



Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2012 May ; 21(0 2): 13–20. doi:10.1002/pds.3248.

Analyzing partially missing confounder information in comparative effectiveness and safety research of therapeutics

Sengwee Toh, ScD¹, Luis A. García Rodríguez, MD, MSc², and Miguel A. Hernán, MD, DrPH^{3,4}

¹Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

²Centro Español de Investigación Farmacoepidemiológica (CEIFE), Madrid, Spain

³Department of Epidemiology, Harvard School of Public Health, Boston, MA

⁴Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA

Abstract

Purpose—Electronic healthcare databases are commonly used in comparative effectiveness and safety research of therapeutics. Many databases now include additional confounder information in a subset of the study population through data linkage or data collection. We described and compared existing methods for analyzing such datasets.

Methods—Using data from The Health Improvement Network and the relation between non-steroidal anti-inflammatory drugs (NSAIDs) and upper gastrointestinal bleeding (UGIB) as an example, we employed several methods to handle partially missing confounder information.

Results—The crude odds ratio (OR) of upper gastrointestinal bleeding was 1.50 (95% confidence interval: 0.98, 2.28) among selective cyclo-oxygenase-2 inhibitor initiators (n = 43,569) compared with traditional non-steroidal anti-inflammatory drug initiators (n = 411,616). The OR dropped to 0.81 (0.52, 1.27) upon adjustment for confounders recorded for all patients. When further considering three additional variables missing in 22% of the study population (smoking, alcohol consumption, body mass index), the OR was between 0.80 and 0.83 for the missing-category approach, the missing-indicator approach, single imputation by the most common category, multiple imputation by chained equations, and propensity score calibration. The OR was 0.65 (0.39, 1.09) and 0.67 (0.38, 1.16) for the unweighted and the inverse probability weighted complete-case analysis, respectively.

Conclusion—Existing methods for handling partially missing confounder data require different assumptions and may produce different results. The unweighted complete-case analysis, the missing-category/indicator approach, and single imputation require often unrealistic assumptions and should be avoided. In this study, differences across methods were not substantial, likely due to relatively low proportion of missingness and weak confounding effect by the three additional variables upon adjustment for other variables.

Corresponding author Sengwee Toh, ScD Department of Population Medicine Harvard Medical School and Harvard Pilgrim Health Care Institute 133 Brookline Ave 6th Floor, Boston, MA 02215, USA darrentoh@post.harvard.edu Tel: +1 617 509 9818 Fax: +1 617 859 8112 .

Conflict of interest: None

Keywords

Comparative effectiveness research; Missing data; Databases; Confounding; Pharmacoepidemiology; THIN

Introduction

Electronic healthcare databases are widely used to assess the comparative effectiveness and safety of therapeutics in real-world settings.^{1,2} However, as these databases are not created for research purposes, data on certain important confounders may not be recorded. For example, administrative claims databases rarely, if ever, include direct measures of cigarette smoking, left ventricular ejection fraction, or depression severity.

Confounder data are sometimes available for a subset of patients. For example, measures of cigarette smoking or left ventricular ejection fraction may be recorded for some patients with electronic health records (EHRs). Administrative claims databases and EHRs can be linked to each other,³ or to device and disease registries,⁴⁻⁷ birth certificates,^{8,9} or survey data¹⁰ to provide additional confounder information that is otherwise not available. However, supplemental information is available only for patients whose records can be found in both data sources and linked successfully.

With a rapid increase in database linkage and the use of EHRs in comparative effectiveness and safety research, many researchers are now dealing with partial missingness of confounder information. Methods that can handle missing data have been described.¹¹⁻¹³

Here we discuss and compare several analytic approaches to handle partially missing confounder data in studies that use electronic healthcare databases. We used the relation between non-steroidal anti-inflammatory drugs (NSAIDs) and upper gastrointestinal bleeding (UGIB) as our example, because the relation is well-known—randomized trials reported a 40-60% lower risk of UGIB for selective COX-2 inhibitors (coxibs) compared with traditional NSAIDs (tNSAIDs)^{14,15}—and because severe confounding is expected in observational studies as coxibs are likely to be preferentially given to patients who have a higher risk of UGIB.

Methods

Data source

We analyzed data from The Health Improvement Network (THIN) database in the United Kingdom,^{16,17} a primary care database of nearly 4 million individuals whose clinical information is recorded by their general practitioner. Available information includes patient demographics; medical diagnoses; free-text comments; referral letters from consultants and hospitalizations; a record of all prescriptions issued; results from clinical examinations and laboratory tests; and additional information such as weight, height, smoking, and alcohol consumption. THIN uses Read Codes to register medical diagnoses and procedures, and a coded drug dictionary based on the Prescription Pricing Authority dictionary to record medications prescribed. The current study was approved by a Multicentre Research Ethics Committee (MREC) in the UK.

Study population

Our source population included 1,810,442 individuals aged 40-84 years between January 1, 2000 and December 31, 2008 with at least five years of enrollment with the general practitioner, at least one year of prospectively recorded information after the first recorded

prescription in the database, and at least one record (e.g., diagnosis) in the year prior to the first day in the study period during which they met all the above criteria (“entry date”).

We identified from the source population all patients with a first prescription of either a tNSAID or a coxib between the entry date and December 31, 2008. We refer to the date of first NSAID prescription as the *index date*. We required eligible patients to have no evidence of any NSAID prescription in the 18 months preceding the index date, and no recorded history of cancer (excluding non-melanoma skin cancer), chronic liver disease, Mallory-Weiss syndrome, coagulopathy, esophageal varices, chronic alcoholism, or bariatric or other surgery resulting in gastrojejunum anastomosis any time before the index date. We further excluded patients who initiated both a coxib and a tNSAID on the index date. The remaining 43,569 coxib and 411,616 tNSAID initiators formed our study cohort. We represent the treatment variable by A (1: coxib initiation, 0: tNSAID initiation).

Each patient in the cohort was followed from the index date until the earliest occurrence of UGIB, 85 years of age, death, 180 days of follow-up, or December 31, 2008. We selected a short follow-up of up to 180 days to minimize exposure misclassification.

Outcome

The validation process of potential UGIB cases has been described previously.¹⁸ Briefly, we first searched for Read Codes that suggest UGIB during the follow-up period, and then reviewed the computerized medical records (after including free-text comments) to confirm the diagnosis. Our initial computer search identified 468 potential cases of UGIB (73 among coxib initiators) during follow-up, of which 183 (25 among coxib initiators) were confirmed as cases after manual review and included in the analysis. The incidence rate of confirmed UGIB per 1,000 person-years was 1.2 for coxib initiators and 0.9 for tNSAID initiators, which was consistent with previous studies.¹⁹⁻²² We represent the outcome variable by Y (1: UGIB, 0: no UGIB).

Potential confounders

We identified the following potential confounders recorded in the entire study cohort during the 12-month period preceding the index date:¹⁹⁻²² age; sex; calendar year of treatment initiation; Charlson comorbidity score; use of gastroprotective drugs, anticoagulants, antiplatelets, and oral steroids; diagnosis of osteoarthritis, rheumatoid arthritis, dyspepsia, complicated and uncomplicated peptic ulcer disease, hypertension, congestive heart failure, and coronary artery disease; and three measures of healthcare utilization (numbers of distinct drugs prescribed, physician visits, and hospitalizations in the prior year). We represent these confounders by the vector X .

We further considered three supplemental variables – smoking, alcohol consumption, and body mass index – that were recorded in only a subset of the study population. About 78% of the study cohort had information for all three of these “lifestyle” variables, which are represented by the vector $L=(L_1, L_2, L_3)$. Both X and L include only baseline variables measured before treatment initiation.

Propensity score analysis

We used propensity score (PS)^{23,24} to adjust for potential confounders. Suppose the values of L variables were known for all patients, we could fit 1) a logistic model for $\Pr[A = 1 \mid X, L]$ to estimate each patient’s PS, i.e., the probability of initiating a coxib conditional on his covariate values, and 2) a logistic model for $\Pr[Y = 1 \mid A, PS]$ to estimate the odds ratio (OR) of UGIB for coxib initiators versus tNSAID initiators conditional on coxib initiation A and the PS (in deciles). This PS analysis would estimate an “intention-to-treat” effect of coxib

initiation on the risk of UGIB (conditional on the PS) compared with tNSAID initiation over the study's follow-up in the entire study population. In this study, we had to handle partial missingness in L before proceeding to performing the PS analysis.

Methods to deal with missing confounders

We describe two ways to handle missing data for L : 1) restricting the outcome model to patients without missing values (complete-case analysis) or 2) fitting the outcome model to all patients after assigning a value to either L or the PS for those with missing values (imputation methods). We analyzed our data under two versions of the complete-case analysis and four versions of imputation.

1. Complete-case analysis

1.1 Unweighted analysis: We defined a missingness indicator M (1: if any of the L variables is missing, 0 otherwise) and performed the PS analysis described above, but only among patients with no missing values ($M=0$). For comparison purposes, we performed an analysis adjusting only for X separately in patients with and without missing values in L .

1.2 Inverse probability (IP) weighted analysis: The IP weight $1/\Pr[M=0 \mid X]$ is the inverse (reciprocal) of the probability of $M=0$ conditional on X .^{25,26} To estimate this weight, we 1) defined a missingness indicator M_j (1 if missing and 0 otherwise, $j=1, 2, 3$) for each of the three variables smoking L_1 , alcohol L_2 , and body mass index L_3 (smoking had the lowest proportion of missing values, BMI the highest; alternative orderings of the three variables did not materially affect the results); 2) fit three logistic models for $\Pr[M_1 = 0 \mid X]$, $\Pr[M_2 = 0 \mid X, M_1 = 0]$, and $\Pr[M_3 = 0 \mid X, M_1 = M_2 = 0]$; 3) calculated the three predicted conditional probabilities for each patient; and 4) multiplied the three predicted probabilities and used the reciprocal of the product as the IP weight for each patient. We then performed a complete-case analysis identical to the one described above, except that each patient was weighted by his estimated IP weight. We used a robust variance estimator to calculate a conservative 95% confidence interval.²⁷ Using either bootstrapping or a variance estimator that explicitly incorporates how the weights were estimated would produce a narrower 95% CI.²⁸

This method attempts to reconstruct the study population without missing values by re-weighting patients with complete information. For example, if a patient had a conditional probability of 0.25 of having no missing values in L , the patient would be assigned a weight of four ($1/0.25$). That is, the patient would represent three other patients with identical X values but whose data are not included in the outcome model due to missing L .

2. Imputation methods

2.1 Missing-category and missing-indicator approach: In the missing-category approach, we created an additional category within each variable in L for patients with missing values. We then conducted the PS analysis described above. In the missing-indicator approach,^{11,29} we estimated the PS via a logistic model for coxib initiation that included the X variables, the missing indicators M_j , and the product terms $L_j(1-M_j)$.

2.2 Single imputation: We replaced the missing values for each L variable by the value of its most common category and conducted the PS analysis described above.

2.3 Multiple imputation by chained equations^{30,31}

Iteration 1: We fit a multinomial logistic regression model for $\Pr[L_1 = j \mid X, A, Y, M_1 = 0]$.^{32,33} The parameter estimates from the model defined a conditional multinomial

distribution, from which we drew values of L_1 to impute them to patients with missing L_1 . We then fit a second multinomial model for $\Pr[L_2 = l \mid X, A, Y, L_1^*, M_2 = 0]$, where L_1^* is the partially imputed smoking status, to impute the missing values for L_2 . Finally, we fit a multinomial model for $\Pr[L_3 = l \mid X, A, Y, L_1^*, L_2^*, M_3 = 0]$ to impute the missing values for L_3 . This first iteration of the procedure resulted in a dataset in which all missing values for variables L were imputed.

Iteration 2: We repeated the above procedure using the imputed data set from the first iteration. We removed the imputed values of L_1 and re-imputed them with a model conditional on A, X, Y, L_2^* , and L_3^* . We then re-imputed the missing values of L_2 and L_3 .

The procedure was repeated until 10 iterations were completed or until stable imputed values were obtained. We then repeated the iterative procedure 10 times to create 10 imputed datasets, in each of which we conducted the PS analysis described above. We then combined the OR estimates from the imputed datasets.³⁴ We performed this multiple imputation analysis using the IVEWare package for SAS software developed by the Survey Research Center, Institute for Social Research, at the University of Michigan (<http://www.isr.umich.edu/src/smp/ive/>).

2.4 PS calibration: This method imputes the value of the PS, rather than the value of the confounders L , in patients with missing values in L .³⁵ Imputing the PS can be conceptualized as a measurement error issue, which can be corrected using regression calibration if the true or “gold-standard” PS, PS_{gs} , can be correctly estimated from an internal or external validation sample.^{36,37} We attempted to create an internal validation sample by randomly selecting 300,000 patients among those with complete information on both X and L with the same age and sex joint distribution as the entire study cohort.

In the entire study cohort, we estimated an “error-prone” PS or PS_{ep} via the logistic model for $\Pr[A = 1 \mid X]$, and then included the estimated PS_{ep} as a linear continuous covariate in the logistic model for $\Pr[Y = 1 \mid A, \widehat{PS}_{ep}] = \beta_0 + \beta_1 A + \beta_2 \widehat{PS}_{ep}$.

In the validation sample, we estimated PS_{gs} via the logistic model for $\Pr[A = 1 \mid X, L]$, and fit the linear regression model $E[PS_{gs} \mid A, \widehat{PS}_{ep}] = \delta_0 + \delta_1 A + \delta_2 \widehat{PS}_{ep}$.

The regression calibration-adjusted estimator for the treatment effect was $\widehat{\eta}_1 = \widehat{\beta}_1 - \widehat{\delta}_1 \widehat{\beta}_2 / \widehat{\delta}_2$. We conducted the analysis using a SAS macro by Spiegelman and Logan, which is publicly available at <http://www.hsph.harvard.edu/spiegelman/blinplus.html>.

Results

Table 1 shows the distribution of baseline characteristics of initiators of coxibs and tNSAIDs ascertained during the 12-month period before the first NSAID prescription. The crude OR of UGIB for coxib initiators vs. tNSAID initiators was 1.50 (95% CI: 0.98, 2.28). The OR was 1.04 (0.68, 1.59) after adjustment for age and sex; 0.98 (0.63, 1.52) upon further adjustment for calendar year of treatment initiation; and 0.84 (0.54, 1.31) after further adjustment for measures of healthcare utilization. When we further adjusted for all remaining confounders in X , the OR was 0.81 (0.52, 1.27) for the entire study cohort; the OR was 0.64 (0.38, 1.07) for the 78% patients of the cohort with complete information on all three lifestyle variables in L , and 1.93 (0.78, 4.74) for patients with missing values on any of the three lifestyle variables.

Table 2 shows the results from different analytic approaches to deal with missing confounder data. The adjusted ORs were 0.65 and 0.67 for the unweighted and IP-weighted complete-case analyses, respectively. In the IP-weighted analysis, the weight had a mean of 1.28 (standard deviation 0.15) and ranged from 1.04 to 2.49. The adjusted ORs ranged between 0.80 and 0.83 for the imputation methods. The 95% CIs from the different methods were overlapping; the 95% CI for any estimate included all other point estimates.

Results from all approaches did not materially change when the PS was included as a continuous variable instead of deciles in the outcome model (as was necessary for the PS calibration approach). The *c*-statistic for the PS model was around 0.80 for all analyses, and the covariates were overall well balanced within PS strata (data not shown).

Discussion

We have reviewed and compared several approaches to deal with partially missing confounder information in electronic healthcare databases. We used the NSAID-UGIB example to illustrate their application to comparative effectiveness and safety research. All these methods require the assumptions of no unmeasured confounding for the effect of treatment on the outcome, and no misspecification of the outcome and PS models.

The missing-category/indicator approach and single imputation by the most common category further require additional assumptions that are generally implausible. In essence, they all assume that patients with missing information on certain variables are unconditionally exchangeable and can be grouped together for analysis. Single imputation by the most common category goes a step further and assumes that patients with missing data are not only comparable with each other, but also with patients with a certain (often arbitrarily chosen) covariate value. Though these methods are easy to implement, they have been shown to produce biased estimates even when patients with and without missing data are unconditionally exchangeable (i.e., data “missing completely at random”).^{11,38-40}

Multiple imputation requires that missingness be unassociated with the outcome conditional on the measured confounders or the corresponding PS (i.e., data are “missing at random”), and that the imputation model for each covariate with missing data be correctly specified.³⁴ The approach has been shown to provide more valid estimates than the missing-indicator approach and single imputation when these assumptions are true.^{11,34,41,42} A recent study that used THIN⁴³ found that patients with missing information on smoking, alcohol consumption, weight, or height differ systematically from the others in terms of comorbidities like cardiovascular disease and chronic obstructive pulmonary disease. Our estimate from multiple imputation would be incorrect if missingness were associated with other prognostic factors that were not included in the analysis. We used a version of multiple imputation that does not require the often unrealistic assumption of joint multivariate normality.^{30,31}

The PS calibration approach is valid under the assumptions that there is an appropriate internal or external validation sample, the linear measurement error model is correctly specified, and the error-prone PS is an appropriate surrogate for the gold-standard PS.^{35,44} The last assumption may be violated if the direction of confounding from the unmeasured or partially measured confounders is in the opposite direction to that from the measured covariates.⁴⁴ This approach may be combined with single imputation of the gold-standard PS based on the parameters of the measurement error model to do away with the need to specify the outcome model through matching or stratifying on the imputed gold-standard PS.⁴⁵

Yet, despite all these differences in the conditions required for valid estimates, we found only small differences across different imputation methods. The reasons might be that the proportion of missingness was relatively low and that the three variables with missing values might not be strong confounders after conditioning on other measured variables. Indeed, the OR adjusted for all potential confounders available in the entire study cohort (0.81) was similar to the ORs that further adjusted for the three lifestyle variables using different imputation approaches (0.80 to 0.83).

Like the imputation methods, the IP-weighted complete-case analysis estimates the effect in the entire study population.^{25,46} It is valid under an additional assumption that the weight models are correctly specified. The unweighted complete-case analysis estimates the effect only among patients without missing values, its results cannot be applied to the entire study population unless the data are missing completely at random. The unweighted complete-cases analysis has been shown to produce more biased estimates compared to other approaches, such as multiple imputation.^{11,39,40}

The point estimates of complete-case analyses and imputation methods were somewhat different, which may be due to random variability (wide and overlapping 95% CIs) or to real differences between patients with complete and incomplete confounder information beyond the information recorded in the database. For example, general practitioners who record patient lifestyle factors—and patients who respond to these questions—might have certain unmeasured characteristics that are associated with the outcome risk. Also, the effect of NSAIDs on UGIB might be modified by certain patient characteristics for which “missingness” is a proxy.

In conclusion, a number of methods are available to deal with missing data in comparative effectiveness and safety studies of therapeutics that analyze electronic healthcare databases. Researchers need to be aware of the underlying assumptions of various methods when choosing among them.

Acknowledgments

The authors would like to thank Ken Kleinman, ScD from Harvard Medical School and Harvard Pilgrim Health Care Institute for his thoughtful comments on an earlier draft of this paper.

Source of financial support: Dr. Toh is partially supported by R03HS019024. Dr. Hernán is partially supported by R01HL080644.

References

1. Suissa S, Garbe E. Primer: administrative health databases in observational studies of drug effects--advantages and disadvantages. *Nat Clin Pract Rheumatol*. 2007; 3:725–732. [PubMed: 18037932]
2. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005; 58:323–337. [PubMed: 15862718]
3. Andrade, SE.; Raebel, MA.; Boudreau, D., et al. Chapter 12: Health Maintenance Organizations/Health plans. In: Strom, BL.; Kimmel, SE.; Hennessy, S., editors. *Pharmacoepidemiology*. 5th ed. Wiley-Blackwell; Chichester, UK: 2012.
4. Doebbeling BN, Wyant DK, McCoy KD, et al. Linked insurance-tumor registry database for health services research. *Med Care*. 1999; 37:1105–1115. [PubMed: 10549613]
5. Bradley CJ, Given CW, Luo Z, et al. Medicaid, Medicare, and the Michigan Tumor Registry: a linkage strategy. *Med Decis Making*. 2007; 27:352–363. [PubMed: 17641138]
6. Hammill BG, Hernandez AF, Peterson ED, et al. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J*. 2009; 157:995–1000. [PubMed: 19464409]

7. Douglas PS, Brennan JM, Anstrom KJ, et al. Clinical effectiveness of coronary stents in elderly persons: results from 262,700 Medicare patients in the American College of Cardiology-National Cardiovascular Data Registry. *J Am Coll Cardiol*. 2009; 53:1629–1641. [PubMed: 19406337]
8. Piper JM, Ray WA, Griffin MR, et al. Methodological issues in evaluating expanded Medicaid coverage for pregnant women. *Am J Epidemiol*. 1990; 132:561–571. [PubMed: 2202203]
9. Andrade SE, Davis RL, Cheetham TC, et al. Medication Exposure in Pregnancy Risk Evaluation Program. *Matern Child Health J*. 2011
10. Lillard LA, Farmer MM. Linking Medicare and national survey data. *Ann Intern Med*. 1997; 127:691–695. [PubMed: 9382381]
11. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol*. 1995; 142:1255–1264. [PubMed: 7503045]
12. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat*. 2007; 61:79–90. [PubMed: 17401454]
13. Faris PD, Ghali WA, Brant R, et al. Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. *J Clin Epidemiol*. 2002; 55:184–191. [PubMed: 11809357]
14. Rostom A, Muir K, Dube C, et al. Gastrointestinal safety of cyclooxygenase-2 inhibitors: a Cochrane Collaboration systematic review. *Clin Gastroenterol Hepatol*. 2007; 5:818–828. 828, e811–815. quiz 768. [PubMed: 17556027]
15. Chen YF, Jobanputra P, Barton P, et al. Cyclooxygenase-2 selective non-steroidal anti-inflammatory drugs (etodolac, meloxicam, celecoxib, rofecoxib, etoricoxib, valdecoxib and lumiracoxib) for osteoarthritis and rheumatoid arthritis: a systematic review and economic evaluation. *Health Technol Assess*. 2008; 12:1–278. iii. [PubMed: 18405470]
16. Lewis JD, Schinnar R, Bilker WB, et al. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf*. 2007; 16:393–401. [PubMed: 17066486]
17. Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality-evaluated database of primary care data. *Inform Prim Care*. 2004; 12:171–177. [PubMed: 15606990]
18. Toh S, García Rodríguez LA, Hernán MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf*. 2011; 20:849–857. [PubMed: 21717528]
19. García Rodríguez LA, Jick H. Risk of upper gastrointestinal bleeding and perforation associated with individual non-steroidal anti-inflammatory drugs. *Lancet*. 1994; 343:769–772. [PubMed: 7907735]
20. García Rodríguez LA, Hernández-Díaz S. The risk of upper gastrointestinal complications associated with nonsteroidal anti-inflammatory drugs, glucocorticoids, acetaminophen, and combinations of these agents. *Arthritis Res*. 2001; 3:98–101. [PubMed: 11178116]
21. García Rodríguez LA, Barreales Tolosa L. Risk of upper gastrointestinal complications among users of traditional NSAIDs and COXIBs in the general population. *Gastroenterology*. 2007; 132:498–506. [PubMed: 17258728]
22. Hernández-Díaz S, García Rodríguez LA. Association between nonsteroidal anti-inflammatory drugs and upper gastrointestinal tract bleeding/perforation: an overview of epidemiologic studies published in the 1990s. *Arch Intern Med*. 2000; 160:2093–2099. [PubMed: 10904451]
23. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
24. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984; 79:516–524.
25. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc*. 1994; 89:846–866.
26. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006; 60:578–586. [PubMed: 16790829]
27. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.

28. Robins, JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran, ME.; Berry, D., editors. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. Springer-Verlag; New York: 1999. p. 95-134.
29. Jones MP. Indicator and stratification methods for missing explanatory variables in multiple linear regression. *J Am Stat Assoc*. 1996; 91:222–230.
30. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*. 1999; 18:681–694. [PubMed: 10204197]
31. Raghunathan TE, Lepkowski JM, Van Hoewyk J, et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*. 2001; 27:85–95.
32. Moons KG, Donders RA, Stijnen T, et al. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006; 59:1092–1101. [PubMed: 16980150]
33. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol*. 2006; 35:1074–1081. [PubMed: 16709616]
34. Rubin, DB. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons; New York, NY: 1987.
35. Sturmer T, Schneeweiss S, Avorn J, et al. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005; 162:279–289. [PubMed: 15987725]
36. 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; 8:1051–1069. discussion 1071-1053. [PubMed: 2799131]
37. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol*. 1990; 132:734–745. [PubMed: 2403114]
38. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol*. 1991; 134:895–907. [PubMed: 1670320]
39. van der Heijden GJ, Donders AR, Stijnen T, et al. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*. 2006; 59:1102–1109. [PubMed: 16980151]
40. Knol MJ, Janssen KJ, Donders AR, et al. Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. *J Clin Epidemiol*. 2010; 63:728–736. [PubMed: 20346625]
41. Little, RJA.; Rubin, DB. *Statistical analysis with missing data*. 2nd Edn. John Wiley & Sons; New York, NY: 2002.
42. Schafer, JL. *Analysis of incomplete multivariate data*. Chapman & Hall; London, UK: 1997.
43. Marston L, Carpenter JR, Walters KR, et al. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf*. 2010; 19:618–626. [PubMed: 20306452]
44. Sturmer T, Schneeweiss S, Rothman KJ, et al. Performance of propensity score calibration--a simulation study. *Am J Epidemiol*. 2007; 165:1110–1118. [PubMed: 17395595]
45. Carroll, RJ.; Ruppert, D.; Stefanski, LA. *Measurement error in nonlinear models*. Chapman and Hall, Ltd; London, United Kingdom: 1995.
46. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat*. 2008;4. Article22.

Key points

- Partially missing confounder information is common in comparative effectiveness and safety research of therapeutics.
- We applied several methods to deal with missing confounder information using data from a primary care electronic medical records database from the United Kingdom.
- Researchers should be aware of the underlying assumptions of various methods for handling missing data when choosing among them.

Table 1

Baseline characteristics of initiators of selective COX-2 inhibitors (coxibs) or non-selective (“traditional”) non-steroidal anti-inflammatory drugs (tNSAIDs) ascertained during the 12-month period before the first NSAID prescription

Characteristics	Patients with no missing supplemental confounder data *		Patients with missing supplemental confounder data *	
	Coxib initiators (n=33,693)	tNSAID initiators (n=320,733)	Coxib initiators (n=9,876)	tNSAID initiators (n=90,883)
	%	%	%	%
Age (years)				
40-44	6.7	18.9	7.9	20.4
45-49	8.2	14.3	9.4	15.9
50-54	10.5	13.9	11.4	14.8
55-59	13.2	13.7	11.8	12.9
60-64	12.6	12.6	11.7	11.0
65-69	13.4	9.7	11.6	8.5
70-74	14.3	8	12.9	7.0
75-79	12.6	5.7	12.7	5.7
80-84	8.6	3.3	10.7	3.9
Male sex				
	35.0	41.3	39.5	50.2
Calendar year of treatment initiation				
2000	5.0	9.7	5.7	12.0
2001	10.9	11.5	13.5	14.1
2002	19.3	10.9	22.1	13.5
2003	24.6	10.3	25.6	12.0
2004	27.3	10.6	23	11.0
2005	2.3	11.9	2.3	11.0
2006	3.4	11.5	2.7	9.6
2007	3.6	11.9	2.7	8.9
2008	3.7	11.7	2.3	7.9
No. of distinct drugs in the prior year				
0-2	22.7	42.2	32	53.3
3-4	17.7	20.5	19.5	19.4
5-7	22.2	18.5	20	15
8	37.3	18.8	28.5	12.4
No. of outpatient visits in the prior year				
0-3	19.5	29.9	34.3	46.1
4-6	20.7	23.5	23	23.3
7-10	22.3	21.0	19.1	16.0
11	37.5	25.6	23.7	14.6

Characteristics	Patients with no missing supplemental confounder data *		Patients with missing supplemental confounder data *	
	Coxib initiators (n=33,693)	tNSAID initiators (n=320,733)	Coxib initiators (n=9,876)	tNSAID initiators (n=90,883)
Hospitalized in the prior year	9.8	8.1	7.2	6.1
Charlson comorbidity score 1	41.4	27.9	33.5	20.4
Prior use of				
Gastroprotective drugs	29.2	11.8	26.2	9.4
Anticoagulants	2.0	0.8	1.7	0.7
Antiplatelets	21.1	12.7	14.4	7.7
Oral steroids	9.8	5.0	9.2	4.2
Diagnosis of				
Osteoarthritis	39.8	21.3	31.8	16.5
Rheumatoid arthritis	3.3	1.2	3.2	1.2
Dyspepsia	4.6	2.1	3.2	1.4
Peptic ulcer disease	0.4	0.1	0.3	0.1
Hypertension	40.5	29.9	26.5	17.0
Congestive heart failure	2.9	1.1	2.5	0.9
Coronary artery disease	15.9	8.3	10.1	4.3
Smoking				
Non smoker	55.9	55.2	35.8	37.6
Current smoker	19.9	21.7	16.4	18.1
Past smoker	24.3	23.1	12.8	12.6
Unknown	--	--	35.0	31.7
Alcohol consumption (drinks/week)				
None	49.6	44.4	14.9	13.6
1-19	33.4	34.6	7.9	8.8
10-19	10.7	12.9	2.5	3.6
20	6.3	8.1	1.8	3.0
Unknown	--	--	73.0	71.0
Body mass index (kg/m²)				
<18.5	1.3	1.1	0.5	0.4
18.5-24.9	33.9	36.7	8.0	8.6
25.0-29.9	39.8	38.8	9.2	9.2

Characteristics	Patients with no missing supplemental confounder data *		Patients with missing supplemental confounder data *	
	Coxib initiators (n=33,693)	tNSAID initiators (n=320,733)	Coxib initiators (n=9,876)	tNSAID initiators (n=90,883)
30.0-34.9	17.6	16.0	4.9	4.6
35	7.4	7.4	2.8	2.4
Unknown	--	--	74.7	74.9

* Supplemental confounding data include information on smoking, alcohol consumption, and body mass index

Table 2

Odds ratios of upper gastrointestinal bleeding during the first 180 days following initiation of selective COX-2 inhibitors vs. non-selective non-steroidal anti-inflammatory drugs, by different analytic approaches to incorporate supplemental confounder data available in a subset of the study cohort

Analytic methods	Number of patients included in the analysis	Adjusted odds ratio * (95% confidence interval)	Standard error of log odds ratio
<i>Complete-case analyses</i>			
Unweighted	354,426	0.65 (0.39, 1.09)	0.27
Inverse probability-weighted	354,426 (outcome/PS model) 455,185 (weight model)	0.67 (0.38, 1.16)	0.28
<i>Imputation methods</i>			
Missing-category approach	455,185	0.81 (0.51, 1.26)	0.23
Missing-indicator approach	455,185	0.80 (0.51, 1.25)	0.23
Single imputation	455,185	0.83 (0.53, 1.30)	0.23
Multiple imputation	455,185	0.82 (0.52, 1.29)	0.23
Propensity score (PS) calibration	455,185 (error-prone PS model) 300,000 (gold-standard PS model)	0.80 (0.50, 1.27)	0.24

* Adjusted via propensity score (PS) for the following variables recorded in the entire study cohort: age; sex; calendar year; Charlson comorbidity score; use of gastroprotective drugs, anticoagulants, antiplatelets, and oral steroids; diagnosis of osteoarthritis, rheumatoid arthritis, upper gastrointestinal symptoms, dyspepsia, complicated or uncomplicated peptic ulcer disease, hypertension, congestive heart failure, and coronary artery disease; and the number of distinct drugs, physician visits, and hospitalization in the prior year. Supplemental covariates recorded only in a subset of study cohort included smoking, alcohol consumption, and body mass index.