining the eligibility of a patient for a clinical trial is a time-consuming and labor-intensive task. The eligibility criteria include a multitude of questions that are based on a patient's medical history and current medical status. Variations in protocol eligibility make it difficult to screen a patient for multiple clinical trials. Studies have even shown that up to 60% of eligible patients are missed by clinicians, resulting in the failure of many clinical trials due to failure to accrue adequately [1,2,3].

The application of artificial intelligence to real-world problems has produced promising results [4]. The work presented in this paper is aimed towards understanding barriers to accrual and increasing the acceptance of the medical expert system called the Moffitt Expedited Accrual Network System (MEANS) into the clinical environment. This expert system is designed to be used as a tool for helping physicians screen patients for clinical trial eligibility. Personal data acquired retrospectively from patients with breast cancer has shown that MEANS has the potential to increase accrual for clinical trials [5]. MEANS allows clinicians to screen any patient for multiple clinical trials at once. We used a participatory design to increase user friendliness and the acceptability of the expert system [3]. In addition, optimization techniques such as data mining and statistical analysis were used to decrease the amount of data entry. We implemented

our expert system for use in the Gastrointestinal Tumor Clinic (GITC) at the H. Lee Moffitt Cancer Center & Research Institute , one of the largest cancer-dedicated hospitals in the United States, to screen patients for phase II clinical trials. This paper presents the challenges of getting people to use such a system, results from its use, and some potential solutions to obstacles encountered.

II Related Work

There were a number of systems created to increase clinical trial accrual. Most of them were developed for a specific medical setting: particular hospitals, certain trials and none was deployed on multi-center basis or for many trials.

The authors of [6,7] describe a system for clinical trial accrual which was deployed in an emergency department. The system works in cycles, scanning patients at a fixed frequency. It checks laboratory values like blood sugar and triggers an alert if an eligible patient is found. The coordinator then contacted these potentially eligible patients and arranged additional tests. The system has been reported to find 10 new patients for one clinical trial. From the technical point of view the success of this real time system for automated accrual is explained by the simplicity of inclusion/exclusion criteria. A few easy to query fields in a patient's medical record were checked in order to alert the coordinator. It is not clear how the system would work in a different medical setting when most of the protocols, for example in cancer-related trials, require more in-depth questions to be answered. In such a case, relying only on easy ones would heighten the alert traffic to an unacceptable level.

The authors of [8] focused their study on an architecture for electronic medical record so that clinical trials accruals can be facilitated by using electronic means. This system is also designed for mass screening: it checks only data elements in a medical record that would signify

patient's eligibility for a trial, avoiding eligibility criteria that require human involvement. The authors emphasized the importance of a solution for legislative barriers to automated clinical trial accrual (HIPAA). Many steps for hiding patients' personal data were taken; notifications about the possible eligibility criteria were sent to the ordering physicians and coordinator, but the possibility of offering participation in a trial was left to the treating physician. The prototype of the system was tested by processed medical records of almost 70 thousand patients but only three patients were accrued as a result. Various reasons to explain this poor result are given. Participation of the ordering physicians is a major one since it is left to the physician to refer a patient to a trial.

The study described in [9] considered a process of facilitating clinical trial accrual from the technical perspective. Based on the information about the medical status of a patient their system was asked to give a list of clinical trials for which a patient is likely to be eligible. Major work was done in encoding the eligibility criteria for those trials and extracting information from a patient record so that it could be compared with the criteria. Encouraging results were obtained from testing the system on retrospective data. Some technical problems were reported as well: not all criteria could be encoded using the system's architecture. Also, the system was not able to process patients whose medical records were incomplete, had missing data.

Work described in [10] was on an architecture for protocol encoding for automated trial accrual. The authors of the paper developed a data model for eligibility criteria for one trial which allowed the handling of such issues as hierarchical relationships and missing data. Another difference of the approach was that it did not output the eligible/ineligible status of a patient but rather a probability value of a patient being eligible. The system utilizes standard medical vocabularies to represent concepts. Medical knowledge was encapsulated in an object-

oriented manner. The system used a Bayesian network for the evaluation process and inferring missing data. The performance of the method was shown on retrospective data from breast cancer research. It needs to be noted again that the study focused on the feasibility of the technical idea of performing mass screening of patients. The implementation was abstracted from the major barrier for such systems, privacy regulations.

In [11] OncoDoc, a commercial decision support system, was used in an experiment aimed at enhancing the accrual to clinical trials. This was done at a major cancer hospital in France, IGR (sees 10,225 new patients per year). They worked with the breast cancer program. They reported no issues with the confidentiality of patient data, probably because this was a commercial tool. They did not restrict usage to any phase of clinical trials during their four-month experiment.

OncoDoc did not completely determine eligibility, but provides general possible matches and further work has to be done to determine eligibility. The authors reported an encouraging 50% increase in accruals based on the use of OncoDoc. Originally, they reported that eight patients had gone on clinical trials and with OncoDoc pointing out potential eligibility that was increased to 12 patients (i.e. 4 were added to trials). The system there worked significantly differently than at the Moffitt Cancer Center. The only patients considered for clinical trials were those that did not have a treatment plan created by the initial clinician they visit. Instead, a committee of 13 considered their case, used OncoDoc and made a decision about treatment/clinical trials. There were 127 patients that fell into this category which is small. Their success provides an indication that MEANS could be effective if regulatory and time cost issues were overcome.

The work in [12,13] presented a clinical trial alert system for a single clinical trial at a

US academic healthcare system. The clinical trial was on diabetes. A unique program was written to extract several fields from patient's electronic health records and when within range alert physicians that they could be eligible for the trial. They were able to double the accruals for the trial. They had eight times the number of physicians referring patients. They did generate a large number of false positive alerts unlike the system described in this paper. The false positives were weeded out by a clinical trial coordinator. Their approach requires that new, specialized program code be written for any new trial. However, they were able to bypass HIPAA because they only needed non-identifiable fields and did not store anything. Further, though they study the physicians because there was minimal impact no consent was required from anyone.

III System Overview

In this section we briefly describe the main purpose of MEANS, as well as system architecture and flow.

A. Main Purpose

The primary goal of the system is for it to serve as the main tool that clinicians use to screen patients for clinical trials not only at the H. Lee Moffitt Cancer Center & Research Institute but also at its affiliated institutions. Since MEANS is a web-based system, no additional software installation is required and it can be accessed from any computer with internet access. Prior to the implementation of MEANS, the screening process was performed manually by physicians, one clinical trial at a time. Enrollment into clinical trials depended mostly on the ability of a physician and their staff to be familiar with the eligibility criteria of the protocol. This being the case, there was the possibility of an eligible patient not being presented with the opportunity to participate in a clinical trial.

Ideally, a physician would use MEANS to screen a patient while that patient is still in the clinic. If a potential match occurred, the physician could simply print out the current informed consent document form from MEANS for the particular trial for which the patient was determined eligible, and then the enrollment process would begin with a presentation of risks and potential benefits. MEANS also allows physicians to print out a detailed list of inclusion and exclusion criteria for all protocols as well as consent forms in English and, when available, Spanish. Figure 1 shows the initial page of the system for the GITC.

H. Lee Moffitt Cancer Center & Research Institute **USF** **University of South Florida**

The Clinical Trial Assignment Expert System: Initial Questions (Version 1.4)
Program: GI

Patient's MEANS ID: 49
 Questions about this page? See instructions below.

Patient age. (Normal value: Greater than 18 years) (Full Text) <input type="text"/> <input type="checkbox"/> Normal <input type="checkbox"/> Defer (Check Units)	Sex (Full Text) <input type="text"/>
Measurable disease (RECIST)? (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer	Life expectancy of > 12 weeks? (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer
Path diagnosis of esophageal SCC or ACA? (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer	Histologically confirmed metastatic colorectal cancer? (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer
Prior therapy for MCRC in the metastatic setting? (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer	Tissue confirmed esophageal adenocarcinoma or adenocarcinoma of the gastroesophageal junction (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer
Metastatic colon or rectal cancer (tissue proven) not suitable for surgery? (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer	Tissue proven pancreas adenocarcinoma? (Full Text) <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Defer

Streamlining Options

<input type="checkbox"/> Select If ALL Patient's Lab Values Within Normal Range	<input checked="" type="checkbox"/> Select To Bypass Protocol Selection Page
--	---

<input type="button" value="HOME PAGE"/> Click to go to home page	<input type="button" value="PROCESS"/> Click to submit your answers to the system
---	---

Figure 1: MEANS Initial Questions Page.

B. System Architecture

MEANS is divided into two main parts: *Patient Assignment* and *Knowledge Entry*

[14,15,16]. The *Knowledge Entry* system has a responsive web-based user interface that allows encoding of clinical protocols for use by the *Patient Assignment* system [17]. To complete the system, we refined the implementation of the probabilistic agent, streamlined the system flow and added a reporting subsystem. Figure 2 shows the detail of the system architecture.

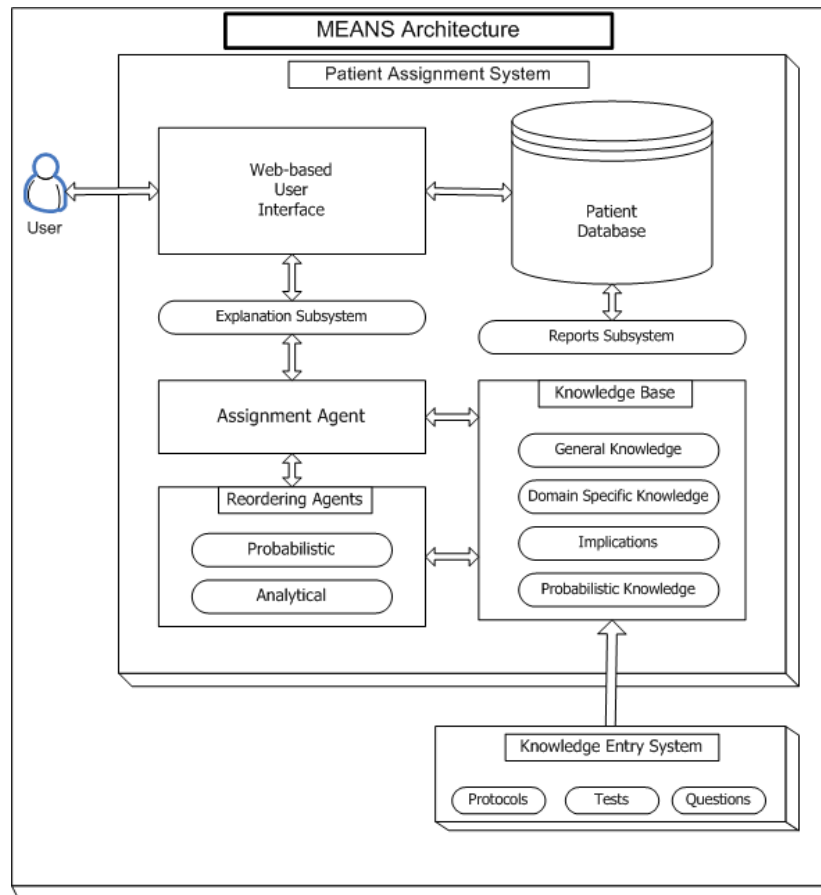


Figure 2: Expert System Architecture.

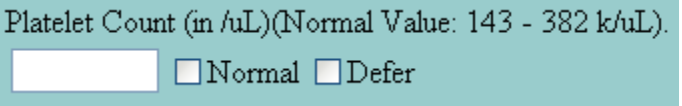
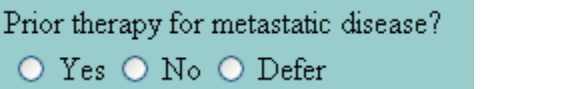
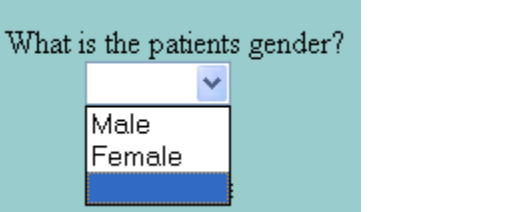
As seen in Figure 2, the knowledge base contains information about clinical trials -- stored as general and domain knowledge -- together with implications and probabilistic knowledge. The probabilistic knowledge is gathered by the system every time an answer is submitted to MEANS. When a sufficient number of patients have been screened, the probabilistic knowledge can be utilized to reorder the questions. The database contains

information from all previously screened patients. It also consists of the trials for which the patients were screened, their current eligibility for screened trials, and their previously provided answers. At any given time the user has the ability to change or delete an answer via the web-based user interface.

C. Question Types

All of the protocols are encoded by utilizing three types of questions: 1) Numeric, 2) Yes/No, and 3) Multiple Choice. A sample of how each question appears in MEANS can be seen in Table 1.

Table I: Types of Questions.

Type	Example
Numeric	
Yes/No	
Multiple Choice	

D. System Flow

The user interacts with the system via a web-based interface. The system presents a page of questions at a time to the user and, based on the responses to the questions, MEANS determines the eligibility of a patient. In order to use the system, the user must first select a patient to be screened. This is accomplished either by the user entering a new patient into the system or looking up a previously entered patient.

Once a patient has been selected, the user is given a set of initial questions. This consists

of one or two questions from each protocol that will most likely rule out the patient as being eligible for the trial, such as a primary cancer site. This would in turn decrease the amount of data entry. The user has the option to defer a question or leave it blank. Once past this opening page, the user can select a trial or set of trials for which to screen a patient. After making this selection, the user is then constantly presented with a page of questions to determine the eligibility of a patient. Every time the user submits answers to the system, the assignment agent determines the eligibility for each clinical trial and assigns the patient one of three eligibility statuses: 1) ELIGIBLE, 2) INELIGIBLE, or 3) MORE. If the status was determined either ELIGIBLE or INELIGIBLE, then the user is presented with a reason for the assessment of each selected trial. If eligibility could not be determined (MORE), the user is offered additional questions in order of priority. If a patient is determined eligible but will not go on trial, the clinician can enter the reason or reasons why the eligible patient chose to decline participation in the trial.

IV Clinical Setting

Before the system could be considered ready for physicians, it had to be properly and securely set up. Since the system is a web-based tool, no installation was required on the physicians' computers and an Internet shortcut was placed on their desktop. The necessary precautions were taken to comply with the *Health Insurance Portability and Accountability Act* (HIPAA) of 1996 to ensure the security and privacy of information.

It was determined HIPAA authorization was required in order for anyone to be entered into the system. The system needs to store subject's personal identification number (ID) so that clinicians may efficiently find partially entered patients. This is the only potentially identifiable piece of information stored, and it is stored in a separate encrypted file along with the MEANS

ID number. This requirement adds an extra layer of complexity, because we needed everyone to sign an authorization form. Furthermore, the recovery of patient data is awkward. This entire authorization process was complex and cumbersome. Of greater concern to us was when a person did not agree to sign the authorization form. Not only will they not be screened by the system but the clinician may determine that they are uninterested in clinical trials, eliminating any possibility for future participation. Not signing the five-page personal data collection authorization form does not necessarily mean they are uninterested in clinical trials altogether. We were not able to investigate this concern.

In October 2006 new regulations introduced by the Institutional Review Board of the University of South Florida made the screening process by the MEANS system more complicated. The HIPAA authorization and IRB consent form were merged, hence the researchers were required to get an IRB-consent from each patient willing to participate in the study. IRB consent is more detailed, has stricter rules and overall requires a person who consents the patients to have research training. Most of the procedures and results described in this paper are based the initial requirement of HIPAA authorization.

Due to the authorization requirement, we could not screen all patients. There were a number of times when the authorization was not obtained because clinicians were busy or (rarely) patients declined. The potential departure from the clinical practice of screening all patients is an unintentional side effect of the HIPAA requirement, but personal data was secure.

Before official introduction of the system to the physicians, several meetings were held to determine how to make the system more physician-friendly. The system was *test driven* by a physician and a nurse with positive results. After that the system was introduced to physicians and nurses in the GITC of the Moffitt Cancer Center on February 7, 2006 and used for

screening patients during a nine-month period.

A. Clinical Trials

During February-November of 2006, the number of all active clinical trials at the Moffitt Cancer Center reached 150 [18]. Since the study was aimed only at phase II clinical trials, we only entered active and open for enrollment phase II clinical trials into MEANS. At the time of the study, the Moffitt Cancer Center had 36 active phase II clinical trials, six (6) of which were conducted in the GIT Clinic.

B. Physical Environment

The GITC at Moffitt Cancer Center is used by several physicians and nurses. As a rule, a physician spends several days a week seeing patients in the clinic. Each physician has a nurse that is assigned to him or her. The nurse works closely with the physician and is also in the clinic while the physician is seeing patients. Since the Moffitt Cancer Center utilizes an electronic system for its medical charts, both the physician and nurse have computer access in the clinic.

When a patient arrives at the clinic, he or she is greeted by front desk personnel before being guided to an examination room. After a physician has seen the patient, he or she is guided out of the clinic, usually by the physician.

V Experiments

MEANS was assigned a protocol study number of MCC 14031 and was subject to the usual rules and regulations. Since the system was part of a study, all of the patients that are screened by the system had to sign a five-page HIPAA informed consent. Only after this form was signed by the patient or guardian may the patient be screened in MEANS.

The patients were asked to consent to normal clinical practice: that is, for an M.D. to

mentally (now aided by MEANS) consider them for a clinical trial. They cannot go on trial without being presented the trial **and** signing and informed consent for that trial.

A. Patient Selection

There are three types of patients in the clinic: *New Patient* (NP), *New Established Patient* (NEP) and *Established Patient* (EP). New Established Patient category is related to patients who have been previously seen by another doctor at the clinic. It would have been ideal to screen every patient that walked through the door for eligibility; however, it was more practical to screen only new patients. Therefore, we decided to focus only on NP and NEP patients.

B. Clinician Involvement

MEANS was designed to serve as a tool for physicians and staff; therefore, the feedback from them as part of the participatory design was crucial to the development and overall acceptance of the system.

During the first week, two researchers spent their time at the GITC showing the physicians and nurses how to use MEANS and answering all questions related to the system.

C. Optimization

In order to decrease the amount of time required for eligibility determination, we used several optimization techniques such as data mining and probabilistic reordering. As seen in Figure 2, implications are part of the knowledge base.

We used WEKA, a data mining software package, to extract the data from MEANS to see if we could find any implications or rules [19,20]. By mining existing data with the Apriori algorithm [21], it was possible to discover implications (association rules) that were not seen before. Such correlations allowed us to make inferences based on existing information and thus

decreased the number of answers needed for eligibility determination. For example an answer of "Male" to the question about the gender of a patient would automatically imply a negative answer to the question about possible pregnancy. We also used previously-gathered statistical information, on which questions made a patient ineligible for a trial for a probability-based reordering of questions.

VI Results

This study focused on the process of how to make an expert system, such as MEANS, more physician-friendly and increase its acceptability in the clinical environment. The system was deployed for use in two clinics in the Moffitt Cancer Center: Gastrointestinal Tumor Clinic (GITC) and Hematology Tumor Clinic. Unfortunately the number of patients from the Hematology Clinic processed by MEANS was only 39, which was too small to draw any significant conclusions. In this section we describe the final results that were achieved during the nine month study period in the GITC only.

A. Clinician Participation

For the duration of our study, we followed the model of participatory design and maintained a constant presence in the GITC. The physicians and nurses were actively involved in screening patients in MEANS.

Due to the HIPAA regulations, a five-page HIPAA authorization form needed to be signed by every patient before they may be screened in MEANS. The original authorization form was placed on file while a copy was given to the patient. We found that during clinic hours physicians and staff were very busy and time was a luxury, especially for physicians.

We had envisioned a process in which a nurse obtained the HIPAA authorization from

the patient and then a physician, after seeing the patient, spent a couple of minutes screening that patient in MEANS.

Since the physicians at the Moffitt Cancer Center specialize in specific cancer programs, they are very familiar with the current clinical trials within their area of interest. Some of the physicians were reluctant to use MEANS for just that reason. In their opinion, they can determine eligibility for all trials for a patient by simply reviewing the patient's medical record. However, this would not be the case for affiliates, who would see a wider range of cancer patients and would not have such in depth knowledge of the clinical protocol eligibility.

We found that during clinic hours, physicians were not always able to screen patients because of time constraints. During such circumstances, physicians would ask their nurses to screen the patients. Nurses would screen a patient by answering as many questions as they could, but if eligibility still could not be determined, a physician would then be asked to complete the screening. At the end of each week a reminder was sent to the physicians, requesting that they screen the patients that could not be screened when the physician was previously in the clinic.

B. Streamlining

One of the major concerns expressed during this study was the amount of time required for a physician in order to enter information into the system. With such feedback from the physicians and nurses, MEANS underwent significant changes.

Screening for eligibility of a clinical trial can be a subjective task. The investigators can take into consideration any borderline cases, whereas the system strictly follows the rules. The final decision about patient eligibility is always physician-investigator determined, so regardless of what the system concludes the eligibility would be reviewed by a human. Based on that observation and the feedback from the physician, it would be best if the system was 95%

accurate and took less time for screening as opposed to 100% accurate and require significantly more time to determine eligibility. Since physicians are pressed for time, they would not use the system if it was too time-consuming.

To reduce the time spent on answering the initial questions in MEANS we introduced a question about the type of cancer the patients have, for example: liver, pancreas, or colorectal cancer. Although it is not a part of the eligibility/ineligibility criteria it allows several trials to be ruled out at once, since trials are usually designed for a particular type of cancer.

We have also discovered that there is a significant difference between the verbiage used by physicians and the legal language in which the protocol inclusion/exclusion criteria are written. The protocol version sometimes can be much lengthier, as seen in Table 2. To conserve the amount of time a physician spends reading the questions, only the shorter version of the questions was displayed on the screen. The full version of a question was accessible by simply placing the mouse cursor over a question or by clicking on the *Full Text* link (Figure 1).

Table II: Yes/No Question Length.

Full Text
Does the patient have histological or cytological confirmed esophageal adenocarcinoma or adenocarcinoma of the gastroesophageal junction? (This refers to tumors at the junction of the esophagus and the stomach, where > 50% of the tumor mass is above the diaphragm)
Short Version
Tissue confirmed esophageal adenocarcinoma or adenocarcinoma of the gastroesophageal junction?

The automatic selection of all encoded clinical trials and *All Labs Normal* options were introduced based on feedback from the physicians. These are shown in Figure 1. The physicians also had the option to individually select a trial. We were informed that the exact value of the

labs was not as important as knowing whether or not the labs were normal. Since all normal lab values were in one column of the lab report, it was easy for the physician to determine if they were normal. It was also much faster to click on the check-box rather than type in the exact value of the lab result.

C. Patient Screening

One of the major hurdles of the HIPAA authorization form was that not all of the NP and NEP patients that come to the clinic were presented with it. As a rule, a nurse sees a patient before a physician. For the purposes of this study the HIPAA authorization was only for patients with a NP and NEP status and sometimes a nurse would miss the opportunity to present a patient with the authorization form before a patient sees the physician. Once that opportunity was missed it was unlikely that authorization would be obtained because the nurse's focus would then have turned to the next patient.

The nursing staff sometimes was making a judgment call about whether or not to present a patient with the authorization form. If a patient was emotionally upset due to information regarding their cancer, then the probability that a patient would be presented with the HIPAA authorization form was low. Occasionally some patients declined to sign the form. Since it was required that we use the standard format for HIPAA authorization, it covered a much greater scope than was needed for this project, and its language was very complex and legalistic. Hence, some of the refusals that we encountered were based on the wording of the authorization. If authorization is not obtained screening can not be performed.

Since the introduction of the system 920 authorizations were obtained, the clinical staff in the GITC screened a total of 510 patients (Table 3). As mentioned earlier, the physicians had a

choice of either screening a patient for all of the encoded trials or picking a subset of trials. A graphical representation of Table 3 is given in Figure 3. We can see that we were able to put 30% of the patients through the screening process.

Table III: Total GI Patients by Category.

Seen:	2074
Signed:	920
Screened:	510
Not Screened:	410
Missed:	1154

We obtained statistics on demographics of patients who went through the GITC during the nine month period. The demographics obtained included to gender, race, and ethnicity of all new and new established patients. It appears that the makeup of the screened population was essentially the same as for typical Moffitt Cancer Center patients with a maximum difference of 8% for the same categories.

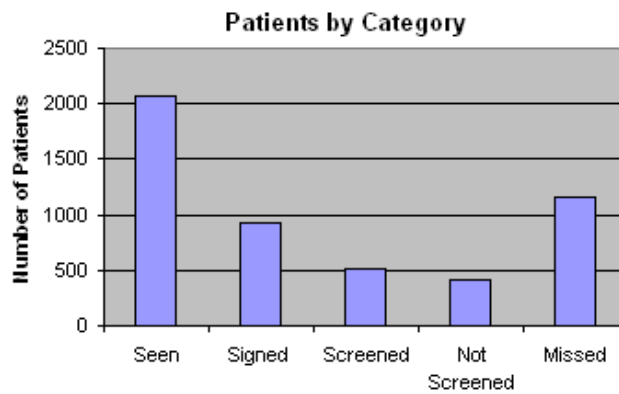


Figure 3: Patients by Category.

During the study we made some observations about the time the doctors spend on screening patients using MEANS. The average number of patients a doctor in the GITC of

Moffitt Cancer Center sees daily is between 12 and 18. About 20% of the patients are new and new-established. Usually it takes about 90 minutes to see a new patient. It takes about two minutes to screen a new and new-established patient using MEANS provided that the doctor is familiar with the patient's condition and doesn't need to lookup additional records. The additional daily time-load for doctors to use MEANS was about 4-8 minutes.

D. Accrual Results

We should note that at this stage of our study we did not find any additional eligible patients that were not found by the physicians. This both surprised and disappointed us. We have a possible explanation for this. As is seen by the accrual statistics in Table 4 and in Figure 4 the average accrual to individual trials in GITC was 1.08 per trial per month during the nine months MEANS was fielded. At the same time MEANS showed an accrual 0.42 for the same period. However, only 510 patients were screened using MEANS out of 2074 which is less than a fourth of the total number. Assuming that the distribution of the patients found by MEANS stays the same we can expect the number of found patients rise more than four times provided that all 2074 patients were screened by MEANS. This would result in 1.78 patients who would go on trials on average which is 0.7 patients per trial per month on average more than in the case when MEANS was not used. According to the demographic statistics the distribution of patients does not change much.

Table IV: Accrual Rate in GITC for years 2005 and 2006.

	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	AVG
9 months before MEANS:	-	-	-	1.2	1.6	0.5	0.67	1.17	1.17	0.67	0.83	0.7	0.95
9 months after MEANS	1.8	1.4	0.8	1.2	1.2	1	2.5	0	0	-	-	-	1.08
MEANS Accruals for 9 months:	2.25	0	0.25	0.5	0.25	0	0.5	0	0	-	-	-	0.42
GITC Accrual Rate in 2005:	0.8	1.67	0.83	1.2	1.6	0.5	0.67	1.17	1.17	0.67	0.83	1	1.01

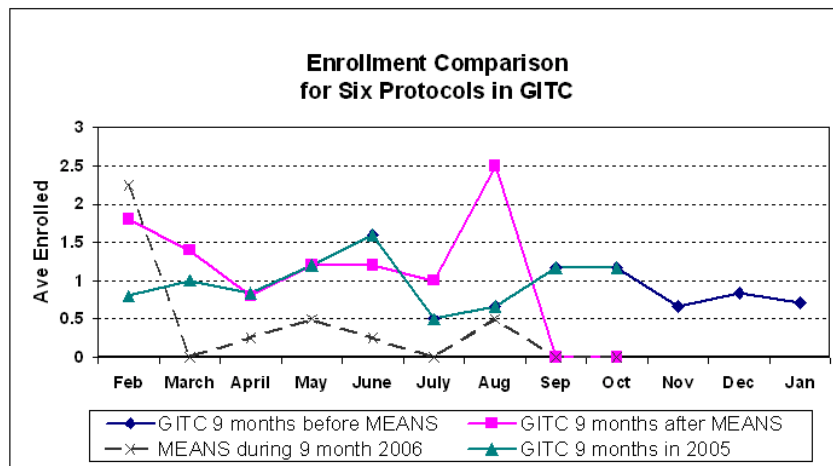


Figure 4: Enrollment Comparison in GITC for years 2005 and 2006. The data shown is for total enrollment in GITC and for patients found by using MEANS.

We should also note that the total number of patients going on trials in GITC is very small. From the statistics from Moffitt Cancer Center GITC we found out that during three-year period only 458 patients went on trial. That number also includes patients from Moffitt's affiliated institutions. Out of 76 trials which were active at that time only four trials accrued 30 or more patients which might make the result of a trial statistically significant. For phase II trials there is not even a single trial which accrued 30 patients during the same period. Such slow

accrual in one of the biggest cancer only hospitals in the US suggests that many trials, in order to be successful, need to be conducted on a multi-center basis. We strongly believe that MEANS may be a useful tool for such collaboration.

VII Additional Streamlining of the Process

Table 3 shows that a significant number of patients who have signed a consent to participate in the screening process using MEANS were not screened. Because of their tight schedule the physicians often cannot allocate time to complete screening those patients using MEANS. That was another reason for the small fraction of screened patients which in combination of small accrual rate for trials in GITC did not allow us to completely evaluate MEANS performance.

In order to further help the physicians in the clinic it was decided that the patients may be screened by researchers for some simple cases. Table 5 shows the results of screening 30 patients over a contiguous period of time using MEANS. The testing period began on June 1st and ended on July 13th 2006. All 30 patients were determined ineligible for all trials by MEANS. Table 5 shows the distribution of numbers of patients who were ruled-out by MEANS according to categories of questions offered by the MEANS interface. The table shows numbers of patients according to these categories in decreasing order. For example the “Type of cancer” questions ruled out 12 patients out of 30. For the three most common cases of question patterns – “Type of Cancer”, “First Page Questions”, and “Advanced Questions (Blood Tests, Chemistry Values)” information was available in the clinic database and it was possible to use this information to screen the patients without the need of doctor's assistance. This was done to see whether we could screen out a number of patients without physicians, leaving only a fraction of patients to be

screened by the physicians.

Table V: Distribution of ruled-out patients according to type of question.

Category of questions	Number of patients affected
Type of cancer	12
First page questions	11
Advanced questions (blood tests, chemistry values)	4
Questions requiring doctor's expertise	3

The patients' information that can be used was not available immediately. We conducted a separate investigation to check availability of information in the patient database. Figure 5 shows the distribution of patients whose clinical records are available for screening purposes. This histogram was created for 130 patients who were chosen randomly during the 9-month period in 2006. According to Figure 5 there were 77.1% of the patients who had information available for screening within two weeks of being seen.

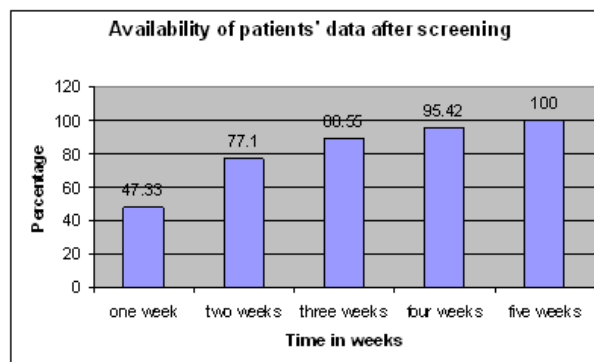


Figure 5: Availability of patients's data in time (shown cumulatively).

Further investigation was aimed to find out how much information about the patients condition can be mined or extracted automatically. Since about 90% of the questions do not require doctor's expertise to be answered we concentrated on them. In the next few paragraphs

we will explain what the formats of the data for patients' records and methods of mining particular types of data were.

The H. Lee Moffitt Cancer Center & Research Institute uses the PowerChart [23] tool as the primary software for accessing patients' records. All of the patients' data including lab-values, questionnaires, consents, clinician notes, etc. is converted into an electronic format and stored in a database with different levels of access. Most of the patient records are available for reading but only authorized personnel have permission to change the records. Figure 6 shows an example of the data as it is represented in the clinical database. All personal information was removed to be compliant with HIPAA regulations. Table 6 shows some of the fields in the PowerChart and the corresponding type of data used to store information in the field. Most of the information about the patients contained in the clinical database is represented by clinician notes and laboratory values.

Results									
Outside Orders/Rx									
Outside Pathology									
Vital Signs									
BP systolic	120								
BP diastolic	75								
Pulse	80								
Respirations	18								
Temperature	37.1 DegC								
BSA (Metrics)	2.15								
Height	180 cm								
Weight	85 kg								
Hemogram									
WBC		* 10.18 k/u						* 9.10 k/uL	
RBC		L 3.88 ml/L						L 3.59 ml/L	
Hemoglobin		L 10.8 g/dL						L 9.8 g/dL	
Hematocrit		L 34.5 %						L 32.4 %	

(a)

REASON FOR CONSULTATION: XXXXX is a XX-year-old man/woman with rectal cancer who was initially diagnosed in year XXXXX. The patient had recurrence of rectal cancer...
 HISTORY OF PRESENT ILLNESS: Patient had surgery...
 ASSESSMENT AND PLAN:

(b)

Figure 6 : Example of information available in the clinical database. (a) Screenshot of PowerChart(R) software showing chemistry values for a particular patient. (b) Example of a clinical note available in the database (all personal information is removed).

Table VI : Data fields available in PowerChart(R) and their corresponding format.

Outside Documentation	Scanned papers (images)
Vital Signs	Lab values (typed text)
Hemogram	Typed text (1 string)
Differential	Typed text (1 string)
Routing chemistry	Typed text (1 string)
Clinician notes	Free text (typed)
Clinical Research Unit	Scanned papers (images)
Imaging reports	Scanned images
Diagnostic reports	Free text (typed)

About 15% of the questions in MEANS that rule out the patients come from comparison of the lab-tests (hemogram, urinalysis, routing chemistry). All these values can be extracted automatically from appropriate fields. But according to the previous tables, most of the questions that rule out a patient are the “type of cancer” questions and questions appearing on the first page of MEANS. Information for those is usually contained in clinician notes and pathology reports in the format of free text. Although, there is no algorithm to accurately extract the meaning of the free text from clinician notes and reports, particular questions can be answered if text-mining techniques are applied [24,25,26,27].

The investigation described in this section was conducted in order to find the possible directions for further streamlining and automating screening process. By our estimation the system has the potential to automatically screen 77% of patients after two weeks. That would leave a significantly smaller number of patients for clinicians to look at - a few possibly not ruled out and any others without data available. MEANS is a tool which helps physicians to find eligible patients for their trials, in the case where a patient is found to be ineligible by MEANS, they would also have their case reviewed by a physician-investigator. From the results of the clinical trial alert system [13], we are encouraged to believe that many patients can be automatically ruled out in a timely manner. Depending on the acceptable time frame only about

20% would need further investigation. Our system, as seen in the results, does not result in many false positives which would be a strength over time.

VIII Discussion and Conclusions

We discussed our final data from the clinical deployment of a medical expert for clinical trial assignment at the H. Lee Moffitt Cancer Center & Research Institute . Past research has shown that physicians are reluctant to use new software due to time constraints. We used the model of participatory design to make our expert system more physician-friendly with the goal that acceptability of the system will increase.

We found that a major hurdle in our research was the five-page HIPAA authorization form that was an additional burden for clinicians. In [28], it was shown that HIPAA resulted in a 73% decrease in patient accrual after implementation. Since we did not work without it, we cannot say how much effect it had, but it clearly had a negative effect as seen in the earlier study.

Using constant feedback from the clinicians allowed us to use various streamlining techniques to improve the system's physician friendliness.

We believe we were unable to adequately evaluate the utility of MEANS because of regulatory issues. The HIPAA rules on data security required us to get permission from all patients for any data entry. This added a layer of complexity because the waiver forms had to be presented, signed, copied and stored. This probably requires a nearly full-time person for each clinical program. Further, this required a physician or other clinical worker to do an extra check to find out whether a HIPAA waiver was signed. This uncertainty undoubtedly made utilizing the system less attractive, as there are significant penalties for violating HIPAA.

We were also concerned that the waiver process might have the negative effect of causing patients to be disinclined toward clinical trials or for physicians not to consider someone for a

clinical trial who declined to have their data entered into MEANS. In our opinion, there was no possible other negative effect on the patient as they could **not** go on any clinical trial without being presented a trial and signing an informed consent for the particular trial. With all data encrypted and only their record number stored a data breach seemed highly unlikely and only the most dedicated person would likely connect any data to a particular individual (and would need to additionally break into the Moffitt Cancer Center database to do so).

We noted that physicians in the clinic are extremely busy and even the relatively modest couple of minutes required to rule out a patient from the available files was quite difficult to find. It is likely possible to extract much of the needed information from the clinical records through automated methods. By our estimation, nearly 80% of the patients could be automatically screened within two weeks of a visit; however as it was not possible for us to evaluate this under HIPAA rules. If MEANS were commercially available software, we understand it would not be subject to some of the restrictions we faced. It might then be fairly evaluated.

The results discussed in this paper pertain only to the physicians in the GITC program at H. Lee Moffitt Cancer Center & Research Institute . We should note that since the physicians at the H. Lee Moffitt Cancer Center & Research Institute specialize in particular types of cancer, they are able to become very familiar with available clinical trials in their specialty.

Authors' Contributions

Halina Greenstien, Amit P. Patha M.D., and Christopher R. Garrett, M.D helped us revise the questions used in the web-based system described by the article. They checked to make sure everything was medically correct. They helped introduce the system into the clinic and liaised with other doctors and nurses. Dr. Garrett also helped with the writing.

Sergiy Fefilyatjev did system development and added protocols and refined protocols in the electronic system. He did a significant amount of the writing.

Li Chen, Shibendra Pobi and Tim V. Ivanovsky did initial system development, further system development and adding of protocols and refining the way questions were presented. Shibendra and Tim did the initial rollout of the system and education of doctors and nurses.

Lawrence O. Hall and Dmitry B. Goldgof directed the project. They developed the concept, directed the implementation of the system, directed the rollout of the system in conjunction with Dr. Garrett and wrote/re-wrote parts of the paper.

Acknowledgment

This work has been supported by the National Institutes of Health under grant #5R21CA108653-02. We would like to thank the physicians at Moffitt Cancer Center who are committed to total cancer care.

Conflicts of interest: None.

Summary Table

Known:	Added:
<ul style="list-style-type: none">• Clinical trial accrual is a slow process and a barrier to the evaluation of new drugs and procedures.• Retrospective studies have shown it may be possible to increase the accrual to clinical trials using an automated system in conjunction with clinical personnel.	<ul style="list-style-type: none">• In the United States, HIPAA regulation is a barrier to research. In this case it required patients to sign a consent to be considered for a clinical trial which in the past has been done routinely as part of clinical practice.• The accrual for trials went up slightly during the time the system was utilized. However, many patients were not actually entered into the system because of issues with available consents and clinicians time.• For true success, it is highly likely that a reminder system which points out potentially eligible patients after mining the data available and medical records, will be utilized more and can substantially speed the accrual to clinical trials.

References

[1] Cyrus A. 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*. 1em plus 0.5em minus 0.4em *Journal of Surgical Oncology*, 50:125-129, 1992.

[2] 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*. 1em plus 0.5em minus 0.4em *Journal of Methods of Information in Medicine*, 32:317-325, 1993.

[3] John H. Gennari, Madhu Reddy, *Participatory design and an eligibility screening tool*. 1em plus 0.5em minus 0.4em In *Proceedings of the American Medical Informatics Association Annual Fall Symposium*, pages 290-294, 2000.

[4] Max Bramer, Frans Coenen, Tony Allen, *Research and Development in Intelligent Systems XXII*. 1em plus 0.5em minus 0.4em In *Proceedings of AI-2005, The Twenty-fifth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 2005.

[5] Eugene Fink, Lawrence O. Hall, Dmitry B. Goldgof, Jeffrey P. Krischer, Bhavesh Goswami, Mathew Boonstra, *Experiments on the automated selection of patients for clinical trials*. 1em plus 0.5em minus 0.4em In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 4541-4545, 2003

[6] A.J. Butte, D.A. Weinstein, I.S. Kohane, *Enrolling patients into clinical trials faster using RealTime Recruiting*. 1em plus 0.5em minus 0.4em In *Proceedings of the AMIA Symposium*, pages 111-115, 2000

[7] D.L. Weiner, A.J. Butte, P.L. Hibberd, G.R. Fleisher, *Computerized recruiting for clinical trials in real Time*. 1em plus 0.5em minus 0.4em *Annals of Emergency Medicine*,

41(2):242-246, 2003

[8] L.B. Afrin, J.C. Oates, C.K. Boyd, M.S. Daniels, *Leveraging of open EMR architecture for clinical trial accrual*. 1em plus 0.5em minus 0.4em In Proceedings of the AMIA Symposium, pages 16-20, 2003

[9] L. Ohno-Machado, S.J. Wang, P. Mar, A.A. Boxwala, *Decision support for clinical trial eligibility determination in breast cancer*. 1em plus 0.5em minus 0.4em In Proceedings of the AMIA Symposium, pages 340-344, 1999

[10] N. Ash, O. Ogunyemi, Q. Zeng, L. Ohno-Machado, *DFinding appropriate clinical trials: evaluating encoded eligibility criteria with incomplete data*. 1em plus 0.5em minus 0.4em In Proceedings of the AMIA Symposium, volume 27, pages 31, 2001

[11] B. Sroussi, J. Bouaud, *Using OncoDoc as a computer-based eligibility screening system to improve accrual onto breast cancer clinical trials*. 1em plus 0.5em minus 0.4em Artificial Intelligence in Medicine, Sep-Oct; volume 29(1-2), pages 153-67, 2003.

[12] P.J. Embi, A. Jain, J. Clark, C.M. Harris, *Development of an Electronic Health Record-based Clinical Trial Alert System to Enhance Recruitment at the Point of Care*. 1em plus 0.5em minus 0.4em In Proceedings of the AMIA Symposium, volume 231, pages 5, 2005

[13] P.J. Embi, A. Jain, J. Clark, S. Bizjack, R.W. Hornung, C.M. Harris, *Effect of a Clinical Trial Alert System on Physician Participation in Trial Recruitment*. 1em plus 0.5em minus 0.4em Archive of Internal Medicine, volume 165, pages 2272-2277, 2005

[14] Sanjukta Bhanja, Lynn M. Fletcher-Heath, Lawrence O. Hall, Dmitry B. Goldgof, Jeffrey P. Krischer, *A Qualitative Expert System for Clinical Trial Assignment*. 1em plus 0.5em minus 0.4em In Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference, pages 84-88, 1998.

[15] Princeton K. Kokku, Lawrence O. Hall, Dmitry B. Goldgof, Eugene Fink, and Jeffrey P. Krischer, *A cost-effective agent for clinical trial assignment*. 1em plus 0.5em minus 0.4em In Proceedings for the IEEE International Conference on Systems, Man and Cybernetics, 2002.

[16] Bhavesh D. Goswami, Lawrence O. Hall, Dmitry B. Goldgof, Eugene Fink, Jeffrey P. Krischer, *Using probabilistic methods to optimize data entry in accrual of patients to clinical trials*. 1em plus 0.5em minus 0.4em In Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems, pages 434-438, 2004.

[17] Savvas Nikiforou, *Selection Of Clinical Trials: Knowledge Representation and Acquisition*. 1em plus 0.5em minus 0.4em Master's thesis, Department of Computer Science and Engineering, University of South Florida, 2002.

[18] H. Lee Moffitt Cancer Center & Research Institute, *Prevention and Treatment: Clinical Trials*. 1em plus 0.5em minus 0.4em ifundefined9 pt <http://www.moffitt.usf.edu>, 2006.

[19] Rakesh Agrawal and Tomasz Imielin'ski and Arun Swami, *Mining association rules between sets of items in large databases*. 1em plus 0.5em minus 0.4em Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pages 207-216, 1993.

[20] The University of Waikato, *Weka 3 - Data Mining with Open Source Machine Learning Software in Java*. 1em plus 0.5em minus 0.4em ifundefined9 pt <http://www.cs.waikato.ac.nz/ml/weka/>, 2007.

[21] Christian Borgelt, Rudolf Kruse, *Induction of Association Rules: Apriori Implementation*. 1em plus 0.5em minus 0.4em Proceedings of the 15th Conference on Computational Statistics, 2002.

[22] Paskett ED ; Cooper MR ; N. Stark, T.C. Ricketts, S. Tropman, T. Hatzell, T. Aldrich, J. Atkins, *Clinical trial enrollment of rural patients with cancer*. 1em plus 0.5em minus 0.4em Cancer Practice, Jan-Feb, pp. 28-35, 2002.

[23] Cerner Corporation, *PowerChart Solution Suite*. 1em plus 0.5em minus 0.4em ifundefined9 pt <http://www.cerner.com>, 2006.

[24] Sophia Ananiadou, John McNaught, *Text Mining for Biology and Biomedicine*. 1em plus 0.5em minus 0.4em Artech House, 2006.

[25] M.A. Hearst, *Untangling Text Data Mining*. 1em plus 0.5em minus 0.4em Proceedings of 37th Annual Meeting of the Association for Computational Linguistics, College Park, MD, 1999, pp.3-10.

[26] R. Arens, *A Preliminary Look into the Use of Named Entity Information for Bioscience Text Tokenization*. 1em plus 0.5em minus 0.4em Proceedings of HLT-NAACL Student Research Workshop, Boston, MA, May 2, 2004, pp.37-42.

[27] P. Srinivasan, *Text Mining: Generating Hypotheses from MEDLINE*. 1em plus 0.5em minus 0.4em Journal of the American Society for Information Science, Vol.55, No.5, 2004, pp.396-413.

[28] M.S. Wolf, C.L. Bennett, *Local perspective of the impact of the HIPAA privacy rule on research*. 1em plus 0.5em minus 0.4em Cancer, Jan 15, pp. 474-9, 2006.