Misclassification in Administrative Claims Data: Quantifying the Impact on Treatment Effect Estimates

Michele Jonsson Funk · Suzanne N. Landi

Published online: 15 October 2014
© Springer International Publishing AG 2014

Abstract Misclassification is present in nearly every epidemiologic study, yet is rarely quantified in analysis in favor of a focus on random error. In this review, we discuss past and present wisdom on misclassification and what measures should be taken to quantify this influential bias, with a focus on bias in pharmacoepidemiologic studies. To date, pharmacoepidemiology primarily uses data obtained from administrative claims, a rich source of prescription data, but susceptible to bias from unobservable factors including medication sample use, medications filled but not taken, health conditions that are not reported in the administrative billing data, and inadequate capture of confounders. Because of the increasing focus on comparative effectiveness research, we provide a discussion of misclassification in the context of an active comparator, including a demonstration of treatment effects biased away from the null in the presence of nondifferential misclassification. Finally, we highlight recently developed methods to quantify bias and offer these methods as potential options for strengthening the validity and quantifying uncertainty of results obtained from pharmacoepidemiologic research.

Keywords Misclassification · Measurement error · Administrative claims · Probabilistic bias analysis · Bayesian bias analysis · Modified maximum likelihood · Multiple imputation for measurement error · Regression calibration · Propensity score calibration · Comparative effectiveness · Pharmacoepidemiology · New user design

Introduction

Administrative claims data offer various advantages for pharmacoepidemiologic research, but limitations are usually acknowledged rather than quantified. We review key findings from simulation studies regarding bias due to misclassification, common sources, and types of misclassification in claims data including recent findings that relate to these, and methods for minimizing and quantifying the impact of misclassification on effect estimates and associated confidence intervals.

Brief Overview of Bias Due to Misclassification

Misclassified Exposures

When an effect exists, simulation studies have shown that absolutely nondifferential misclassification of a binary exposure will, on average, bias the estimate of effect towards the null [1]. Researchers often rely on this knowledge and make the assumption that the observed effect is, at worst, an underestimate of the true effect given nondifferential misclassification of the exposure. The exceptions to this rule of thumb are less well known. If the exposure is polytomous (i.e., is categorized in more than two levels), the bias from nondifferential misclassification could be in either direction [2]. Nor is bias toward the null guaranteed if the misclassification is related to errors in other variables or if there are other sources of systematic error (e.g., selection bias or confounding) in exposure classification [3]. In the unlikely event that each level of exposure is misclassified beyond what would be expected if exposure had been randomly assigned, the bias could remove or even reverse (i.e., beyond the null) the appearance of an association. Of course, exposure misclassification that differs by outcome status (differential) can bias in any direction.

M. Jonsson Funk (✉) · S. N. Landi
Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, CB 7521, Chapel Hill, NC 27599-7521, USA
e-mail: mfunk@unc.edu

Misclassified Outcomes

In most cases, nondifferential misclassification of a binary outcome will also result in bias toward the null. However, there are some exceptions to this general rule where disease misclassification is not expected to produce bias. Given perfect specificity (i.e., no false positives) and nondifferential misclassification of disease due only to low sensitivity, the risk ratio will be unbiased in expectation; however, the risk difference will be biased toward the null to the degree of the sensitivity (e.g., if the sensitivity is 80 %, the expected risk difference value will be 80 % of the true risk difference). In this situation, odds ratios and rate ratios will also be biased toward the null. If the outcome is rare in all exposure groups, the odds ratio will approximate the risk ratio and the estimate will be unbiased [3]; similarly, if the impact of misclassification on person-time is negligible, the rate ratio will be unbiased [4].

Misclassified Confounders

Misclassification of a confounder leads to an estimated measure of effect that is biased toward the crude measure, if the misclassification is nondifferential with regard to both the exposure and the outcome and there are no other sources of error. The direction of the bias or “residual confounding” can be either away or toward the null [2]. For a binary confounder, the independent nondifferential misclassification will cause a bias in the direction of the confounding because the ability to control fully for the confounder has been reduced, resulting in a partially controlled result that lies between the crude and true values. This partial control is limited to situations where there is no qualitative interaction between the exposure and the confounder [5]. The most extreme effects of confounder misclassification are observed when the study exposure variable is weakly associated with the outcome in comparison to a stronger confounder [6, 7].

Expected Bias vs. Bias of a Single Study

The conditions under which the direction of any given bias is predictable are limited and, in practice, are impossible to guarantee. Investigators need to be aware that conventional wisdom regarding the expected bias due to misclassification in various scenarios reflects an *average* estimated across multiple study repetitions; thus, an estimate from a single study may not follow the direction of bias according to these rules [1, 8]. In a simulation study of nondifferential misclassification, the mean result across many trials was biased toward the null, as expected, but the estimates from the individual trials were biased both away from and toward the null [1]. This highlights the importance of quantifying the impact of

misclassification in each study rather than relying on the expected direction of bias.

Prescription Medications

Arguably, one of the strengths of using administrative claims to evaluate medication effects is the relatively complete nature of data regarding prescription fills. These data derive from insurance claims for medications that are filled by the patient at a community-based pharmacy. These data are generally superior to self-reported medication use (which is susceptible to recall bias) [9, 10]. In some cases, these data are also more accurate than records of physician-ordered prescriptions (which may include medications that are never obtained by the patient) [11, 12]. Nonetheless, there are a variety of circumstances in which these pharmacy claims may not reflect the actual medication exposure of patients.

Non-Users Misclassified as Users

This type of misclassified exposure status includes patients with prescriptions that are filled but never taken, those initially taken and then discontinued, and those taken PRN (as needed) or intermittently. One common approach to minimizing the impact of these misclassified individuals is to require evidence of a second prescription fill within a fixed period of time to increase the likelihood that patients are actually taking the medication [13, 14]. This necessitates starting follow-up at the second fill to avoid introducing immortal time, thus limiting the ability to study short-term effects [15]. We discuss the implications of imperfect identification of *when* medications are started and stopped in the section on misclassified duration of use below.

Users Misclassified as Non-Users

In the setting of administrative claims data, this type of misclassification occurs when patients pay for prescription medications out of pocket (including \$4 generics [16–19]), receive samples [20•], or are hospitalized (as inpatient medications are typically included in the bundled payment). For administrative databases that include only those medications on a formulary (such as in Canada), there is also potentially important misclassification of exposure to specific medications in a class that are not included on the formulary. A recent study set in Canada noted a dramatic increase in the number of reported prescriptions for thiazolidinediones (TZDs) which corresponded to a change in policy providing for an automated prior-authorization process for this diabetes medication, suggesting that perhaps 20 % of prior TZD exposure was misclassified as non-use prior to the policy change [21]. There

are also instances in which medications are available both with and without a prescription (e.g. analgesics, proton-pump inhibitors, antihistamines) [22]. Patients who obtain these medications over-the-counter would also be misclassified as non-users according to the insurance claims data.

The scenarios in which differential misclassification would affect users of a medication are less clear, although we can imagine that, e.g., in the US Medicare data, individuals who have more complicated medical conditions are more likely to enter the ‘donut hole’ when they become responsible for all prescription costs. These individuals would be at greater risk of experiencing outcomes such as mortality and hospitalization, and would also be more likely to obtain a prescription from a \$4 generic list and pay out of pocket if they did not expect to accrue sufficient additional prescription costs during the remainder of the benefit year to qualify for catastrophic coverage. Thus, the sensitivity with which truly exposed individuals would be correctly classified as exposed might differ by outcome status.

Duration of Use Misclassified

Misclassification of the timing of an event – new use of a therapy, or the occurrence of the outcome – has received little attention, perhaps because much of the literature on misclassification deals with settings in which the data can be represented in the form of a 2×2 table. But it is the rare analysis in pharmacoepidemiology that conforms to this structure. More often, the timing of exposures, outcomes, and covariates are complex and the analyst must carefully evaluate the sequencing of these to assure that the proper temporal relationships between them are maintained.

The new user design is intended to synchronize patients with respect to the duration of the treatment and ensure that covariates have not been affected by prior exposure [23]. In the case where patients receive free medication samples from the prescribing physician for a new prescription, there is an initial period of use that is misclassified as preceding initiation using claims data. Among patients who go on to file a prescription claim, there is selection bias [24] (non-responders and those with early adverse effects would be less likely to fill) and a difference in the duration of exposure compared to true new users. In one recent paper, 13.4 % of ‘new’ users of a branded statin had lab values suggesting that they were prevalent users, whereas there was no indication that those identified as new users of generic statins were prevalent users [20••].

A similar scenario could occur in which patients paid out of pocket for medication while awaiting special authorization. In a simulation study of this form of misclassified person-time, Gamble et al. found substantial bias of the hazard ratio for mortality for new users of metformin versus sulfonylureas as

an increasing proportion of users of metformin were misclassified as non-users while awaiting special authorization [25]. There is also evidence suggesting that the days-supply associated with the prescription, which is used to determine periods of continuous use and discontinuation (for an as-treated analysis) is not uniformly recorded. In Ontario, investigators studied the pattern of days-supply for osteoporosis medications and found that those filled for patients in a long-term care facility were substantially shorter than those for community-dwelling patients [26].

In addition, the effects of the treatment often vary considerably with time [27], and by misidentifying the start of therapy the comparison groups may not reflect the same duration of treatment. These problems would be most pronounced when the duration of follow-up is relatively short, the hazard is not constant, and the extent of misclassification differs between the groups. While within-subject study designs such as the case-crossover [28] and self-controlled case series [29, 30] minimize bias due to confounding by time-invariant characteristics and comorbidities, they remain susceptible to bias due to misclassification of exposure.

Clinical Outcomes

The ability to study rare clinical outcomes in a very large, population-based sample is a potential strength of claims data, but likewise a source of concern due to potential misclassification. Outcomes such as death are considered reliable in some data sources while they are only observed when they occur in the hospital in other data sources [31]. Clinical events may be acute and result in hospitalization (such as hip fracture) or chronic with or without specific clinical interventions (such as type II diabetes). The degree of misclassification of these outcomes can vary considerably.

Medical procedures are considered reliable in billing data given the close relationship and regulated nature of billings for procedures and physician payment. ICD-9 procedure codes, used by hospitals to bill for the facility component of charges, are not sufficiently specific in many instances, while CPT codes, used by physicians to bill for their services, are more specific.

The importance of validating clinical outcomes has been appreciated since the early days of studies conducted using health care claims data as a means by which to assure that a highly specific outcome definition was devised. These are routinely based on an algorithm that may include multiple instances of a given diagnosis code on unique service dates, prescription fills, claims from an inpatient setting, and/or specific procedures to maximize the specificity of the outcome [32••].

Misclassification of outcomes can occur differentially by exposure status by virtue of the fact that individuals who receive a prescription may receive more intensive health screenings and monitoring than patients who are not receiving medication or are receiving a different medication. These might include differences in health-seeking behaviors (screening or diagnostic workups for suspected health problems) [33], more frequent lab testing for potential liver or kidney damage if the medication is suspected to increase risk [34], or use of follow-up colonoscopies after selected types of radiation [35, 36]. This would decrease the proportion of individuals who have the outcome who are incorrectly classified as unaffected among patients with the exposure.

Confounders

Misclassification in the setting of claims data is a significant concern in light of the fact that the absence of a diagnosis or related procedures in claims during a specified time period is taken to indicate the absence of the condition. Patients who do not have healthcare encounters will not generate evidence of their conditions, and those with significant co-morbidity may not have evidence of common, less serious conditions (such as hypertension) when they are under active treatment (CABG) [37]. Typically, studies using insurance claims data define a baseline period during which individuals must be continuously enrolled [38]. But a recent simulation study suggested that using all available data to define confounders may better control confounding than restricting to a uniform time period [39•]. The robustness of this finding under a variety of conditions is still being established, but it serves as a challenge to reconsider the status quo.

The Special Case of Comparative Effectiveness Research (CER)

As questions were being raised about the use of placebo-controlled trials when effective treatment alternatives were available [40], so did pharmacoepidemiologists begin to recognize the value of active comparators in the setting of non-experimental research on medication safety and effectiveness. The comparison of two active agents has made pharmacoepidemiologic studies less susceptible to biases due to confounding by indication, healthy user bias, confounding by frailty, and other sources of unmeasured confounding [41]. In addition, biases due to misclassification of confounders and outcomes (described above) are likely less pronounced with an active comparator. That said, there are several aspects of comparative effectiveness studies which make them particularly susceptible to bias due to misclassification including the comparison of two active treatments,

modest effect sizes that are clinically meaningful, the value of absolute measures of effect (such as the risk difference), and the extreme precision that comes from analyzing large datasets.

Comparing Active Treatments

In studies of comparative effectiveness in which two active treatments are being compared, there are at least three (and possibly more) levels of exposure: non-user, user of medication A, and user of medication B. Misclassification of individuals who were truly exposed to medications A and B would place individuals in the non-user category, not in the other category of exposure. Misclassification of this type could result in estimates that are toward or away from the null, even though there are only two levels of exposure being analyzed.

We present a hypothetical example in Table 1 in which each of the medications increases the risk of the outcome two-fold ($RR=2.0$) compared to unexposed individuals. (An Excel version of the spreadsheet is available at <http://www.unc.edu/~mfunk> to facilitate exploration of alternative scenarios.) We apply nondifferential misclassification of each medication with the unexposed group and assume that there is no misclassification of users between the two medication groups. We observe the classic finding that nondifferential misclassification biases the relative risk (RR) toward the null in two hypothetical ‘studies’ comparing each active medication to a group of non-users. But because the degree of bias toward the null is not uniform across the two studies, the resulting head-to-head comparison of medication A versus medication B is biased away from the null – an apparent 20 % increase in risk ($RR=1.2$) whereas the true effect is null. In the example, we applied different degrees of sensitivity and specificity to each medication study, but the bias we observed is not dependent on that. (Using uniform sensitivity and specificity across the medications, the bias increases to a relative risk of 1.4.) Rather, the difference in the prevalence of the medications in the population or the presence of differential specificity of the exposure each lead to bias in the relative effect comparing the medications to each other.

Modest Effect Sizes

Many important differences in safety and effectiveness of active treatments are in the range of 20 to 40 % [42–44]. The potential for bias to obscure a clinically relevant difference (or create the appearance of a difference where there is none) is heightened in this context. Modest effect sizes are particularly susceptible to the effects of residual confounding due to misclassified covariate data. In light of the potential for bias due to exposure misclassification that could be in any direction, this is a setting in which validation studies and

Table 1 Hypothetical example of studies in which individuals exposed to one of two drugs are each compared with non-users, or compared with each other in the presence of nondifferential exposure misclassification.

	Truth		Observed Data	
Drug A vs. Nonusers	Drug A ₁	Drug A ₀	Drug A ₁	Drug A ₀
Y=1	2,000	5,000	1,950	5,050
Y=0	8,000	45,000	9,050	43,950
Total	10,000	50,000	11,000	49,000
Risk	0.20	0.10	0.18	0.10
OR	2.25		1.88	
RR	2.00		1.72	
RD	0.10		0.07	
Drug B vs. Nonusers	Drug B ₁	Drug B ₀	Drug B ₁	Drug B ₀
Y=1	200	5,000	280	4,920
Y=0	800	45,000	1,620	44,180
Total	1,000	50,000	1,900	49,100
Risk	0.20	0.10	0.15	0.10
OR	2.25		1.55	
RR	2.00		1.47	
RD	0.10		0.05	
Drug A vs. Drug B	Drug A ₁	Drug B ₁	Drug A ₁	Drug B ₁
Y=1	2,000	200	1,950	280
Y=0	8,000	800	9,050	1,620
Total	10,000	1,000	11,000	1,900
Risk	0.20	0.20	0.18	0.15
OR	1.00		1.25	
RR	1.00		1.20	
RD	0.00		0.03	

Drug A vs. non-users: sensitivity 0.85, specificity 0.95. Drug B vs. non-users: sensitivity 0.90, specificity 0.98. All are non-differential with respect to disease (Y) status. No individuals exposed to Drug A are misclassified as exposed to Drug B or vice versa. OR: odds ratio; RR: risk ratio; RD: risk difference. Y=1: experienced the outcome of interest; Y=0: did not experience the outcome of interest.

quantifying the impact on estimates and uncertainty are particularly important.

Absolute Effect Measures

The choice of effect measures in CER also increases concern about bias due to misclassification. While relative effect measures remain dominant, there is growing recognition that absolute measures are important, particularly in terms of communicating the relevance of the findings to patients [45, 46]. Achieving near perfect specificity in the outcome classification may allow us to claim that the relative effect estimate is unlikely to be considerably biased, but the estimated risks and risk differences will still be under-estimated if the sensitivity is not perfect unless further analysis is used to correct for the non-perfect sensitivity of the outcome definition.

Very Large Study Sizes

Analyses of claims data are powerful and allow us to examine rare outcomes. Very large sample sizes which may give the appearance of precision, making a very small increase or decrease (e.g., HR=1.06, 95 % CI 1.01, 1.11 %) appear statistically significant, or a null effect seem to exclude any possibility of a protective or harmful effect (e.g., 95 % CI 0.96, 1.04). In the presence of misclassification, these confidence intervals misrepresent the true uncertainty about the estimate.

Because of the very nature of comparative effectiveness research, quantifying the extent of these errors and adjusting the effect estimates and their confidence intervals is particularly important. Various methods for doing so have been developed and are discussed in the following section.

Quantifying Impact and Adjusting Estimates

In this section, we highlight several methods that can be used to quantify the effect of bias due to misclassification. These methods are summarized in Table 2. Here we focus on understanding how these might be best applied in the setting of treatment effect estimation using claims data.

Simple Bias Analysis

This method is the easiest to implement, but also has the most limited potential for use in the setting of pharmacoepidemiology. It re-allocates the observed, tabled data to the underlying ‘true’ tabled data using point estimates for (possibly differential) sensitivity and specificity or positive and negative predicted values. The corrections can be applied to categorized exposure, outcome, or covariates. Corrections can be implemented based only on expert opinion or estimates from the literature – an advantage in the setting where validation data are not available. This approach would be suitable for the analysis of short-term outcomes (such as in-hospital mortality) where all individuals are followed for a consistent period of time, but it is not suitable for outcomes that are partially censored (time-to-event). It does not account for error in the estimation of the sensitivity and specificity in the adjusted effect estimates, and it does not simultaneously control for other covariates. Lash et al. (2009) [8] provide an excellent, in-depth discussion of this approach as well as an Excel spreadsheet.

Probabilistic Bias Analysis

This approach is essentially an iterative version of the simple bias analysis which uses a distribution of values for sensitivity and specificity (or positive and negative predictive values) combined with a Monte Carlo process to produce a

Table 2 Review of methods for quantifying bias due to misclassification

Method	Capabilities	Application	Example
Simple bias analysis Sources: Rothman et al. 2008 [3], Lash et al. 2009 [8], Greenland 1996 [59], Schneeweiss 2006 [60] Code: Lash et al. 2009 provide downloadable Excel spreadsheets to conduct simple bias analyses for various misclassification situations [8].	<ul style="list-style-type: none"> Assumes independence of other sources of bias Assumes certainty of bias parameters Evaluates one fixed bias parameter Can use sensitivity and specificity, or positive and negative predictive value Can be used when only summary data are available Cannot directly incorporate random error Cannot simultaneously adjust for other measured confounders Cannot be used in the setting of a time-to-event outcome Able to account for misclassification in multiple analytic variables (i.e. exposure, outcome, covariate) through matrix adjustment 	Variable type Dichotomous, polytomous Analytic variable Exposure, outcome, covariate Data source(s) Internal validation sample, external validation study, prior literature, expert opinion	Fink and Lash 2003 [61] conducted a simple bias analysis in a birth cohort in Massachusetts to explore a range of sensitivity values for maternal smoking reported on the birth certificate, based on results of prior validation studies. Jurek and Greenland 2013 [62] simultaneously considered misclassification of exposure (maternal smoking) and outcome (clefting) using the matrix adjustment method of bias analysis.
Probabilistic bias analysis Sources: Rothman et al. 2008 [3], Lash and Fink 2003 [63], Fox et al. 2005 [48], Lash et al. 2009 [8] Code: Fox et al. 2005 and Lash et al. 2009 provide a SAS macro, sample dataset, and example code for probabilistic sensitivity analysis [8, 48].	<ul style="list-style-type: none"> Uses a specified probability distribution for each bias parameter Allows specification of a range of validated values for the bias parameters Can use sensitivity and specificity, or positive and negative predictive value Interval estimate combines systematic error with random error Can incorporate multiple sources of bias (i.e., misclassification, selection bias, and confounding) Can be conducted using summary data or record level data Can be applied at the record level to evaluate misclassified exposure data for a time-to-event outcome and/or include adjustment for measured confounders 	Variable type Dichotomous, polytomous Analytic variable Exposure, outcome, covariate Data source(s) Internal validation sample, external validation study, prior literature or study results, expert opinion	Ahrens et al. 2012 [47••] conducted probabilistic bias analysis for misclassification of the exposure, prenatal influenza vaccination, in a Cox regression model. Lash et al. 2014 [64•] compared bias analysis populations (full cohort, case cohort, and weighted full cohort) in the Penn MOMS cohort for exposure misclassification of pre-pregnancy BMI on the birth certificate using a validation substudy involving medical record review, in order to determine which strategy would be the most computationally intensive.
Bayesian bias analysis Sources: Rothman et al. 2008 [3], Chu et al. 2006 [65], McCandless et al. 2007 [66] Code: MacLehose et al. 2009 provide WinBUGS code in the Web Appendix (ArticlePlus) for Bayesian bias analysis [50].	<ul style="list-style-type: none"> Incorporates prior distributions for the bias parameters Analysis allows for examination of the entire posterior distribution, from which estimates are drawn Bayesian formulation forces use of prior information about any parameters in the analysis (i.e. exposure, measured confounders), not just parameters in the bias model Allows specification of a range of validated values for the bias parameters Combines uncertainty in bias parameters with random error of regression coefficients 	Variable type Dichotomous, polytomous Analytic variable Exposure, outcome, covariate Data source(s) Internal validation sample, external validation study, prior literature or study results, expert opinion	MacLehose et al. 2009 [50] examined exposure misclassification of self-reported maternal smoking data using prior distributions of sensitivity and false positive rates from the control population of a previous case-control study. Keil et al. 2014 [67•] used Bayesian methods to conduct a sensitivity analysis of imidacloprid exposure with an estimated range of values for sensitivity and specificity, as the true extent of exposure misclassification was unknown.

Table 2 (continued)

Method	Capabilities	Application	Example
Modified maximum likelihood Sources: Magder and Hughes 1997 [68], Neuhaus 1999 [69], Lyles et al. 2011 [70] Code: Edwards et al. 2014 provides SAS code in the Web Appendix for the maximum likelihood estimation method [53•].	<ul style="list-style-type: none"> • Uses values of sensitivity and specificity • Uses the expectation-maximization algorithm • Can accommodate differential misclassification by exposure and uncertainty around the estimates of sensitivity and specificity 	Variable type Dichotomous, polytomous Analytic variable Exposure, outcome Data source(s) Internal validation sample, external validation study, prior literature or study results, expert opinion	Edwards et al. 2014 [51•] examined outcome misclassification of lung cancer death in a cohort of South Carolina textile workers, using a Poisson regression model to estimate a rate ratio Shebl et al. 2012 [71] examined misclassified hepatitis C virus (outcome) in a cohort of pregnant women using a range of sensitivity and specificity values in a Poisson regression model.
Multiple imputation for measurement error (MIME) Sources: Cole et al. 2006 [52], Bang et al. 2013 [72•] Code: Cole et al. 2006 provide SAS code to perform multiple imputation measurement error correction in Appendix 3 [52].	<ul style="list-style-type: none"> • Method of external adjustment that requires an internal validation sample • Requires outcome assessment in the validation sample • Can accommodate differential misclassification with respect to the outcome and uncertainty around the estimates of sensitivity and specificity • Does not require surrogacy like other methods (i.e., propensity score calibration) 	Variable type Dichotomous, polytomous, continuous Analytic variable Exposure, outcome, covariate Data source(s) Internal validation study,	Edwards et al. 2013 evaluated outcome misclassification of ocular herpes simplex virus with multiple imputation using patient self-report and physician diagnosis (gold-standard) data from the Herpetic Eye Disease Study randomized trial [73].
Regression calibration Sources: Rosner et al. 1989 [53], Spiegelman et al. 1997 [54], Bang et al. 2013 [72•] Code: SAS macro %blinplus and related documentation are available to implement correction for measurement error in logistic regression [55].	<ul style="list-style-type: none"> • Method of external adjustment that requires a validation sample • Estimates a model in the gold-standard validation sample and uses regression parameters to correct the naïve model • Can be used for multiple explanatory variables • Corrects point and interval estimates of effect • Assumes that measurement error is not severe • Assumes that the measurement in the validation study is a proper surrogate for the true exposure or covariate • For use in the Cox model, measurement error must be additive and nondifferential 	Variable type Dichotomous, polytomous, continuous Analytic variable Exposure, covariate Data source(s) Internal validation study, external validation sample	Murphy et al. 2013 applied regression calibration for measurement error of the exposure, dietary consumption of dairy products, using two sources of data: dietary questionnaires and a 24-h dietary recall provided by a subset of the cohort [74].
Propensity score calibration Sources: Stürmer et al. 2005 [57], Stürmer et al. 2007 [75], Lunt et al. 2012 [56] Code: Stürmer et al. 2005 provide SAS program code to implement propensity score calibration in Appendix 1 [57].	<ul style="list-style-type: none"> • Method of external adjustment that requires a validation sample • Does not require outcome assessment in the validation sample • Accounts for the joint effect of misclassified confounders available only in the validation study • Strong surrogacy assumption 	Variable type Dichotomous, polytomous, continuous Analytic variable Covariate Data source(s) External validation sample, Internal validation study – note: MIME preferred	Toh et al. 2012 analyzed data from The Health Improvement Network in the United Kingdom, including an internal validation sample of 300,000 randomly selected patients with complete covariate information, to estimate the relationship between NSAIDs and gastrointestinal bleeding [76•].

distribution of estimates adjusted for misclassification. The credible intervals from this analysis can reflect the uncertainty around the validation measures. This method can also be applied at the record level so that misclassified exposures can be evaluated while controlling for measured covariates in the setting of a time-to-event outcome [47••], making this an excellent choice for use in pharmacoepidemiology studies. This method is also described in detail by Lash et al. (2009) [8] in Chapter 8, and Fox et al. (2005) [48] provide a SAS macro to facilitate application of this method.

Bayesian Bias Analysis

Bayesian bias analysis is similar to the probabilistic bias analysis, but with the addition of prior distributions for all parameters – not just those for misclassification. Like probabilistic bias analysis, random error is reflected in the adjusted effect estimates. In most cases, this method does not outperform probabilistic bias analysis [49]. The more complex implementation (in terms of software and programming) makes the Bayesian approach less attractive as a general method for application in analyses of claims data, although code for applying this method using BUGS has been published by MacLehose et al. 2009 in an online appendix [50].

Modified Maximum Likelihood

This method uses the full data (rather than tabled data) to fit a modified maximum likelihood that forces the sensitivity and specificity to be less than perfect. This method has been demonstrated with dichotomous and polytomous exposures and outcomes, including outcomes that are Poisson distributed to estimate the rate ratio. This method would be suitable for analyses in which follow-up time varies between individuals (for estimating rates rather than risks) and the hazard is approximately constant. Edwards et al. 2014 provide sample SAS code for this method [51•].

Multiple Imputation for Measurement Error (MIME)

In this approach, the true value of the misclassified variable is treated as partially missing data. The gold standard measures from an internal validation sample are used to fit a model for the imperfect data, and multiple datasets with imputed values for the misclassified variable are created. The effect estimates from analyses of these datasets are then combined to account for the variability introduced through the imputation. This approach would be well suited to analyses of claims data in which the exposure or covariate are misclassified and the outcome has been ascertained during differing amounts of follow-up time. Cole et al. provide SAS code for implementing MIME [52].

Regression Calibration

This method is best suited to the setting in which a continuous variable (exposure or covariate) is measured with error. Regression calibration can take advantage of multiple, imperfect measures of a characteristic (such as blood pressure) in the absence of a single gold standard measure [53, 54]. This approach has been used extensively in the field of nutritional epidemiology, but could be useful in pharmacoepidemiology studies in which lab results are available. A SAS macro is provided by Logan and Spiegelman for the correction of measurement error in the context of logistic regression [55].

Propensity Score Calibration

Propensity score calibration addresses covariate misclassification and measurement error by treating the propensity score as having been estimated with error. By fitting a ‘gold standard’ propensity score using the covariate data that is measured without error (in addition to the data measured with error) in a validation study, one can adjust the error prone propensity score values. Like regression calibration, propensity score calibration requires a surrogacy assumption. Surrogacy is, however, less likely to hold for the propensity score than for a mismeasured covariate [56]. Given the prominence of propensity score analyses in the pharmacoepidemiology field, this is a natural extension of the analytic methods used in many studies. Stürmer et al. provide SAS code in an online Appendix [57].

Challenges

Accessible Methods to Account for Misclassification in Complex Data

The relatively few published papers in which methods for accounting for misclassification have been applied tend to ‘cluster’ around the authors who originally published these methods. Clustering of examples for methods with the original method ‘creator’ on the paper suggest that implementation remains a significant challenge to widespread adoption of these methods. In the context of claims data, hundreds of covariates (many of which are presumably measured with error) related to thousands of individuals pose difficult logistical problems for applying these methods and presenting an integrated view of the effect of misclassification. The use of directed acyclic graphs to examine potential sources of bias due to misclassification and measurement error may allow the analyst to identify the variables of greatest concern (exposure, outcome, and/or specific confounders) so that efforts to quantify their effects can be targeted [58].

In Search of a Gold Standard for Prescription Exposures

There has also been an assumption that prescription claims data were sufficiently reliable that there was little concern for misclassification of exposure. Compared to self-reported data on medication use, often even retrospective, these data are likely more reliable. But they are not infallible, as shown by Li et al. Lauffenberger et al. and others [16, 20••]. It is not yet clear what the gold-standard measure for prescription medications use should be in light of research showing that administrative claims, physician orders, medical records, pharmacy records, and self-report are all subject to some degree of error.

Methods for Censored Outcomes

In the context of pharmacoepidemiologic analyses, follow-up time is typically censored which necessitates the use of methods such as Poisson, Kaplan-Meier lifetables, or Cox proportional hazards regression to estimate the treatment effect. Some of the methods noted here allow the investigator to account for misclassification of the exposure or covariates in this setting. The challenge of possibly misclassified outcomes – including the time at which the outcome occurred – is not as tractable.

Further Development of Methods to Handle Misclassified Person-Time

Analytic methods for addressing misclassified data are not yet able to adjust easily for errors in the timing of treatment initiation. This poses a challenge for studies in which patients may be identified as new users, but are actually prevalent users. Recent work by Ahrens et al. [47••] points to one way forward, but this approach has not yet been applied to the claims setting. Further development of this method would enable more thoughtful investigation of the impact of errors in the identification of treatment initiation and discontinuation which are particularly important given the time-varying nature of medication effects. In addition, valid methods are needed to adjust estimates and confidence intervals from self-controlled study designs given that the outcomes of interest in this setting are typically acute, and therefore, misclassified duration of use would be more problematic.

Conclusions

While it is common practice in pharmacoepidemiology to conduct and report the results of sensitivity analyses that examine the influence of many of the assumptions and decisions made during the design and conduct of the study, we

found few examples in the literature of sensitivity analyses that quantified the impact of misclassification. Perhaps the greatest challenge in this area is to acknowledge and then quantify the imperfect nature of claims data in spite of this status quo. Particularly with the rise of comparative effectiveness research, we cannot rely on nondifferential misclassification of the exposure to bias effect estimates toward the null. Rather than speculate about the effects of misclassification, we can and should be quantifying the impact on estimates and the uncertainty around them more accurately than we are currently doing using confidence intervals based on sampling error alone.

Acknowledgments We would like to acknowledge thoughtful suggestions from Jess Edwards and Alex Keil. During the development of this manuscript, Dr. Jonsson Funk was supported by grant number K02HS017950 from the Agency for Healthcare Research and Quality and grant number R01HL118255 from the National Institutes of Health (NIH), National Heart Lung and Blood Institute (NHLBI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. Suzanne Landi was supported by the NIH/NICHD T32HD052468 Reproductive Perinatal Pediatric Training Grant.

Compliance with Ethics Guidelines

Conflict of Interest M Jonsson Funk and SN Landi both declare no conflicts of interest.

Human and Animal Rights and Informed Consent All studies by the authors involving animal and/or human subjects were performed after approval by the appropriate institutional review boards. When required, written informed consent was obtained from all participants.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
 - Of major importance
1. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol*. 2005;34(3):680–7.
 2. Brenner H. Bias due to non-differential misclassification of polytomous confounders. *J Clin Epidemiol*. 1993;46(1):57–63.
 3. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
 4. Poole C. Exceptions to the rule about nondifferential misclassification (abstract). *Am J Epidemiol*. 1985;122:508.
 5. Ogburn EL, VanderWeele TJ. On the nondifferential misclassification of a binary confounder. *Epidemiology*. 2012;23(3):433–9.
 6. Marshall JR, Hastrup JL. Mismeasurement and the resonance of strong confounders: uncorrelated errors. *Am J Epidemiol*. 1996;143(10):1069–78.
 7. Savitz DA, Baron AE. Estimating and correcting for confounder misclassification. *Am J Epidemiol*. 1989;129(5):1062–71.

8. Lash TL, Fox MP, Fink AK. In: Gail M, Krickeberg K, Samet J, Tsiatis A, Wong W, editors. Applying quantitative bias analysis to epidemiologic data. New York: Springer; 2009.
9. Sackett DL. Bias in analytic research. *J Chron Dis*. 1979;32(1–2): 51–63.
10. Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol*. 1986;123(4):670–6.
11. Raebel MA, Ellis JL, Carroll NM, Bayliss EA, McGinnis B, Schroeder EB, et al. Characteristics of patients with primary non-adherence to medications for hypertension, diabetes, and lipid disorders. *J Gen Intern Med*. 2012;27(1):57–64.
12. Esposito D, Schone E, Williams T, Liu S, Cybulski K, Stapulonis R, et al. Prevalence of unclaimed prescriptions at military pharmacies. *J Manag Care Pharm*. 2008;14(6):541–52.
13. Hong JL, Meier CR, Sandler RS, Jick SS, Stürmer T. Risk of colorectal cancer after initiation of orlistat: matched cohort study. *BMJ*. 2013;347:f5039.
14. Karve S, Cleves MA, Helm M, Hudson TJ, West DS, Martin BC. An empirical basis for standardizing adherence measures derived from administrative claims data among diabetic patients. *Med Care*. 2008;46(11):1125–33.
15. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf*. 2007;16(3):241–9.
16. Lauffenburger JC, Balasubramanian A, Farley JF, Critchlow CW, O'Malley CD, Roth MT, et al. Completeness of prescription information in US commercial claims databases. *Pharmacoepidemiol Drug Saf*. 2013;22(8):899–906.
17. Harris Teeter. Generic Savings Club Quick Reference Matthews, NC: Harris Teeter, Inc.; 2012 [cited 2014 July 25]. Available from: http://www.harristeeter.com/files/docs/2012_GenericSavingsClubQuickReference_v2.pdf
18. Target. \$4 generic drugs listed by condition Minneapolis, MN: Target Corporation; 2014 [cited 2014 July 25]. Available from: <http://www.target.com/pharmacy/generics-condition>.
19. Walmart. Retail Prescription Program Drug List Bentonville, AR: Wal-Mart Stores, Inc.; 2014 [cited 2014 July 25]. Available from: http://i.walmartimages.com/i/if/hmp/fusion/customer_list.pdf.
20. Li X, Stürmer T, Brookhart MA. Evidence of sample use among new users of statins: implications for pharmacoepidemiology. *Med Care*. 2014. *The authors use administrative claims data to examine a patient population of statin users who had undergone LDL testing. The distribution of LDL results prior to the first dispensed prescription were used to estimate statin sample use as a first course therapy. The results provided strong evidence of sample use prior to a new prescription fill, indicating drug initiation misclassification.*
21. Gamble JM, Johnson JA, Majumdar SR, McAlister FA, Simpson SH, Eurich DT. Evaluating the introduction of a computerized prior-authorization system on the completeness of drug exposure data. *Pharmacoepidemiol Drug Saf*. 2013;22(5):551–5.
22. CHPA. Ingredients & Dosages Transferred From Rx-to-OTC Status (or New OTC Approvals) by the Food and Drug Administration Since 1975 Washington, DC: Consumer Healthcare Products Association; 2014 [cited 2014 July 25]. Available from: <http://www.chpa.org/SwitchList.aspx>.
23. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158(9):915–20.
24. Moride Y, Abenham L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol*. 1994;47(7):731–7.
25. Gamble JM, McAlister FA, Johnson JA, Eurich DT. Quantifying the impact of drug exposure misclassification due to restrictive drug coverage in administrative databases: a simulation cohort study. *Value Health*. 2012;15(1):191–7.
26. Burden AM, Huang A, Tadrous M, Cadarette SM. Variation in the days supply field for osteoporosis medications in Ontario. *Arch Osteoporos*. 2013;8(1–2):128.
27. Guess HA. Exposure-time-varying hazard function ratios in case-control studies of drug effects. *Pharmacoepidemiol Drug Saf*. 2006;15(2):81–92.
28. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol*. 1991;133(2):144–53.
29. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol*. 1996;143(11):1165–73.
30. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*. 1995;51(1):228–35.
31. Alonso A, Bengtson LG, MacLehose RF, Lutsey PL, Chen LY, Lakshminarayan K. Intracranial hemorrhage mortality in atrial fibrillation patients treated with dabigatran or warfarin. *Stroke*. 2014.
32. Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:90–9. *This initiative systematically reviewed the literature for algorithms to identify 19 different acute clinical events including cerebrovascular events, congestive heart failure, depression, and seizures, among others. The specific results for each condition are published in this supplemental issue of Pharmacoepidemiology and Drug Safety.*
33. Li X, Chen Y, Gokhale M, Chandler J, McNeill A, Girman CJ, et al. Radiographic and endoscopic diagnostic workup around initiation of oral bisphosphonates. 30th International Conference on Pharmacoepidemiology & Therapeutic Risk Management. [Conference Abstract]. In press 2014.
34. Hong JL, Jonsson Funk M, Lund JL, Pate V, Stürmer T. Differential health care utilization in metformin versus sulfonylureas users pre- and post-initiation. 30th International Conference on Pharmacoepidemiology & Therapeutic Risk Management. [Conference Abstract]. In press 2014.
35. Sheets NC, Goldin GH, Meyer AM, Wu Y, Chang Y, Stürmer T, et al. Intensity-modulated radiation therapy, proton therapy, or conformal radiation therapy and morbidity and disease control in localized prostate cancer. *JAMA*. 2012;307(15):1611–20.
36. Meyer A, Godley PA, Chen R. Radiation therapy modalities for prostate cancer—reply. *JAMA*. 2012;308(5):450.
37. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med*. 2008;358(8):771–83.
38. FDA. Best Practices for Conducting and Report Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. Silver Spring, MD: U.S. Department of Health and Human Services Food and Drug Administration, 2013.
39. Brunelli SM, Gagne JJ, Huybrechts KF, Wang SV, Patrick AR, Rothman KJ, et al. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. *Pharmacoepidemiol Drug Saf*. 2013;22(5):542–50. *Via simulation, the authors compare approaches to using historical data on a dichotomous covariate in administrative claims data when availability differs among subjects: using all available historical data versus using data from a fixed look-back window.*
40. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med*. 1994;331(6):394–8.
41. Stürmer T, Jonsson Funk M, Poole C, Brookhart MA. Nonexperimental comparative effectiveness research using linked healthcare databases. *Epidemiology*. 2011;22(3):298–301.
42. Dormuth CR, Hemmelgarn BR, Paterson JM, James MT, Teare GF, Raymond CB, et al. Use of high potency statins and rates of admission for acute kidney injury: multicenter, retrospective observational analysis of administrative databases. *BMJ*. 2013;346:f880.

43. Yan YL, Qiu B, Hu LJ, Jing XD, Liu YJ, Deng SB, et al. Efficacy and safety evaluation of intensive statin therapy in older patients with coronary heart disease: a systematic review and meta-analysis. *Eur J Clin Pharmacol*. 2013;69(12):2001–9.
44. Gutierrez J, Ramirez G, Rundek T, Sacco RL. Statin therapy in the prevention of recurrent cardiovascular events: a sex-based meta-analysis. *Arch Intern Med*. 2012;172(12):909–19.
45. Fortin JM, Hirota LK, Bond BE, O'Connor AM, Col NF. Identifying patient preferences for communicating risk estimates: a descriptive pilot study. *BMC Med Inf Decis Making*. 2001;1:2.
46. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. *JAMA*. 2004;291(19):2359–66.
47. Ahrens K, Lash TL, Louik C, Mitchell AA, Werler MM. Correcting for exposure misclassification using survival analysis with a time-varying exposure. *Ann Epidemiol*. 2012;22(11):799–806. *This application of probabilistic bias analysis suggests a method for handling time-varying exposures, adjusted for measured confounders, in the context of a time-to-event outcome.*
48. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34(6):1370–6.
49. MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian? *Epidemiology*. 2012;23(1):151–8.
50. MacLehose RF, Olshan AF, Herring AH, Honein MA, Shaw GM, Romitti PA, et al. Bayesian methods for correcting misclassification: an example from birth defects epidemiology. *Epidemiology*. 2009;20(1):27–35.
51. Edwards JK, Cole SR, Chu H, Olshan AF, Richardson DB. Accounting for outcome misclassification in estimates of the effect of occupational asbestos exposure on lung cancer death. *Am J Epidemiol*. 2014;179(5):641–7. *This paper uses the modified maximum likelihood method for misclassification of lung cancer death, and provides helpful SAS code using PROC NLMIXED in the Web Appendix for this method.*
52. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol*. 2006;35(4):1074–81.
53. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med*. 1989;8(9):1051–69. *discussion 71–3.*
54. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr*. 1997;65(4 Suppl):1179s–86s.
55. Logan R, Spiegelman D. The SAS %BLINPLUS Macro Boston, MA: Harvard School of Public Health; 2012 [cited 2014 July 25]. Available from: <http://www.hsph.harvard.edu/donna-spiegelman/software/blinplus-macro/>.
56. Lunt M, Glynn RJ, Rothman KJ, Avorn J, Stürmer T. Propensity score calibration in the absence of surrogacy. *Am J Epidemiol*. 2012;175(12):1294–302.
57. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005;162(3):279–89.
58. Hernán MA, Cole SR. Invited commentary: causal diagrams and measurement bias. *Am J Epidemiol*. 2009;170(8):959–62.
59. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25(6):1107–16.
60. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*. 2006;15(5):291–303.
61. Fink AK, Lash TL. A null association between smoking during pregnancy and breast cancer using Massachusetts registry data (United States). *Cancer Causes Control*. 2003;14(5):497–503.
62. Jurek AM, Greenland S. Adjusting for multiple-misclassified variables in a study using birth certificates. *Ann Epidemiol*. 2013;23(8):515–20.
63. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003;14(4):451–8.
64. Lash TL, Abrams B, Bodnar LM. Comparison of bias analysis strategies applied to a large data set. *Epidemiology*. 2014;25(4):576–82. *The analysis explored three separate strategies for probabilistic bias analysis to evaluate computational intensity and applicability to the desktop computing environment.*
65. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. *Ann Epidemiol*. 2006;16(11):834–41.
66. McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med*. 2007;26(11):2331–47.
67. Keil AP, Daniels JL, Hertz-Picciotto I. Autism spectrum disorder, flea and tick medication, and adjustments for exposure misclassification: the CHARGE (Childhood Autism Risks from Genetics and Environment) case-control study. *Environ Health*. 2014;13(1):3. *The authors utilized a Bayesian approach to quantifying exposure misclassification where exposure was assessed retrospectively.*
68. Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. *Am J Epidemiol*. 1997;146(2):195–203.
69. Neuhaus J. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*. 1999;86(4):843–55.
70. Lyles RH, Tang L, Superak HM, King CC, Celentano DD, Lo Y, et al. Validation data-based adjustments for outcome misclassification in logistic regression: an illustration. *Epidemiology*. 2011;22(4):589–97.
71. Shebl FM, El-Kamary SS, Shardell M, Langenberg P, Dorgham LS, Maguire JH, et al. Estimating incidence rates with misclassified disease status: a likelihood-based approach, with application to hepatitis C virus. *Int J Infect Dis*. 2012;16(7):e527–31.
72. Bang H, Chiu YL, Kaufman JS, Patel MD, Heiss G, Rose KM. Bias correction methods for misclassified covariates in the Cox Model: comparison of five correction methods by simulation and data analysis. *J Stat Theory Pract*. 2013;7(2):381–400. *The authors evaluate different methods to address measurement error/misclassification in the Cox proportional hazards regression model using simulation, including regression calibration and multiple imputation.*
73. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for misclassified outcomes in binary regression models using multiple imputation with internal validation data. *Am J Epidemiol*. 2013;177(9):904–12.
74. Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G, et al. Consumption of dairy products and colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *PLoS One*. 2013;8(9):e72715.
75. Stürmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med Care*. 2007;45(10 Supl 2):S158–65.
76. Toh S, Garcia Rodríguez LA, Hernán MA. Analyzing partially missing confounder information in comparative effectiveness and safety research of therapeutics. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 2:13–20. *This paper is among the first to utilize the propensity score calibration method, among other methods, to evaluate partially missing confounder information in an electronic health database.*