Vaccine 31S (2013) K41–K61

Contents lists available at ScienceDirect

Vaccine

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

Review

A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data

Cecilia P. Chung[a,*], Patricia Rohan[b], Shanthi Krishnaswami[c], Melissa L. McPheeters[d]

[a] Division of Rheumatology, Vanderbilt University School of Medicine, 1161 21st Avenue South, D-3100, Medical Center North, Nashville, TN 37232-2358, USA
[b] Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, Food and Drug Administration, WOC1 Building, Room 454S, 1401 Rockville Pike, Rockville, MD 20852-1428, USA
[c] Vanderbilt Evidence-based Practice Center, Vanderbilt University Medical Center, Suite 600, 2525 West End Avenue, Nashville, TN 37203-1738, USA
[d] Vanderbilt Evidence-based Practice Center and Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Suite 600, 2525 West End Avenue, Nashville, TN 37203-1738, USA

ARTICLE INFO

Article history:
Received 7 December 2012
Received in revised form 15 February 2013
Accepted 26 March 2013

Keywords:
Rheumatoid arthritis
Validation
Administrative database
ICD-9
Positive predictive value
Algorithm

ABSTRACT

Purpose: To review the evidence supporting the validity of billing, procedural, or diagnosis code, or pharmacy claim-based algorithms used to identify patients with rheumatoid arthritis (RA) in administrative and claim databases.
Methods: We searched the MEDLINE database from 1991 to September 2012 using controlled vocabulary and key terms related to RA and reference lists of included studies were searched. Two investigators independently assessed the full text of studies against pre-determined inclusion criteria and extracted the data. Data collected included participant and algorithm characteristics.
Results: Nine studies reported validation of computer algorithms based on International Classification of Diseases (ICD) codes with or without free-text, medication use, laboratory data and the need for a diagnosis by a rheumatologist. These studies yielded positive predictive values (PPV) ranging from 34 to 97% to identify patients with RA. Higher PPVs were obtained with the use of at least two ICD and/or procedure codes (ICD-9 code 714 and others), the requirement of a prescription of a medication used to treat RA, or requirement of participation of a rheumatologist in patient care. For example, the PPV increased from 66 to 97% when the use of disease-modifying antirheumatic drugs and the presence of a positive rheumatoid factor were required.
Conclusions: There have been substantial efforts to propose and validate algorithms to identify patients with RA in automated databases. Algorithms that include more than one code and incorporate medications or laboratory data and/or required a diagnosis by a rheumatologist may increase the PPV.

Contents

Abbreviations: A, algorithm; ACR, American College of Rheumatology; AMI, acute myocardial infarction; ANA, antinuclear antibody; anti-TNF, anti-tumor necrosis factor; anti-CCP, anti-cyclic citrullinated peptide; CI, confidence interval; CD, Crohn's disease; CPT, current procedural terminology; DB, database; DMARD, disease-modifying antirheumatic drug; DMBA, Deseret Mutual Benefits Administration; EDC, estimated date of conception; EMR, electronic medical records; GHS, Geisinger health system; HCPCS, Health Care Financing Administration Common Procedure Coding System; HZ, herpes zoster; IBD, inflammatory bowel disease; ICD, International Classification of Diseases; JRA, juvenile-onset rheumatoid arthritis; KPNC, Kaiser Permanente Northern California; MEDECHO, Maintenance et Exploitation des Donnees pour l'Etude de la Clientele Hospitaliere; MTX, methotrexate; NDC, National Drug Code; NJ, New Jersey; NMSC, non-melanoma skin cancer; NPV, negative predictive value; NSAID, non-steroidal anti-inflammatory drug; OSHPD, Office of Statewide Health Planning and Development; PA, Pennsylvania; PACE, Pennsylvania Assistance Contract for the Elderly; PPV, positive predictive value; PsA, psoriatic arthritis; RA, rheumatoid arthritis; RAMQ, Regie de l'assurance maladie du Quebec; RF, rheumatoid factor; ROC, receiver operating characteristic (curve area); RX, prescription; SLE, systemic lupus erythematosus; THR, total hip replacement; VA, Veterans Affairs/Administration; VAMC, Veterans Affairs Medical Center; VISN, Veterans Integrated Services Network.

* Corresponding author. Tel.: +1 615 322 4746; fax: +1 615 322 6248.
E-mail addresses: c.chung@vanderbilt.edu (C.P. Chung), shanthi.krishnaswami@vanderbilt.edu (S. Krishnaswami), melissa.mcpheeters@vanderbilt.edu (M.L. McPheeters).

0264-410X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.vaccine.2013.03.075

1. Introduction

Mini-Sentinel, a pilot project sponsored by the United States Food and Drug Administration (FDA), aims to inform and facilitate the development of an active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products [1]. Mini-Sentinel is one facet of the Sentinel Initiative, an FDA effort to develop a national electronic system that will complement existing methods of safety surveillance.

To support this goal, Mini-Sentinel uses administrative and claims data to examine relationships between medical product exposures and health outcomes [1,2]. This serves to refine safety signals and facilitate active surveillance of adverse events potentially related to medical products. A first step in developing the Sentinel system is to understand the validity of algorithms (i.e., combinations of billing, procedural, or diagnosis codes, or pharmacy claims) for identifying health outcomes of interest in administrative data. Mini-Sentinel program collaborators selected health outcomes of interest using an expert elicitation process through which investigators developed a list of candidate outcomes based on input from global vaccine safety experts. A panel of 5 vaccine experts then prioritized the list via an iterative process using criteria including clinical severity, public health importance, incidence, and relevance [2].

A relationship between vaccination and autoimmune diseases such as Guillain Barré, multiple sclerosis, and type 1 diabetes has been suggested primarily based on series of individual case safety reports [3]. In rheumatology, a case series of patients developing rheumatoid arthritis (RA) after hepatitis B vaccination [4], suggested that vaccines may precipitate rheumatic autoimmune diseases, although controversy remains since previous studies failed to confirm that association [5,6] and did not find evidence for other vaccines including tetanus and influenza [7].

RA is a common disease that affects 1% of the population [8]. Patients with RA die prematurely and are at increased risk of multiple comorbidities, including infections [9]. The last two decades have brought significant changes in the management of patients with RA. These are summarized in early treatment and tight control of inflammation with the use of traditional disease-modifying antirheumatic drugs (DMARDs) and/or new biologic agents. The goal is to achieve low disease activity or remission.

Current guidelines from the American College of Rheumatology (ACR) review the use of traditional DMARDs, biologic agents, monitoring for side effects, tuberculosis screening, and the need for vaccinations in patients starting or receiving DMARDs or biologic agents [10]. However, optimal clinical use of these drugs requires accurate determination of the risks associated with their use. In rheumatology, many studies are focused on the safety of traditional and biologic DMARDs. To facilitate these studies, investigators have developed algorithms to identify patients with RA. These strategies include using multiple diagnosis codes or sets of codes and medications to define the presence of a disease.

The goal of this project was to identify algorithms used to detect RA using administrative data sources and to describe the performance characteristics of these algorithms as reported by the studies in which they were used.

2. Materials and methods

A detailed description of the methods for the project can be found in the accompanying paper by McPheeters et al. [11]. Briefly, we searched the MEDLINE database via the PubMed interface using the strategies outlined in Appendix A. We also checked the reference lists of included studies for additional relevant citations. The search strategy was developed by building on prior Mini-Sentinel approaches to searching [12]. We expanded those approaches and tested the need to incorporate additional methods, including searches of Google Scholar. This last approach did not yield any relevant citations beyond the traditional MEDLINE search.

We limited searches to the last 21 years (1991 to September 2012) and required that included studies at the abstract review stage evaluate rheumatoid arthritis and use an administrative database reporting data from the United States or Canada. The first step required that two reviewers independently determine that an abstract did not meet criteria in order to exclude the study from further review.

Second, two investigators independently assessed the full text of those studies fulfilling the abstract review criteria, with disagreements resolved via a third investigator or discussion to reach consensus. We required that studies meet the abstract review criteria and also clearly define an algorithm to identify cases of RA. We tracked whether studies reported validation of the algorithm (e.g., via chart review or independent diagnosis).

One investigator extracted data regarding the study population, outcome studied, algorithms used, validation procedure, and validity statistics. A second investigator independently verified the accuracy of the data extracted. The first author independently extracted methodologic data including elements such as the population sampled and sampling methods, methods for locating cases, and methods for validating the accuracy of diagnoses in cases located to inform the writing of the report. We summarized results of studies qualitatively and report key characteristics below.

3. Results

We identified 1218 non-duplicate citations with potential relevance; of these, 580 required full-text review. Of these, 99 studies met our inclusion criteria, and of these, nine reported methods for confirming cases identified and reported the number of cases confirmed (Fig. 1). These studies are the focus of this report (Table 1). Table 2 provides definitions for each code used in these nine studies. The other studies meeting overall inclusion criteria provided algorithms but no discussion of confirmation or validation methods and are therefore summarized in Table 3. The studies describing case confirmation were conducted using a range of data sources, including the Veterans Affairs (VA) system, tertiary care hospitals, Medicare data, and insurance company databases.

Ng et al. assessed the validity of an algorithm for identifying RA cases using VA data from one hospital in Houston, Texas [13]. The investigators then confirmed potential cases using chart review of a random sample of records stratified by several criteria. To establish the initial dataset of potential cases, the investigators sought patients with at least two ICD-9 codes of 714 (no extensions