REVIEW ARTICLES

Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data

Eric I. Benchimol[a,b,c,d,e,f,g,*], Douglas G. Manuel[a,h,i,j,k], Teresa To[a,c,d], Anne M. Griffiths[b,e], Linda Rabeneck[a,d,l], Astrid Guttmann[a,d,e,m]

[a]The Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada
[b]Division of Gastroenterology, Hepatology and Nutrition, The Hospital for Sick Children, Toronto, Ontario, Canada
[c]Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada
[d]Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
[e]Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada
[f]Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada
[g]Department of Pediatrics, University of Ottawa, Ottawa, Ontario, Canada
[h]Department of Family Medicine, University of Ottawa, Ottawa, Ontario, Canada
[i]Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Ottawa, Canada
[j]Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[k]Statistics Canada, Ottawa, Ontario, Canada
[l]Department of Medicine, University of Toronto, Toronto, Ontario, Canada
[m]Division of Paediatric Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

Abstract

Background and Objectives: Validation of health administrative data for identifying patients with different health states (diseases and conditions) is a research priority, but no guidelines exist for ensuring quality. We created reporting guidelines for studies validating administrative data identification algorithms and used them to assess the quality of reporting of validation studies in the literature.

Methods: Using Standards for Reporting of Diagnostic accuracy (STARD) criteria as a guide, we created a 40-item checklist of items with which identification accuracy studies should be reported. A systematic review identified studies that validated identification algorithms using administrative data. We used the checklist to assess the quality of reporting.

Results: In 271 included articles, goals and data sources were well reported but few reported four or more statistical estimates of accuracy (36.9%). In 65.9% of studies reporting positive predictive value (PPV)/negative predictive value (NPV), the prevalence of disease in the validation cohort was higher than in the administrative data, potentially falsely elevating predictive values. Subgroup accuracy (53.1%) and 95% confidence intervals for accuracy measures (35.8%) were also underreported.

Conclusions: The quality of studies validating health states in the administrative data varies, with significant deficits in reporting of markers of diagnostic accuracy, including the appropriate estimation of PPV and NPV. These omissions could lead to misclassification bias and incorrect estimation of incidence and health services utilization rates. Use of a reporting checklist, such as the one created for this study by modifying the STARD criteria, could improve the quality of reporting of validation studies, allowing for accurate application of algorithms, and interpretation of research using health administrative data. © 2011 Elsevier Inc. All rights reserved.

Keywords: Health administrative data; Misclassification bias; Diagnostic accuracy; Sensitivity and specificity; Predictive values; Health services research; Epidemiology

1. Introduction

Health services and epidemiologic research are best conducted with population-level data. This helps ensure the appropriate estimation of incidence and prevalence rates, the minimization of referral bias, and the overall generalizability of the study conclusions to the population of interest. Because prospective clinical registries and retrospective

What is new?

- Significant deficits exist in the validation and reporting of algorithms used to identify patients within health administrative data.

- Misclassification error represents an important form of bias in research using health administrative databases.

- The modified Standards for Reporting of Diagnostic accuracy criteria reported here can be used to improve the quality of reporting in studies of health state (disease or conditions) identification validation.

- Future efforts should address criteria for conduct and reporting of research using health administrative data.

chart review comprising a representative sample or all residents of a jurisdiction are impractical, health administrative data are an alternative for population-based chronic disease surveillance, outcomes research, and health services research. Health administrative data are defined as information passively collected, often by government and health care providers, for the purpose of managing the health care of patients [1], and are a subtype of automated health care data [2]. Examples include physician billing databases (such as those managed by government in single-payer health systems or by health maintenance organizations [HMOs]), and hospital discharge record databases. Accuracy of the diagnostic codes used to identify patients within these data depends on multiple factors including database quality, the specific condition being identified, and the validity of the codes in the patient group. A large gradient in data quality exists, with some databases being of higher quality than others [3]. Isolated diagnostic codes associated with physician billing records have been shown to be accurate to identify patients with some chronic diseases [4,5] but not others [6−9]. Since chronic diseases usually require multiple contacts with the health system to diagnose, a single-visit diagnostic code is often insufficient to accurately identify patients with the disease. The validity of codes is also dependent on the patient group being studied. For instance, the accuracy of diagnostic codes or combinations of codes (algorithms) varies across age groups because of variable use of the health system [6,7,10]. As such, validation of algorithms used to identify patients with different health states (including acute conditions, chronic diseases, and other health outcomes) is essential to avoid misclassification bias [11], which may threaten the internal validity and interpretation of study conclusions. For example, assessment of health services utilization in a cohort of patients with a chronic disease contaminated by large

number of healthy residents falsely labeled as having a chronic disease would underestimate the burden of the disease on the health system or the quality and performance of the health system. Similarly, assessment of incidence of the disease in the cohort would overestimate risk to the population. Although the validation of administrative data coding has been identified as a priority in the health services research by an international consortium [3], the complete and accurate reporting of algorithm validation research is equally important to appropriate application. The growing availability of administrative data for research coupled with the expense, privacy concerns, and complex methodologies required to validate identification algorithms have resulted in algorithms being applied to these databases by researchers not involved in their initial validation. As such, minimum quality criteria for the conduct and reporting of algorithm validation studies would benefit scientists using these algorithms and consumers of the research on which these algorithms rely.

The purpose of this study was to appraise all studies that validated algorithms to identify patients with different health states within the administrative data with newly developed consensus criteria for the reporting of studies that validate health administrative data algorithms, based on the Standards for Reporting of Diagnostic accuracy (STARD) initiative [12]. In so doing, we aimed to identify strengths and weaknesses in the methods of such validation studies to improve the future reporting of research using health administrative data.

2. Methods

2.1. Development of validation study quality checklist

Algorithms to identify patients with different health states may be considered a type of diagnostic test applied to health administrative data, and markers of diagnostic accuracy are often reported in studies validating algorithms against reference standards. As such, we modified the criteria published by the STARD initiative for the accurate reporting of studies using diagnostic tests [12] to evaluate included studies. Four experts (E.I.B., D.M., T.T., and A.G.) in research using these data modified the STARD criteria using an iterative approach to create a 40-point data collection tool for evaluation of studies (Table 1). A fillable version of the data collection tools is available in Supplemental Data 1 and can be used to assess the algorithm validation literature for completeness of reporting.

2.2. Systematic review: search strategy and selection criteria

An online database literature search was performed for human studies, without language restrictions, using the following databases: MEDLINE (National Library of Medicine [NLM], Bethesda; January 1950 to June 2009) and