Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2011 August ; 20(8): 849–857. doi:10.1002/pds.2152.

Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: An application using electronic medical record data

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

¹Department of Population Medicine, Harvard Medical School/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—A semi-automated high-dimensional propensity score (hd-PS) algorithm has been proposed to adjust for confounding in claims databases. The feasibility of using this algorithm in other types of healthcare databases is unknown.

Methods—We estimated the comparative safety of non-steroidal anti-inflammatory drugs (NSAIDs) and COX-2 inhibitors regarding the risk of upper gastrointestinal bleeding (UGIB) in The Health Improvement Network (THIN), an electronic medical record (EMR) database in the UK. We compared the adjusted effect estimates when the confounders were identified using expert knowledge or the semi-automated hd-PS algorithm.

Results—Compared to the 411,616 traditional NSAID initiators, the crude odds ratio (OR) of UGIB was 1.50 (95% confidence interval: 0.98, 2.28) for the 43,569 selective COX-2 inhibitor initiators. The OR dropped to 0.81 (0.52, 1.27) upon adjustment for known risk factors for UGIB that are typically available in both claims and EMR databases. The OR remained similar when further adjusting for covariates – smoking, alcohol intake, and body mass index – that are not typically recorded in claims databases (OR 0.81; 0.51, 1.26), or an additional 500 empirically identified covariates using the hd-PS algorithm (OR 0.78; 0.49, 1.22). Adjusting for age and sex plus 500 empirically identified covariates produced an OR of 0.87 (0.56, 1.34).

Conclusions—The hd-PS algorithm can be implemented in pharmacoepidemiologic studies that use primary care EMR databases such as THIN. For the NSAID-UGIB association for which major confounders are well-known, further adjustment for covariates selected by an automated algorithm had little impact on the effect estimate.

Keywords

Confounding; Databases; Pharmacoepidemiology; Propensity score analysis; THIN

Corresponding author: Sengwee Darren Toh, Sc.D., Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Healthcare 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

Introduction

Electronic healthcare databases, such as administrative claims and electronic medical record (EMR) databases, are widely used to study the health effects of medical products.^{1,2} Because most of these databases are not compiled for research purposes, data on some important confounders may not be recorded. As a result, observational studies that use electronic healthcare databases are often criticized for their inability to control for confounding bias.¹⁻³

Although many confounders are not directly recorded in some electronic healthcare databases, it may be argued that some recorded variables (e.g., medical diagnosis and drug use) are proxies for unrecorded ones and thus could be used to adjust for confounding. The challenge is how to empirically identify those proxies out of the thousands of variables available in these databases.

One possible way to address this challenge is to identify adjustment variables via an automated search. This approach could be especially helpful in administrative claims databases in which many confounders are not directly measured. Recently, a semi-automated search procedure was applied to examine several well-known effects, including the effect of non-steroidal anti-inflammatory drugs (NSAIDs) on the risk of upper gastrointestinal bleeding (UGIB) in an administrative claims database.⁴ The authors concluded that the procedure was helpful to identify adjustment variables.

In contrast to claims databases, EMR databases typically include information on more potential confounders (at least for a subset of individuals). Further, the clinical information encoded in variables present in both types of data sources might vary quantitatively. It is therefore unclear whether automated or semi-automated search procedures are helpful for comparative effectiveness and safety research of medical products using EMRs. Here we estimate the effect of NSAIDs on UGIB using an EMR database widely used for pharmacoepidemiologic research. We compare the effect estimates when the confounders are chosen using expert knowledge only, an automated search, and both.

Methods

Data source

We conducted a cohort study to estimate the effect of NSAIDs on the incidence of UGIB using The Health Improvement Network (THIN) database in the United Kingdom.^{5,6} THIN is a population-based EMR database of close to 4 million individuals whose clinical information is recorded by their general practitioner. Among other items, the recorded information includes patients' demographics; medical diagnoses; free-text comments arising from patients' visit to the general practitioner; referral letters from consultants and hospitalizations; a record of all prescriptions issued by their general practitioner; results from clinical examinations and laboratory tests; and other additional information such as weight, height, smoking and alcohol consumption.

THIN uses Read codes (www.connectingforhealth.nhs.uk/terminology/readcodes) 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.

Source population

The study period was from January 1, 2000 to December 31, 2008. Our source population included 1,810,442 individuals aged 40–84 years with at least 5 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., medication, diagnosis) in the year prior to the first day in the study period they met all above criteria (“entry date”).

Study population

We identified all individuals with a first prescription of either a non-selective “traditional” NSAID (tNSAID) or a COX-2 inhibitor (coxib) in the source population between the entry date and December 31, 2008. tNSAIDs included aceclofenac, acemetacin, diclofenac, diflunisal, etodolac, fenbufen, fenclofenac, fenoprofen, feprazone, flufenamic acid, flurbiprofen, ibuprofen, indomethacin, indoprofen, ketoprofen, ketorolac, lornoxicam, mefenamic acid, meloxicam, nabumetone, naproxen, piroxicam, sulindac, suprofen, tenoxicam, tiaprofenic acid, tolfenamic acid, and tolmetin; coxibs included celecoxib, etoricoxib, lumiracoxib, parecoxib, rofecoxib, and valdecoxib. We refer to the date of first NSAID prescription as the *index date*. We required eligible individuals to have no evidence of any NSAID prescription in the 18 months preceding the index date, and no recorded history of any of the following conditions before the index date: cancer (excluding non-melanoma skin cancer), chronic liver disease, Mallory-Weiss syndrome, coagulopathy, esophageal varices, chronic alcoholism, or bariatric or other surgery resulting in gastrojejunal anastomosis. Individuals with a diagnosis of any of these conditions during the follow-up period were also excluded from the analysis because their diagnosis almost surely implied that the condition was present at baseline (<1% of individuals in each study cohort were in this situation and their inclusion did not affect the results shown below). We further excluded patients who initiated both a coxib and a tNSAID on the index date.

The remaining coxib and tNSAID initiators formed our study cohort. Each individual in the cohort was followed from the index date until the earliest occurrence of one of the following endpoints: UGIB (see below), 85 years of age, death, 180 days of follow-up, or December 31, 2008. We selected a short follow-up of 180 days to minimize current exposure misclassification during the follow-up period.

Definition, ascertainment and validation of outcome

We first used a computer search to identify potential cases with a Read code suggesting UGIB during the follow-up. We then manually review the computerized records, including anonymized free text comments, of all potential cases with us being blinded to all drug exposures.⁷ Individuals who were confirmed to meet all following criteria through manual review were included as cases of UGIB in our analysis: (1) the specific site of bleeding or perforation originated in the stomach or duodenum, (2) the lesion type was erosion, gastritis, duodenitis or peptic ulcer, and (3) the patient had been referred to a consultant/specialist or was hospitalized. All others were defined as non cases.

Confounders

Table 1 shows the data typically available in EMR databases like THIN and in claims databases. The classification is necessarily vague because many types of EMR and claims databases exist. For example, some EMR databases include only ambulatory information and/or start collecting data when the individual joins a particular health plan. Further, a variable available in both claims and EMR databases might vary quantitatively between these two types of data sources. We classified the potential confounders used in our analysis as covariates typically available in (1) both claims and EMR databases, and (2) EMR databases but not claims databases. In our analysis, smoking, alcohol consumption, and body mass index were in the latter group.

Statistical analysis

As in the study that used an administrative claims database,⁴ we used an “intention-to-treat” approach to compare the risk of UGIB between coxib initiators and tNSAID initiators, regardless of their treatment duration during the follow-up period. We used a logistic model to estimate the crude odds ratio (OR) of UGIB for coxib vs. tNSAID initiation and its 95% confidence interval (CI). Because UGIB is a rare condition over a short period (180 day-risk less than 1%), the OR is an approximation to the risk ratio. We fitted separate logistic models that, in addition to the indicator for coxib initiation, included the propensity score (in deciles) estimated by different sets of covariates identified by different strategies (see below). The propensity score^{8,9} is the probability of initiating a coxib; we estimated it via a separate logistic model for coxib initiation that included a set of covariates ascertained on or during the 12-month baseline period before the index date. A non-parametric bootstrap approach¹⁰ with 500 samples resulted in similar 95% CIs for the OR of UGIB (results not shown).

Methods for confounder selection

We used three different strategies to select the potential confounders used to estimate the propensity score.

1. **Expert knowledge only.** The confounders are identified a priori using subject-matter knowledge only, as commonly done in observational studies. The potential confounders used in previous analyses^{11–18} were age; sex; calendar year of treatment initiation; number of distinct drugs used, physician visits, and hospitalizations in the prior year; Charlson comorbidity score;¹⁹ prior use of gastroprotective drugs, anticoagulants, antiplatelets, and oral steroids; history of osteoarthritis, rheumatoid arthritis, dyspepsia, complicated and uncomplicated peptic ulcer disease, hypertension, congestive heart failure, and coronary artery disease; smoking; alcohol drinking; and body mass index. Table 2 shows how we categorized these covariates. We conducted separate analyses with and without the last three “lifestyle” variables, which are not typically available in claims databases. A previous study⁴ adjusted for nursing home residence and race/ethnicity, which are not systematically collected in our EMR database.
2. **Semi-automated covariate selection.** The covariates selected a priori through expert knowledge are supplemented with covariates selected via an automated procedure. We followed the procedure referred to as the semi-automated high-dimensional propensity score (hd-PS) by Schneeweiss et al.⁴
3. **Automated covariate selection.** Identical to the second strategy, except that only age and sex were included a priori.

For strategies 2 and 3, we used the SAS program provided by Schneeweiss et al at <http://www.drugapi.org/downloads/index.php>. Briefly, the hd-PS algorithm (1) requires the identification of data dimensions or categories of covariates (e.g., inpatient diagnoses, outpatient drug dispensings), (2) defines covariates using the codes within each dimension (e.g., International Classification of Diseases, 9th Revision, Clinical Modification [ICD-9-CM] codes for identifying diagnoses), (3) ranks candidate covariates by their recurrence (the frequency that the codes are recorded for each individual during the baseline period), (4) ranks covariates by their potential for confounding control based on the bivariate associations of each covariate with the treatment and with the outcome, (5) selects a pre-specified number of covariates (e.g., 500 per individual) for adjustment, and (6) estimates the propensity score using the selected covariates plus any pre-specified covariates.

We arranged our data structure into six datasets corresponding to the following data dimensions: (a) demographics and other pre-specified covariates, (b) inpatient diagnoses, (c) inpatient procedures, (d) outpatient diagnoses, (e) outpatient procedures, and (f) outpatient drug use. We used the first character of the Read codes to classify the codes into diagnoses and procedures. Read codes that start with “A” to “Z” usually denote diagnoses, while those that start with “3” to “8” usually represent procedures performed. We conducted secondary analyses that did not differentiate between diagnosis and procedure codes, and that used only the first three characters of the diagnosis codes. We also performed an analysis restricted to the years before 2005.

Results

The study cohort included 43,569 initiators of coxibs and 411,616 initiators of tNSAIDs. Our initial computer search identified 468 potential cases of UGIB (73 among coxib users) during the follow-up period, of which 183 (25 among coxib initiators) were confirmed as cases after manual review. The incidence rate of UGIB (per 1,000 person-years) was 1.2 for coxib initiators and 0.9 for tNSAID initiators, which was consistent with previous studies.^{11–18} During the 180-day follow-up, the average duration of treatment use was 59 days for coxib initiators and 42 days for tNSAID initiators; 89% of the coxib initiators and 95% of the tNSAIDs initiators stopped using their medications or switched to the other class of NSAIDs before the end of follow-up.

As expected, there was considerable confounding by indication, as individuals whose baseline characteristics and medical history suggest a higher UGIB risk were more likely to initiate a coxib (Table 2).

The unadjusted OR of UGIB for coxib vs. tNSAID initiators was 1.50 (95% CI: 0.98, 2.28). Both confounding by indication and longer duration of coxib use might have contributed to the elevated risk among coxib initiators. The OR was 1.04 (0.68, 1.59) after adjustment for age and sex; 0.98 (0.63, 1.52) after further adjustment for calendar year of treatment initiation; and 0.84 (0.54, 1.31) after further adjustment for the intensity of healthcare utilization summarized by the number of distinct drugs used, physician visits and hospitalizations in the prior year.

Under Strategy 1, we adjusted for investigator-identified covariates only. When selecting covariates that are typically available in both claims and EMR databases, the adjusted OR was 0.81 (0.52, 1.27). Figure shows the distribution of the estimated propensity score of coxib and tNSAID initiators of this analysis. There was substantial overlap in the distribution of the propensity score in the two groups. We present the *c*-statistic in Table 3 for informative purposes only; note that a high *c*-statistic is neither necessary nor sufficient for control of confounding.^{21,22)}

Under Strategy 2, we then further adjusted for 500 empirically identified covariates using the hd-PS algorithm; the adjusted OR was 0.78 (0.49, 1.22). The adjusted ORs did not materially change whether or not we adjusted for the three lifestyle variables typically available in EMR databases only (Table 3).

Finally, under Strategy 3, we adjusted only for age and sex, plus 500 empirically identified covariates using the hd-PS algorithm. The OR was 0.87 (0.56, 1.34). The top 3 covariates that were judged to introduce most bias based on their prevalence and bivariate associations with the exposure and outcome were outpatient diagnosis of menorrhagia (Read code K592000), outpatient “procedure” of foreign travel advice (Read code 67E..00; classified as a preventive procedure), and prescription of acetaminophen with codeine. Adding calendar

year, and intensity of healthcare utilization to the list of pre-specified covariates made the OR drop to 0.81 (Table 3).

The results were similar in the analyses that did not differentiate diagnosis and procedure Read codes, or that used the first three characters of the Read codes, when we categorized the propensity score into 20 groups instead of 10, or when we adjusted for the propensity score as a continuous variable (not shown). Restricting the analyses to the period before 2005 did not materially change the point estimates (not shown).

Discussion

We estimated a reduction in the 180-day risk of UGIB for initiators of coxib compared with initiators of tNSAIDs in THIN, an ambulatory, population-based EMR database. We demonstrated that implementing the semi-automatic hd-PS algorithm for confounder selection and adjustment was (i) feasible in an EMR database, and (ii) had a negligible impact in our particular application.

The risk reduction in UGIB risk for coxib initiators versus tNSAID initiators was 19% when adjusting for all investigator-identified covariates, and 16% when only adjusting for age, sex, calendar year and the intensity of healthcare utilization summarized by the number of distinct drugs used, physician visits and hospitalizations in the prior year. A previous study suggested that these summary variables, in particular the number of distinct drugs used, performed as well or better than other comorbidity scores in predicting the risks of clinical outcomes.¹⁹ Our findings suggest that these variables may need to be incorporated a priori in pharmacoepidemiologic analyses.

Further adjustment for 500 empirically identified covariates had a small impact on the point estimates, and the 95% CIs overlapped extensively. The estimated risk reduction was 21–22% when the 500 covariates were added to all investigator-identified variables, and 19% when added to age, sex, calendar year and the intensity of healthcare utilization. The version of the hd-PS algorithm used for this analysis did not create summary utilization variables automatically. We have suggested to the authors to include this feature in future versions of the software.

Our estimates suggested a greater benefit of coxib initiation than those from a previous study that applied the hd-PS algorithm to the U.S. Medicare claims database.⁴ In that study the estimated reduction in the risk of severe gastrointestinal complication was 6% after adjustment for investigator-identified covariates, and 12% when the hd-PS algorithm was used. The greater benefit in our analyses may be due to better adjustment for confounding in EMR databases, differences in definition and ascertainment of outcome, differences between the U.S. Medicare and UK study populations (including their distributions of duration of treatment), random variation, or a combination of these. Both our study and the previous study used the same statistical approach, length of “wash-out” period to define new use, length of follow-up, and period used to define and ascertain baseline covariates.

The hd-PS algorithm may be a step towards the development of reliable methods for automatic selection of confounders. The algorithm might be especially helpful in settings in which little is known about the determinants and effects of treatment. However, the hd-PS algorithm lacks theoretical justification²⁴ and has several limitations. The approach may select covariates that are not related with the outcome or are “colliders”, which may increase the variance or introduce bias;^{4,25,26} it ranks the potential confounders based on their crude bivariate associations with treatment and outcome rather than on the associations adjusted for known confounders; and it deals only with binary exposures and covariates (except pre-specified covariates).

Another limitation of the hd-PS algorithm is the difficulty to extend the current version of the algorithm to time-varying treatments and confounders. The estimated hd-PS can be used to adjust for confounding through stratification, matching, regression, or inverse probability weighting,^{8,23} but only for confounding due to baseline variables. As a result, the method forces investigators to conduct either the observational analog of an “intention-to-treat” analysis with a baseline treatment variable (i.e., treatment initiation), or the observational analog of an “as-treated” analysis that does not appropriately adjust for time-varying confounders. The adequacy of the “intention-to-treat” approach is questionable when, as in our example, individuals spend a substantial proportion of the follow-up not taking the treatment that they initiated at baseline, even more so when this proportion is differential between the two groups compared (e.g., greater proportion of exposure among coxib users).

Meta-analyses of randomized controlled trials reported intention-to-treat estimates of 40–60% lower risk of UGIB among coxib initiators compared with tNSAID initiators.^{27,28} One might argue that the difference in the intention-to-treat estimates between our study and randomized trials is due to differences in treatment duration, which make the intention-to-treat estimates generally non-comparable. A better comparison between observational and randomized studies would involve “per-protocol” analyses that appropriately adjust for measured predictors of treatment duration by, say, inverse probability weighting.^{29–31}

Future methodologies for automatic confounder selection in electronic healthcare databases might outperform expert knowledge. However, those methodologies will face the same formidable challenges as the hd-PS algorithm. Accurate identification of some confounders of specific treatment-outcome pairs in these databases may involve complex algorithms with multiple diagnosis, procedure, and drug codes.⁷ Any automatic procedure will need to allow for selection of a set of validated codes that collectively defines a clinical event or important confounder. Further, the information recorded in electronic healthcare databases may not be accurate and there may be systematic coding errors. With hundreds of covariates involved in the hd-PS algorithm, data quality check becomes challenging.

Our study has several limitations. First, we used the NSAID-UGIB association as an example; it is unclear whether our findings regarding the hd-PS algorithm apply to other treatment-outcome pairs. Second, the number of UGIB cases was small, which resulted in wide 95% CIs and did not allow us to perform analyses stratified by baseline covariates (e.g., indication for treatment). Third, THIN records drugs prescribed and therefore does not necessarily capture drugs actually consumed. This may introduce bias due to treatment misclassification, but may also improve confounding adjustment for intention-to-treat estimates if the indications for prescription are better known and recorded than the reasons why individuals decide not to take a prescribed treatment. Finally, use of over-the-counter medications, such as some tNSAIDs, is not fully captured in THIN. Although patients over 60 years of age in the UK receive free prescriptions, the average duration of tNSAID use in our study would be underestimated if some of the initiators paid for their treatment after initiation; however, out-of-pocket purchases are likely to be rare.

In conclusion, our study shows that the semi-automated hd-PS algorithm can be implemented in pharmacoepidemiologic studies that use primary care databases such as THIN. In a setting in which major confounders are probably well-identified via expert knowledge, we showed that further adjustment for automatically selected covariates had little impact on the effect estimate.

Acknowledgments

Source of financial support: This study was partially funded by NIH grant R01 HL080644

We thank Drs. Sebastian Schneeweiss, MD, ScD and Jeremy Rassen, ScD for their helpful comments on an earlier draft of this paper. We also thank Dr. Rassen for his advice on applying the high-dimensional propensity score SAS macro, Oscar Fernández Cantero for creating the analytic datasets, and Roger Logan, PhD for additional programming support. Miguel Hernán was partly funded by NIH grant R01 HL080644.

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. Rubin DB. On the limitations of comparative effectiveness research. *Stat Med*. 2010; 29:1991–1995. discussion 1996–1997. [PubMed: 20683890]
4. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009; 20:512–522. [PubMed: 19487948]
5. 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]
6. 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]
7. García Rodríguez LA, Ruigomez A. Case validation in research using large databases. *Br J Gen Pract*. 2010; 60:160–161. [PubMed: 20202361]
8. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
9. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984; 79:516–524.
10. Wasserman, L. All of Nonparametric Statistics. Springer; New York: 2006.
11. 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]
12. Gutthann SP, García Rodríguez LA, Raiford DS. Individual nonsteroidal antiinflammatory drugs and other risk factors for upper gastrointestinal bleeding and perforation. *Epidemiology*. 1997; 8:18–24. [PubMed: 9116088]
13. 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]
14. 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]
15. García Rodríguez LA, Hernández-Díaz S. Relative risk of upper gastrointestinal complications among users of acetaminophen and nonsteroidal anti-inflammatory drugs. *Epidemiology*. 2001; 12:570–576. [PubMed: 11505178]
16. García Rodríguez LA, Hernández-Díaz S. Risk of uncomplicated peptic ulcer among users of aspirin and nonaspirin nonsteroidal antiinflammatory drugs. *Am J Epidemiol*. 2004; 159:23–31. [PubMed: 14693656]
17. Lanas A, García Rodríguez LA, Arroyo MT, et al. Risk of upper gastrointestinal ulcer bleeding associated with selective cyclo-oxygenase-2 inhibitors, traditional non-aspirin non-steroidal anti-inflammatory drugs, aspirin and combinations. *Gut*. 2006; 55:1731–1738. [PubMed: 16687434]
18. 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]
19. Schneeweiss S, Seeger JD, Maclure M, et al. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001; 154:854–864. [PubMed: 11682368]

20. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol.* 2006; 163:262–270. [PubMed: 16371515]
21. Weitzen S, Lapane KL, Toledano AY, et al. Weaknesses of goodness-of-fit tests for evaluating propensity score models: the case of the omitted confounder. *Pharmacoepidemiol Drug Saf.* 2005; 14:227–238. [PubMed: 15386700]
22. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med.* 2007; 26:734–753. [PubMed: 16708349]
23. Rosenbaum PR. Model-based direct adjustment. *J Am Stat Assoc.* 1987; 82:387–394.
24. Joffe MM. Exhaustion, automation, theory, and confounding. *Epidemiology.* 2009; 20:523–524. [PubMed: 19525688]
25. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol.* 2006; 163:1149–1156. [PubMed: 16624967]
26. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004; 15:615–625. [PubMed: 15308962]
27. 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]
28. 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]
29. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat.* 2008; 4:Article 22.
30. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008; 19:766–779. [PubMed: 18854702]
31. Danaei G, García Rodríguez LA, Logan R, et al. Observational data for comparative effectiveness: An emulation of randomized trials to estimate the effect of statins on primary prevention of coronary heart disease. *Stat Methods Med Res.* (in press).

Take home message

- We confirmed the feasibility of using a semi-automated confounding adjustment procedure, the high-dimensional propensity score (hd-PS) algorithm, in The Health Improvement Network (THIN), a primary care electronic medical record database in the United Kingdom.
- There was substantial confounding for the effect of non-steroidal anti-inflammatory drugs on upper gastrointestinal bleeding. Age, sex, calendar year, and summary measures of healthcare utilization appeared to account for much of the measured confounding.
- Compared with adjustment for investigator-identified variables only, adjustment using the hd-PS algorithm (with only age and sex included a priori) was closer to the unadjusted estimate.
- Compared with adjustment for investigator-identified variables only, further adjustment using the hd-PS algorithm had little impact on the estimates.

Figure.

The distribution of the estimated propensity score, i.e., the probability of initiating a coxib, for initiators of selective COX-2 inhibitors (coxibs) and non-selective (“traditional”) non-steroidal anti-inflammatory drugs (tNSAIDs) *

* The propensity score model included age; sex; calendar year; the number of distinct drugs used, physician visits, and hospitalization in the prior year; Charlson comorbidity score; prior use of gastroprotective drugs, anticoagulants, antiplatelets, and oral steroids; history of osteoarthritis, rheumatoid arthritis, upper gastrointestinal symptoms, dyspepsia, complicated or uncomplicated peptic ulcer disease, hypertension, congestive heart failure, and coronary artery disease. Table 2 shows how these covariates were categorized.

Table 1

Variables typically available in claims and electronic medical record databases

	Claims database (e.g., U.S. Medicare)	Electronic medical record database (e.g., The Health Improvement Network, UK)
Medication use		
Prescribed	No	Yes
Dispensed	Yes	No
Demographics		
Age	Yes	Yes
Sex	Yes	Yes
Race/ethnicity	Variable	Variable
Enrollment information	Yes	Variable
Medical diagnoses	Yes	Yes
Healthcare utilization		
Visits to primary care physicians	Yes	Yes
Visits to specialists	Yes	Yes
Visits to emergency departments	Yes	Yes
Hospitalizations	Yes	Yes
Laboratory tests ordered	Yes	Yes
Financial information (e.g., copay)	Yes	Variable
Other clinical data		
Vital signs (e.g., blood pressure)	No	Yes, at least in a subset
Results of laboratory tests	No	Yes, at least in a subset
Lifestyle factors		
Smoking	No	Yes, at least in a subset
Alcohol intake	No	Yes, at least in a subset
Body mass index	No	Yes, at least in a subset

Table 2

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

<i>Characteristics</i>	<i>%</i>		<i>OR (95% CI) of Coxib initiation*</i>
	<i>Coxib initiators (n=43,569)</i>	<i>tNSAID initiators (n=411,616)</i>	
Age (years)			
40–44	7.0	19.2	Reference
45–49	8.5	14.6	1.48 (1.40, 1.55)
50–54	10.7	14.1	1.77 (1.69, 1.86)
55–59	12.9	13.5	2.05 (1.96, 2.15)
60–64	12.4	12.2	2.10 (2.01, 2.21)
65–69	13.0	9.5	2.54 (2.42, 2.67)
70–74	14.0	7.8	3.06 (2.91, 3.22)
75–79	12.6	5.7	3.67 (3.48, 3.86)
80–84	9.1	3.4	4.08 (3.85, 4.32)
Male sex	36.1	43.3	0.78 (0.77, 0.80)
Calendar year of treatment initiation			
2000	5.1	10.2	Reference
2001	11.5	12.1	1.95 (1.84, 2.05)
2002	19.9	11.5	3.71 (3.53, 3.90)
2003	24.9	10.7	5.19 (4.94, 5.45)
2004	26.2	10.6	5.61 (5.35, 5.90)
2005	2.3	11.7	0.42 (0.38, 0.45)
2006	3.2	11.1	0.64 (0.59, 0.68)
2007	3.4	11.2	0.67 (0.62, 0.71)
2008	3.4	10.9	0.69 (0.64, 0.74)
Number of distinct drugs used in the prior year			
0–2	12.4	25.6	Reference
3–4	19.0	26.3	1.24 (1.20, 1.29)
5–7	24.5	23.9	1.39 (1.35, 1.45)
≥ 8	44.0	24.2	1.66 (1.59, 1.73)
Number of outpatient visits in the prior year			
0–3	17.0	25.2	Reference
4–6	21.3	25.6	1.05 (1.02, 1.08)
7–10	23.2	22.5	1.02 (0.99, 1.06)
≥ 11	38.4	26.8	1.02 (0.99, 1.06)
Hospitalized in the prior year	9.3	7.8	0.96 (0.92, 1.00)

<i>Characteristics</i>	<i>%</i>		<i>OR (95% CI) of Coxib initiation *</i>
	<i>Coxib initiators (n=43,569)</i>	<i>tNSAID initiators (n=411,616)</i>	
Charlson comorbidity score ≥ 1	39.8	26.3	1.12 (1.09, 1.15)
Prior use of			
Gastroprotective drugs	31.2	15.0	2.40 (2.34, 2.47)
Anticoagulants	1.9	0.8	1.23 (1.13, 1.35)
Antiplatelets	19.9	11.8	0.84 (0.81, 0.87)
Oral steroids	10.0	5.0	1.23 (1.18, 1.28)
History of			
Osteoarthritis	41.8	22.2	1.56 (1.52, 1.59)
Rheumatoid arthritis	3.4	1.2	1.76 (1.65, 1.89)
Dyspepsia	4.5	2.0	1.06 (1.00, 1.13)
Peptic ulcer disease	0.4	0.1	1.45 (1.18, 1.78)
Hypertension	37.5	27.2	0.98 (0.96, 1.01)
Congestive heart failure	2.8	1.1	0.95 (0.88, 1.02)
Coronary artery disease	14.6	7.5	1.09 (1.05, 1.14)
Smoking			
Non smoker	51.5	51.5	Reference
Current smoker	19.1	20.9	1.04 (1.01, 1.07)
Past smoker	22.1	21.2	1.00 (0.97, 1.03)
Unknown	7.4	6.4	0.92 (0.87, 0.96)
Alcohol drinking (drinks/week)			
None	41.7	37.6	Reference
1–9	27.6	28.9	0.95 (0.92, 0.97)
10–19	8.1	10.9	0.95 (0.91, 0.98)
≥ 20	5.3	6.9	0.97 (0.92, 1.02)
Unknown	16.5	15.7	1.12 (1.08, 1.17)
Body mass index (kg/m²)			
<18.5	1.1	0.9	1.07 (0.96, 1.19)
18.5–24.9	28.0	30.5	Reference
25.0–29.9	32.9	32.3	1.03 (1.00, 1.06)
30.0–34.9	14.7	13.5	1.06 (1.03, 1.10)
≥ 35	6.4	6.3	1.04 (0.99, 1.09)

<i>Characteristics</i>	<i>%</i>		<i>OR (95% CI) of Coxib initiation</i> [*]
	Coxib initiators (n=43,569)	tNSAID initiators (n=411,616)	
Unknown	16.9	16.5	1.06 (1.01, 1.10)

OR: odds ratio; CI: confidence interval

^{*} Adjusted for all other covariates in the table.

Table 3

Odds ratios and their 95% confidence intervals of upper gastrointestinal bleeding during the first 180 days after initiation of selective COX-2 inhibitors (coxibs) vs. non-selective (“traditional”) non-steroidal anti-inflammatory drugs

Covariates included in the propensity score model	Odds ratio (95% confidence interval)	c-statistic of the propensity score model
None	1.50 (0.98, 2.28)	--
Strategy 1: Investigator-identified covariates only		
Age, sex	1.04 (0.68, 1.59)	0.66
+ Calendar year	0.98 (0.63, 1.52)	0.77
+ Number of distinct drugs used, physician visits and hospitalizations in the prior year	0.84 (0.54, 1.31)	0.79
+ Covariates typically available in both claims and EMRs [*]	0.81 (0.52, 1.27)	0.80
+ Covariates typically available in EMRs [†]	0.81 (0.51, 1.26)	0.80
Strategy 2: Investigator-identified covariates plus 500 empirically identified covariates		
Covariates typically available in both claims and EMRs [*]	0.78 (0.49, 1.22)	0.81
Covariates typically available in EMRs [†]	0.79 (0.50, 1.25)	0.81
Strategy 3: Basic covariates plus 500 empirically identified covariates		
Age, sex	0.87 (0.56, 1.34)	0.74
+ Calendar year	0.81 (0.51, 1.27)	0.81
+ Number of distinct drugs used, physician visits and hospitalization in the prior year	0.81 (0.52, 1.28)	0.81

^{*} In addition to age; sex; calendar year; and the number of distinct drugs used, physician visits, and hospitalization in the prior year, the covariates further included Charlson comorbidity score; prior use of gastroprotective drugs, anticoagulants, antiplatelets, and oral steroids; history of osteoarthritis, rheumatoid arthritis, upper gastrointestinal symptoms, dyspepsia, complicated or uncomplicated peptic ulcer disease, hypertension, congestive heart failure, and coronary artery disease. Table 2 shows how these covariates were categorized.

[†] Included the covariates above, plus smoking, alcohol intake, and body mass index. Table 2 shows how these covariates were categorized.