International Journal of Medical Informatics 86 (2016) 62–70

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.ijmijournal.com

Computer-assisted expert case definition in electronic health records

Alexander M. Walker [a,*], Xiaofeng Zhou [b], Ashwin N. Ananthakrishnan [c], Lisa S. Weiss [b], Rongjun Shen [b], Rachel E. Sobel [b], Andrew Bate [b], Robert F. Reynolds [b]

[a] WHISCON, Newton, MA 02466, USA
[b] Epidemiology, Research and Development, Worldwide Safety and Regulatory, Pfizer, New York, NY 10017, USA
[c] Division of Gastroenterology, Crohn's and Colitis Center, Massachusetts General Hospital, Boston, MA 02114, USA

ARTICLE INFO

Article history:
Received 19 March 2015
Received in revised form 13 October 2015
Accepted 15 October 2015

Keywords:
Health outcomes
Case definition
Electronic health records
Insurance claims data
Safety surveillance

ABSTRACT

Purpose: To describe how computer-assisted presentation of case data can lead experts to infer machine-implementable rules for case definition in electronic health records. As an illustration the technique has been applied to obtain a definition of acute liver dysfunction (ALD) in persons with inflammatory bowel disease (IBD).

Methods: The technique consists of repeatedly sampling new batches of case candidates from an enriched pool of persons meeting presumed minimal inclusion criteria, classifying the candidates by a machine-implementable candidate rule and by a human expert, and then updating the rule so that it captures new distinctions introduced by the expert. Iteration continues until an update results in an acceptably small number of changes to form a final case definition.

Results: The technique was applied to structured data and terms derived by natural language processing from text records in 29,336 adults with IBD. Over three rounds the technique led to rules with increasing predictive value, as the experts identified exceptions, and increasing sensitivity, as the experts identified missing inclusion criteria. In the final rule inclusion and exclusion terms were often keyed to an ALD onset date. When compared against clinical review in an independent test round, the derived final case definition had a sensitivity of 92% and a positive predictive value of 79%.

Conclusion: An iterative technique of machine-supported expert review can yield a case definition that accommodates available data, incorporates pre-existing medical knowledge, is transparent and is open to continuous improvement. The expert updates to rules may be informative in themselves. In this limited setting, the final case definition for ALD performed better than previous, published attempts using expert definitions.

1. Introduction

Expert clinical review has permitted researchers to bring subject-matter expertise to bear on the problem of disease definition in health records. The risks of unsupervised expert review are poor standardization and reproducibility, together with high cost and slow turnaround.

Algorithms for creating rules for case definition from clinical records vary in the degree to which they incorporate expert knowledge. Machine-learning algorithms may use encoded data to discriminate expert-adjudicated cases from noncases, and so rely on domain knowledge through the expert's initial creation of a "gold standard." The result can be effective discrimination, but with opaque decision rules. For hepatic disorders, Duh and colleagues demonstrated these characteristics in administrative HMO files and the corresponding clinical records using neural networks [1,2]. Algorithm-derived rules can be intelligible to the user and relatively simple to comprehend. Partitioning, for example, can produce decision trees that display intelligible rules, particularly when the classification applies to the information available at a single point in time [3]. Regression techniques can underlie data-driven algorithms for creating interpretable rules. Ananthakrishnan et al. for example included terms from natural language processing (NLP) for a regression analysis for identifying Crohn's disease in electronic outpatient records [4]. However derived, to the extent that classification methods are accessible to human interpretation, they can be checked against domain knowledge.

Carrell et al. have incorporated intelligibility of a candidate rule to examine the disagreements between a definition based on NLP extracts of medical texts and expert review of the same material. Their procedure for developing a case definition allowed modifications to a machine-derived rule so that it would replicate the

* Corresponding author at: WHISCON, 275 Grove Street, Newton, MA 02466, USA.
E-mail address: AlecWalker@WHISCON.com (A.M. Walker).

http://dx.doi.org/10.1016/j.ijmedinf.2015.10.005
1386-5056/

experts' judgment [5]. Such a cycle of development and expert checking can continue until the concordance between a machine-implementable rule and expert adjudication in new test cases reaches a pre-specified level or level of stability, or it may go on, in continuous adaptation.

Algorithms for deriving the rules for case definition can also employ medical expertise implicitly through the use of standard nosologies. Murff and colleagues for example found that rules that map terms extracted from patient notes into SNOMED (Systemized Nomenclature of Medicine; see http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html) could outperform expert-defined patient safety indicators for detecting adverse outcomes after surgery in the Veterans Affairs records [6].

Although rules for identifying health outcomes and their onset have often relied on already-encoded terms [7], the free text in medical records is now widely available, particularly at the level of single institutions.[6,8]. An early example of computerized review of text notes for drug safety surveillance relied on data generated in 1995–1996 at the Brigham and Women's Hospital in Boston, where Honigman and colleagues used text-mining to identify adverse drug reactions [9].

Clinical expert readers take the relative timing of elements of care as self-evident components of disease inference from electronic records. As an example, we found previously that the sequence "diagnosis of duodenal ulcer followed by endoscopy procedure followed by diagnosis of gastritis" in insurance data almost never corresponded to confirmed duodenal ulcer on medical record review, whereas a sequence that inverted the temporal order of the two diagnostic codes almost always indicated the presence of chart-confirmed duodenal ulcer [10]. This experience and other similar ones of trying to improve rules to define health outcomes suggested the wisdom of giving clinical experts a chance to identify and record patterns of care that might point to targeted health events [11–13].

The work that we report here is part of a larger research program, in which the overall goal was to assess the likely additional contribution of NLP-derived terms when structured data were already available. It was necessary to develop a definition that could include both NLP and structured terms, but for which neither the terms nor the structure would be specified *a priori*. The target clinical entity was acute liver dysfunction (ALD). Other than the two experts' consensus opinion, the study did not employ a gold standard of clinical truth.

We chose the setting of inflammatory bowel disease (IBD) because we have an interest in whether automated drug safety surveillance can be conducted in populations with IBD, and we chose liver dysfunction because it is a common complication of IBD and its treatments [14,15].

This report describes computer-assisted elicitation of rules for case definition from experts, using candidate cases drawn from medical records. The technique differs from machine-learning methods in that there is no encoded inference; there are only procedures for eliciting repeatable classifications from the experts. The technique differs from the expert consensus that has been the standard in the medical research using electronic health data. While consensus methods can be systematized, one very common approach involves teaming up persons with prior experience, domain expertise and knowledge of the data to create a definition that seems to them intuitively correct [16].

We use the example of liver dysfunction only to illustrate that definitions that emerge from a close clinical reading of the data can be easy to create and understand and may be highly predictive of true case status. They can be derived in a manner that adds consistency and formal structure to the expert rules.

2. Methods

2.1. Data

Humedica, a division of the health-data firm Optum, provided de-identified data for this study. Humedica offers services to large health care providers in the course of which it extracts data from clients' EHRs, insurance claims, prescribing records and practice management data.

From 2007 through 2012 there were approximately 70,000 patients with codes for IBD in the Humedica data. From the de-identified records of care, Humedica provided us structured data as well as NLP-derived data that were extracted from clinical notes. Humedica derives NLP items from text entries that correspond primarily to terms in two large dictionaries, SNOMED and MedDRA (Medical Dictionary for Regulatory Activities; see http://www.meddra.org/). Each NLP item consists of a concept such as "nausea" together with attributes derived from the immediate sentence context and from the location of the observation in the medical record. Under this extraction system the text "Patient complains of intermittent severe nausea for five days" might be recorded as entries of the form Nausea (Attribute), where the attributes are the intermittency, severity and duration of the complaint. Additional features capture the location in the record (e.g. the "Subjective" or "S" section of a SOAP note) and important contextual mentions such as family history or denial, should they be present. To guide the downloading of NLP terms into a set useful for the present purposes, we specified a range of clinical terms that might appear in medical records and that could be related to IBD, ALD, chronic liver disease or their causes.[1] The list was developed in discussion among the research team and with clinicians who had treated patients with IBD and ALD.

2.2. Human subjects

Humedica converted patient and provider identifiers to non-identifying study code numbers before providing us with files. The work did not require IRB approval under HIPAA as the Humedica data are not individually identifiable.

2.3. IBD study population

We identified persons whose records included codes for IBD on at least two different dates, who did not have codes for chronic liver disease (including primary sclerosing cholangitis, alcoholic and viral hepatitis, identified from an *a priori* list of codes), and who additionally met minimal utilization metrics. See Table 1 for details.

Eligible individuals entered analysis when at age 18 or older they had any evidence of a healthcare encounter, and they exited the study immediately after their last encounter.

2.3.1. Technique for developing a rule for case definition

The technique for obtaining a rule for case definition for ALD consisted of the following steps

(1) Create a machine-implementable candidate rule of an enriched population of candidate cases.

[1] We sought the following terms, with all their variants and associated attributes: abdomen, abuse, acholic, addiction, alcohol, anorexia, appetite, ascites, asterixis, cirrhosis, confusion, dark urine, dependence, encephalopathy, fatigue, fetor, hepatitis, icterus, itch, jaundice, liver, malaise, mentation, nausea, pruritus, transaminases and names of specific tests of liver function, infection or inflammation.