Journal of Clinical Epidemiology 84 (2017) 114–120

Journal of Clinical Epidemiology

Bootstrap imputation with a disease probability model minimized bias from misclassification due to administrative database codes

Carl van Walraven*

Epidemiology & Community Medicine, University of Ottawa
Ottawa Hospital Research Institute, ASB1-003, 1053 Carling Ave., Ottawa, Ontario K1Y 4E9, Canada
Institute for Clinical Evaluative Sciences

Accepted 27 January 2017; Published online 4 February 2017

Abstract

Objective: Diagnostic codes used in administrative databases cause bias due to misclassification of patient disease status. It is unclear which methods minimize this bias.

Study Design and Setting: Serum creatinine measures were used to determine severe renal failure status in 50,074 hospitalized patients. The true prevalence of severe renal failure and its association with covariates were measured. These were compared to results for which renal failure status was determined using surrogate measures including the following: (1) diagnostic codes; (2) categorization of probability estimates of renal failure determined from a previously validated model; or (3) bootstrap methods imputation of disease status using model-derived probability estimates.

Results: Bias in estimates of severe renal failure prevalence and its association with covariates were minimal when bootstrap methods were used to impute renal failure status from model-based probability estimates. In contrast, biases were extensive when renal failure status was determined using codes or methods in which model-based condition probability was categorized.

Conclusion: Bias due to misclassification from inaccurate diagnostic codes can be minimized using bootstrap methods to impute condition status using multivariable model–derived probability estimates.

Keywords: Misclassification bias; Information bias; Observation bias; Bootstrap; Categorization; Health administrative data

1. Introduction

Bias in clinical research can arise from many different sources [1]. The ultimate goal of all research is to minimize bias so that the results of a particular study most closely represent the truth.

An important type of bias (which has been variously termed misclassification bias [2,3], information bias [4], and observation bias [4]) results from assigning incorrect values to exposures, covariates, or outcomes. This bias is prominent in administrative database research in which a large majority of studies use diagnostic or procedural codes to determine the presence or absence of categorical variables [5]. Because these codes never perfectly reflect the entities they replace, studies using diagnostic and procedural codes to determine exposure, covariate, or outcome status will always have some degree of misclassification bias.

Several studies have tried to minimize this bias using multivariate models to determine disease status. In this method, multiple variables within a health administrative database are used to predict the probability that the condition of interest truly exists for a particular patient. This probability provides more information about condition status than the mere presence or absence of particular diagnostic codes. However, to proceed with their analyses, researchers must choose a threshold to classify a patient's condition status based on the model-derived probability estimate. In the published literature, these thresholds can be selected in several different ways including a 50% threshold [6], based on a receiver operating characteristic (ROC) curve analysis [7,8], using a ''double-threshold'' analysis [9], or with multiple imputation techniques [8,10].

No published study has directly quantified the amount of misclassification bias using these different methods to determine disease status. This study compared the amount of bias associated with determining disease status using

Funding: C.v.W. is supported by a University of Ottawa Department of Medicine Clinician Scientist Chair.

* Corresponding author. Tel.: +1-613-761-4903; fax: +1-613-761-5492.

E-mail address: carlv@ohri.ca

http://dx.doi.org/10.1016/j.jclinepi.2017.01.007
0895-4356/

What is new?

- Using diagnostic codes to define patient cohorts will bias results because of misclassification.
- Bias in the measurement of disease prevalence or its association with covariates was minimized when disease status was imputed using a disease-probability model and bootstrap methods.
- Misclassification bias due to diagnostic codes can be minimized using bootstrap methods to impute disease status from an accurate disease probability model.

diagnostic codes, a model-derived disease probability estimate that was categorized using a variety of methods, or the imputation of disease status using bootstrap techniques and a model-derived disease probability estimate.

2. Methods

2.1. Determining true severe renal failure (i.e., gold standard) status

This study used data from a previously conducted analysis of 100,000 randomly selected adults (derivation cohort n = 49,926, validation cohort n = 50,074) hospitalized between 2002 and 2008 at a teaching hospital in Ottawa, Canada [11]. In that study, renal function was measured in each person using the abbreviated Modified Diet in Renal Disease formula to estimate the glomerular filtration rate (GFR) using each in-hospital serum creatinine [12]. Patients were classified with severe renal failure if they had two or more consecutive GFRs less than 30 mL/min/1.73 m^2. This is the definition used by the National Kidney Foundation Kidney Disease Outcomes Quality Initiative guidelines to define chronic kidney disease [13,14]. In the present study, this classification determined the ''true'' value (or the ''gold-standard'' value) for severe renal failure status to which all other measures were compared.

Patients with only one serum creatinine measure during their admission were classified with severe kidney disease if its GFR was less than 30 mL/min/1.73 m^2. This was done to avoid incorrectly classifying patients with only one serum creatinine during their admission without severe renal failure (even if that measure was severely elevated). We reasoned that the risk of misclassification was lower if we classified such patients as having severe renal failure rather than without. All other patients, including those without any creatinine measures, were classified without severe kidney disease.

2.2. Severe renal failure model

This model was derived in the previous study on 49,926 randomly selected people and internally validated on the remaining 50,074 [11]. Candidate variables were created using health administrative data retrieved from the discharge abstract database. These variables included the following: patient factors (age, sex, and all Elixhauser comorbidities using International Classification of Disease [ICD] codes cited by Quan [15]); hospitalization factors (admission urgency, admitting service, intensive care unit treatment, surgical procedures, hospital survival status, and length of stay); and renal failure-specific codes (dialysis-related diagnoses and procedures, the most common acute diagnoses causing renal dysfunction, and manifestations of renal dysfunction). Fractional polynomials were used to identify best transformations for noncategorical covariates. In the validation population, the model was highly discriminative (c-statistic 0.937) and well calibrated (Hosmer–Lemeshow statistic 48.5). For each patient, this model used values for the covariates in the model to return the probability that the patient had severe renal failure.

2.3. Surrogate measures of severe renal failure status

ICD-10 codes were used as surrogate indicators of severe renal failure status in the validation cohort (n = 50,174). Patients were coded with severe renal failure if they were assigned any of the codes listed in Appendix A at www.jclinepi.com. Using true severe renal failure (defined previously) as the gold standard, these codes had a sensitivity of 71.0%, specificity of 96.3%, and a positive predictive value of 60.1%.

Surrogate severe renal failure status was also determined using probability estimates generated by the severe renal failure model. Three methods in the literature have been used to transform such probability estimates to a categorical condition status including the following:

a) Patients with a predicted probability of severe renal failure of 0.5 or greater were classified with severe renal failure [6], and all other patients were classified without severe renal failure (this was termed the ''≥50% method'').
b) An ROC analysis [16] was conducted to identify the predicted probability with the minimal linear distance to perfect accuracy; all patients whose predicted probability exceeded this threshold were deemed to have severe renal failure [7,8], and all other patients were deemed to not have severe renal failure (this was termed the ''ROC method'').
c) A ''double-threshold'' analysis which limits the analysis to patients with more extreme estimated disease probabilities (with the assumption being that such patients are more likely to have or not have the disease, thereby producing a study cohort with more accurate disease classification [9]). The ''double-