Journal of Biomedical Informatics 46 (2013) 795–804

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Developing an expert panel process to refine health outcome definitions in observational data

Brent I. Fox [a,*], Joshua C. Hollingsworth [a], Michael D. Gray [b], Michael L. Hollingsworth [c], Juan Gao [a], Richard A. Hansen [a]

[a] Auburn University, Harrison School of Pharmacy, Department of Pharmacy Care Systems, 020 Foy Hall, Auburn, AL 36849, USA
[b] Hewlett-Packard Laboratories, 1501 Page Mill Road, MS 1140, Palo Alto, CA 94304, USA
[c] Auburn University, Office of Information Technology, 300 Lem Morrison Drive, Auburn, AL 36849, USA

ARTICLE INFO

Article history:
Received 16 October 2012
Accepted 14 May 2013
Available online 13 June 2013

Keywords:
Administrative claims
Expert panel
Observational
Dashboard
Health outcome of interest

ABSTRACT

Objectives: Drug safety surveillance using observational data requires valid adverse event, or health outcome of interest (HOI) measurement. The objectives of this study were to develop a method to review HOI definitions in claims databases using (1) web-based digital tools to present de-identified patient data, (2) a systematic expert panel review process, and (3) a data collection process enabling analysis of concepts-of-interest that influence panelists' determination of HOI.

Methods: De-identified patient data were presented via an interactive web-based dashboard to enable case review and determine if specific HOIs were present or absent. Criteria for determining HOIs and their severity were provided to each panelist. Using a modified Delphi method, six panelist pairs independently reviewed approximately 200 cases across each of three HOIs (acute liver injury, acute kidney injury, and acute myocardial infarction) such that panelist pairs independently reviewed the same cases. Panelists completed an assessment within the dashboard for each case that included their assessment of the presence or absence of the HOI, HOI severity (if present), and data contributing to their decision. Discrepancies within panelist pairs were resolved during a consensus process.

Results: Dashboard development was iterative, focusing on data presentation and recording panelists' assessments. Panelists reported quickly learning how to use the dashboard. The assessment module was used consistently. The dashboard was reliable, enabling an efficient review process for panelists. Modifications were made to the dashboard and review process when necessary to facilitate case review. Our methods should be applied to other health outcomes of interest to further refine the dashboard and case review process.

Conclusion: The expert review process was effective and was supported by the web-based dashboard. Our methods for case review and classification can be applied to future methods for case identification in observational data sources.

1. Introduction

Adverse drug events (ADEs) continue to be a common problem leading to significant morbidity, mortality, and financial costs [1–7]. Serious limitations exist for two frequently used ADE detection methods: clinical trials and post-marketing surveillance. Clinical trials use narrowly defined populations that, while relevant to determining medication efficacy, are usually too small to identify ADEs to a meaningful extent. Post-marketing surveillance mechanisms rely on spontaneous ADE reporting by clinical researchers, health care professionals, and the public. Both methods provide limited value in identifying ADEs. Accordingly, the Institute of Medicine (IOM) called for systematic use of automated health care databases from a variety of settings to actively monitor drug safety and efficacy [8]. Congress subsequently mandated that the FDA collaborate with a variety of groups to implement the IOM's recommendation [9].

The FDA Sentinel Initiative was borne out of Congress' mandate, and is focused on monitoring medical products throughout their entire life cycle by using data from large, disparate electronic data sources [10]. To this end, researchers are exploring approaches and methods for the use of large observational data sources for active safety surveillance, including the Mini-Sentinel project [11] and the Observational Medical Outcomes Partnership (OMOP) [12]. These initiatives are founded on the belief that active surveillance can be performed using administrative claims and electronic health record (EHR) data. These observational data sources offer

* Corresponding author. Fax: +1 334 844 8307.
E-mail address: foxbren@auburn.edu (B.I. Fox).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.05.006

several advantages: (1) the ability to analyze data from multiple sources covering a large number of subjects, (2) the data are collected as a routine part of the provision of care and do not rely on the additional time and effort required to submit a spontaneous report, (3) the sheer volume of data provides access to millions of records, better reflecting actual medication-related usage than demonstrated in a clinical trial, and (4) the current national focus on EHR adoption [13] suggests the amount of available data will only increase.

While observational data offer new opportunities, their use is not without limitations. One particularly concerning challenge is the ability to accurately identify health outcomes of interest (HOIs). Administrative claims data, for example, are designed for reimbursement, not clinical care documentation. When conducting research with claims data, the lack of standards has resulted in outcome definitions that rely on billing codes such as ICD-9-CM diagnosis codes or CPT procedure codes which may not accurately reflect a patient's clinical status or care delivery because of problems including sloppy coding, up-coding (e.g., assigning procedural billing codes that commands higher reimbursement), or coding that reflects clinical work-up to rule out a diagnosis [14]. As is evident from the literature, when these codes are used in research, additional problems are introduced by the multitude of possible combinations of codes that are included to define an HOI [15]. For example, one study might define acute myocardial infarction as any ICD-9-CM diagnosis code beginning with 410, while another study might include only 410.2 and 410.4, and a third might require one of these eligible diagnosis codes plus a relevant coronary artery bypass graft procedures (e.g., CPT codes 33510-33536). Variability in which codes comprise definitions will impact measurement.

If observational data are to be used reliably for active drug safety surveillance, approaches to identifying HOIs must be improved. To first determine the variability across definitions, OMOP funded two independent systematic literature reviews to detail how studies of observational data defined 10 example HOIs [15]. They concluded that large variability exists in the literature, and no single definition of the HOIs they reviewed demonstrated clear superiority. As a result, a library of competing, and in most cases hierarchical, definitions for HOI measurement was developed [16]. Further research is needed to identify best practices for measuring HOIs in observational data.

We conducted methodological work to better understand how HOI definitions in the OMOP library compare, and to explore how these definitions might be refined. To do this, we developed a web-based dashboard to facilitate expert panel review of patient cases with competing HOI definitions. To ensure patient confidentiality, data needed to be under our control and securely accessed. Additionally, panelists needed to review the data from disparate locations with dynamic filtering and sorting capabilities. We also needed an efficient mechanism to collect panelists' assessments and opinions for subsequent analysis, as well as the ability to conduct mediated disagreement resolution sessions that allowed simultaneous review of panelists' evaluations and patient data. After considering our needs and available options, existing methods such as manual chart review and surveys were not deemed suitable.

Expert panelists were presented observational data from a sample of patient cases identified by competing HOI definitions, and were asked to provide opinions on whether they believed these patient data were consistent with having the HOI. Through dual, independent panelist review and a mediated consensus process for cases of disagreement, we were able to classify cases and create a modeling dataset for studying HOI measurement.

In this paper, we present our review process and the Web dashboard we created to efficiently present large volumes of data to panelists, capture panelists' assessments, and resolve disagreements. Specifically, we describe the development and implementation of: (1) a systematic case review and consensus-building process; (2) a Web-based dashboard to present de-identified claims and laboratory data; and (3) a process to collect case review data in a manner enabling identification and analysis of concepts of interest that influenced panelists' determination of presence/absence of specific HOIs.

2. Material and methods

2.1. Data sources and structure

The Truven MarketScan Lab Database (MSLR), licensed by OMOP within their research lab on the Amazon EC2 cluster, provided the observational data source for this study. These data represent inpatient, outpatient, and pharmacy claims from approximately 1.5 million privately insured patients from 2003 to 2007. The claims data are supplemented with laboratory results.

The MSLR data were transformed to the OMOP Common Data Model (CDM), which is a data transformation that allows consistent coding to be used across various types of observational data sources [17]. Data transformed to the CDM are amenable to use with publicly available OMOP tools for selecting cohorts (e.g., RICO [16]) and producing standardized summaries (e.g., OSCAR [18] and NATHAN [19]). Details on these tools can be found in the references cited. The CDM provides a mechanism for combining multiple occurrences of diagnosis codes and prescription drugs to define eras of conditions or drug treatments which simplifies the complex data presentation. For our data presentation, we relied on the following data tables from the CDM: condition eras (e.g., combining conditions with start and stop dates into eras), procedure occurrences, laboratory observations, visit occurrences, and drug eras (e.g., continuous periods of drug filling with start and stop dates).

2.2. Patient sample

Patients were sampled for competing HOI definitions using the HOIs of acute liver injury (4 of 7 definitions), acute kidney injury (4 of 4 definitions), and myocardial infarction (3 of 4 definitions) [16]. For each HOI, we sampled cohorts of patients using a modified version of OMOP's Person-level Exploratory Data Review of Outcomes (PEDRO). The PEDRO algorithm identified patients that met one or more of our HOI definitions by assigning an identifier (Concept ID) to the patient record. The HOI definition-specific Concept ID was attached to the patient's record with the first calendar date when all of the conditions for group inclusion for an HOI definition were met. This first occurrence date when a patient became part of the HOI cohort was recorded as the patient's index date (time zero). For purposes of strict data de-identification, all elements of data were anchored to this index date, with time preceding the index date representing negative values and time following the index date representing positive values.

Due to the hierarchical HOI definition relationships, the patient cohort extraction process yielded patients meeting more than one HOI definition option. For example, acute liver injury used a hierarchical definition where the broadest definition only includes diagnosis codes, a secondary definition includes both diagnosis and procedure codes, and a third definition includes diagnosis codes, procedures codes, and laboratory values. Patients meeting the definition that includes diagnosis, procedures, and laboratory values would also meet the previous two definitions. Because sample sizes were smaller for the most restrictive definitions and we did not want to inadvertently exclude patients meeting the most restrictive HOI definitions, we randomly sampled (without