Journal of Biomedical Informatics xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: A case study of azathioprine

Vishal N. Patel [a,b,*], David C. Kaelber [a,c,d,e,f]

[a] Center for Clinical Informatics Research and Education, The MetroHealth System, Case Western Reserve University, Cleveland, OH, United States
[b] Center for Proteomics and Bioinformatics, The MetroHealth System, Case Western Reserve University, Cleveland, OH, United States
[c] Departments of Information Services, The MetroHealth System, Case Western Reserve University, Cleveland, OH, United States
[d] Department of Internal Medicine, The MetroHealth System, Case Western Reserve University, Cleveland, OH, United States
[e] Department of Pediatrics, The MetroHealth System, Case Western Reserve University, Cleveland, OH, United States
[f] Departments of Epidemiology and Biostatistics, The MetroHealth System, Case Western Reserve University, Cleveland, OH, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Clinical research informatics
De-identified data
Systems biology
Postmarketing drug surveillance
Pharmacosurveillance
Azathioprine

ABSTRACT

Objective: To demonstrate the use of aggregated and de-identified electronic health record (EHR) data for multivariate post-marketing pharmacosurveillance in a case study of azathioprine (AZA).
Methods: Using aggregated, standardized, normalized, and de-identified, population-level data from the Explore platform (Explorys, Inc.) we searched over 10 million individuals, of which 14,580 were prescribed AZA based on RxNorm drug orders. Based on logical observation identifiers names and codes (LOINC) and vital sign data, we examined the following side effects: anemia, cell lysis, fever, hepatotoxicity, hypertension, nephrotoxicity, neutropenia, and neutrophilia. Patients prescribed AZA were compared to patients prescribed one of 11 other anti-rheumatologic drugs to determine the relative risk of side effect pairs.
Results: Compared to AZA case report trends, hepatotoxicity (marked by elevated transaminases or elevated bilirubin) did not occur as an isolated event more frequently in patients prescribed AZA than other anti-rheumatic agents. While neutropenia occurred in 24% of patients (RR 1.15, 95% CI 1.07–1.23), neutrophilia was also frequent (45%) and increased in patients prescribed AZA (RR 1.28, 95% CI 1.22–1.34). After constructing a pairwise side effect network, neutropenia had no dependencies. A reduced risk of neutropenia was found in patients with co-existing elevations in total bilirubin or liver transaminases, supporting classic clinical knowledge that agranulocytosis is a largely unpredictable phenomenon. Rounding errors propagated in the statistically de-identified datasets for cohorts as small as 40 patients only contributed marginally to the calculated risk.
Conclusion: Our work demonstrates that aggregated, standardized, normalized and de-identified population level EHR data can provide both sufficient insight and statistical power to detect potential patterns of medication side effect associations, serving as a multivariate and generalizable approach to post-marketing drug surveillance.

1. Introduction

Azathioprine (AZA), a purine analog widely used in solid organ transplants and autoimmune disorders, is known to induce a spectrum of toxicities. While bone marrow suppression [1,2] and hepatotoxicity [3] – thought to be dose-dependent effects due to the accumulation of 6-thioguanine metabolites – are the most widely recognized side effects, dermatologic [4] and renal effects [5] have also been reported. Predictions of adverse reactions can be guided by thiomercaptopurine methyltransferase (TPMT) activity [6,7], though TPMT activity has poor sensitivity as a test for toxicity [8]. As a supplement to measuring TPMT activity, case reports of AZA's toxicities help document the clinical course of side effects, yet the likelihood of a particular organ system's involvement is often established on an ad hoc basis from a survey of the literature. Consequently, side effects with a poor showing in the literature are presumed to be rare, though they may, in fact, be common. More generally, tracking side effects via case reports or reporting databases provides inaccurate and incomplete estimates of incidence.

While case reports and TPMT activity help to substratify patients into cohorts likely to develop adverse reactions, clinical

Abbreviations: AZA, Azathioprine; EHR, electronic health record.

* Corresponding author. Address: Center for Proteomics and Bioinformatics, 10900 Euclid Ave., BRB 9th Floor, Case Western Reserve University, Cleveland, OH 44106, United States.
E-mail address: vishal.patel2@case.edu (V.N. Patel).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.10.009

awareness can be further heightened by assembling the frequent sets of clinical patterns that arise in patients prescribed AZA. Studying multivariate side effects allows for a syndromic approach to identifying adverse drug reactions, which can point to clinical patterns for gauging a subgroup's risk of developing complications. While knowledge of disease presentations and syndromes was classically acquired through prolonged clinical observations, the emergence of electronic health data sources now allow for the possibility of mining such information through retrospective data analysis, based on data captured as part of routine clinical care. In particular, network models have proven useful in mining and organizing patterns of clinical phenomena, with recent applications in the assembly of disease comorbidity networks based on frequencies of ICD-9 (International Classification of Diseases, ninth revision) code co-occurrence [9,10]. These networks of side effects provide a global overview of disease-disease associations and may be useful for lifetime risk calculations, though they have limited utility in informing clinical practice as they lack temporal associations and specific clinical variables. Additionally, there may be significant issues with the accuracy of ICD-9 codes. In the area of pharmacosurveillance, research has classically focused on signal detection from voluntary reporting databases, which are limited by the biases inherent in voluntary submissions of adverse drug event [11]. Methodologically, pharmacosurveillance has largely focused on the extraction of bivariate – i.e. drug-event – signals [11,12], though studies have recently emerged that focus on detecting multivariate patterns through association rule mining [13,14].

In this work, we aim to (1) quantify the incidence of size effects associated with AZA, (2) compare the incidence of side effects associated with AZA to other similar drugs, and (3) combine network-based approaches with traditional pharmacosurveillance signal detection methods to discover patterns of multiple events associated with AZA that can be clinically informative. As opposed to discovering events in reporting databases, we used a database of aggregated, standardized, normalized, and de-identified population level EHR data from over 10 million patients. Rather than using ICD-9 codes, we use laboratory and vital sign measurements – reliable fields in most EHR systems [15] – to assess the presence of side effects. The use of laboratory and vital sign values from aggregated EHR data avoids the biases of spontaneous reporting systems and the problems with provider-entered ICD-9 codes, while simultaneously increasing our power to detect rare event associations.

2. Materials and methods

2.1. Database description

De-identified data was obtained using the Explore application of the Explorys platform (Explorys, Inc.), which places a health data gateway (HDG) server behind the firewall of each participating healthcare organization. After collecting data from a variety of health information systems – electronic health records (EHRs), billing systems, laboratory systems, etc. – the HDG maps the data to informatics ontologies, standardizing and normalizing measurements. Next, the data from each participating healthcare organization is passed into a data grid. A web application allows each healthcare organization to search and analyze the aggregated, standardized, normalized, and de-identified population level data.

At the time of this study, the aggregated data grid contained information on more than 10 million patients from multiple, distinct healthcare systems with different EHRs; the EHR serves as the primary medical record within participating institutions. All data used were de-identified to meet Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health (HITECH) Act standards. Therefore, this work was deemed not to be human studies research by the Institutional Review Board of the MetroHealth System. Business affiliation agreements were in place between all participating healthcare systems and Explorys Inc. regarding contribution of EHR data and the use of these de-identified data. Unified Medical Language System (UMLS) ontologies were used to map EHR data to facilitate searching and indexing. Diagnoses, findings, and procedures were mapped into the systematized nomenclature of medicine-clinical terms (SNOMED-CT) hierarchy [16]. Prescription medication orders were mapped to RxNorm [17]. Laboratory test observations were mapped to logical observation identifiers names and codes (LOINC), established by the Regenstrief Institute [18]. At the time of analysis, the application contained 14,580 records of patients who had ever been prescribed AZA.

2.2. Side effect network analysis

Our analysis focused on end-organ dysfunction known to be implicated with AZA, as well as non-specific side effects, such as hypertension, fever, and cell lysis. To study side effect patterns of AZA associated with particular organ systems, we used key reference ranges for lab values as proxies of organ function (Table 1) [19].

Patients were selected who had normal lab values within 90 days prior to being prescribed AZA, and an abnormal measurement within 90 days after being prescribed AZA. As the actual administration or consumption of a medication is often not recorded in EHRs, especially for outpatient medications, we used medication orders as a proxy for drug administration (i.e. the term "prescribed" refers to the date at which the order for "azathioprine" appeared in the patient's EHR). For neutropenia and neutrophilia, we extended the selection window to patients with a normal neutrophil count within 365 days prior to AZA prescription. To avoid inflating the significance of creatinine measurements due to subgroup effects from pre-existing renal dysfunction, we excluded all patients with an ICD-9 or American Medical Association Common Procedural Terminology (CPT) code mapped to any of the following SNOMED terms: renal impairment, renal failure syndrome, history of kidney transplant, or renal transplant (procedure).

2.3. Control cohort

Identification of patients suffering from various side effects is subject to the myriad biases of dealing with EHR data. For instance, an elevation in blood pressure observed after prescribing AZA may be due to external factors (e.g. environment, other drugs) or errors in measurement. Similarly, an elevation in liver enzymes may simply be a phenomenon associated with the underlying autoimmune disease rather than AZA. To address these issues and to calculate the significance of AZA-induced side effects, we assembled a control cohort of patients who experienced abnormal values when administered one of 11 other anti-rheumatic drugs (Table 2). The control drugs were identified by first selecting those drugs tagged with the SNOMED (Systemized Nomenclature of Medicine) code "anti-rheumatic agents." This unfiltered list of 42 drugs contains agents with both frequent and infrequent side effect profiles, and the inclusion of drugs with infrequent side effects would artificially inflate the significance of side effects associated with AZA. Apart from identifying statistically significant side effects, we also sought to identify clinically relevant side effects. In this regard, it was important to account for the prevalence of "common" side effects, e.g. headache, to judge the relevance of side effects associated with AZA. Thus, to avoid statistical bias and produce clinically relevant

Please cite this article in press as: Patel VN, Kaelber DC. Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: A case study of azathioprine. J Biomed Inform (2013), http://dx.doi.org/10.1016/j.jbi.2013.10.009