Biomedical Informatics Investigator

Peter L. ELKIN[a,1], Sarah MULLIN[a], Sylvester SAKILAY[a]
[a]*Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, SUNY, New York, USA*

Abstract: The BMI Investigator is a computer human interface built in .Net which allows simultaneous query of structured data such as demographics, administrative codes, medications (coded in RxNorm), laboratory test results (coded in LOINC) and formerly unstructured data in clinical notes (coded in SNOMED CT). The ontology terms identified using SNOMED are all coded as either positive, negative or uncertain assertions. They are then where applicable built into compositional expressions and stored in both a graph database and a triple store. The SNOMED CT codes are stored in a NOSQL database, Berkley DB, and the structured data is stored in SQL using the OMOP / OHDSI format. The BMI investigator also lets you develop models for cohort selection (data driven recruitment to clinical trials) and automated retrospective research using genomic criteria and we are adding image feature data currently to the system. We performed a usability experiment and the users identified some usability flaws which were used to improve the software. Overall, the BMI Investigator was felt to be usable by subject matter experts. Next steps for the software are to integrate genomic criteria and image features into the query engine.

Keywords: Clinical Research Informatics, Ontology, Recruitment to clinical trials, Automated retrospective research, clinical genomic trial recruitment

1. Introduction

Semantic Interoperability is a long held goal of the field of biomedical informatics. (1) (2) This requires formal representation of the knowledge in the clinical record(3). We describe our effort to use and validate a semantically interoperable interface and system to automate retrospective research, to enhance our ability to author clinical predication rules and our ability to perform data driven recruitment to clinical trials. (4) (5)

Many authors have written about semantic interoperability and ISO TC 215 TS 17117 describes the value and composition of nomenclatures and terminologies that enable semantic interoperability. (6, 7) The Springer series book Terminology and Terminological Systems guides one through the principles of semantic interoperability and the nomenclatures and tools available to help one achieve that goal. (8, 9)

Today we have standards such as SNOMED CT which represents general medicine in a description logic based terminology. (10) (11) (12) (13) RxNorm or the ATC remain the standard for drug terminologies in both the US and Europe respectively. Elkin and Brown published a drug semantics from the US physicians' desk reference (online as the Daily Med), which provides in codified form the indications, contraindications, and adverse reactions for all drugs which can be used for clinical decision support. (14)

LOINC is an open source terminology which began as a code set for laboratory test results. By utilizing these standards on our primary data we have developed an application which can query across Clinical, Genomic and Image data and enable fully automated retrospective research. (15)

[1] Corresponding Author, Peter L. Elkin, MD, Department of Biomedical Informatics, University at Buffalo, 77 Goodell Street, Buffalo, NY 14203

2. Methods

The data for the BMI Investigator is stored in OMOP / OHDSI format with a Berkley DB NOSQL database. The medications are all coded with RxNorm and the labs are coded with LOINC. The Berkley DB database holds SNOMED CT codes that are parsed out of the patients' clinical notes. The data is stored by patient, document, section, subsection, problem, paragraph, sentence, compositional expression, then named entity and polarity. We code the polarity of each entity as a positive, negative or uncertain assertion, explicitly using the HTP-NLP system. (16, 17) These are then formed into compositional expressions where possible and this data is stored in a triple store.

The BMI Investigator application was written in .Net and was created using the user-centered design development method. (18, 19) We tested the system on a population of 212,343 patients in our outpatient practices at the UBMD practice plans. The data for this trial was from 2010 to 2015. The data used in the system was judged by the IRB to be IRB Exempt #587570.

Inclusion Criteria: All patients 18 years old or older

In our development process, we had Clinical Informatics Fellows and Biomedical Informatics Masters' and PhD students use the system. We observed them using the system and asked them to describe their experience using the think aloud method. We paid particular attention to the understandability of the screen ques and the results.

The system allows users to use Boolean logic and parentheses to construct their queries. It also allows subqueries so that one can define a population and then ask questions of the population. The users do not ever see a code and do not have to know anything about the information model or the ontologies in use to use the system. When the input string or parts of the input string have no map to our ontologies they are searched as a keyword search. The system allows one to save intermediate queries, reuse them, add to them and import them for reuse. Once created these models can be run in a batch mode.

Genomic data is presented as gene abnormalities that are used in clinical medicine and polymorphisms that have been identified are stored in a separate set of tables and they are also used to match to our patients who are included in the precision oncology project. (20) We add image features which are stored matrices and vectors extracted from images using image data analysis tools developed at UB. These act as separate Boolean connected search criteria. Datasets can then be exported in a csv format for further analysis and reporting.

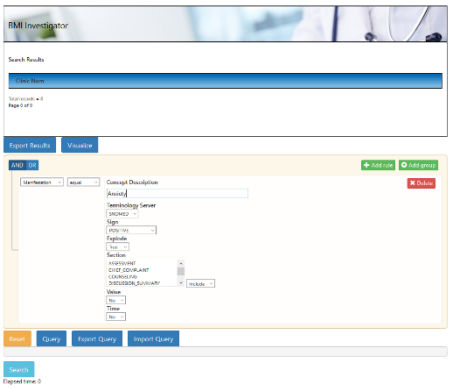

Figure 1: BMI Investigator Data Entry

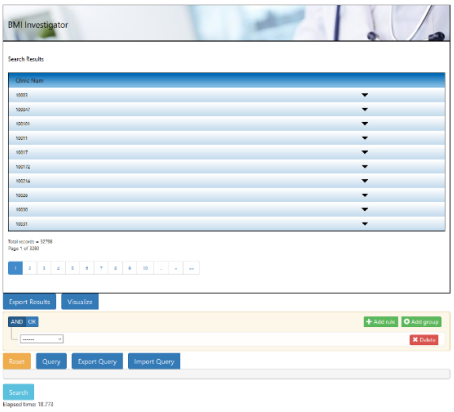

Figure 2: Results Page (right)

We report the results of the Usability study were 8 participants used the system under supervision going through the same scenarios. (21) (22, 23) Each participant was asked to identify the relative risk of Obstructive Sleep Apnea comparing patients who have Rosacea and those that do not have Rosacea. This is a complex task that requires four queries to accomplish. Each student asked to set up the problem as two ratios that could then be compared using a Pearson Chi-Square test. The students were asked how easy was the software to use? How easy was the software to learn? Could you design a more intuitive interface?

3. Results:

The system has a simple interface. Where researchers enter what they want to query and the results are returned almost always in less than a minute. Users enter into simple search line what they are interested in looking for in their query. They specify which ontology if they want to use. They specify if they are looking for positive, negative, uncertain or not mentioned cases. They specify whether they want the ontology terms exploded (the reflexive transitive closure on subsumption) or not. They specify if they want to limit the search to certain sections of the clinical note or not. The user specifies if there is a value that they are looking for or range of values and units or not. Then they specify any time constraints on the query (perhaps you want to recruit patients over one time period who meet the inclusion / exclusion criteria and then follow some outcome sometime in the future.

Results come back quickly and in this case we are looking at patients who have anxiety in the practice and we can see that there are 32,798 patients reporting anxiety in our dataset (see figure xx). We display that about twice as many women report anxiety as men (See figure 3). You can also see the age distribution of our anxious patients (See Figure 4).

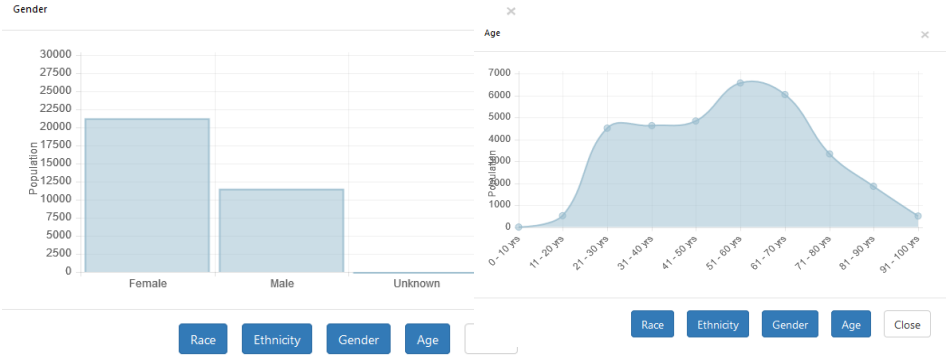

Figure 3: Male / Female Rate of Anxiety Reporting Figure 4: Age distribution of Anxiety Reporting

The usability study showed that all clinicians and students using the system (four clinicians and four graduate students) were able to use the system easily. They all understood the screen ques. One clinician, a pediatrician felt that he would like more direct control of the codes in the query. One graduate student believed that she would like more statistical tools integrated with the system. All students were able to accomplish the task and to obtain the correct answer. The survey was a five point likert scale and all participants thought the system was easy to learn intuitive (score 5/5) and all participants thought that they system was easy to use (score 5/5), three participants thought that they could design a better interface (score 4.2/5).

4. Discussion:

This application allows investigators to do clinical, genomic, image research using deep phenotyping to better understand their patient populations. We believe that this is one method for discovering how best to implement precision medicine into the practice. Research using tools like the BMI Investigator will allow us to speed translation of research to the bedside by finding patients who would be amenable to target based therapies.

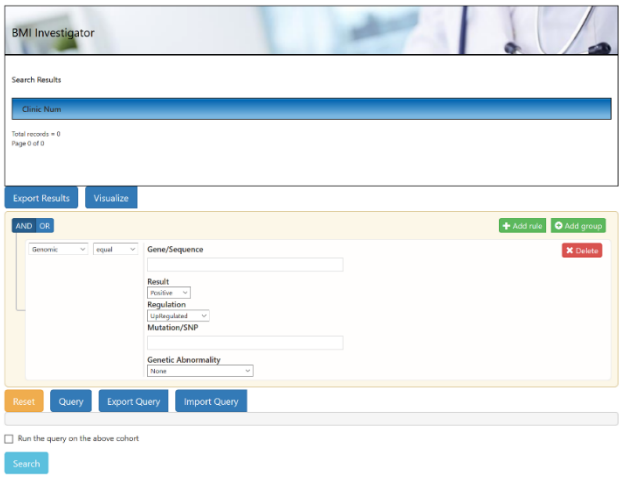

Figure 5: Genomic Data entry

The system has now had genomic data incorporated into its search engine and we are working on image feature recognition next (See Figure 5).

The BMI Investigator is a usable system that hides much of its complexity from the user. This allows one to employ more advanced ontological techniques without an army of ontologists being present to assist uses. That said, one does have to know how to ask a good critical question to use the system effectively.

The BMI Investigator is a usable tool that facilitates clinical genomic and image analytic research in a single interface. The usability study is leading to changes in the design of the BMI Investigator Interface. The system is available with attribution for academic use.

We have used the system to generate clinical prediction rules for clinical decision support to improve the practice of medicine. It is also being used to teach logic and clinical decision support to Biomedical Informatics students.

The next steps are to integrate searching by Image features. This tool can be used to automate retrospective research, to create models for data driven recruitment to clinical trials and to plan prospective clinical trials. This is done by seeing what each additional inclusion or exclusion criteria would do to the rate of recruitment to a prospective clinical trial.

Acknowledgements: This work has been supported in part by grants from NIH NLM T15LM012595, and NCATS UL1TR001412. This study was funded in part by the NCI and the Department of Veterans Affairs through the BD-STEP program, and through a grant from the VA's MAVERIC research group.

References:

1. Elkin PL, Bailey KR, Chute CG. A randomized controlled trial of automated term composition. Proc AMIA Symp. 1998:765-9.
2. Elkin PL, Froehling D, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Bailey K, et al. Aequus communis sententia: defining levels of interoperability. Stud Health Technol Inform. 2007;129(Pt 1):725-9.
3. Elkin PL, Brown SH, Chute CG. Guideline for health informatics: controlled health vocabularies--vocabulary structure and high-level indicators. Stud Health Technol Inform. 2001;84(Pt 1):191-5.
4. Elkin PL, Ruggieri AP, Brown SH, Buntrock J, Bauer BA, Wahner-Roedler D, et al. A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM. Proc AMIA Symp. 2001:159-63.
5. Elkin PL, Brown SH, Lincoln MJ, Hogarth M, Rector A. A formal representation for messages containing compositional expressions. Int J Med Inform. 2003;71(2-3):89-102.
6. Elkin PL, Brown SH, Carter J, Bauer BA, Wahner-Roedler D, Bergstrom L, et al. Guideline and quality indicators for development, purchase and use of controlled health vocabularies. Int J Med Inform. 2002;68(1-3):175-86.
7. Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. Stud Health Technol Inform. 2010;155:14-29.
8. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA. 2011;306(8):848-55.
9. Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. Ann Intern Med. 2012;156(1 Pt 1):11-8.
10. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006;13(3):277-88.
11. Brown SH, Husser CS, Wahner-Roedler D, Bailey S, Nugent L, Porter K, et al. Using SNOMED CT as a reference terminology to cross map two highly pre-coordinated classification systems. Stud Health Technol Inform. 2007;129(Pt 1):636-9.
12. Matheny ME, Fitzhenry F, Speroff T, Green JK, Griffith ML, Vasilevskis EE, et al. Detection of infectious symptoms from VA emergency department and primary care clinical documentation. Int J Med Inform. 2012;81(3):143-56.
13. Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, et al. Using SNOMED CT to represent two interface terminologies. J Am Med Inform Assoc. 2009;16(1):81-8.
14. Elkin PL, Carter JS, Nabar M, Tuttle M, Lincoln M, Brown SH. Drug knowledge expressed as computable semantic triples. Stud Health Technol Inform. 2011;166:38-47.
15. Sinha S, Jensen M, Mullin S, Elkin PL. Safe Opioid Prescription: A SMART on FHIR Approach to Clinical Decision Support. Online J Public Health Inform. 2017;9(2):e193.
16. Schlegel DR, Crowner C, Lehoullier F, Elkin PL. HTP-NLP: A New NLP System for High Throughput Phenotyping. Stud Health Technol Inform. 2017;235:276-80.
17. Ceusters W, Elkin P, Smith B. Negative findings in electronic health records and biomedical ontologies: a realist approach. Int J Med Inform. 2007;76 Suppl 3:S326-33.
18. Beuscart-Zephir MC, Elkin P, Pelayo S. Human factors engineering for clinical applications. Stud Health Technol Inform. 2006;124:685-90.
19. Ammenwerth E, Talmon J, Ash JS, Bates DW, Beuscart-Zephir MC, Duhamel A, et al. Impact of CPOE on mortality rates--contradictory findings, important messages. Methods Inf Med. 2006;45(6):586-93.
20. Husser CS, Buchhalter JR, Raffo OS, Shabo A, Brown SH, Lee KE, et al. Standardization of microarray and pharmacogenomics data. Methods Mol Biol. 2006;316:111-57.
21. Pelayo S, Anceaux F, Rogalski J, Elkin P, Beuscart-Zephir MC. A comparison of the impact of CPOE implementation and organizational determinants on doctor-nurse communications and cooperation. Int J Med Inform. 2013;82(12):e321-30.
22. Elkin PL. Human Factors Engineering in HI: So What? Who Cares? and What's in It for You? Healthc Inform Res. 2012;18(4):237-41.
23. Beuscart-Zephir MC, Elkin P, Pelayo S, Beuscart R. The human factors engineering approach to biomedical informatics projects: state of the art, results, benefits and challenges. Yearb Med Inform. 2007:109-27.
24. Chopra G, Kaushik S, Elkin PL, Samudrala R. Combating Ebola with Repurposed Therapeutics Using the CANDO Platform. Molecules. 2016;21(12).