NIH Public Access
Author Manuscript
Epidemiology. Author manuscript; available in PMC 2011 April 14.

Published in final edited form as:
Epidemiology. 2009 July ; 20(4): 512–522. doi:10.1097/EDE.0b013e3181a663cc.

High-dimensional propensity score adjustment in studies of treatment effects using health care claims data

Sebastian Schneeweiss, **Jeremy A. Rassen**, **Robert J. Glynn**, **Jerry Avorn**, **Helen Mogun**, and **M. Alan Brookhart**
Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine Brigham and Women's Hospital/Harvard Medical School

Abstract

Background—Adjusting for large numbers of covariates ascertained from patients' health care claims data may improve control of confounding, as these variables may collectively be proxies for unobserved factors. Here we develop and test an algorithm that empirically identifies candidate covariates, prioritizes covariates, and integrates them into a propensity-score-based confounder adjustment model.

Methods—We developed a multi-step algorithm to implement high-dimensional proxy adjustment in claims data. Steps include (1) identifying data dimensions, e.g. diagnoses, procedures, and medications, (2) empirically identifying candidate covariates, (3) assess recurrence of codes, (4) prioritizing covariates, (5) selecting covariates for adjustment, (6) estimating the exposure propensity score, and (7) estimating an outcome model. This algorithm was tested in Medicare claims data, including a study on the effect of Cox-2 inhibitors on reduced gastric toxicity compared to nonselective nonsteroidal anti-inflammatory drugs (NSAIDs).

Results—In a population of 49,653 new users of Cox-2 inhibitors or nonselective NSAIDs, a crude relative risk (RR) for upper GI toxicity (RR = 1.09 [95% confidence interval = 0.91–1.30]) was initially observed. Adjusting for 15 predefined covariates resulted in a possible gastroprotective effect (0.94[0.78–1.12]). A gastroprotective effect became stronger when adjusting for an additional 500 algorithm-derived covariates (0.88[0.73–1.06]). Results of a study on the effect of statin on reduced mortality were similar. Using the algorithm adjustment confirmed a null finding between influenza vaccination and hip fracture (1.02[0.85–1.21]).

Conclusion—In typical pharmacoepidemiologic studies, the proposed high-dimensional propensity score resulted in improved effect estimates compared with adjustment limited to predefined covariates, when benchmarked against results expected from randomized trials.

Large health care utilization databases are frequently used to estimate the causal effect of prescription drugs on health outcomes.[1] Health care utilization data reflect routine practice, are large enough to study rare drug effects, and avoid the delays common in the collection of primary data.[2,3] Despite their importance, studies of pharmacoepidemiologic claims data have been criticized for the incompleteness of information on potential confounders such as markers of clinical disease severity, laboratory results, functional status, body mass index, smoking status, and over-the-counter medication use. Such factors may lead to selective prescribing, which may in turn result in biased estimates of the association between drugs and health outcomes.[4] Longitudinal claims data contain information about patient health

---

Address for correspondence: Sebastian Schneeweiss, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street (suite 3030), Boston, MA 021205. Phone: (617) 278-0937, Fax: (617) 232-8602, schneeweiss@post.harvard.edu.

status and confounding beyond what is normally used in pharmacoepidemiologic research. We have explored the utility of this additional information with the help of a computer algorithm examining all health care claims data.

Proxy adjustment

Longitudinal health care claims data can be understood and analyzed as a set of proxies that indirectly describe the health status of patients. This status is presented through the lenses of health care providers recording their findings and interventions via coders and operating under the constraints of a specific health care system.[5] Quite often, several levels of proxies, which we call chains of proxies, are involved. For example, the health state of a patient can be (a) assessed through the dispensing of a drug that was (b) prescribed by a physician who made a diagnosis in a (c) patient who came forward for medical care and (d) presented certain symptoms (Figure 1). Such a chain of proxies is influenced by access to care,[6] severity of the condition, diagnostic ability of the physician, preference for one drug over another,[7] the patient's ability to pay the medication co-payment,[8] and the accurate recording of the dispensed medication. Here, the chain of proxies leads to a reasonable interpretation that the patient indeed had a condition that troubled the patient enough to see the physician, and that was severe enough for the physician to treat and for the patient to pay a co-payment for the medication. Medical evidence and treatment options were weighed in several steps along the way. These are not observable in claims data, but collectively they resulted in a measurable action.

The measured action in this case had a clear interpretation—the prescribed medication addressed a specific condition—but such interpretations are not always possible. In fact, we cannot determine the exact interpretation in most cases but, at the same time, an exact interpretation is not required for effective confounder adjustment. For example, old age serves a proxy for comorbidity, frailty, cognitive decline, and many other factors. Adjusting for a perfect surrogate of an unmeasured factor is equivalent to adjusting for the factor itself.[9] The degree to which a surrogate is related to an unobserved or imperfectly observed confounder is proportional to the degree to which adjustment can be achieved.[10,11]

If we could measure a battery of proxies, we would increase the likelihood that in combination they are a good overall proxy for relevant unobserved confounding factors. Using a large number of proxy covariates for propensity score estimation and then estimating the average causal treatment effect conditional on deciles of the propensity score may result in improved control for confounding in epidemiologic studies of treatment effects using claims data compared with models that have fewer covariates. Some authors have explored the use of very large propensity score models in nonrandomized assessment of treatment effects.[12,13] In some studies, large propensity score models resulted in better control of confounding than estimating the propensity score with fewer covariate information.[14,15,16] A major challenge remains, however, to identify a very large pool of potential covariates that can be implemented in claims data, and then to identify which are influential enough in the treatment/disease relationship to include in an analysis.

We propose an algorithm that identifies a large number of covariates in claims databases, eliminates covariates with very low prevalence and minimal potential for causing bias, and then uses propensity score techniques to adjust for a large number of target covariates. The approach will be illustrated using three pharmacoepidemiologic studies of intended treatment effects in elderly patients, including statin use and reduced risk of death, use of selective Cox-2 inhibitors and reduced risk of GI complications, and influenza vaccination and hip fractures.

Methods

We describe a generic algorithm that identifies a large number of target covariates in claims databases and selects covariates for propensity score adjustment to minimize residual confounding. We present seven steps to achieve high-dimensional propensity score adjustment using health care claims databases. These steps assume that the cohort, exposure, and outcome have already been defined.

1. Specify data sources

A wide range of databases of health care utilization data ("claims") is available for use in pharmacoepidemiology.[3] Each database is arranged in specific ways using a variety of classifications to code diagnoses (e.g. International Classification of Diseases [ICD]-8 through ICD-10), procedures (e.g. Current Procedural Terminology, Canadian Classification of Proceddures, ICD-9-Clinical Modification), or medications (e.g. National Drug Codes, American Hospital Formulary Services, Anatomical Therapeutic Chemical Classification). Beyond these basic data dimensions and coding systems, many more data dimensions can be found in such databases. Some databases provide additional dimensions such as laboratory results, other electronic medical record information, and accident registries.

We propose an algorithm that is independent of the specific data source as long as the source's data dimensions can be identified. In Figure 2 we provide a flow diagram using a typical example of data dimensions available in US Medicare claims data linked to medication use data. First, a temporal window must be defined in which baseline covariates will be identified. A frequent choice is 6 or 12 months preceding the initiation of the study or comparison drug.[2] The recording of diagnoses and procedures is correlated with the frequency of health care encounters. Therefore, longer baseline periods increase the number of encounters and therefore yield more covariate information.[2]

The most basic patient information always available to typical databases is age, sex and calendar time. We assume that given their ubiquity, these demographic covariates will always be adjusted for.

Additional covariates can then be identified from the various data dimensions, but it is first necessary to identify variables that should not be part of covariate adjustment. While it is generally recommended to include many covariates in a propensity score regression model, in specific cases researchers may exclude variables from covariate adjustment.[17] Surrogates for the exposure (i.e. covariates that are strong correlates of the study exposure but not associated with the outcome) will not only increase standard errors but may also increase bias—and should therefore not be included in propensity score analyses.[18,19] Bias can also occur through the inclusion of so-called "collider" variables, although this bias is generally thought to be weak.[20,21] In our example study comparing statin initiation with glaucoma drug initiation, diagnostic codes for glaucoma should not be included in a propensity score because of their close correlation with treatment choice. [22,23] At this stage of the procedure, such codes can be identified and removed from the dimension data input to the algorithm. We have developed a screening tool for such covariates as part of the algorithm that will help investigators identify and remove such covariates (eAppendix 1, http://links.lww.com).

2. Identify candidate empirical covariates

Within each of p data dimensions (e.g. outpatient diagnostic ICD codes, inpatient procedure codes, and drugs dispensed) codes were sorted by their prevalence. Prevalence was measured as the proportion of patients having a specific code at least once during a 6-month baseline period. Since the prevalence of a binary factor is symmetrical around 0.5, we subtracted all prevalence estimates larger than 0.5 from 1.0. The top n most prevalent codes were identified as candidate empirical covariates. If fewer than 100 patients were identified with a covariate, the covariate was dropped.

The prevalence of each code (and therefore its empirical ranking) depends on the granularity of the coding; ICD-9 codes are hierarchical such that each additional digit provides more detail of the diagnosis. Considering the fourth or fifth digit of the ICD-9 code will reduce the prevalence of the code in the data but may be a better proxy for the underlying confounder. We initially set the granularity to 3 digits for ICD-9 data for this illustration, since every system using ICD-9 records at least 3 digits. Granularity decisions need to be considered for all data dimensions, including medication coding. Depending on the application, therapeutic class may be sufficiently detailed and in other settings individual drugs or even drug dose or preparations may be required. We chose the individual drug level for the base-case algorithm.

3. Assess recurrence

For the top *n* most prevalent codes in each data dimension, we assessed how frequently that code was recorded for each patient during the baseline period. We divided each code into three binary variables: code occurred ≥ 1 times, ≥ median number of times, and ≥ 75th percentile number of times. A code that appeared above the 75th percentile number of times would have a “true” value for all three recurrence variables. If any of the values were equal, the variable representing the higher cutpoint was dropped. For a data structure with *p* data dimensions this results in up to $p*n*3$ covariates.

4. Prioritize covariates

If we now wish to combine information from all *p* data dimensions to reduce the total number of covariates, we need to consider that the average prevalence of codes can be quite different among dimensions. From our experience, prevalence of procedure codes, including codes for simple office visits, have a higher prevalence than drug dispensings. Simply combining these tables and picking the top *k* prevalent candidate covariates would down-weight the importance of medication-dispensing in controlling for confounding. Further, Brookhart et al.[24] showed that including patient characteristics in the propensity score that are associated with the exposure but not the outcome will increase variance of the estimator with no improvement in confounding control, and in some situations can actually introduce confounding. We therefore decided to prioritize covariates across data dimensions by their potential for controlling confounding that is not conditional on exposure and other covariates. Because we are exclusively dealing with binary covariates, the confounded or apparent relative risk (ARR) is a function of the imbalance in prevalence of a binary confounding factor among exposed (P_{C1}) and unexposed (P_{C0}) subjects as well as the independent association between a confounder and the study outcome (RR_{CD}) [25]:

$$ARR = RR \times \frac{P_{C1}(RR_{CD}-1)+1}{P_{C0}(RR_{CD}-1)+1}, \quad \text{if } RR_{CD} \geq 1$$

$$ARR = RR \times \frac{P_{C1}\left(\frac{1}{RR_{CD}}-1\right)+1}{P_{C0}\left(\frac{1}{RR_{CD}}-1\right)+1}, \quad \text{if } RR_{CD} < 1$$

The fraction on the right side of the equation is the multiplicative bias term, Bias_M. We then sorted all $p*n*3$ covariates by the magnitude of $|\log(Bias_M)|$ in descending order. We chose multiplicative bias assessment because a bias term on the absolute risk scale ($Bias_A = |RD_{CE} \times RD_{CD}|$) would implicitly down-weight the association between a confounder and outcome if the outcome event rate is small but the prevalence of the exposure high, a typical occurrence in pharmacoepidemiologic cohort studies. The covariate prioritization is illustrated for binary variables since our algorithm to generate target covariates exclusively creates binary variables.

5. Select covariates

Once this prioritization of covariates was accomplished, we included the top k covariates from step 4, which could be as large as $p*n*3$ when including all candidate covariates. Our base case settings were $p = 8$ and $n = 200$ resulting in 4,800 candidate covariates. We selected the top $k = 500$ binary empirical covariates (about 10 %) for inclusion in the propensity score modeling.

In addition to these k binary empirical covariates, we included covariates that should always be adjusted for if available, including d binary demographic covariates age, sex, race (available in Medicare data), and calendar year. In addition to these, we allowed the investigator to force l binary, categorical, or numeric pre-defined covariates into the propensity score model, based on context knowledge regarding the specific study question.

In a subanalysis we explored the impact of adjusting for 2-way interaction terms. Of the k empirical covariates, we selected the 20 highest priority empirical covariates and computed multiplicative 2-way interactions among those covariates and with the demographic and predefined covariates, resulting in another $(20+d+l)*(20+d+l)/2$ covariates.

6. Estimate exposure propensity score

Using multivariate logistic regression, a propensity score was estimated for each subject as the predicted probability of exposure conditional on all $d + l + k$ covariates.

7. Estimate propensity score-adjusted outcome models

We grouped subjects into propensity score deciles and used multivariate logistic regression analyses to model the study outcome as a function of exposure and indicator terms for decile of propensity score. In addition to an adjusted estimate we computed a standardized mobility-ratio (SMR)-weighted estimate using weights of 1 for subjects in the study drug group and the odds of the propensity score (PS) for members of the comparison group $[PS/(1-PS)]$. SMR-weighted estimates provide treatment effect estimates among the treated. As the output of Step 6 includes each subject’s propensity score, other ways to utilize propensity scores in the outcome estimation may be applied, including matching, inverse probability of treatment weighting, or modelling the propensity score as continuous variable.[26]

The high dimensional propensity score algorithm is implemented as a SAS macro available at www.drugepi.org.

Example data sources and study cohorts

All three study cohorts were drawn from a population of patients aged 65 years and older enrolled in both Medicare and the Pennsylvania Pharmaceutical Assistance Contract for the Elderly (PACE) programs between 1995 and 2002. PACE is a state pharmaceutical benefits program with incomes below $14,000 for individuals and below $17,200 for couples; its data have been frequently used for pharmacoepidemiologic studies.[27,28] All prescription drugs commercially available in the US during the study period were fully covered by PACE, requiring a nominal copayment of $6. Prescription drug information was assessed based on pharmacy claims from PACE with detailed and highly accurate information[29,30] on drug name, dosage, quantity, and date of dispensing.

Study exposures and outcomes

Example cohort 1—Initiation of nonselective NSAID use vs. selective Cox-2 inhibitor use was defined if an eligible beneficiary filled at least one prescription for an NSAID between 1 January 1999 and 31 December 2002 but did not use any NSAID during the 18 months prior to the index date. The index date was the first date an NSAID prescription was filled.[31] The follow-up period included the 180 days after the initiation of therapy.

The study outcome of severe gastrointestinal (GI) complication was defined as either a hospitalization for GI hemorrhage or peptic ulcer disease complications including perforation (coded as ICD-9 discharge diagnoses 531.x, 532.x, 533.x, 534.x, 535.x or 578.x in the first or second position or a physician service code for GI hemorrhage). These definitions were validated in 1,762 patients in a hospital discharge database, with a composite positive predictive value (PPV) of 90%.[32] We expected to find a moderate protective effect of Cox-2 inhibitors on GI complications,[33,34,35] which may be concealed by confounding.[36]

Example cohort 2—The initial exposure status of statin use, nonuse, or comparator drug use as determined from pharmacy claims was carried forward until censoring after 1 year or death, whichever came first. We analyzed the extent to which patients classified as non-users started statins during follow-up and how many statin users discontinued use, using a gap of 90 or more days in addition to the dispensed supply without statin use as the definition for statin discontinuation.

We then used Medicare claims data to ascertain time to death. Death information from Medicare records is routinely cross-checked with Social Security data. Subjects were censored at the end of 365 days after drug initiation or disenrollment from the pharmacy assistance program. We expected to find a moderate protective effect of statins on mortality in older adults (RR about 0.85)[37] that may be exaggerated by confounding.[38]

Example cohort 3—For the previous examples we hypothesized protective effects of the drug therapy with confounding going either toward the null (example 1) or away from the null (example 2). We added a third example with a strong prior hypothesis of a null association, based on context knowledge. This is the relationship between influenza vaccination in elderly people and the risk of hip fracture. A typical pre-flu season (1 October through 31 December 1996) was selected to assess the exposure to influenza vaccine, and the next 4 months (January–April 1997) were the follow-up period. Patients with prior hip fractures of bisphosphonate use for the treatment of osteoporosis were excluded. Patients were censored after the occurrence of the study endpoint, death or dis-enrollment.

Overall analytic strategy

In order to make the comparisons among models that contained a varying number of covariates as fair as possible, we used the available covariates for propensity score estimation and then adjusted the respective outcome models (logistic regression for examples 1 and 2 and Cox proportional hazard regression for example 3) for deciles of the estimated propensity score. We report the numbers of covariates entered into each propensity score model as well as its c-statistic of model discrimination.

Results

Population characteristics of the two sample cohort studies are presented in Tables 1 and 2. 32,042 subjects initiated selective Cox-2 inhibitors. The 17,611 nonselective NSAID initiators were older and had more comorbidities and more risk factors for GI complications.

The NSAID initiators had 185 GI complication in 180 days (1.1%), and the Cox-2 initiators had 367 events (1.2%). Compared with 14,889 glaucoma drug initiators, the 21,233 initiators of statin therapy were younger, more likely to have cardiovascular risk factors, have more health care utilization, and about equal numbers of comorbidities. The statin initiators had 784 deaths in 1 year (3.7%); the glaucoma drug initiators had 955 deaths (6.4%).

In a traditional multivariable analysis comparing Cox-2 inhibitors and NSAIDs (Table 3) we observed no association with GI complications (RR=0.94; 95% confidence interval [CI] 0.78–1.12) which was slightly reduced from an unadjusted RR of 1.09, suggesting that additional adjustment for residual confounding would move the relative-risk further towards a protective effect.

Considering statin initiators vs. non-statin using initiators of glaucoma drugs (Table 4), we observed a strongly reduced risk of 1-year mortality (RR=0.80; 0.70–0.90) which is closer to the expected results from RCTs in elderly people then an unadjusted analysis (RR=0.56), suggesting that additional adjustment for residual confounding would move the relative risk further towards the null.

In the first two example studies, we observed several trends regarding the performance of the high-dimensional propensity score algorithms:

1. Adding the high-dimensional propensity scores to the predefined covariates moved the point estimate in the expected direction (Cox-2 inhibitors toward a more protective effect, statins toward a less protective effect), consistent with RCT findings.
2. Results of the high-dimensional propensity score alone were identical to the second decimal place compared with the high-dimensional propensity score combined plus predefined covariates (Models 5 and 5a in Tables 3 and 4). SMR-weighted propensity score outcome models resulted in effect estimates of 0.77 (0.67–0.88) for Cox-2 inhibitors and GI complication, and 0.64 (0.59–0.70) for statins and death.
3. First stage c-statistics quantifying the degree of exposure prediction did not consistently correlate with the changes in effect estimates.
4. Including 500 rather than 200 covariates in the high-dimensional propensity score appeared to move the effect estimate little.
5. More finely granulated diagnostic codes (4 digit ICD-9 vs. 3 digit ICD-9) appeared to move the effect estimate little.
6. Dropping the recurrence assessment appeared to move effect estimates slightly away from the expected direction.
7. Including 2-way interactions appeared to leave the effect estimate unchanged or move it slightly away from the expected direction.
8. Selecting covariates based only on their prevalence without any further covariate prioritization moved effect estimates slightly away from the expected direction in the Cox-2 inhibitor example but not in the statin example.

Bootstrapped 95% confidence intervals based on 1000 samples were very similar to the base case algorithm (Model 5) in both example studies (0.73–1.06 in Table 3 and 0.76–0.98 in Table 4).

For the third example study, on the relationship between influenza vaccination in seniors and hip fractures, we expected a null association with strong contextual support of that null hypothesis. In a cohort of 147,583 patients, 42% received influenza vaccination and we observed 710 hip fractures (0.5%). Due to confounding we observed a protective effect in an unadjusted analysis (RR = 0.93; 0.80–1.08). After adjustment for demographic factors and the high-dimensional propensity score, the effect was entirely explained (RR = 1.02; 0.85–1.21).

Discussion

We hypothesized that high-dimensional proxy adjustment based on propensity score techniques could reduce residual confounding in claims databases of treatment effects. To explore this hypothesis, we developed a generic algorithm that identifies a large number of target covariates and selects covariates for propensity score adjustment to facilitate high-dimensional propensity score adjustment. In the example studies of drug-outcome relationships, we found that the application of the high-dimensional propensity score algorithm produced results closer to the expected findings based on randomized trials, compared with propensity score adjustment that uses a more limited number of investigator pre-defined covariates. We further found that some components of the algorithm were more important than others in our example studies. The covariate prioritization as well as the assessment and adjustment of recurrent coding of health services seem to be important contributors to the algorithm in our examples.

The algorithm's main strength rests on the exploitation of information that is usually untapped in epidemiologic analyses of health care utilization databases. It also includes a variable selection component to limit the number of adjusted covariates to an arbitrary number (500 in our base case) because in theory a number of covariates larger than the number of study subjects could be generated. Hirnao and Imbens[16] have developed a propensity score variable selection algorithm that is based on comparing the t-statistics of the entire propensity score regression model (t_{prop}) with those of individual covariates (t_k). Such test-based approaches to variable selection have been criticized for their dependency on study size and potential for bias.[39] This approach also seems impractical if the number of candidate covariates is very large (e.g. the 4,800 in our example) which can provide a challenge to fitting the entire PS regression model even in large datasets.

Brookhart et al.[40,41] found that the inclusion of variables that predict only exposure but not outcome can result in larger standard errors in small studies and, if residual confounding exists, can increase bias. These effects were not evident here, probably because of the large sample size and the modest associations between individual covariates and exposure. However, the possibility that an empirically generated variable could increase bias and variance represents the primary limitation of the algorithm. Users of this methodology should remove covariates that are a priori expected to be strong predictors of exposure but not likely to be related to the outcome. An example of such a covariate would be the history of glaucoma as a strong predictor for glaucoma treatment (but not death) that was removed in our second example cohort. Further research is needed to consider ways in which automatically generated claims-based covariates could generate collider bias, and how they could be identified and removed.

Variable selection techniques may result in falsely narrow standard errors.[42] We therefore applied bootstrapping to estimate 95% confidence intervals[43] and found very similar confidence intervals compared with the simple logistic regression of the base case algorithm. This is not surprising as we did not apply any confounder selection algorithms that resulted in multiple tests of exposure-outcome associations, including change-in-estimate, forward or backward selection.[39] Instead, we did a preliminary screen of candidate confounders by estimating unconditional associations of individual potential confounders with the exposure and then separately with the outcome. We then ranked covariates with regard to their potential for being a confounder and included these candidate covariates up to a predefined maximum number. We observed fairly weak unconditional associations of individual factors.

The present study is an empirical comparison of methods without a true gold standard. We used randomized trial findings to set specific expectations regarding the treatment effect estimates, but it ultimately remains unanswerable whether our high-dimensional propensity score algorithm reached that goal fully. Simulation studies are unlikely to clarify the performance of this algorithm because it is inherently empirical and relies on data-generating mechanisms that will vary from study to study, and thus are difficult to pre-specify. The strength of the high-dimensional propensity score is rather that it does not make any assumptions about data quality, quantity, and interpretation. Further validation of the algorithm is possible by replicating findings that are expected based on randomized trial findings, including our example of a null association in which the high-dimensional propensity score algorithm eliminated all confounding. A specific point of concern is the performance of a covariate prioritization strategy that considers the association of each factor with the study outcome if outcomes are rare. At some point the prioritization rule may miss potentially important confounders by chance. While this is a theoretically important point it needs to be seen in light of the fact that the proposed high-dimensional proxy adjustment will be used in addition to adjustment for factors specified by the investigator.

Further work may provide improved ways of selecting covariates or using the covariates in the analysis. For example, optimization of the variable selection algorithm may be possible by considering the association between the candidate covariate and the exposure, conditional on either a set of predefined covariates or the entire list of selected covariates. It is also possible that algorithms based on cross-validation could prove to be useful for covariate selection.[44,45] Finally, we have considered the use of selected covariates in an analysis that depends on correct specification of the propensity score model. Doubly robust approaches are based on an assumed model of both the exposure and outcome, but are consistent if only one of the models is correctly specified.[18] The use of the selected covariates in the setting of a doubly robust estimator may improve the performance of the algorithm. However, since the outcome model must be more parsimonious, a separate list of covariates would need to be generated for the outcome model – perhaps those with particularly strong outcome associations.

It is too early to conclude that the proposed algorithm or variations thereof will be able to substitute for existing confounder adjustment strategies in claims data analyses, although in our limited examples the algorithm performed better than standard techniques. Practical advantages are that the algorithm can be run efficiently on a large scale, it reduces investigator and programming time substantially, and it reduces programming errors and potential mischaracterization of covariate definitions or adjustments without a loss of validity. This last point might be of particular practical advantage in studies pooling multiple claims databases.

In conclusion, in some typical pharmacoepidemiologic studies of treatment effects, the proposed proxy adjustment via high-dimensional propensity scores generated effect estimates closer to randomized trial findings, compared with standard covariate adjustment of predefined covariates. Further replication will be necessary in a variety of settings to assess the value of this approach.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: Supported by grants from the National Institute of Mental Health (RO1-MH078708), and the National Institute on Aging (RO1-AG021950, RO1-AG023178, RO1-AG018833, K25-AG027400).

References

1. Arana A, Rivero E, Egberts TCG. What do we show and who does so? An analysis of the abstracts presented at the 19th ICPE. Pharmacoepidemiol Drug Saf. 2004; 13:S330–S331.
2. Schneeweiss S, Avorn J. Using health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005; 58:323–37. [PubMed: 15862718]
3. Strom BL, Carson JL. Use of automated databases for Pharmacoepidemiology research. Epidemiol Rev. 1990; 12:87–107. [PubMed: 2286228]
4. Walker AM. Confounding by indication. Epidemiology. 1996; 7:335–336. [PubMed: 8793355]
5. Schneeweiss S. Understanding secondary databases: a commentary on “Sources of bias for health state characteristics in secondary databases”. J Clin Epidemiol. 2007; 60:648–50. [PubMed: 17573976]
6. Anderson RM. Revisiting the behavioral model and access to medical care: Does it matter? J Health Social Behav. 1995; 36:1–10.
7. Schneeweiss S, Glynn RJ, Avorn J, Solomon DH. A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors. J Clin Epidemiol. 2005; 58:98–102. [PubMed: 15649677]
8. Roblin DW, Platt R, Goodman MJ, Hsu J, Nelson WW, Smith DH, Andrade SE, Soumerai SB. Effect of increased cost-sharing on oral hypoglycemic use in five managed care organizations: how much is too much? Med Care. 2005; 43:951–9. [PubMed: 16166864]
9. Wooldridge, JM. Econometric Analysis of Cross Section and Panel Data. MIT Press; Cambridge, MA: 2001.
10. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol. 1980; 112:564–9. [PubMed: 7424903]
11. Greenland S, Robins JM. Confounding and misclassification. Am J Epidemiol. 1985; 122:495–506. [PubMed: 4025298]
12. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. Pharmacoepidemiol Drug Saf. 2005; 14:465–76. [PubMed: 15651087]
13. McAfee AT, Ming EE, Seeger JD, Quinn SG, Ng EW, Danielson JD, Cutone JA, Fox JC, Walker AM. The comparative safety of rosuvastatin: a retrospective matched cohort study in over 48,000 initiators of statin therapy. Pharmacoepidemiol Drug Saf. 2006; 15:444–53. [PubMed: 16761308]
14. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. Med Care. 2007; 45(Supl 2):S143–8. [PubMed: 17909373]
15. Patrick AR, Schneeweiss S, Glynn RJ, Brookhart A, Avorn J, Sturmer T. Applying propensity score analyses to clinical outcomes studies on the effects of statins on acute MI, hip fracture, and short term mortality in seniors. Under review.
16. Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Serv Outcomes Res Method. 2001; 2:259–78.
17. Judkins DR, Morganstein D, Zador P, Piesse A, Barrett B, Mukhopadhyay P. Variable selection and ranking in propensity scoring. Stat Med. 2007; 26:1022–33. [PubMed: 16708347]
18. Van der Laan, M.; Robins, JM. Unified Methods for Censored Longitudinal Data and Causality. Springer; New York: 2003.

19. Robins JM, Mark SD, Newey WK. Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders. Biometrics. 1992; 48:479–95. [PubMed: 1637973]
20. Brookhart MA, Schneeweiss S, Rothman K, Glynn RJ, Avorn J, Sturmer T. Variable selection in propensity score models. Am J Epidemiol. 2006; 163:1149–56. [PubMed: 16624967]
21. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003; 14:300–6. [PubMed: 12859030]
22. Schneeweiss S, Patrick AR, Sturmer T, Brookhart MA, Avorn J, Maclure M, Rothman K, Glynn RJ. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. Medical Care. 2007; 45:S131–42. [PubMed: 17909372]
23. Patrick AR, Schneeweiss S, Glynn RJ, Brookhart A, Avorn J, Sturmer T. Applying propensity score analyses to clinical outcomes studies on the effects of statins on acute MI, hip fracture, and short term mortality in seniors.
24. Brookhart MA, Schneeweiss S, Rothman K, Glynn RJ, Avorn J, Sturmer T. Variable selection in propensity score models. Am J Epidemiol. 2006; 163:1149–56. [PubMed: 16624967]
25. Bross IDJ. Spurious effects from an extraneous variable. J Chronic Dis. 1966; 19:637–647. [PubMed: 5966011]
26. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006; 59:437–47. [PubMed: 16632131]
27. Schneeweiss S, Glynn RJ, Avorn J, Solomon DH. A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors. J Clin Epidemiol. 2005; 58:98–102. [PubMed: 15649677]
28. Solomon DH, Schneeweiss S, Glynn RJ, Kiyota Y, Levin R, Mogun H, Avorn J. The relationship between selective COX-2 inhibitors and acute myocardial infarction. Circulation. 2004; 109:2068–73. [PubMed: 15096449]
29. McKenzie DA, Semradek J, McFarland BH, Mullooly JP, McCamant LE. The validity of medicaid pharmacy claims for estimating drug use among elderly nursing home residents: The Oregon experience. J Clin Epidemiol. 2000; 53:1248–57. [PubMed: 11146272]
30. West S, Savitz DA, Koch G, Strom BL, Guess HA, Hartzema A. Recall accuracy for prescription medications: self report compared with database information. Am J Epidemiol. 1995; 142:1103–12. [PubMed: 7485055]
31. Ray W. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003; 158:915–20. [PubMed: 14585769]
32. Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. Epidemiology. 1996; 7:101–4. [PubMed: 8664388]
33. Moore RA, Derry S, Makinson GT, McQuay HJ. Tolerability and adverse events in clinical trials of celecoxib in osteoarthritis and rheumatoid arthritis: systematic review and meta-analysis of information from company clinical trial reports. Arthritis Res Ther. 2005; 7:R644–65. [PubMed: 15899051]
34. Watson DJ, Harper SE, Zhao PL, Quan H, Bolognese JA, Simon TJ. Gastrointestinal tolerability of the selective cyclooxygenase-2 (COX-2) inhibitor rofecoxib compared with nonselective COX-1 and COX-2 inhibitors in osteoarthritis. Arch Intern Med. 2000; 160:2998–3003. [PubMed: 11041909]
35. Eisen GM, Goldstein JL, Hanna DB, Rublee DA. Meta-analysis: upper gastrointestinal tolerability of valdecoxib, a cyclooxygenase-2-specific inhibitor, compared with nonspecific nonsteroidal anti-inflammatory drugs among patients with osteoarthritis and rheumatoid arthritis. Aliment Pharmacol Ther. 2005; 21:591–8. [PubMed: 15740543]
36. Schneeweiss S, Solomon DH, Wang PS, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective COX-2 inhibitors and non-selective NSAIDs: an instrumental variable analysis. Arthritis Rheum. 2006; 54:3390–8. [PubMed: 17075817]

37. Roberts CG, Guallar E, Rodriguez A. Efficacy and safety of statin monotherapy in older adults: a meta-analysis. Gerontol A Biol Sci Med Sci. 2007; 62:879–87.
38. Glynn RJ, Schneeweiss S, Wang P, Levin R, Avorn J. Selective prescribing can lead to overestimation of the benefits of lipid-lowering drugs. J Clin Epidemiol. 2006; 59:819–28. [PubMed: 16828675]
39. Greenland S. Invited commentary: Variable selection versus shrinkage in the control of multiple confounders. Am J Epidemiol. 2008; 167:523–9. [PubMed: 18227100]
40. Lefebrve G, Delaney JAC, Platt RW. Impact of mis-specification of the treatment model on estimates from a marginal structural model. Stat Med. 2008 e-published ahead of print.
41. Fu AZ, Li L. Thinking of having a higher predictive power for your first-stage model in propensity score analysis? Think again. Health Serv Outcomes Res Method. 2008 epub ahead of print.
42. Freedman, DA.; Navidi, W.; Peters, SC. On the impact of variable selection in fitting regression equations. In: Dijlestra, TK., editor. On model uncertainty and its statistical implications. Berlin, Germany: Springer; 1988. p. 1-16.
43. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, and what? Stat Med. 2000; 19:1141–64. [PubMed: 10797513]
44. van der Laan, MJ.; Dudoit, S. Unified cross-validation methodology for selection among estimators and a general crossvalidated adaptive epsilon-net estimator: finite sample oracle inequalities and examples. (UC Berkeley Division of Biostatistics Working Paper Series, paper 130). Berkeley, CA: Division of Biostatistics, University of California, Berkeley; 2003. (http://www.bepress.com/ucbbiostat/paper130)
45. Brookhart MA, van der Laan MJ. A semiparametric model selection criterion with applications to the marginal structural model. Comput Stat Data Anal. 2006; 50:475–98.

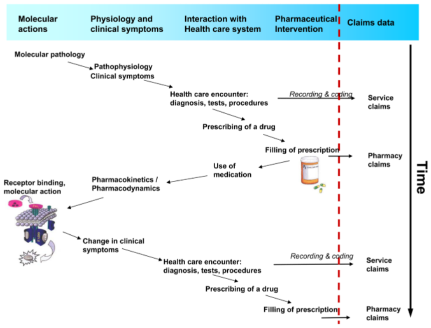
Molecular actions
Physiology and clinical symptoms
Interaction with Health care system
Pharmaceutical Intervention
Claims data
Molecular pathology
Pathophysiology
Clinical symptoms
Health care encounter: diagnosis, tests, procedures
Recording & coding
Service claims
Prescribing of a drug
Filling of prescription
Pharmacy claims
Use of medication
Pharmacokinetics / Pharmacodynamics
Receptor binding, molecular action
Change in clinical symptoms
Health care encounter: diagnosis, tests, procedures
Recording & coding
Service claims
Prescribing of a drug
Filling of prescription
Pharmacy claims
Time

Figure 1.

Proxies in health care utilization databases.

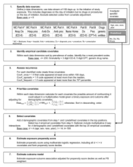
1. Specify data sources
2. Identify empirical candidate covariates
3. Assess recurrence
4. Prioritize covariates
5. Select covariates
6. Estimate exposure propensity score
7. Estimate outcome model

Figure 2.
Flow chart for basic high-dimensional propensity score algorithm.

Table 1

Characteristics of 49,653 initiators of selective COX-2 inhibitors or nonselective NSAIDs as defined during 6 months prior to first medication use.

	Initiators of Cox-2 selective NSAIDs (n=32,042)	**Initiators of nonselective NSAIDs (n=17,611)**	
	No.(%)	**No.(%)**	**OR(95% CI)**
Age 75 years or older	24,079(75)	11,496(65)	1.61(1.55–1.67)
Female sex	27,528(86)	14,293(81)	1.42(1.35–1.49)
Race: white	30,583(95)	15,808(90)	2.39(2.23–2.57)
black	1,133(4)	1,580(9)	0.37(0.34–0.40)
other	326(1)	223(1)	0.80(0.68–0.95)
Charlson comorbidity score >= 1	24,343(76)	12,521(71)	1.29(1.23–1.34)
Use of >4 distinct drugs in prior year	24,120(75)	11,852(67)	1.48(1.42–1.54)
>4 physician visits in prior year	22,919(72)	11,363(65)	1.38(1.33–1.44)
Hospitalized in prior year	9,804(31)	4,591(26)	1.25(1.20–1.30)
Nursing home resident	2,671(8)	996(6)	1.52(1.41–1.64)
Prior use of gastroprotective drugs	8,785(27)	3,600(20)	1.47(1.41–1.54)
Prior use of warfarin	4,252(13)	1,153(7)	2.18(2.04–2.34)
Prior use of oral steroids	2,800(9)	1,373(8)	1.13(1.06–1.21)
History of OA	15,549(49)	5,898(33)	1.87(1.80–1.95)
History of RA	1,602(5)	476(3)	1.90(1.71–2.10)
History of peptic ulcer disease	1,189(4)	426(2)	1.55(1.39–1.74)
History of gastrointestinal hemorrhage	551(2)	196(1)	1.55(1.32–1.83)
History of hypertension	23,332(76)	12,363(70)	1.14(1.09–1.18)
History of congestive heart failure	9,727(30)	4,328(25)	1.34(1.28–1.40)
History of coronary artery disease	5,266(16)	2,603(15)	1.13(1.08–1.19)

Table 2

Characteristics[a] of 36,122 initiators of statin drugs or initiators of glaucoma drugs as defined during 6 months prior to first medication use.

	Initiators of statins[a] (n=21,233)	**Initiators of glaucoma drugs[a] (n=14,889)**	**OR (95% CI)**
Demographic Variables			
Age (years); mean (SD)	75.8 (6.0)	80.4 (6.8)	0.90 (0.89–0.90)
Female sex	17,205 (81.0)	12,329 (82.8)	0.89 (0.84–0.94)
Race: white	19,675 (92.7)	13,355 (89.7)	1.45 (1.35–1.56)
black	1,319 (6.2)	1,369 (9.2)	0.65 (0.61–0.71)
other	239 (1.1)	165 (1.1)	1.02 (0.83–1.24)
Cardiovascular conditions			
Comorbidity score; mean (SD)	1.9 (2.0)	1.9 (2.1)	1.00 (0.99–1.02)
Prior MI	2,124 (10.0)	478 (3.2)	3.35 (3.03–3.71)
Prior CABG or PTCA	1,627 (7.7)	146 (1.0)	8.38 (7.07–9.94)
Angina	7,828 (36.9)	3,807 (25.6)	1.70 (1.62–1.78)
Atrial fibrillation	2,486 (11.7)	1,872 (12.6)	0.92 (0.87–0.98)
Cardiovascular diagnoses; mean (SD)	4.8 (4.4)	3.5 (3.4)	1.09 (1.08–1.10)
Cardiovascular system symptoms	2,804 (13.2)	1,474 (9.9)	1.39 (1.30–1.48)
Chest pain	7,055 (33.2)	3,664 (24.6)	1.52 (1.45–1.60)
Complications of heart disease	2,774 (13.1)	1,649 (11.1)	1.21 (1.13–1.29)
Conduction disorders	1,311 (6.2)	778 (5.2)	1.19 (1.09–1.31)
Heart failure	6,008 (28.3)	4,349 (29.2)	0.96 (0.91–1.00)
Coronary atherosclerosis	9,550 (45.0)	4,351 (29.2)	1.98 (1.89–2.07)
Diabetes	7,714 (36.3)	4,903 (32.9)	1.16 (1.11–1.22)
Hyperlipidemia	14,984 (70.6)	4,190 (28.1)	6.12 (5.85–6.41)
Hypertension	16,692 (78.6)	10,907 (73.3)	1.34 (1.28–1.41)
Ischemic heart disease	10,680 (50.3)	5,104 (34.3)	1.94 (1.86–2.03)
Palpitations	940 (4.4)	492 (3.3)	1.36 (1.21–1.52)
Peripheral vascular disease	4,991 (23.5)	3,648 (24.5)	0.95 (0.90–0.99)
Stroke, transient ischemic attack	3,781 (17.8)	1,911 (12.8)	1.47 (1.39–1.56)
Other Comorbid Conditions			
Alzheimer disease	785 (3.7)	1,019 (6.8)	0.52 (0.48–0.58)
Cancer	4,907 (23.1)	3,833 (25.7)	0.87 (0.83–0.91)
Depression	1,163 (5.5)	877 (5.9)	0.93 (0.85–1.01)
COPD	3,965 (18.7)	2,825 (19.0)	0.98 (0.93–1.03)
Prior hip fracture (frailty marker)	75 (0.4)	157 (1.1)	0.33 (0.25–0.44)
Osteoporosis (frailty marker)	1,733 (8.2)	1,497 (10.1)	0.80 (0.74–0.86)
Urinary tract infection (frailty marker)	3,251 (15.3)	2,455 (16.5)	0.92 (0.87–0.97)
Renal disease	961 (4.5)	590 (4.0)	1.15 (1.04–1.28)
Parkinson’s disease	233 (1.1)	281 (1.9)	0.58 (0.48–0.69)

	Initiators of statins[a] **(n=21,233)**	**Initiators of glaucoma drugs**[a] **(n=14,889)**	**OR (95% CI)**
Use of medications			
No. drugs taken; mean (SD)	7.9 (5.0)	8.2 (5.1)	0.99 (0.99–1.00)
Cardiovascular	17,866 (84.1)	11,477 (77.1)	1.58 (1.50–1.66)
Loop diuretic use	3,561 (16.8)	2,797 (18.8)	0.87 (0.83–0.92)
NSAIDs	6,125 (28.8)	4,748 (31.9)	0.87 (0.83–0.91)
Estrogen	1,174 (5.5)	674 (4.5)	1.23 (1.12–1.36)
Use of health care services			
Preventive care	14,580 (67.9)	9,878 (66.3)	1.11 (1.06–1.16)
Hospitalization w/cardiac diagnosis	4,177 (19.7)	1,378 (9.3)	2.40 (2.25–2.56)
Electrocardiogram	13,273 (62.5)	8,589 (57.7)	1.22 (1.17–1.28)
Hospitalization	7,068 (33.3)	4,318 (29)	1.22 (1.17–1.28)
Lab test ordered	14,879 (70.1)	9,715 (65.2)	1.25 (1.19–1.30)
Lipid test ordered	8,554 (40.3)	3,245 (21.8)	2.42 (2.31–2.54)
Nursing home stay	825 (3.9)	855 (5.7)	0.66 (0.60–0.73)
No office visits; mean (SD)	9.3 (6.3)	9.8 (6.8)	0.99 (0.98–0.99)
With cardiovascular diagnosis	7.8 (7.5)	6.5 (6.9)	1.03 (1.02–1.03)

[a] No. (%), except otherwise indicated.

MI indicates myocardial infarction; CABG, coronary artery bypass graft; PTCA, percutaneous transluminal coronary angioplasty; COPD, chronic obstructive pulmonary disease.

Table 3

Variations in covariate adjustment and relative risk estimates for the association of selective Cox-2 inhibitors and GI complications within 180 days of first medication use (n=49,653)

Model #	Covariates included in propensity score model	Number of covariates adjusted	Variables tested per data source	Data source granularity	Covariate prioritization algorithm	c-statistic of PS model	Outcome model RR(95% CI)
1	Unadjusted					-	1.09 (0.91–1.30)
2	Age, sex, race, year[a]	*d*=4				0.61	1.01 (0.84–1.21)
3	+ predefined covariates (Table 1)	*d*=4; *l*=14				0.66	0.94 (0.78–1.12)
4	+ empirical covariates	*d*=4;*l*=14;*k*=200	*n*=200	3-digit ICD	Bias$_{mult}$	0.69	0.86 (0.72–1.04)
5[b]	+ empirical covariates	*d*=4;*l*=14;*k*=500	*n*=200	3-digit ICD	Bias$_{mult}$	0.71	0.88 (0.73–1.06)
5b	Only demographics + empirical covariates	*d*=4; *k*=500	*n*=200	3-digit ICD	Bias$_{mult}$	0.71	0.87 (0.72–1.05)
6	+ empirical covariates	*d*=4;*l*=14;*k*=200	*n*=200	3-digit ICD	Bias$_{add}$	0.68	0.87 (0.72–1.05)
7	+ empirical covariates	*d*=4;*l*=14;*k*=500	*n*=200	3-digit ICD	Bias$_{add}$	0.70	0.87 (0.72–1.05)
8	+ empirical covariates	*d*=4;*l*=14;*k*=200	*n*=200	4-digit ICD	Bias$_{mult}$	0.69	0.89 (0.74–1.07)
9	+ empirical covariates	*d*=4;*l*=14;*k*=500	*n*=200	4-digit ICD	Bias$_{mult}$	0.71	0.85 (0.70–1.03)
10	+ empirical covariates	*d*=4;*l*=14;*k*=200	*n*=200	3-digit ICD, 6-digit AHFS[c]	Bias$_{mult}$	0.69	0.87 (0.72–1.05)
11	+ empirical covariates	*d*=4;*l*=14;*k*=500	*n*=200	3-digit ICD, 6-digit AHFS[c]	Bias$_{mult}$	0.71	0.87 (0.72–1.06)
12	+ empirical covariates without recurrence assessment (drop step 3 of hd-PS algorithm)	*d*=4;*l*=14;*k*=200	*n*=200	3-digit ICD	Bias$_{mult}$	0.70	0.89 (0.73–1.07)
13	+ empirical covariates without recurrence assessment (drop step 3 of hd-PS algorithm)	*d*=4;*l*=14;*k*=500	*n*=200	3-digit ICD	Bias$_{mult}$	0.71	0.91 (0.75–1.10)
14	+ empirical covariates with 2-way interactions	*d*=4;*l*=14; *k*=200 + 38^2	*n*=200	3-digit ICD	Bias$_{mult}$	0.70	0.86 (0.71–1.04)
15	+ empirical covariates with 2-way interactions	*d*=4;*l*=14; *k*=500 + 38^2	*n*=200	3-digit ICD	Bias$_{mult}$	0.71	0.90 (0.74–1.09)
16	+ empirical covariates without prioritization (drop step 4 of hd-PS algorithm)	*d*=4;*l*=14; *k*=7*100	*n*=100	3-digit ICD	None	0.71	0.91 (0.75–1.10)

[a] Age (65–74; 75 +); race (white, black, other); year = year of cohort entry (1999, 00–01, 02–03)

[b] Model 5 represents the Base Case model for a high-dimensional propensity score as outlined in Figure 2

[c] The 6-digit AHFS (American Hospital Formulary Service) code aggregates individual drugs to drug classes, e.g. beta blockers, ACE inhibitors. d = number of demographic covariates; l = number of predefined covariates; n = number of variables considered per data source base on prevalence ranking; k = number of covariates eventually included in the propensity score estimation. hd–PS indicates high-dimensional propensity score.

Table 4

Variations in covariate adjustment and relative risk estimates of the association of statin initiation with 1-year mortality (n=36,122)

Model #	Covariates included in propensity score model	Number of covariates adjusted	Variables tested per data source	Data source granularity	Covariate prioritization algorithm	c-statistic of PS model	Outcome model Relative risk (95% CI)
1	Unadjusted						0.56 (0.51–0.62)
2	Age, sex, race, year[a]	d=4				0.70	0.77 (0.69–0.85)
3	+ predefined covariates (Table 2)	d=4;l=42				0.82	0.80 (0.70–0.90)
4	+ empirical covariates	d=4;l=14;k=200	n=200	3-digit ICD	$\text{Bias}_{\text{mult}}$	0.86	0.86 (0.76–0.98)
5[b]	+ empirical covariates	d=4;l=14;k=500	n=200	3-digit ICD	$\text{Bias}_{\text{mult}}$	0.87	0.86 (0.76–0.98)
5b	Only demographics + empirical covariates	d=4; k=500	n=200	3-digit ICD	$\text{Bias}_{\text{mult}}$	0.86	0.89 (0.78–1.02)
6	+ empirical covariates	d=4;l=14;k=200	n=200	3-digit ICD	Bias_{add}	0.85	0.85 (0.75–0.96)
7	+ empirical covariates	d=4;l=14;k=500	n=200	3-digit ICD	Bias_{add}	0.86	0.88 (0.77–1.00)
8	+ empirical covariates	d=4;l=14;k=200	n=200	4-digit ICD	$\text{Bias}_{\text{mult}}$	0.85	0.86 (0.76–0.97)
9	+ empirical covariates	d=4;l=14;k=500	n=200	4-digit ICD	$\text{Bias}_{\text{mult}}$	0.87	0.87 (0.76–0.99)
10	+ empirical covariates	d=4;l=14;k=200	n=200	3-digit ICD, 6- digit AHFS[c]	$\text{Bias}_{\text{mult}}$	0.86	0.85 (0.75–0.97)
11	+ empirical covariates	d=4;l=14;k=500	n=200	3-digit ICD, 6- digit AHFS[c]	$\text{Bias}_{\text{mult}}$	0.87	0.86 (0.75–0.98)
12	+ empirical covariates without recurrence assessment (drop step 3 of hd-PS algorithm)	d=4;l=14;k=200	n=200	3-digit ICD	$\text{Bias}_{\text{mult}}$	0.86	0.83 (0.73–0.950
13	+ empirical covariates without recurrence assessment (drop step 3 of hd-PS algorithm)	d=4;l=14;k=500	n=200	3-digit ICD	$\text{Bias}_{\text{mult}}$	0.87	0.87 (0.76–1.000
14	+ empirical covariates with 2-way interactions	d=4;l=14; k=200 + 67^2	n=200	3-digit ICD	$\text{Bias}_{\text{mult}}$	0.87	0.83 (0.73–0.950
15	+ empirical covariates with 2-way interactions	d=4;l=14; k=500 + 67^2	n=200	3-digit ICD	$\text{Bias}_{\text{mult}}$	0.88	0.85 (0.75–0.970
16	+ empirical covariates without prioritization (drop step 4 of hd-PS algorithm)	d=4;l=14; k=7*100	n=100	3-digit ICD	None	0.88	0.88 (0.77–1.00)

[a] Age (65–74; 75 +); race (white, black, other); calendar time = year of cohort entry (1995–97, 98–00, 01–02).

[b] Model 5 represents the Base Case model for a high-dimensional propensity score as outlined in Figure 2.

[c] The 6-digit AHFS (American Hospital Formulary Service) code aggregates individual drugs to drug classes, e.g. beta blockers, ACE inhibitors. d = number of demographic covariates; l = number of predefined covariates; n = number of variables considered per data source base on prevalence ranking; k = number of covariates eventually included in the propensity score estimation.