Vaccine 31S (2013) K2–K6

Contents lists available at ScienceDirect

Vaccine

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

Review

Methods for systematic reviews of administrative database studies capturing health outcomes of interest

Melissa L. McPheeters[a,b,*], Nila A. Sathe[b], Rebecca N. Jerome[c], Ryan M. Carnahan[d]

[a] Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Suite 600, 2525 West End Avenue, Nashville, TN 37203-1738, USA
[b] Vanderbilt Evidence-Based Practice Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Suite 600, 2525 West End Avenue, Nashville, TN 37203-1738, USA
[c] Eskind Biomedical Library and Department of Biomedical Informatics, Vanderbilt University Medical Center, 2209 Garland Avenue, Nashville, TN 37232, USA
[d] Department of Epidemiology, University of Iowa College of Public Health, S437 CPHB University of Iowa, 105 River Street, Iowa City, IA 52242, USA

ARTICLE INFO

Article history:
Received 3 December 2012
Received in revised form 8 June 2013
Accepted 17 June 2013

Keywords:
Administrative data
Sensitivity
Specificity
Sensitivity
Positive predictive value
International Classification of Diseases

ABSTRACT

This report provides an overview of methods used to conduct systematic reviews for the US Food and Drug Administration (FDA) Mini-Sentinel project, which is designed to inform the development of safety monitoring tools for FDA-regulated products including vaccines. The objective of these reviews was to summarize the literature describing algorithms (e.g., diagnosis or procedure codes) to identify health outcomes in administrative and claims data. A particular focus was the validity of the algorithms when compared to reference standards such as diagnoses in medical records. The overarching goal was to identify algorithms that can accurately identify the health outcomes for safety surveillance. We searched the MEDLINE database via PubMed and required dual review of full text articles and of data extracted from studies. We also extracted data on each study's methods for case validation. We reviewed over 5600 abstracts/full text studies across 15 health outcomes of interest. Nearly 260 studies met our initial criteria (conducted in the US or Canada, used an administrative database, reported case-finding algorithm). Few studies ($N = 45$), however, reported validation of case-finding algorithms (sensitivity, specificity, positive or negative predictive value). Among these, the most common approach to validation was to calculate positive predictive values, based on a review of medical records as the reference standard. Of the studies reporting validation, the ease with which a given clinical condition could be identified in administrative records varied substantially, both by the clinical condition and by other factors such as the clinical setting, which relates to the disease prevalence.

Contents

Abbreviations: CINAHL, Cumulative Index of Nursing and Allied Health; FDA, US Food and Drug Administration; ICD, International Classification of Diseases; *N*, number; NA, not applicable; NPV, negative predictive value; NR, value not reported or data needed to calculate value not reported; PPV, positive predictive value; RA, rheumatoid arthritis; Se, sensitivity; Sp, specificity; SLE, systemic lupus erythematosus.

* Corresponding author at: Vanderbilt Evidence-Based Practice Center, Women's Health Research, Institute for Medicine and Public Health, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 600, 6th Floor, Nashville, TN 37203-1738, USA. Tel.: +1 615 936 8317; fax: +1 615 936 8291.
E-mail addresses: melissa.mcpheeters@vanderbilt.edu (M.L. McPheeters), nila.sathe@vanderbilt.Edu (N.A. Sathe), rebecca.jerome@vanderbilt.edu (R.N. Jerome), ryan-carnahan@uiowa.edu (R.M. Carnahan).

0264-410X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.vaccine.2013.06.048

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, including vaccines. Mini-Sentinel is one facet of the Sentinel Initiative, an FDA effort to develop a national system using electronic healthcare data that will complement existing methods of safety surveillance. This system largely relies on administrative claims data.

In order to conduct vaccine safety research in administrative data effectively, methods for identifying events of interest need to be accurate. This may include using diagnosis or procedural codes – or combinations of codes – as indications that a clinical event has occurred. Therefore, this project aimed to identify existing studies in which specific codes or sets of codes typically used for administrative purposes (e.g., International Classification of Diseases, Ninth revision [ICD-9] codes for diagnoses or procedures) are able to capture clinical events (health outcomes of interest) accurately.

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 five vaccine experts then prioritized the list via an iterative process and using criteria including clinical severity, public health importance, incidence, and relevance [1,2].

Of the 23 health outcomes of interest selected for the initial list relevant to vaccine safety (Table 1), one had already been addressed by evidence reviews conducted for drug safety surveillance. Reports on some of these and other health outcomes of interest were published, along with a paper on the methods used to develop the prior reports [2]. We completed an additional 15 reports on the health outcomes of interest identified in Table 1. This paper outlines updates to the methods used to develop the previously published reviews for the Mini-Sentinel program [2] and discusses additional lessons learned in developing this latest round of reviews.

2. Overview of methods

2.1. Search strategy and resources

We sought to improve existing search strategies employed in prior Mini-Sentinel evidence reviews. The previous search strategy used combinations of controlled vocabulary terms and keywords

Table 1
Health outcomes of interest addressed.

ICD-9 disease group	Health outcome	New evidence review conducted?
Endocrine, nutritional, and metabolic		
	Type 1 diabetes	No; deferred due to projected volume of evidence
Blood		
	Idiopathic thrombocytopenic purpura	No; already well-characterized
	Henoch Schönlein purpura	Yes; completed
Mental		
	Tics	Yes; completed
Nervous system		
	Febrile seizures	No; completed previously for Mini-Sentinel drug safety planning
	Afebrile seizures	No; completed previously for Mini-Sentinel drug safety planning
	Guillain–Barré syndrome	No; already well-characterized
	Bell's palsy	Yes; completed
	Transverse myelitis	Yes; completed
	Acute disseminated encephalomyelitis	Yes; completed
	Optic neuritis	Yes; completed
	Uveitis	Yes; completed
	Brachial neuritis	Yes; no relevant studies identified
	Narcolepsy	Yes; no relevant studies identified
Circulatory system		
	Myocarditis and pericarditis	Yes; completed
	Kawasaki disease	Yes; completed
Respiratory system		
	Bronchospasm	Yes; completed
Digestive system		
	Intussusception	No; already well-characterized
Pregnancy		
	Spontaneous abortion and stillbirth	Yes; completed
Musculoskeletal system		
	Systemic lupus erythematosus	Yes; completed
	Rheumatoid arthritis and juvenile rheumatoid arthritis	Yes; completed
Congenital anomalies		
	Birth defects	No; deferred due to projected volume of evidence
Injury and poisoning		
	Anaphylactic shock (anaphylaxis) or acute systemic allergic reaction	No; completed previously for Mini-Sentinel drug safety planning[a]

Completed, review published as part of current supplement.

[a] Review published as part of *The U.S. Food and Drug Administration's Mini-Sentinel Program*. Pharmacoepidemiol Drug Saf 2012;21(Suppl 1):1–303. The review is also available at: http://mini-sentinel.org/methods/outcome_identification/default.aspx.

to identify the following core concepts: drug adverse events or other studies likely to contain validation of an outcome measure (example terms include "pharmaceutical preparations/adverse effects" and "Sensitivity and Specificity"); administrative or claims database studies (example terms include "insurance database" and "Medicare"); and the health outcome of interest (terms specific to each outcome).

To assess the feasibility of reusing the search strategies developed for previous reports [2], we examined citations meeting inclusion criteria for all previous project reports. We identified the citations in these reports that were only retrieved by handsearching or Google Scholar. We further assessed whether these items were available in PubMed but missed by the existing search strategy to explore whether modifications to the existing search strategy may have allowed identification of these items. This post hoc analysis revealed that most items would be accessible via PubMed if errors in subject term indexing that prevented their identification (e.g., very broad terms used, data terms omitted) were corrected. We also examined whether searching the Cumulative Index of Nursing and Allied Health (CINAHL) or EMBASE databases would identify useful citations; there was complete overlap with the PubMed retrieval, with these two databases yielding no unique relevant citations. We concluded that use of PubMed alone was appropriate and reasonable for our searches.

For each of the current topics, we designed keyword and subject search queries in PubMed for each of the conditions; which we then combined with the updated search strategy developed previously for the project. Table 2 provides the search strategy template. To identify the most recent literature on each condition; we further complemented this literature search by keyword searching of the not-yet-indexed portion of PubMed. In addition to searches of PubMed; we scanned the reference lists of included studies for potentially relevant citations.

2.2. Screening/inclusion and exclusion

We required that studies addressed the health outcome of interest (Table 1) and used an administrative database (i.e., databases including codified diagnostic or procedural data) containing data from individuals receiving health care in the United States or Canada. We restricted studies to those conducted in the US or Canada in order to align with methods used in prior Mini-Sentinel reviews, and because the results of the reviews needed to be applicable to the US health care system. We further required that studies clearly report the algorithm used to identify potential cases. We did not require that studies report validation of cases identified, although in areas where there was ample research, we prioritized studies with validation for our reporting.

We created databases using the search strategies described above and uploaded all results into the DistillerSR systematic review software. Two investigators independently reviewed the full text of each study against our inclusion/exclusion criteria. A senior epidemiologist was available to adjudicate discrepancies between reviewers.

One investigator also extracted data regarding the study using the evidence table template. The evidence table included data regarding the country of conduct and time period of a study; the data source (e.g., hospital system database, provincial health data); characteristics of the sample of cases identified from the population; the clinical event under study (i.e., incident or prevalent cases of the health outcome of interest); the algorithm used to locate cases; operational definition for a case; procedures used to validate potential cases (e.g., medical record review); and validation statistics as reported in the study (positive predictive value, negative predictive value, sensitivity, specificity). We extracted information about the diagnostic or procedural codes used in the algorithm as well as other parameters including age restrictions, timing of visits, concomitant prescriptions, etc. A second investigator independently verified the accuracy of the data extracted, with disagreements resolved through discussion to reach consensus.

Table 2
Search strategy template.

#1	Condition terms – controlled vocabulary terms and keywords (e.g., myocarditis[mh] OR myocarditis[tw] OR pericarditis[tw] OR pericarditis[mh] OR pleuropericarditis[tw] OR myocarditides[tw])
#2	("Diseases Category/epidemiology"[mh] OR "Validation Studies"[pt] OR "Validation Studies as Topic"[mh] OR "Sensitivity and Specificity"[mh] OR "Predictive Value of Tests"[mh] OR "Reproducibility of Results"[mh] OR "Predictive Value"[tw])
#3	("Outcome Assessment"[All] OR "insurance database"[All] OR "insurance databases"[All] OR "Data Warehouse"[All] OR "ICD-9"[All] OR "international statistical classification"[All] OR "international classification of diseases"[All] OR "ICD-10"[All] OR "Database Management Systems"[mh] OR "Medical Records Systems, Computerized"[mh] OR "CPT"[All] OR "Current procedural terminology"[All] OR "drug surveillance"[All] OR ("claims"[tw] AND "administrative"[tw]) OR ("data"[tw] AND "administrative"[tw]) OR "Databases, Factual"[mh] OR "Databases as topic"[mh] OR "Medical Record Linkage"[mh] OR "ICD-9-CM"[All] OR "ICD-10-CM"[All] OR "database" OR "registered persons database" OR "Medical Records Systems" OR "Population Surveillance" OR "Data Collection" or "Automatic Data Processing" OR "Incidence" [mh] OR "Medical Records" OR "Patient Discharge" OR "Hospital Records")
#4	("Editorial"[pt] OR "Meta-Analysis"[pt] OR "Comment"[pt] OR "case reports"[pt] OR "Review"[pt])
#5	#2 AND #3 NOT #4 AND eng[la] AND humans[mh] AND 1991:2012[dp]
#6	#1 AND #5

All, all fields; tw, textword; la, language; mh, medical subject heading; pt, publication type.

2.3. Analysis

One reviewer, typically the lead author of a review, extracted data on the study's reporting of methods used to identify and confirm cases. Table 3 outlines the questions used for this assessment. A second investigator reviewed the results of this "checklist" and commented on points as needed. The results of this assessment were used to inform the analysis of the studies described in each review. Where possible and not already reported, we also calculated 95% confidence intervals for performance characteristics (positive predictive value, etc.) included in the studies.

3. Results

Over 5600 abstracts/full text studies were reviewed across all conditions to identify 257 studies meeting our criteria. The number of studies available by clinical topic varied substantially, with no studies identified for narcolepsy or brachial neuritis and nearly 100 for rheumatoid arthritis (RA) or juvenile rheumatoid arthritis (Table 4). Of these studies, only 9 reported validation of the algorithm used. Similarly, a number of studies of systemic lupus erythematosus (SLE) have been conducted ($N = 50$) though again validation is reported in only 12. We also identified a number of administrative database studies addressing bronchospasm, which we defined broadly to include acute asthma exacerbation and wheezing. Among 38 studies meeting our initial criteria, only 2 reported validation of the case-finding algorithm.

The most common approach to validation was to calculate positive predictive values, based on a review of medical records as the gold standard. In the studies reporting validation, the ease with which a given clinical condition can be identified in administrative records varies substantially, both by the clinical condition and by

Table 3
Questions used to assess reporting of case confirmation methods.

Question	Considerations
Were incident or prevalent cases sought?	• Incident cases (initial diagnosis of condition) are stronger for surveillance
For studies that identified acute/incident cases, did the study identify a disease-free baseline period required for the diagnosis to be considered a new condition?	• Studies should specify the duration of time (e.g., 6 months, 12 months) prior to the diagnosis in which an individual could not have a diagnosis of the condition of interest in order for the individual to be considered an incident case
Were multiple codes used in the algorithm utilized to identify cases?	• The study should clearly indicate exactly which ICD, CPT, etc. codes were used, the data sources they were used to search, and, if codes were validated, exactly which ones • Use of multiple codes does not necessarily mean lower quality of evidence; in some cases, multiple codes may be appropriate
Were codes identified as primary or secondary diagnosis codes, or both?	Diagnostic setting may affect the generalizability/applicability of the codes
Is the sample representative of the larger population (e.g., mix of demographic factors, insurance status, hospital/practice types, etc.) or representative of a more narrowly defined group (e.g., all with specific risk factors, from one hospital or practice, etc.)?	• The study sample may have implications for results and applicability to surveillance activities. A description of the study sample demographics (e.g. age, gender, race breakdown), administrative data source, and inclusion/exclusion criteria should be provided • Narrowly defined populations should be explicitly noted and implications addressed
Was the data source the study used for validation clearly defined?	• The study should clearly indicate the data source (this is often medical records)
Did the study validate all cases or a random or convenience sample of cases?	• The study should indicate whether a sample was used, and if so, how the sample was selected. • Was the sampling frame appropriate?
Were the methods used to validate cases clearly described (medical record review including review of objective data such as lab findings, record review solely noting listing of the diagnosis in the chart, review by expert clinician, etc.)?	• The study should clearly indicate the data source • Ideally studies will use a rigorous validation technique to be certain cases accurately reflect diagnoses
How many records (or other validation source) were sought but could not be obtained?	• A high number of unobtainable records may indicate sampling problems or problems with the data source • Did the study make attempts to account for issues with the data source?
If multiple codes were used, did the study validate each one or the group of codes together?	• Ideally studies will validate each code individually in order to assess how well each code performed
Does the study report PPV, NPV, or sensitivity?	• Negative predictive value may provide a better indication of true cases

CPT, common procedural terminology; ICD, International Classification of Diseases; NPV, negative predictive value; PPV, positive predictive value.

other factors such as the clinical setting, which relates to the disease prevalence.

4. Discussion

The positive predictive value provides an estimate of the proportion of cases identified with the given algorithm that are true cases. It does not provide an assessment of cases missed, and most studies either did not have the population data or the resources to assess false negatives. Clinically, this means that the best algorithms identified in these studies will be designated as such because they identify few false positives; in other words, they do not overdiagnose cases. With the exception of a few conditions (e.g., SLE) there is little evidence available to identify algorithms that ensure that no cases are missed. One could assume that algorithms that are developed to be highly sensitive and thus less specific (i.e., have the highest rates of false positives) are least likely to have missed cases, at least in the event that the algorithm developed to enhance sensitivity also includes all elements of the more specific algorithm. However, the trade-off between sensitivity and specificity or positive predictive value needs to be empirically assessed. In the SLE example, an algorithm was identified that had a specificity of 72.5% but a sensitivity of 98.2%, indicating that one could be quite confident that few cases would be missed. There is a trade-off, however, between identifying all possible cases and ensuring that those identified are accurate – a second SLE study with a specificity of 99.9% had only 42% to 68% sensitivity. Thus, the choice of algorithm in practice will be highly dependent on the goals for its use.

The variation in detectability of health outcomes in administrative data likely has multiple reasons, including, for example, the reality that individuals with certain conditions such as early spontaneous abortion may not seek medical care and therefore may not be represented in administrative data. The clinical setting was influential – for example, in both the studies of RA and SLE, analyses in databases limited to rheumatology clinics produced higher PPVs than those in the general population. Given the higher prevalence of the clinical conditions in these clinics, this finding is not unexpected, but affects the applicability of the results to future research. Most vaccines are typically administered routinely to all patients in a given age group, and safety surveillance thus involves evaluation of large general populations. Investigators using algorithms tested in populations with specific clinical conditions cannot expect the same PPV or other performance characteristics when applying them in a general population. Similarly, several studies also found that the addition of a requirement around who made the diagnosis (e.g., that it be a rheumatologist) or around a combination of diagnostic and pharmacy data (e.g., for a disease modifying antirheumatic drug) could increase PPV.

In a number of cases, there is a clear need for a validation study to be conducted and the currently available literature is inadequate for identifying an acceptable algorithm. This includes, at a minimum, those conditions for which there were no or very few studies. For the other clinical conditions, where no data are available to calculate sensitivity, specificity or negative predictive value, additional studies should also be conducted. We suggest that, across the board, individuals wishing to apply the algorithms identified in these studies should carefully assess the applicability to their available data sources in order to have realistic expectations about the performance of the approaches.

Table 4
Overview of studies retrieved for each review.

Health outcome	Abstracts/full studies reviewed (*N*)	Studies meeting criteria (*N*)	Studies with validation described (*N*)	Range of PPV, NPV, Se, Sp (%)
ADEM	27	2	0	NA
Bell's Palsy	124	6	2	PPV: 81–84 NPV: NR Se, Sp: NR
Brachial neuritis	5	0	0	NA
Bronchospasm	677	38	2	PPV: 41.8–94 NPV: 65–75.9 Se: 27–56 Sp: 91.6–99
Henoch Schönlein purpura	55	1	0	NA
Kawasaki disease	175	22	6	PPV: 74–86 NPV: NR Se, Sp: NR
Myocarditis and pericarditis	196	9	4	PPV: 0 NPV: NR Se, Sp: NR
Narcolepsy	34	0	0	NA
Optic neuritis	76	2	1	NA[a]
Rheumatoid arthritis/Juvenile rheumatoid arthritis	1218	99	9	PPV: 2.5–97 NPV: 77–100 Se: 51–100 Sp: 55–97
SLE	658	50	12	PPV: 49–100 NPV: 99.1 Se: 11.8–98.2
Stillbirth and spontaneous abortion	1924	14	3	PPV: 99–100 NPV: NR Se: 38–86 (unweighted) Sp: NR
Tics	160	4	0	NA
Transverse myelitis	47	3	3	PPV: 62.1–75.7 NPV: NR Se: NR Sp: NR
Uveitis	300	7	3	PPV: 24–52.1 NPV: Se: NR Sp: NR

N, number; NA, not applicable; NPV, negative predictive value; NR, value not reported or data needed to calculate value not reported; PPV, positive predictive value; Se, sensitivity; Sp, specificity.

[a] These studies described using chart review to confirm cases, but none reported the number of cases actually confirmed.

Acknowledgments

The authors gratefully acknowledge the contributions and input of the Mini-Sentinel team and investigators.

Contributors: All authors declare that they have participated in (1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for important intellectual content, (3) final approval of the version submitted. *Conflict of interest statement*: The authors have no conflicts to declare. *Funding*: Mini-Sentinel is funded by the Food and Drug Administration (FDA) through Department of Health and Human Services (HHS) Contract Number HHSF223200910006I. The views expressed in this document do not necessarily reflect the official policies of the Department of Health and Human Services, nor does mention of trade names, commercial practices, or organizations imply endorsement by the US government. FDA staff reviewed articles prior to publication but had no role in study design or conduct.

References

[1] Lieu TA, Nguyen MD, Ball R, Martin DB. Health outcomes of interest for evaluation in the post-licensure rapid immunization safety monitoring program. Vaccine 2012;30(April (18)):2824–30.

[2] Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf 2012;21(January (Suppl 1)):82–9.