

## Trial Prospector: Matching Patients with Cancer Research Studies Using an Automated and Scalable Approach

Satya S. Sahoo<sup>1,\*</sup>, Shiqiang Tao<sup>1,\*</sup>, Andrew Parchman<sup>2,3,\*</sup>, Zhihui Luo<sup>1</sup>, Licong Cui<sup>1</sup>, Patrick Mergler<sup>3,4</sup>, Robert Lanese<sup>4</sup>, Jill S. Barnholtz-Sloan<sup>4,5</sup>, Neal J. Meropol<sup>2-4</sup> and Guo-Qiang Zhang<sup>1,4</sup>

<sup>1</sup>Division of Medical Informatics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA. <sup>2</sup>Division of Hematology and Oncology, School of Medicine, Case Western Reserve University, Cleveland, OH, USA. <sup>3</sup>University Hospitals Case Medical Center, Seidman Cancer Center, Cleveland, OH, USA. <sup>4</sup>Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA. <sup>5</sup>Institute for Computational Biology, School of Medicine, Case Western Reserve University, Cleveland, OH, USA. \*Authors contributed equally to the paper.

**ABSTRACT:** Cancer is responsible for approximately 7.6 million deaths per year worldwide. A 2012 survey in the United Kingdom found dramatic improvement in survival rates for childhood cancer because of increased participation in clinical trials. Unfortunately, overall patient participation in cancer clinical studies is low. A key logistical barrier to patient and physician participation is the time required for identification of appropriate clinical trials for individual patients. We introduce the Trial Prospector tool that supports end-to-end management of cancer clinical trial recruitment workflow with (a) structured entry of trial eligibility criteria, (b) automated extraction of patient data from multiple sources, (c) a scalable matching algorithm, and (d) interactive user interface (UI) for physicians with both matching results and a detailed explanation of causes for ineligibility of available trials. We report the results from deployment of Trial Prospector at the National Cancer Institute (NCI)-designated Case Comprehensive Cancer Center (Case CCC) with 1,367 clinical trial eligibility evaluations performed with 100% accuracy.

**KEYWORDS:** clinical trial, gastrointestinal cancer, clinical oncology, patient recruitment, clinical decision support system

**CITATION:** Sahoo et al. Trial Prospector: Matching Patients with Cancer Research Studies Using an Automated and Scalable Approach. *Cancer Informatics* 2014;13:157–166  
doi: 10.4137/CIN.S19454.

**RECEIVED:** August 17, 2014. **RESUBMITTED:** September 29, 2014. **ACCEPTED FOR PUBLICATION:** October 4, 2014.

**ACADEMIC EDITOR:** JT Efrid, Editor in Chief

**TYPE:** Methodology

**FUNDING:** This work was supported by the NIH Clinical & Translational Science Collaborative (CTSC) of Cleveland (UL1TR000439) and the NCI Case CCC (3P30CA043703) grants. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

**COMPETING INTERESTS:** Authors disclose no potential conflicts of interest.

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

**CORRESPONDENCE:** [neal.meropol@case.edu](mailto:neal.meropol@case.edu), [gq@case.edu](mailto:gq@case.edu)

Paper subject to independent expert blind peer review by minimum of two reviewers. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

### Introduction

Cancer is one of the leading causes of death in the world with 7.6 million cancer deaths reported in 2008 by a GLOBOCAN survey.<sup>1</sup> Furthermore, the World Health Organization reports that the incidence of cancer cases is increasing.<sup>2,3</sup> Clinical trials are essential to make advances in cancer treatment and improve patient outcomes. Leading national organizations, such as the American Cancer Society and National Comprehensive Cancer Network, recognize clinical trials as a key component of high-quality care.<sup>4,5</sup> Clinical trials must enroll an adequate number of

patients in order to evaluate the effectiveness of new treatments.<sup>6</sup> A recent report by Stiller et al.<sup>7</sup> found that survival rates for childhood cancer at the population level have dramatically improved in Britain from 1978 to 2005 because of increased participation in clinical trials. Unfortunately, the participation of adult cancer patients in clinical trials is very low; in the United States, estimates suggest that less than 5% of patients take part.<sup>8,9</sup>

Multiple barriers to clinical trial enrollment have been described with one major barrier being the time and effort required by physicians to identify active trials, review



eligibility criteria, and then match them with specific patient data.<sup>6,10,11</sup> Many studies have characterized patient and provider barriers to clinical trial participation, and recommended measures to improve patient awareness, access to trial information, and attitudes toward participation in clinical studies.<sup>8,9,11</sup> In addition, difficulty for physicians in accessing trial information and protocol availability can also impede patient enrollment.<sup>12</sup> Although National Cancer Institute (NCI)-designated comprehensive cancer centers have high levels of provider motivation and study access, trial enrollment in these centers also rarely exceeds 15% for racial and ethnic minorities.<sup>13</sup> An important practical barrier that is being increasingly reported by physicians is a lack of time available for patient ascertainment and consent for clinical trials.<sup>14</sup> A common clinical practice for a physician is to manually screen potential clinical trials for a patient with additional time spent by research staff to collect patient data from multiple sources and coordinate enrollment.<sup>12,15,16</sup> This is an extremely labor-intensive process that

1. increases the overall cost of clinical trials, whether conducted in community settings or academic medical centers;
2. fails to leverage electronic health record (EHR) systems to automate different phases of patient screening; and
3. introduces errors or leads to missed opportunities for enrollment as patient load and the number of trials increase.

Informatics-based approaches that incorporate formal representation of eligibility criteria, data from EHR systems, and accurate trial matching algorithms are important for automating many phases of the trial recruitment workflow and efficiently matching trials with patients.<sup>15,17</sup> In addition, default matching of all visiting patients to available trials by an automated matching tool can help improve the participant diversity across economic and social categories.<sup>12,18</sup> Automated matching tools can also be integrated with existing data management systems used in health care institutions (eg, patient scheduling systems) to leverage available resources while streamlining the trial recruitment process.<sup>16,19</sup>

In this paper, we introduce a scalable clinical trial eligibility-screening tool called Trial Prospector that has been deployed at the University Hospitals Seidman Cancer Center (UHSCC) of the NCI-designated Case Comprehensive Cancer Center (Case CCC) at Case Western Reserve University (CWRU), Cleveland, OH, USA.

### Related Work: Computational Approaches to Match Clinical Trial Eligibility Criteria with Patient Data

Automated tools for clinical trial recruitment have been used in many medical specialties, including cardiology,<sup>16</sup> family medicine,<sup>20</sup> and Parkinson disease.<sup>21</sup> Several trial matching

systems have been implemented in the cancer domain, such as caMatch,<sup>22</sup> the Agreement on Standardized Protocol Inclusion Requirements for Eligibility (ASPIRE),<sup>23</sup> and the Virginia Commonwealth University clinical trial system based on the OnCore tool (Forte Research Systems, Madison, WI, USA).<sup>18</sup> Many of these tools have been developed for accessing the institutional data warehouse to screen patients for clinical trials<sup>24</sup> or for creating an alert system for physicians.<sup>16,17</sup> The OncoLink and Cancer.gov software tools have used the Web infrastructure to facilitate greater accessibility of trial information with additional features to connect potential recruits with the study investigators.<sup>25</sup> OncoLink is based on the National Colorectal Cancer Research Alliance (NCCRA) database and matched patients based on their demography and related information,<sup>25</sup> but it entailed significant effort in terms of manual entry of patient characteristics. The ASPIRE system used a set of disease codes, such as breast cancer and diabetes, to augment trial eligibility data to facilitate “keyword”-based search for matching trials at the City of Hope National Medical Center.<sup>23</sup>

However, the existing systems have limited access to the complete patient information, such as the latest laboratory test results, and are not integrated with the clinical systems used in routine patient care.<sup>26</sup> Further, existing tools have limited support for structured entry of trial information, interactive user interfaces (UIs) that allow clinicians to review the matching results and re-execute the matching process with updates to patient records. In addition to development of trial matching tools, there has been extensive research in creating formal, computable representation of eligibility trial specifications that can be used together with electronic representation of patient data in EHR systems.<sup>15</sup> A review by Weng et al.<sup>15</sup> considers three aspects of formal representation for trial information: (a) expression language for trial criteria<sup>27</sup> (eg, Arden Syntax<sup>28</sup> and GELLO<sup>29</sup>), (b) encoding of diagnosis information in trials (eg, Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT)<sup>30</sup> and Logical Observation Identifier Names and Codes (LOINC) vocabulary<sup>31</sup>), and (c) representation of patient information (eg, Health Level 7 Reference Information Model<sup>32</sup>). Recent work has found development of template set of eligibility criteria is feasible after comparing the representation of eligibility criteria in ClinicalTrials.gov and existing industry standards.<sup>33</sup> An important challenge for computational representation of trial information is the lack of suitable interfaces for entering eligibility criteria. The Trial Prospector system, which is described in the next section, includes the “Trial Builder” visual interface to facilitate structured entry of trial information.

### Methods

Trial Prospector has been designed for easy integration with the existing clinical workflow and to facilitate patient recruitment in ongoing clinical trials using:



1. structured entry of clinical trial eligibility criteria using an intuitive visual interface;
2. automatically integrated patient data from multiple existing data sources (eg, laboratory test results, diagnosis information, and demographic data); and
3. a scalable matching algorithm for near real-time evaluation of patient data against active clinical trials.
4. An interactive visual interface to review matches results with detailed description of exclusion conditions (both satisfied exclusion criteria and unsatisfied inclusion criteria) that disqualify a patient for specific trials.

Trial Prospector was developed in close collaboration with medical oncologists using an agile software engineering approach for rapid prototyping and iterative implementation of features based on user feedback. Using the Ruby-on-Rails technology stack, Trial Prospector was developed as a Web-based integrated environment (Fig. 1 illustrates the Trial Prospector architecture and its integration with external data sources). The Trial Prospector tool was reviewed by the UHSCC information technology support (ITS) team for data management practices to ensure compliance with the appropriate UHSCC policies on patient data access, privacy, and security. The following sections describe the four modules of Trial Prospector that

have been developed to support the patient recruitment workflow.

**Trial Builder: intuitive interface for composing clinical trial eligibility.** A key component of Trial Prospector is creation and maintenance of a clinical trial eligibility database. Formal representation of clinical trial eligibility criteria is an active area of research and has led to development of various approaches, including the use of clinical trial authoring tools.<sup>15,34–36</sup> However, there are no standard data entry tools for composing executable eligibility criteria that include simple expressions as well as complex nested eligibility expressions. A complex eligibility criterion expression may include conditional statements, for example, “aspartate aminotransferase (AST/SGOT) to be less than five times of upper limit of normal if diagnosis is hepatocellular carcinoma (HCC).” These expressions may be composed of simple constructs that are connected by different types of connectives, such as “if, then” or “if, then, else.” The Trial Builder module has been developed by extending an existing visual UI called the Visual Aggregator and Explorer (VIS-AGE)<sup>37</sup> to support structured entry of eligibility criteria.

*A library of eligibility criteria variables.* The Trial Builder module is prepopulated with a library of eligibility variables to represent various kinds of values used in defining an eligibility criterion. The current version of the library consists of 32 data elements that are classified into five categories:

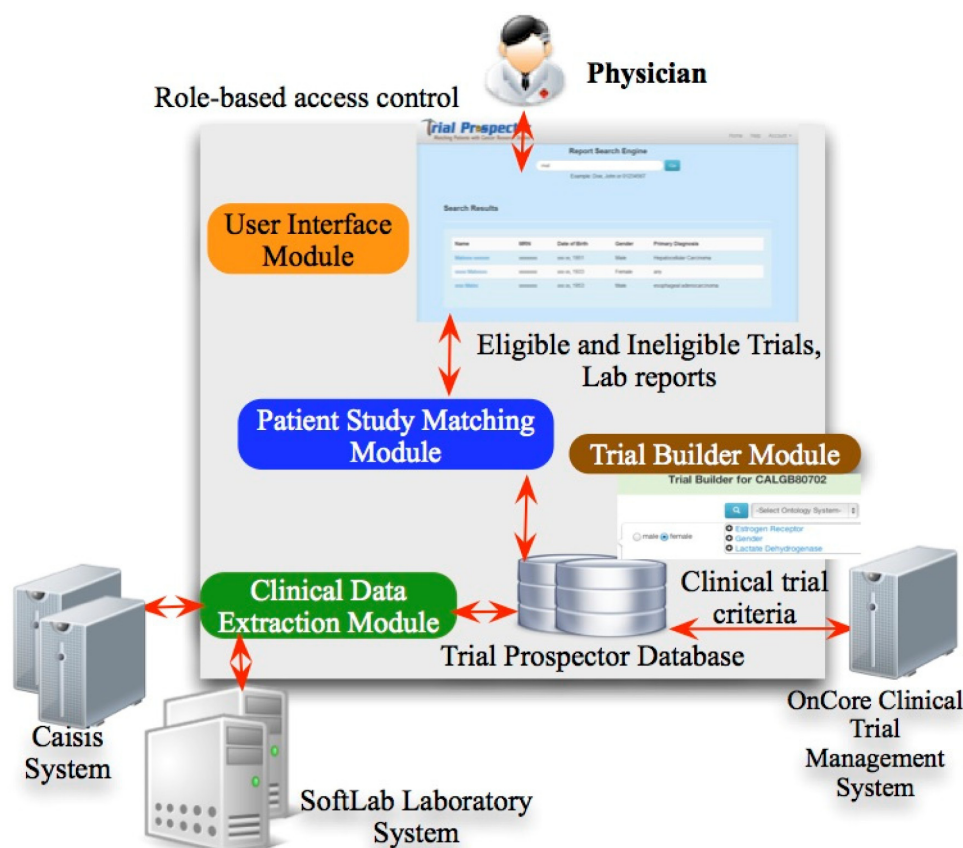


Figure 1. Architecture of Trial Prospector illustrating interface with three external data sources.



1. *Demographic variables:* For example, date of birth, gender, age, and ethnicity.
2. *Social status variables:* For example, education level.
3. *Behavior history variables:* For example, current tobacco use and past tobacco use.
4. *Laboratory test variables:* For example, AST/SGOT, alanine aminotransferase (ALT/SGPT), international normalized ratio (INR), prothrombin time (PT), activated partial thromboplastin time (aPTT), and absolute neutrophil count (ANC).
5. *Health status variables:* For example, presence of metastasis; tumor, node, metastasis (TNM) stage, and stage group.

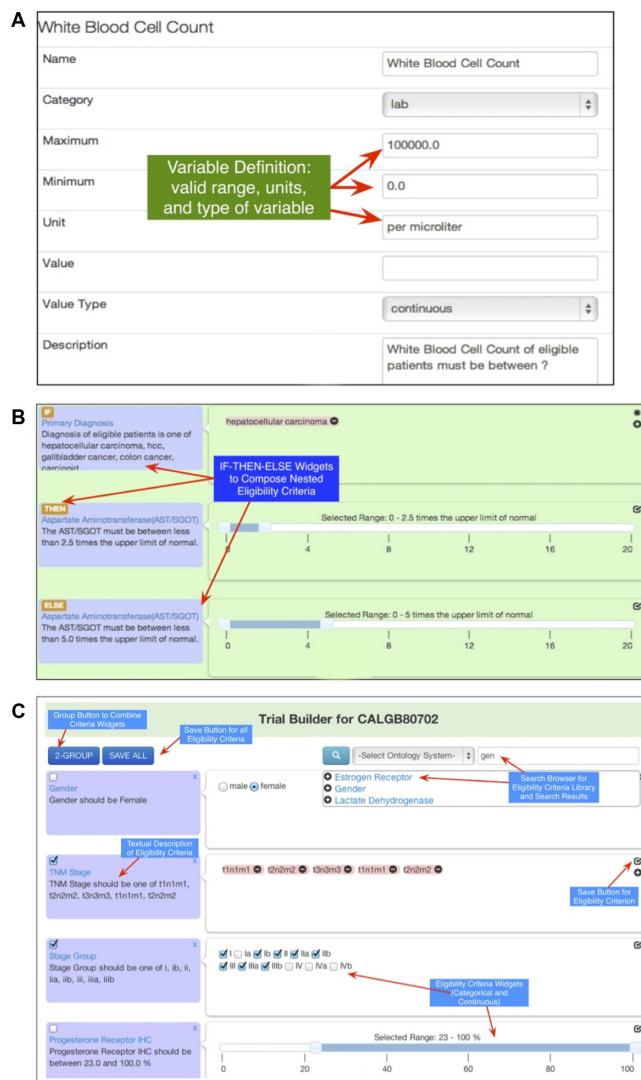
Each variable in the eligibility library includes the variable identifier, the set of valid values, the minimum and maximum values, the unit of measurement, description about the subject of an eligibility criterion, and the “variable type.” Trial Builder uses four categories of variable types, namely, boolean (eg, gender), categorical (eg, race), dynamic categorical (eg, primary diagnosis), and continuous (eg, white blood cell (WBC) count). Figure 2A illustrates an example eligibility variable called “WBC count” that is defined to be a continuous value type with minimum 0 and maximum 100,000 units per microliter.

**Trial Builder UI.** A user (clinical trial investigator) can compose an eligibility checklist for each clinical trial using a multistep process that is supported by the subcomponents of Trial Builder UI (Fig. 2):

1. *Trial criteria variable selector:* Users can select eligibility variables from the Trial Builder library using the “keyword”-based search function in the interface. For example, “Gender” can be retrieved using the search term “...gen...” (Fig. 2C).
2. *Composing complex eligibility expressions:* Eligibility variables can be added to a given clinical trial using the “add” icon in the front (Fig. 2C). Trial Builder creates customized “widgets” corresponding to the specific category of an eligibility variable (eg, “continuous” or “categorical”). Users can specify values for the eligibility variables through the widgets and complete the composition of eligibility criteria for clinical trials.
3. *Grouping together eligibility variables:* Complex criteria expressions can be defined using the “if-then” and “if-then-else” connectives (Fig. 2B). Eligibility criteria in a clinical trial are selectable, and selected eligibility criteria can be grouped as a complex eligibility expression using “Group” button (Fig. 2C).

The output of Trial Builder is stored in a structured format to facilitate easier matching with patient data.

**Real-time extraction and integration of multisource patient data.** A key requirement for an effective trial matching



**Figure 2.** The Trial Builder module: (A) UI to define a new eligibility variable, (B) use of conditional constructs to compose nested eligibility criteria, and (C) composition of a clinical trial CALGB80702 (a phase III trial for resected stage III colon cancer with eight eligibility criteria).

tool is access to the most recent patient data by effectively interfacing with existing hospital data management systems. The trial prospector clinical data extractor (CDE) module retrieves patient demography information, primary diagnosis, TNM classification, metastasis status, stage group, and the three most recent laboratory reports in real time from the following existing systems in UHSCC:

1. *Caisis Clinical Data Management System:* Caisis is an open-source Web-based clinical data management (CDM) system<sup>38</sup> that is used at the UHSCC to store patient demography, disease findings, staging information, and related data across multiple clinical management disease teams.
2. *SoftLab Laboratory Information System:* SoftLab is a proprietary patient laboratory information management (LIM) system<sup>39</sup> and is used at UHSCC to store all blood-based diagnostic tests ordered by physicians that

are performed at any laboratory within the UH health system. The laboratory test results are retrieved from the SoftLab LIM system using an application called Clearview Ace<sup>40</sup> in the CDE module, which complies with the Health Insurance Portability and Accountability Act (HIPAA) data security and privacy regulations implemented in UHSCC.

The CDE module ignores invalid data elements, such as unreported or cancelled tests, to retrieve the three most recent test results for each patient together with metadata for the test results, such as test date, test code, test order number, ordering physician, test result, test range, and test unit. The CDE module includes a dashboard interface to display both new requests and status of all past requests (Fig. 3). The extracted patient data and trial protocol information are used as inputs to the matching module.

### Scalable matching of patients with research studies.

Trial Prospector uses a scalable matching algorithm in the patient study matching (PSM) module to evaluate the eligibility of a patient for a given trial (Fig. 4 illustrates the matching steps implemented in the algorithm). The algorithm includes conversion of laboratory test values to an appropriate unit that allows comparison of patient data with trial eligibility values. For example, the LIM system may store the WBC count per liter (L), while the trial protocols may specify the WBC count per microliter ( $\mu\text{L}$ ). After data conversion, the algorithm parses and stores the trial constraints as inverted indices indexed on trial identifiers, which also include nested eligibility criteria such as “conditional trial constraints.” Conditional trial constraints define an appropriate value for a laboratory test in terms of patient diagnosis or demography information. For example, the alkaline phosphatase value for a phase II trial is defined to be less than 2.5

**Input:** Trial Protocols

**Output:** Eligible and Ineligible Trials with Explanation

1. Load trial protocol and patient records into memory;
2. Create inverted indices for each constraint in trial protocol;
3. Parse and tokenize patient record into separate fields;
4. **For** each patient field record **do**
5.     Match primary diagnosis of patient with trial constraint, record error;
6.     Compare patient TNM value with trial constraint, record error;
7.     Compare patient Stage value with trial constraint, record error;
8.     Compare patient Age value with trial constraint, record error;
9.     Compare patient Metastasis value with trial constraint, record error;
10.    **For** each patient laboratory test result **do**
11.     Convert units of patient test to match trial constraint units;
12.     Check nested trial constraint requirements and compute new value;
13.     Compare computer patient value with trial constraint, record error;
14.    **End**
15.    Store eligible and ineligible trial list in indexed array with exclusion reason
16. **End**

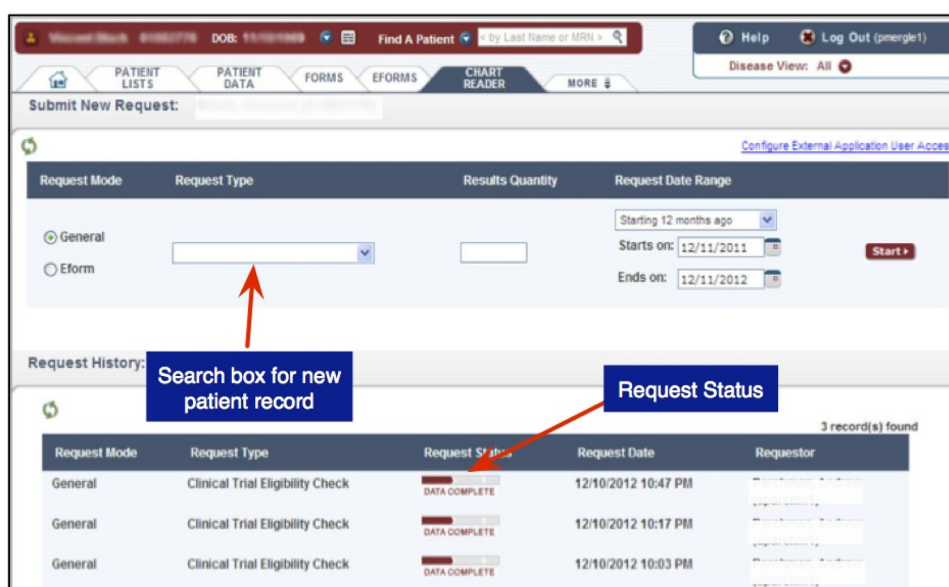
**Figure 4.** The scalable matching algorithm for identifying clinical trials for a patient.

or 5 times the upper limit of normal alkaline phosphatase test value if the primary diagnosis for the patient is HCC. The use of customized data structures enables the PSM algorithm to perform fast lookup of multiple trial constraints and enables both lexical as well as numerical comparisons with patient data.

The algorithm identifies a patient to be eligible for a clinical trial if the following two conditions are satisfied:

1. The patient demography, primary diagnosis, TNM, metastasis, and stage group information match the eligibility criteria.
2. All the values of at least one laboratory test report, among the three most recent reports of a patient, satisfy all the eligibility criteria of the trial.

Thus, the algorithm ensures that the maximum possible set of eligible clinical trials is identified for a given patient. The physician makes the final decision whether to



**Figure 3.** The CDE module interface implemented by extending the Caisis tool interface.

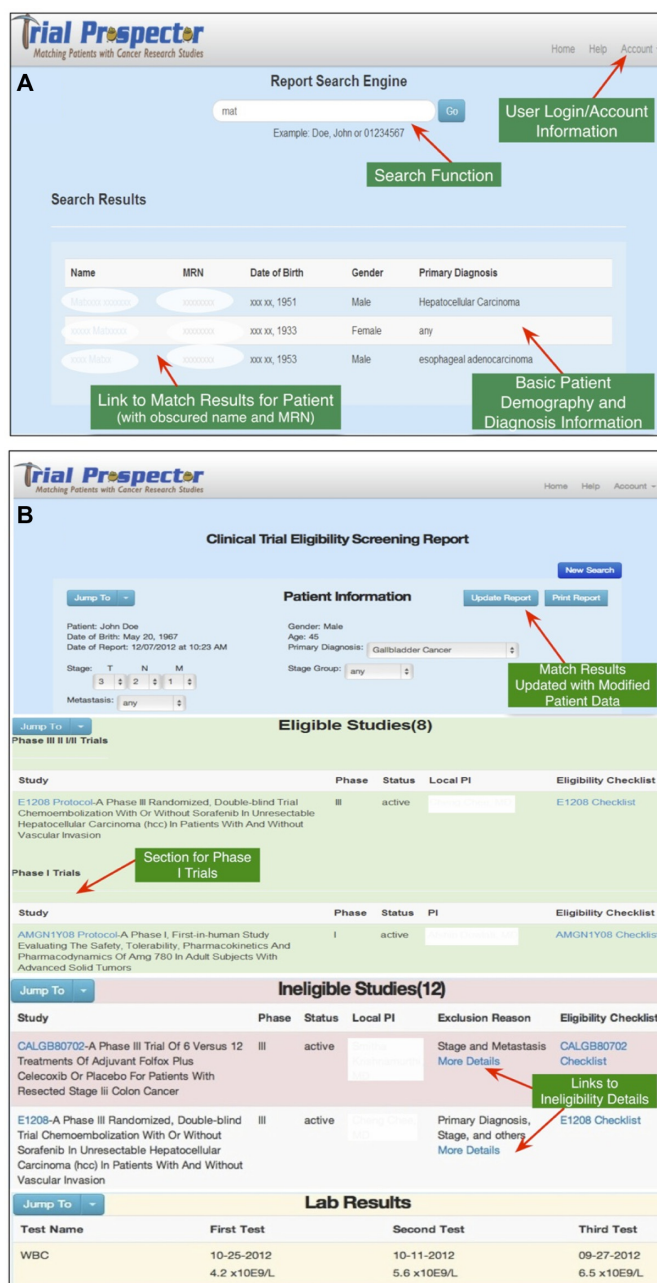


offer trial enrollment to the patient or not. The algorithm keeps track of the laboratory test results that do not match the trial criteria and makes them available for easy review by the physician. This feature ensures that patients are not excluded from participating in a trial based on one specific test result, which may be a transient event. To ensure scalability of the PSM module with increasing number of trials and patients, the algorithm uses a “greedy” approach to prune the search space by removing trials that fail to match a patient record from all subsequent comparisons with the patient data.

**Interactive UI for match results.** A Web browser-based UI module was developed in close consultation with the end-user physicians. The physicians can search for patients by their name (first name, last name) or Medical Record Number (MRN) (Fig. 5A) shows a screenshot of the initial search feature) and select an appropriate patient record. The results of the matching algorithm for the selected patient record with links to trial protocols are displayed in the UI module using four visual facets (Fig. 5B) displays screenshots of the four facets):

1. *Patient information:* The basic patient information, including demography, diagnosis, stage information, and metastasis status, are displayed in this facet. The physician may modify some of the values, for example, “stage group,” and the matching results are updated in real time (using the “update report” button).
2. *Eligible trials:* This facet lists the clinical trials that the patient is qualified for enrollment. It includes the study details, such as descriptive title of the study, the study phase, the status of the study, the principal investigator of the study, and hyperlinks to the clinical trial documents for manual review by the physician.
3. *Ineligible trials:* The clinical trials for which the patient is ineligible are listed in this facet. A brief description of the reason for ineligibility is also displayed with a link to a detailed description of the cause for ineligibility, which can be selected by the physician for additional review.
4. *Laboratory test results:* The three most recent patient laboratory reports describing the type of test, the test value, and the appropriate units of the test are displayed in this facet. This information allows the physician to quickly review and reconfirm the eligibility or ineligibility of a patient for a given trial.

*Explanation of ineligibility of a patient for specific trials.* Many decision support systems, reasoning tools, and question answering systems include a feature for representing explanations for a given decision or result, which is also called as “proof metadata.”<sup>41</sup> Unlike many existing matching tools, the PSM module of Trial Prospector not only enumerates the eligible and ineligible trials for a patient but also describes the specific reasons for exclusion from a trial. This allows physicians



**Figure 5.** The Trial Prospector UI. (A) Results of search performed using first name of patient. (B) The four facets constituting the Trial Prospector interactive UI displaying match results.

to efficiently review the match results and identify transient reasons for trial ineligibility. For example, a previously healthy patient may have been recently hospitalized for dehydration and acute kidney injury with an elevated creatinine of 2 mg/dL, which would lead to the patient’s exclusion from a trial based on the elevated creatinine. However, when the physician reviews the report describing the last three laboratory test results in Trial Prospector, they would recognize that the elevated creatinine is an “outlier value” compared to the patient’s baseline value. Thus, the “exclusion reasons” section of the UI module will allow the physician to still consider the patient for the trial (Fig. 6).

## Results

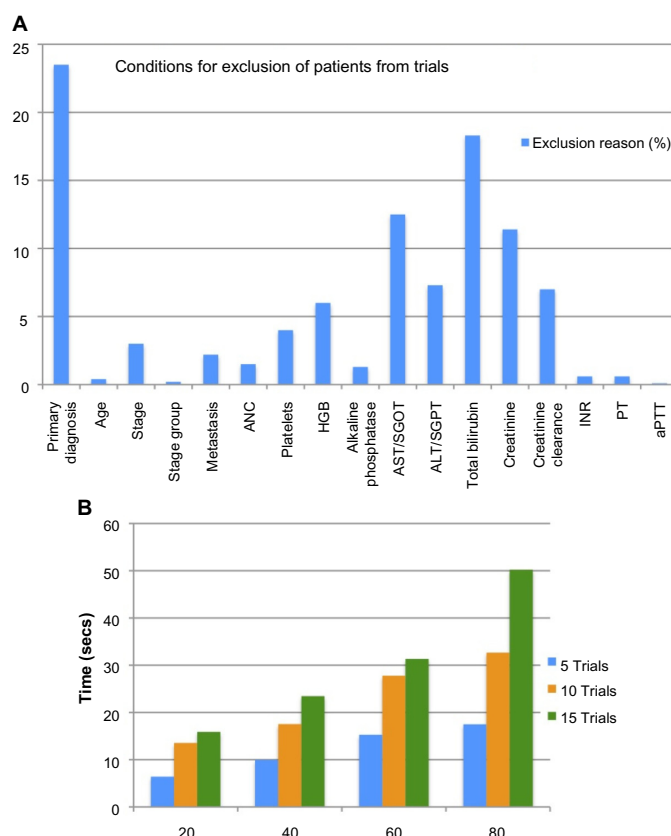
In order to assess the clinical utility of Trial Prospector, it was deployed at UHSCC and used by physicians in gastrointestinal oncology subspecialty clinics. We report the results of a pilot deployment of Trial Prospector between December 2012 and January 2013. During this time period, a Trial Prospector report was generated for each new patient visit and integrated into the clinician workflow. The UH Institutional Review Board (IRB) approval was obtained prior to performing the pilot deployment, and all data were managed in a HIPAA-compliant manner. Prior to the deployment of Trial Prospector, a control survey was administered to physicians to assess baseline characteristics. Once the control group was completed, physicians were provided training on the use of Trial Prospector. They were then given web-based access to Trial Prospector and asked to utilize it during new patient evaluations. Physicians were asked to complete a survey after each patient encounter and a summary-experience survey at the end of the pilot program. The surveys were created and administered using the research electronic data capture (REDCap) application.<sup>42</sup>

**Patient eligibility.** During the development and pilot phase, a total of 85 Trial Prospector reports were generated. There were 15 GI/Phase I clinical trials that were included for matching purposes. A total of 1,367 clinical trials matching computations were performed by Trial Prospector, and a mean of  $7.19 \pm 3.41$  eligible trials was identified per patient. Likewise, a mean of  $8.89 \pm 3.87$  ineligible trials was identified per patient. Figure 7A displays the list of conditions that resulted in a patient being excluded from a clinical trial. The most common reasons for ineligibility were primary diagnosis and labs. A physician performed manual review of each report, and evaluation found that the matching algorithm was 100% accurate for the eligibility criteria included in this pilot study.

**User evaluation survey.** A total of 11 medical oncologists (6 attending physicians and 5 oncology fellows) participated in the pilot program. Trial prospector was deployed at the point of care for 60 new patient visits in the gastrointestinal

Test Name	Trial Requirement	Patient Data	Reference Range	Test Date
Primary Diagnosis	colon cancer	Adenoma		
Creatinine Clearance	to be calculated if cr is out of range and needs to be >= 60	36.65	97-137	07-05-2012
Creatinine Clearance	to be calculated if cr is out of range and needs to be >= 60	36.65	97-137	05-20-2012
Creatinine	<= 1.5x upper limit of normal	1.8 mg/dL	0.51-1.17	08-05-2012
Creatinine	<= 1.5x upper limit of normal	1.8 mg/dL	0.51-1.17	07-05-2012

**Figure 6.** The detailed explanation for exclusion of a patient from specific trial.



**Figure 7.** The results from deployment of Trial Prospector at the Seidman Cancer Center. (A) Distribution of exclusion conditions for patients. (B) Results of performance evaluation of Trial Prospector over increasing number of patients and trials.

oncology subspecialty clinics. During the study, a total of 14 control surveys, 60 patient-specific surveys, and 11 summary-experience surveys were completed with the physicians and oncology fellows. The results of the surveys are depicted in Tables 1 and 2 (these data were presented in part (poster presentation) at the Annual Meeting of the American Society of Clinical Oncology in June 2013).

Oncologists reviewed the Trial Prospector report at the point of care during 95% of the new patient visits. Trial Prospector complemented the existing workflow to assess patient eligibility with 70% of participating oncologists spending 0–5 minutes, about 20% of the oncologists spending 6–10 minutes, and 10% of the oncologists spending 11–15 minutes. Furthermore, physicians reported that it saved time in identifying trials during 57.1% of the patient visits. Trial Prospector was a useful tool in the clinic with 72.7% of participating oncologists stating that Trial Prospector made it easier to find clinical trials for their patients. The majority of physicians found Trial Prospector to be accurate (75.9%), visibly pleasing (90.9%), and easy to use (100%). Clinical trial enrollment was similar between the two groups with 12.5% of the control group and 10.5% of the Trial Prospector group enrolling in a trial. Meaningful comparison is limited by the small sample sizes. Overall, 81.8% of the participating physicians would

**Table 1.** Physician survey results for the control and Trial Prospector groups.

	CONTROL (14)		TRIAL PROSPECTOR (60)	
	YES	NO	YES	NO
1. When thinking about treatment options for this patient, did you consider clinical trials?	8 (57.1%)	6 (42.9%)	40 (66.7%)	20 (33.3%)
2. Will you need to spend additional time outside of clinic hours to identify clinical trials for this patient?	0	8 (100%)	2 (5%)	38 (95%)
3. Did you discuss clinical trial opportunities with this patient?	5 (62.5%)	3 (37.5%)	15 (37.5%)	25 (62.5%)
4. Did this patient enroll on a clinical trial?	1 (12.5%)	7 (87.5%)	4 (10.5%)	34 (89.5%)
5. Did you review a TP report for this patient?	–	–	29 (72.5%)	11 (27.5%)
6. To your knowledge was the information provided by TP accurate?	–	–	22 (75.9%)	7 (24.1%)
7. Did TP save you time in identifying potential clinical trials?	–	–	16 (57.1%)	12 (42.9%)
8. Would you recommend utilizing TP for eligibility screening?	–	–	25 (89.3%)	3 (10.7%)

**Table 2.** A summary of the user experience survey.

	YES	NO
1. Did TP allow you to spend more time with your patients?	2 (18.2%)	9 (81.8%)
2. Did TP make it easier to find eligible clinical trials for your patients?	8 (72.7%)	3 (27.3%)
3. Did TP help you communicate with patients about clinical trials?	2 (18.2%)	9 (81.8%)
4. Did TP make it easier to review protocols and eligibility checklists?	9 (81.8%)	2 (18.2%)
5. Is TP easy to use and navigate?	11 (100%)	0
6. Is the TP interface visibly pleasing?	10 (90.9%)	1 (9.1%)
7. Would you recommend TP to other Physicians?	9 (81.8%)	2 (18.2%)

recommend Trial Prospector to other physicians for clinical trial eligibility screening.

Free text comments were solicited from users. The physicians reported that they liked the ease of use, auto-population of data, and detailed report of why a patient was ineligible for certain trials.<sup>43</sup> Several areas for improvement were also suggested, including the use of more eligibility criteria and stricter matching criteria by the matching algorithm. It was also suggested that phase I trials be automatically excluded from the report when the patient is being considered for adjuvant therapy. Overall, these results indicate that Trial Prospector is a feasible, accurate, and effective means to identify clinical trials for individual patients in a busy outpatient oncology clinic.

**Performance scalability evaluation.** Trial Prospector is a scalable tool that efficiently computes matching reports over increasing number of both clinical trials and patients. Figure 7B shows the performance of Trial Prospector as the number of patients increase from 20 to 80 with 5, 10, and 15 trials. To support larger patient populations and number of trials, we propose to implement the CDE and PSM modules using parallel computing infrastructure to enhance the performance of both data extraction and matching.

## Discussion

Trial Prospector is a practical, end-to-end automated tool that can be seamlessly incorporated into the clinical workflow and, thus, facilitate the enrollment of patients on clinical trials. We demonstrated that Trial Prospector increases accessibility to protocol details for physicians and, thus, reduces the workload of research staff through automated data extraction, integration, and matching processes. The feedback from physicians strongly encouraged further deployment of Trial Prospector. To facilitate wider deployment of Trial Prospector with support for multiple types of cancer, some additional features will be implemented during the next phase of development.

**Integration with patient scheduling system and trial data entry.** At present, a list of patients is manually entered into Trial Prospector using the CDE module interface by the research staff. This is a potential bottleneck in supporting patients across all oncology clinics, which can be addressed through integration with the patient scheduling system. The next phase of Trial Prospector development will include integration with the hospital scheduling system (eg, Athena).<sup>44</sup> Further, we propose to incorporate terms from the Eligibility Rule Grammar and Ontology (ERGO) project<sup>45</sup> to facilitate representation of trial information in a machine-readable format. We are also upgrading the clinical trial eligibility interface to improve ease of use for entering clinical trials information as new studies are activated or study amendments occur.

**Use of NCI Thesaurus for matching primary diagnosis.** There is significant variability in the use of terminology for reporting cancer diagnosis, which makes it difficult to accurately compare diagnosis term with the trial specification. For example, the trial may specify the primary diagnosis constraint as “pancreatic cancer,” but the patient record may use the term “adenocarcinoma of the pancreas.” Hence, a standard lexical comparison of the two primary diagnosis terms will result in ineligibility of the patient for the trial. This issue can be addressed by an ontology, such as the NCI Thesaurus, which classifies “adenocarcinoma of the pancreas” as a specific type of “pancreatic cancer.” We propose to use the Web Ontology





Language (OWL2)<sup>46</sup> version of the NCI Thesaurus, which is available from the NCI Enterprise Vocabulary Services (EVS), in the PSM module for semantic matching.

The complete NCI Thesaurus is a large ontology with 60,000 concepts, which is not required for the current scope of Trial Prospector in the UHSCC gastrointestinal cancer clinic. We will extract a segment of the NCI Thesaurus class hierarchy rooted at “ncit: GastroIntestinal\_Carcinoma,” which includes 193 classes representing various subtypes of gastrointestinal cancer using an automated script. Using the open-source OWLAPI java-based framework<sup>47</sup> together with open-source OWL reasoners (eg, Pellet<sup>48</sup> or HermiT<sup>49</sup>), the PSM module will integrate the NCI Thesaurus into the primary diagnosis phase of the matching algorithm. The subsumption reasoning algorithm used in the PSM module will be based on the standard OWL2 semantics.<sup>50</sup>

**Limitations.** The user survey may not accurately capture clinical trial enrollment because some patients may have enrolled in a trial during a subsequent visit. Additional studies with larger patient numbers are required to accurately measure the effect of Trial Prospector on clinical trial enrollment and analyze other factors, such as patient characteristics and clinical trial availability, that may bear on treatment decision making. We plan to conduct a more comprehensive survey about the role of Trial Prospector after its deployment in all oncology clinics in UHSCC. Finally, we did not identify patients recruited through Trial Prospector who would not have been recruited otherwise; a prospective randomized clinical trial of Trial Prospector vs. standard management will be necessary to evaluate the effectiveness of Trial Prospector in affecting patient enrollment.

## Conclusion

We described an automated tool that matches patients with ongoing clinical trials at the point of care. Physicians used this tool to facilitate patient enrollment in active clinical trials without disrupting existing clinical workflows. Trial Prospector features four loosely coupled modules to support end-to-end data management: (a) structured entry of trial eligibility criteria, (b) automatically extracting patient information from the CDM and LIM systems, (c) an algorithm that uses a scalable approach for matching patient data with eligibility criteria, and (d) an interactive UI for matching results that support updates to patient information for real-time rematch with trial data. In a pilot study of Trial Prospector, 1,367 clinical trial matching tests were performed. The feedback from the users, consisting of attending physicians and clinical fellows, has been very positive with strong recommendation for wider deployment of Trial Prospector.

## Acknowledgments

We acknowledge the contributions of Joseph Teagno, James Warfe, Christopher Opper, and Dawn Miller in the development of Trial Prospector.

## Author Contributions

Conceived and designed the experiments: NJM, GQZ, SSS, ST, PM. Analyzed the data: AP, ST, PM, GQZ, NJM, SSS. Wrote the first draft of the manuscript: SSS, GQZ. Contributed to the writing of the manuscript: NJM, AP, JSBS, PM, ZL, LC, RL. Agree with manuscript results and conclusions: SSS, ST, AP, ZL, LC, PM, RL, JSBS, NJM, GQZ. Jointly developed the structure and arguments for the paper: SSS, GQZ, NJM, AP, PM. Made critical revisions and approved the final version: NJM, JSBS, GQZ, SSS. All authors reviewed and approved the final manuscript.

## REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010;127:2893–917.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69–90.
3. World Health Organization. The Global burden of disease: 2004 update. Report. Geneva. 2008.
4. American Cancer Society. Available at [http://www.cancer.org/treatment/treatmentsandsideeffects/clinicaltrials/index].0000. Accessed on August 16, 2014.
5. National Comprehensive Cancer Network. Available at [http://www.nccn.org/patients/resources/clinical\_trials/default.aspx].0000. Accessed on August 16, 2014.
6. Lara PN Jr, Higdon R, Lim N, et al. Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment. *J Clin Oncol*. 2001;19(6):1728–33.
7. Stiller CA, Kroll ME, Pritchard-Jones K. Population survival from childhood cancer in Britain during 1978–2005 by eras of entry to clinical trials. *Ann Oncol*. 2012;23(9):2464–9.
8. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA*. 2004;291(22):2720–6.
9. Al-Refai WB, Vickers SM, Zhong W, Parsons H, Rothenberger D, Habermann EB. Cancer trials versus the real world in the United States. *Ann Surg*. 2011;254(3):438–42.
10. Meropol NJ, Buzaglo JS, Millard J, et al. Barriers to clinical trial participation as perceived by oncologists and patients. *J Natl Compr Canc Netw*. 2007;5(8):655–64.
11. Denicoff AM, McCaskill-Stevens W, Grubbs SS, et al. The National Cancer Institute-American Society of Clinical Oncology Cancer Trial Accrual Symposium: summary and recommendations. *J Oncol Pract*. 2013;9:267–76.
12. Butte A, Weinstein D, Kohane I. Enrolling patients into clinical trials faster using RealTime Recruiting. *Proc AMIA Symp*. 2000;2000:111–5.
13. Coalition of Cancer Cooperative Groups (CCCG). *Baseline study of patient accrual onto publicly sponsored US cancer clinical trials: an analysis conducted for the Global Access Project of the National Patient Advocate Foundation*. Philadelphia, PA: Coalition of Cancer Cooperative Groups; 2006.
14. Breitfeld PP, Weisburd M, Overhage JM, Sledge G Jr, Tierney WM. Pilot study of a point-of-use decision support tool for cancer clinical trials eligibility. *J Am Med Inform Assoc*. 1999;6(6):466–77.
15. Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. *J Biomed Inform*. 2010;43(3):451–67.
16. Weng C, Batres C, Borda T, et al. A real-time screening alert improves patient recruitment efficiency. *AMIA Annu Symp Proc*. 2011;2011:1489–98.
17. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a clinical trial alert system on physician participation in trial recruitment. *Arch Intern Med*. 2005;165(19):2272–7.
18. Penberthy L, Brown R, Puma F, Dahman B. Automated matching software for clinical trials eligibility: measuring efficiency and flexibility. *Contemp Clin Trials*. 2010;31(3):207–17.
19. Cohen E, Belkora J, Tyler J, et al. Adoption, acceptability, and accuracy of an online clinical trial matching website for breast cancer. *J Med Internet Res*. 2012;14(4):e97.
20. Pressler TR, Yen PY, Ding J, Liu J, Embi PJ, Payne PR. Computational challenges and human factors influencing the design and use of clinical research participant eligibility pre-screening tools. *BMC Med Inform Decis Mak*. 2012;12:47.
21. Ashburn A, Pickering RM, Fazakarley L, Ballinger C, McLellan DL, Fitton C. Recruitment to a clinical trial from the databases of specialists in Parkinson's disease. *Parkinsonism Relat Disord*. 2007;13(1):35–9.
22. Eea C. caMATCH: a patient matching tool for clinical trials. In: caBIG Annual Meeting. Bethesda, MD. 2005.



23. Niland J. ASPIRE: Agreement on Standardized Protocol Inclusion Requirements for Eligibility. 2007.
24. Kamal J, Pasuparthi K, Rogers P, Buskirk J, Mekhjian H. Using an information warehouse to screen patients for clinical trials: a prototype. *AMLA Annu Symp Proc.* 2005;2005:1004.
25. Metz JM, Coyle C, Hudson C, Hampshire M. An internet-based cancer clinical trials matching resource. *J Med Internet Res.* 2005;7(3):e24.
26. Tassignon JP, Sinackevich N. Speeding the critical path. *Appl Clin Trials.* 2004;13:42.
27. Tu SW, Peleg M, Carini S, et al. A practical method for transforming free-text eligibility criteria into computable criteria. *J Biomed Inform.* 2011;44(2):239–50.
28. Hripscak G. Writing Arden Syntax Medical Logic Modules. *Comput Biol Med.* 1994;24(5):331–63.
29. Sordo M, Boxwala AA, Ogunyemi O, Greenes RA. Description and status update on GELLO: a proposed standardized object-oriented expression language for clinical decision support. *Stud Health Technol Inform.* 2004;107(pt 1):164–8.
30. SNOMED Clinical Terms. Available at [[http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)] 0000. Accessed on August 16, 2014.
31. Huff SM, Rocha RA, McDonald CJ, et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. *J Am Med Inform Assoc.* 1998;5(3):276–92.
32. Jenders RA, Sujansky W, Broverman CA, Chadwick M. Towards improved knowledge sharing: assessment of the HL7 Reference Information Model to support medical logic module queries. *Proc AMLA Annu Fall Symp.* 1997;1997:308–312.
33. Bhattacharya S, Cantor MN. Analysis of eligibility criteria representation in industry-standard clinical trial protocols. *J Biomed Inform.* 2013;46(5):805–13.
34. Rubin DL, Gennari JH, Srinivas S, et al. Tool support for authoring eligibility criteria for cancer trials. *Proc AMLA Symp.* 1999;1999:369–73.
35. Olasav B, Sim I. RuleEd, a web-based semantic network interface for constructing and revising computable eligibility rules. *AMLA Annu Symp Proc.* 2006;2006:1051.
36. Gennari J, Sklar D, Silva J. Cross-tool communication: from protocol authoring to eligibility determination. *Proc AMLA Symp.* 2001;2001:199–203.
37. Zhang GQ, Siegler T, Saxman P, et al. VISAGE: a query interface for clinical research. *AMLA Jt Summits Transl Sci Proc.* 2010; ;2010:76–80.
38. Fearn PA, Regan K, Sculli F, et al. A Chronological Database as Backbone for Clinical Practice and Research Data Management. In: 16th Annual IEEE Symposium on Computer-Based Medical Systems: 2003; 2003: 9–15; New York (NY).
39. SCC Soft Computer's LIS solution (SoftLab). Available at [<http://www.softcomputer.com/products-services/laboratory/softlab/>] 0000. Accessed on August 16, 2014.
40. Cleaview Technologies. Available at [<http://www.clearviewti.com/>], 0000. Accessed on August 16, 2014.
41. Sizov S. What Makes You Think That? The Semantic Web's Proof Layer. *IEEE Intell Syst.* 2007;22(6):94–9.
42. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377–81.
43. Parchman AJ, Zhang GQ, Mergler P, et al. Trial prospector: an automated clinical trials eligibility matching program. *J Clin Oncol.* 2013;(suppl; abstr 6538).
44. athenaCommunicator. Available at [<http://www.athenahealth.com/our-services/athenaCommunicator/patient-communication.php>], 0000 Accessed on August 16, 2014.
45. ERGO. A Template-Based Expression Language for Encoding Eligibility Criteria. Available at [[http://128.218.179.58:8080/homepage/ERGO\\_Technical\\_Documentation.pdf](http://128.218.179.58:8080/homepage/ERGO_Technical_Documentation.pdf)], 0000 Accessed on August 16, 2014.
46. Hitzler P, Krötzsch M, Parsia B, et al. OWL 2 Web Ontology Language Primer. W3C Recommendation. World Wide Web Consortium W3C; 2009.
47. Horridge M, Bechhofer S. The OWL API: a Java API for OWL Ontologies. *Semantic Web J.* 2011;2(1):11–21.
48. Sirin E, Parsia B, Cuenca Grau B, et al. Pellet: a practical OWL-DL reasoner. *Semantic Web J.* 2007;5(2):51–3.
49. Glimm B, Horrocks I, Motik B, et al. Optimising Ontology Classification. *International Semantic Web Conference (ISWC)*; . Shanghai, China: Springer; 2010.
50. Motik B, Patel-Schneider PF, Grau BC. OWL 2 Web Ontology Language Direct Semantics. W3C Recommendation. World Wide Web Consortium W3C; 2012.