



Published in final edited form as:

Med Care. 2010 June ; 48(6 0): S114–S120. doi:10.1097/MLR.0b013e3181d8e3e3.

Confounding control in healthcare database research: challenges and potential approaches

MA Brookhart^{1,2}, T Stürmer², RJ Glynn¹, J Rassen¹, and S Schneeweiss¹

¹Department of Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital & Harvard Medical School, Boston, MA.

² Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina at Chapel Hill, NC.

Abstract

Epidemiologic studies are increasingly used to investigate the safety and effectiveness of medical products and interventions. Appropriate adjustment for confounding in such studies is challenging because exposure is determined by a complex interaction of patient, physician, and healthcare system factors. The challenges of confounding control are particularly acute in studies using healthcare utilization databases where information on many potential confounding factors is lacking and the meaning of variables is often unclear. We discuss advantages and disadvantages of different approaches to confounder control in healthcare databases. In settings where considerable uncertainty surrounds the data or the causal mechanisms underlying the treatment assignment and outcome process, we suggest that researchers report a panel of results under various specifications of statistical models. Such reporting allows the reader to assess the sensitivity of the results to model assumptions that are often not supported by strong subject-matter knowledge.

Keywords

confounding; unmeasured confounding; propensity scores; variable selection; pharmacoepidemiology

Introduction

Epidemiologic studies are increasingly used to investigate the safety and effectiveness of medical products and procedures as they are used in routine care in unselected patient populations. One of the principal problems of such studies is confounding—a systematic difference between a group of patients exposed to an intervention and a chosen comparator group. Confounding in studies of medical treatments often arises when the factors that influence physician treatment decisions and patient medication use are also independent determinants of health outcomes. Statistical approaches such as multivariable outcome models and propensity score methods can be used to remove the confounding effects of such factors if they are captured in the data. These approaches require that all confounding factors are accurately measured and their effects on the exposure or outcome are correctly modeled. However, because the processes that determine treatment choices and outcomes are often complex and poorly understood, it may be difficult to correctly specify the necessary

statistical models. Furthermore, in studies based on healthcare utilization databases, many potential confounding factors are missing or poorly measured and the meaning of measured variables may be unclear.¹

We describe some potential sources of confounding in studies of medication use and outcomes. We then review some features of healthcare databases that complicate confounding control. Finally, we consider some theoretical aspects of confounding from the perspective of causal graph theory^{2,3} and describe how statistical control for a variable can either increase or decrease bias in the presence or absence of a true treatment-outcome association. In light of these theoretical issues and the limitations of healthcare databases, we consider the strengths and limitations of two different approaches to confounder control in healthcare databases.

Sources of Confounding in Studies of Medications and Healthcare Services

The use of a medical intervention by patients is determined by system-, physician- and patient-level variables that may often interact in complex and poorly understood ways. For example, physicians' treatment decisions may be based on an evaluation of the patient's health status and prognosis, the physician's past experience with the medication, or an assessment of the patient's ability and willingness to take a medication as prescribed. Patients may initiate and remain adherent to a new therapeutic regimen because of their health beliefs, preferences about treatment, their trust in the physician, out-of-pocket cost of the medication, or perceptions of their disease risk and the benefits of treatment. Treatment initiation and adherence may also depend on a patient's physical and cognitive abilities. Patient and physician variables that determine use of a treatment may directly affect health outcomes, or be related to them through indirect pathways. From this process, several sources of bias can result.

Confounding by indication/disease severity—A common and often intractable form of confounding arises from good medical practice: physicians' tendency to prescribe medications to and perform procedures on patients who are most likely to benefit. Because it is often difficult to assess medical indications and underlying disease severity and prognosis, confounding by indication often makes medications appear to cause outcomes they are meant to prevent.^{4,5} For example, statins, lipid-lowering drugs, reduce risk of cardiovascular events in patients with cardiovascular risk factors. Thus these drugs tend to be prescribed to patients perceived to be at increased cardiovascular risk. Incomplete control of cardiovascular risk factors can make statins appear to cause rather than prevent cardiovascular events.

Confounding by functional status and cognitive impairment—Patients who are functionally impaired (defined as having difficulty performing daily activities of living) may be less able to visit a physician or pharmacy and therefore may be less likely to collect prescriptions and receive preventive healthcare services. This phenomenon could exaggerate the benefit of prescription medications, vaccines, and screening tests. For example, functional status appeared to be a strong confounder in studies of both the effect of NSAIDs

and the influenza vaccine on all-cause mortality in the elderly.⁶⁻⁸ A similar form of confounding could result from differences in cognitive functioning among elderly patients.

Selective prescribing and treatment discontinuation of preventive medications in frail and very sick patients

—Patients who are perceived by a physician to be close to death or who face serious medical problems may be less likely to be prescribed preventive medications. Similarly, patients may decide to discontinue preventive medications when their health deteriorates. These phenomena arise through different processes but have the same effect: decreased exposure to preventive medication in patients close to death. One or both of these mechanisms may be responsible for the substantially decreased mortality risk observed among elderly users of statins and other preventive medications compared with apparently similar non-users.^{9, 10} Patients with certain chronic diseases or patients who take many medications may also be less likely to be prescribed a potentially beneficial medication over concern about drug-drug interactions or metabolic problems. For example, patients with end-stage renal disease are less likely to receive medications for secondary prevention after myocardial infarction.¹¹

The healthy user/ adherer bias—Patients who initiate a preventive medication may be more likely than other patients to engage in other healthy, prevention-oriented behaviors. For example, patients who start a preventive medication may be more likely to seek out preventive healthcare services, exercise regularly, moderate their alcohol consumption, and avoid unsafe and unhealthy activities. Incomplete adjustment for such behaviors can make use of preventive medications spuriously associated with reduced risk of a wide range of adverse health outcomes. Similarly, patients who adhere to treatment also may be more likely to engage in other healthy behaviors.^{12, 13} Strong evidence of this “healthy adherer” effect comes from a meta analysis of randomized controlled trials where adherence to placebo was been found to be associated with reduced mortality risk.¹⁴ This is clearly not an effect of the placebo but is rather due to characteristics of patients who take a medication as prescribed. The healthy adherer bias is also evident in studies that reported associations between statin adherence and an increased use of preventive healthcare services and a decreased risk of accidents.^{15,16}

Access to healthcare—Patients may vary substantially in their ability to access healthcare. For example, patients who live in rural areas may have to drive long distances to receive specialized care. Other obstacles to accessing healthcare include cultural factors (e.g., trust in medical system), economic factors (e.g., ability to pay), immigration status, and institutional factors (e.g., prior authorization programs, restrictive formularies), all of which may have some direct or indirect effects on study outcomes.

The limitations of healthcare utilization databases

The data commonly used for epidemiologic studies of medical products and procedures in routine care typically come from existing healthcare utilization data (e.g., claims data from Medicare or private insurance companies) or electronic medical records databases. Such databases often contain information on large, heterogeneous populations of patients that are geographically dispersed. However, because the data were not collected as part of designed

study, many variables that the researcher might wish to have access to remain unrecorded.¹ Furthermore, the meaning of the data elements that are available is often unclear. In healthcare claims data, inpatient and outpatient diagnosis codes can be difficult to interpret. Although principal discharge diagnosis codes for certain medical conditions such as hip fracture, myocardial infarction, major surgical procedures, and most cancers are reliably recorded,¹⁷ codes for many other medical conditions are coded with limited accuracy.¹⁷ Furthermore, a single diagnosis code depends not only on the presence of a particular condition but also on a provider's use of the codes. For example, psychiatric diagnosis codes may be used routinely by mental healthcare providers, but omitted by other medical providers. The absence of a diagnosis code may not indicate the absence of a particular condition, but rather the presence of a more serious condition that pushes the milder condition off the diagnosis list or discharge abstract. For example, a diagnosis of depression would be expected to be associated with decreased mortality risk among elderly patients because it would be less likely to be coded in patients with other serious medical conditions. Studies of hospital discharge abstract data have found that secondary diagnoses such as diabetes, angina, and history of myocardial infarction are paradoxically associated with improved in-hospital mortality.¹⁸ This appears to reflect an under-coding of chronic comorbidities in patients who die in the hospital.

Information may also be missing in informative ways. For example, data on erythropoiesis stimulating agent (ESA) use by end-stage renal disease patients with anemia is reported to Medicare by outpatient dialysis clinics, but not by hospitals. This can create spurious associations between ESA exposure and morbidity and mortality because hospitalized patients will appear to be getting smaller ESA doses.^{19, 20} In these same patients, hematocrit laboratory values must be reported to Medicare to justify reimbursement for ESA claims. Therefore, ESRD patients with high hematocrit who do not need ESAs as well as critically ill patients who are having treatment withdrawn have missing hematocrit lab values. This creates informative patterns of missingness that may be problematic if hematocrit level is being used as a confounder, subgroup identifier, or outcome variable.

Data from electronic medical records may contain clinical measurements and some lifestyle variables, but these are also selectively recorded and often inaccurate. For example, history of smoking in the General Practice Research Database has only a 60% sensitivity.¹⁵

This discussion is not exhaustive and is only meant to represent some of the known issues with healthcare databases.

Causal graphs: a theoretical approach to confounder control

For confounding control, the researcher wishes to select a set of adjustment variables that, appropriately included in a statistical model, will yield an unbiased and efficient estimate of a causal effect of interest. Unfortunately, the property of being a confounder is an issue of causality and is not statistically identifiable (testable) in data.²¹⁻²³ Because a confounder cannot be identified using statistical criteria, no model selection approach can guarantee unbiased estimates, even in very large studies.

The theory of causal directed acyclic graphs provides a theoretical framework for using subject-matter knowledge to identify minimal sets of variables that must be included in a statistical model to eliminate confounding bias.^{2, 3} These graphs require that the analyst specify the causal relations between every relevant variable for a given problem; only the exposure-outcome relation can be in doubt. Causal graph theory has been helpful in understanding and describing potential bias in many empirical studies. For example, these approaches have helped to reveal source of bias in studies of infant birth weight.²⁴ Unfortunately, in many studies of medical interventions, the available subject-matter knowledge is inadequate to specify with any degree of certainty the causal connections between variables that determine exposure or outcome. This problem is worsened in the setting of healthcare database research where many variables are not directly measured and the meaning of diagnosis codes is not always clear.

Variable selection: over-adjustment versus under-adjustment

In the absence of a causal graph that determines the set of variables that one should include in a statistical model, researchers must consider each variable individually using some combination of subject-matter knowledge and statistical criteria to decide whether to include the variable in the model. It is well known that the failure to adjust for an important confounder (or its correlates) will result in a biased estimate of an exposure effect. Therefore “under-adjustment” is a clear threat to the validity of studies using healthcare databases. However, it is less appreciated that certain variables can introduce or increase bias when included in a statistical model. We review the types of variables that can lead to “over-adjustment” bias.

Intermediate variables—Adjustment for variables that are affected by treatment can introduce bias in an estimate of the total effect of treatment.²⁵⁻²⁷ This bias can be avoided by selecting appropriate study designs, such as the new user design in which the covariates are measured prior to the start of exposure.²⁸ Prevalent user studies and case-control studies may be more subject to this bias as there is often not a clear temporal ordering of covariates and exposure. When adjustments need to be made for time-varying covariates, for example in the setting of outcome-related treatment changes, appropriate statistical models, such as marginal structural models, are needed.^{29,30}

Bias resulting from statistical control of an instrumental variable (Z-bias)—Inclusion of variables that are strongly related to the exposure, but unrelated to the outcome (i.e., instrumental variables), can increase the variance and bias of an estimated exposure effect when added to a statistical model. The increase in bias happens when there exists residual bias due to uncontrolled confounding. We term these variables “Z-variables” as Z is often used to denote an instrumental variable. We term the bias induced by the inclusion of such variables “Z-bias.” The bias increases as the amount of uncontrolled confounding increases and the strength of the association between the Z-variable and exposure increases.³¹

Z-bias has been evident in simulation studies,³² has been discussed theoretically,²² and has been demonstrated analytically.³¹ It may seem counterintuitive that a variable that only

affects exposure could increase bias when included in a statistical model. However, consider that within strata of the exposure a Z -variable will be related to the outcome via a pathway through the unmeasured confounder (Figure 1). Any plausible instrumental variable could potentially introduce Z -bias in the presence of uncontrolled confounding. For example, if there are large differences in medical practice between physicians, hospitals, or geographic regions – three variables used in prior studies as instruments³³⁻³⁸ -- including variables representing these differences could plausibly introduce Z -bias. Similarly, if there are strong secular trends in medication use and time is not related to any uncontrolled confounding variables,³⁹⁻⁴² adjustment for calendar time could increase bias. Because some unmeasured confounding is likely in studies based on healthcare utilization data, Z -bias may be common.

Collider-stratification bias (M-bias)—In Figure 2, we depict a situation in which adjustment for a non-confounding variable can increase bias even in the absence of uncontrolled confounding. Here, adjustment for C , which is a common effect of X_1 and X_2 , but not independently related to exposure or outcome, creates bias if both X_1 and X_2 are excluded from the statistical model. Such a variable is called a “collider” as pathways from X_1 and X_2 collide at C . Adjustment for C creates an association between X_1 and X_2 and opens a path from the exposure to the outcome and therefore could create a spurious association between the exposure and outcome.^{43, 44} From the shape of the causal graph, this bias has also been termed M -bias.⁴³ For example, in an observational study of the effect of anti-depressant use and lung cancer, let X_1 be depression, C be cardiovascular disease, and X_2 be smoking status. Here no adjustment is necessary to obtain unbiased estimates, but adjustment for cardiovascular disease without adjustment for smoking status or depression could introduce bias in the estimated effect of anti-depressant use on lung cancer. M -bias has been shown to be generally smaller than confounding bias.⁴³

When included in a statistical model, any of these bias-increasing variables may be predictive of the outcome and shift the estimated exposure effect and therefore be identified as a possible confounders by most variable selection strategies.

Practical strategies for variable definition and selection

Given the complex determinants of treatment use, the limitations of the available data, and theoretical issues of confounder control, we now consider some approaches to controlling confounding in studies based on healthcare databases.

A priori defined covariates—The usual approach to confounding control in healthcare database research is to define an *a priori* set of covariates. Variable selection strategies may then be used to select which of these to include in a statistical model or recently proposed shrinkage estimators could be used to model their effect.⁴⁵ These covariates typically include demographic characteristics, history of major medical comorbidities, measures of overall comorbidity (such as the Charlson⁴⁶ and Romano scores⁴⁷), history of medication use, history of acute care hospitalizations, and measures of health care system use, such as numbers of physician visits.

The advantage of an *a priori* approach is that covariate definitions are typically based on subject matter knowledge such that the analyst can attempt to adjust for all relevant medical

conditions that might affect the outcome or the use of treatment. Also, given the relatively limited set of covariates included, the analyst can assess each one for its potential to introduce *Z*-bias or *M*-bias by drawing causal diagrams. The major limitation of this approach is that the typical set of covariates will not include many of the potentially important confounding factors discussed previously that are not directly observable in healthcare claims data. Another practical problem is the possibility of missing or mis-defining a medical condition as a result of changes in the use of diagnostic codes.

Non-parsimonious or “proxy” adjustment—Seeger et al suggested that medical claims may serve as proxies for important unmeasured variables in hard-to-predict ways.⁴⁸ In a study of statin use and cardiovascular outcomes, they found that certain healthcare utilization variables, such as frequency of lipid tests ordered and physician visits, were strong predictors of statin initiation and appeared also to be strong confounders. In their analysis, a simple propensity score model that omitted these covariates appeared to yield a more biased estimate of the exposure effect than a larger propensity score model.⁴⁸ Although frequency of lipid testing does not directly affect cardiovascular risk, it could be viewed as a proxy for concern about disease risk. Therefore, frequency of testing may be associated with other risk-modifying behaviors or to underlying risk. Other healthcare claims may serve as proxies for other important unmeasured confounders. For example, the use of oxygen canisters could be a proxy for failing health and functional impairment, and adherence to preventive medications may be a proxy for important health-related behaviors, such as regular exercise.

This idea has motivated the construction of very large propensity score models,^{49, 50} that provide a way to statistically control for large numbers of covariates when exposure is common but the outcomes is rare.⁵¹ For example, Johannes et al constructed a propensity score model from healthcare claims that considered as candidate variables the 100 most frequently occurring procedures, diagnoses, and outpatient medications.⁵² In a recent paper, Schneeweiss et al proposed an algorithm to create a very large set of covariates defined from healthcare utilization data. The algorithm prioritizes covariates for inclusion in a propensity score model based on their association with both the exposure and outcome.⁵³ In several example analyses where the true effect of an exposure was approximately known, the algorithm appeared to perform as well as or better than approaches based on *a priori* defined covariates.⁵³

The approach of defining covariates prior to exposure, will not “over-adjust” by including intermediate variables. However, the procedure could introduce bias by identifying variables that could lead to *M*-bias or variables that behave like instrumental variables and lead to *Z*-bias. Therefore, the analyst should attempt to remove such variables from the set of identified variables. For example, variables that are strong predictors of exposure but have no obvious relation to the outcome should be considered potential sources of *Z*-bias.

Combined Approaches—The empirically-driven covariate definition algorithm just described and the *a priori*-defined covariate selection represent poles on a continuum. Certainly some combination of the two approaches could be helpful. For example, the automated procedure could be used to identify a set of potentially important confounders

that are then manually examined and selected for inclusion in the model. The resulting set of potentially important confounders could then be added to a model that included a set of *a priori* defined covariates. This more limited use of empirically-derived covariates may reduce confounding while simultaneously reducing the risk of including variables that could increase bias. One downside of this approach is that one may end up with a large number of covariates that need to be evaluated, leading to investigator exhaustion if multiple exposures and outcomes need to be studied.⁵⁴ However, such an approach could also be used to identify important potential confounders, such as frequency of lipid testing, that could be used in future studies.

The particular approach that one adopts for healthcare database research will depend on the researchers' substantive knowledge and whether one is more concerned about missing a potentially important confounder and thus under-adjusting *versus* including a harmful covariate and thus over-adjusting. In healthcare database research under-adjustment may often be the greater concern.

Uncontrolled confounding

Regardless of the approach that one adopts to control confounding in healthcare database research, there usually exists the possibility of some bias due to unmeasured confounders. Often researchers assess the sensitivity of analyses to assumptions about hypothetical unmeasured confounders.⁵⁵⁻⁵⁸ When information is available on a single potential unmeasured confounder, for example from a sub-sample or from external data, it is possible to remove residual confounding attributable to the unmeasured confounder from the effect estimates.^{59, 60} When external data are available on multiple unmeasured confounders, the method of propensity score calibration may be used to adjust estimates.^{6, 61} Instrumental variable methods may also be used to estimate causal treatment effects in the presence of unmeasured confounders, provided a valid instrumental variable is available.⁶²

Conclusion: Providing transparency through sensitivity analyses

Given the complexity of the underlying medical, sociological, and behavioral processes that determine exposure to medical products and interventions as well as the limitations of typical healthcare databases, there will often exist substantial uncertainty about how one should specify necessary statistical models to control confounding. We will typically not know with certainty whether a variable is a confounder, an instrumental variable, a potential source of *M*-bias, or simply irrelevant. Therefore, we suggest that researchers report results under different specifications of the statistical models. This allows the reader to evaluate the robustness of the findings to arbitrary modeling decisions made by the analyst.

We suggest that a primary analytic approach be chosen before data are examined and then variations of that analysis performed and described in the paper or reported in an appendix. For example, in secondary analyses one could consider a high-dimensional specification or one could remove variables that might plausibly lead to over-adjustment bias. If a researcher is agnostic about which variables to put in the model, it would be reasonable to pool results from different model specifications using meta-analysis or Bayesian methods. Such approaches would enable confidence intervals to represent more faithfully uncertainty about

effects of interest. This would allow clinicians, patients, regulators, and policy makers to make the most appropriate decisions regarding the use of medical products and procedures in light of the available evidence.

Acknowledgments

We acknowledge helpful comments from Stephen Cole, Leah Sirkus, and Brian Sauer.

References

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005; 58:323–37. [PubMed: 15862718]
2. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995; 82:669–88.
3. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999; 10:37–48. [PubMed: 9888278]
4. Walker AM. Confounding by indication. *Epidemiology*. 1996; 7:335–6. [PubMed: 8793355]
5. J LFB, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol*. 2009
6. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005; 162:279–89. [PubMed: 15987725]
7. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *Int J Epidemiol*. 2006; 35:337–44. [PubMed: 16368725]
8. Jackson LA, Nelson JC, Benson P, et al. Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors. *Int J Epidemiol*. 2006; 35:345–52. [PubMed: 16368724]
9. Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology*. 2001; 12:682–9. [PubMed: 11679797]
10. Glynn RJ, Schneeweiss S, Wang PS, Levin R, Avorn J. Selective prescribing led to overestimation of the benefits of lipid-lowering drugs. *J Clin Epidemiol*. 2006; 59:819–28. [PubMed: 16828675]
11. Winkelmayer WC, Levin R, Setoguchi S. Associations of kidney function with cardiovascular medication use after myocardial infarction. *Clin J Am Soc Nephrol*. 2008; 3:1415–22. [PubMed: 18614775]
12. White HD. Adherence and outcomes: it's more than taking the pills. *Lancet*. 2005; 366:1989–91. [PubMed: 16338439]
13. Brookhart MA, Patrick AR, Dormuth C, et al. Adherence to lipid-lowering therapy and the use of preventive health services: an investigation of the healthy user effect. *American journal of epidemiology*. 2007; 166:348–54. [PubMed: 17504779]
14. Simpson SH, Eurich DT, Majumdar SR, et al. A meta-analysis of the association between adherence to drug therapy and mortality. *Bmj*. 2006; 333:15. [PubMed: 16790458]
15. Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf*. 2004; 13:437–41. [PubMed: 15269927]
16. Dormuth CR, Patrick AR, Shrank WH, et al. Statin Adherence and Risk of Accidents. A Cautionary Tale. *Circulation*. 2009
17. Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health*. 1992; 82:243–8. [PubMed: 1739155]
18. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992; 267:2197–203. [PubMed: 1556797]
19. Bradbury BD, Wang O, Critchlow CW, et al. Exploring relative mortality and epoetin alfa dose among hemodialysis patients. *Am J Kidney Dis*. 2008; 51:62–70. [PubMed: 18155534]

20. Bradbury BD, Brookhart MA, Winkelmayr WC, et al. Evolving statistical methods to facilitate evaluation of the causal association between erythropoiesis-stimulating agent dose and mortality in nonexperimental research: strengths and limitations. *Am J Kidney Dis.* 2009; 54:554–60. [PubMed: 19592144]
21. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol.* 1986; 15:413–9. [PubMed: 3771081]
22. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology.* 2001; 12:313–20. [PubMed: 11338312]
23. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol.* 2002; 155:176–84. [PubMed: 11790682]
24. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight “paradox” uncovered? *Am J Epidemiol.* 2006; 164:1115–20. [PubMed: 16931543]
25. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992; 3:143–55. [PubMed: 1576220]
26. Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol.* 2002; 31:163–5. [PubMed: 11914314]
27. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* 2009; 20:488–95. [PubMed: 19525685]
28. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol.* 2003; 158:915–20. [PubMed: 14585769]
29. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000; 11:550–60. [PubMed: 10955408]
30. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians’ Health Study. *Am J Epidemiol.* 2002; 155:1045–53. [PubMed: 12034583]
31. Bhattacharya J, Vogt WB. Do Instrumental Variables Belong in Propensity Scores? NBER Working Paper Series. 2007
32. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *American journal of epidemiology.* 2006; 163:1149–56. [PubMed: 16624967]
33. Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. *J Clin Epidemiol.* 1999; 52:7–12. [PubMed: 9973068]
34. Johnston SC, Henneman T, McCulloch CE, van der Laan M. Modeling treatment effects on binary outcomes with grouped-treatment variables and individual covariates. *Am J Epidemiol.* 2002; 156:753–60. [PubMed: 12370164]
35. Solomon DH, Schneeweiss S, Glynn RJ, Levin R, Avorn J. Determinants of selective cyclooxygenase-2 inhibitor prescribing: are patient or physician characteristics more important? *American Journal of Medicine.* 2003; 115:715–20. [PubMed: 14693324]
36. Brooks, JM.; Chrischilles, EA.; Scott, SD.; Chen-Hardee, SS. Instrumental variables evidence for stage II patients from Iowa. Vol. 38. *Health Services Research*; 2003. Was breast conserving surgery underutilized for early stage breast cancer?; p. 1385-402.[erratum appears in *Health Serv Res.* 2004 Jun;39(3):693]
37. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology.* 2006; 17:268–75. [PubMed: 16617275]
38. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *Jama.* 2007; 297:278–85. [PubMed: 17227979]
39. Juurlink DN, Mamdani MM, Lee DS, et al. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med.* 2004; 351:543–51. [PubMed: 15295047]

40. Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med.* 2008; 27:1539–56. [PubMed: 17847052]
41. Cain LE, Cole SR, Greenland S, et al. Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable. *Am J Epidemiol.* 2009; 169:1124–32. [PubMed: 19318615]
42. Mamdani M, Juurlink DN, Kopp A, Naglie G, Austin PC, Laupacis A. Gastrointestinal bleeding after the introduction of COX 2 inhibitors: ecological study. *BMJ (Clinical research ed.)* 2004; 328:1415–6.
43. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology.* 2003; 14:300–6. [PubMed: 12859030]
44. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol.* 2010
45. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiol.* 2008; 167:523–9. discussion 30-1. [PubMed: 18227100]
46. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987; 40:373–83. [PubMed: 3558716]
47. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol.* 1993; 46:1075–9. discussion 81-90. [PubMed: 8410092]
48. 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:S143–8. [PubMed: 17909373]
49. Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983; 70:41–55.
50. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med.* 1997; 127:757–63. [PubMed: 9382394]
51. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol.* 2003; 158:280–7. [PubMed: 12882951]
52. Johannes CB, Koro CE, Quinn SG, Cutone JA, Seeger JD. The risk of coronary heart disease in type 2 diabetic patients exposed to thiazolidinediones compared to metformin and sulfonylurea therapy. *Pharmacoepidemiol Drug Saf.* 2007; 16:504–12. [PubMed: 17245800]
53. Schneeweiss S, Rassen J, Glynn RJ, Brookhart MA. High dimensional propensity score adjustment. *Epidemiology.* 2009; 20:512–22. [PubMed: 19487948]
54. Joffe MM. Exhaustion, automation, theory, and confounding. *Epidemiology.* 2009; 20:523–4. [PubMed: 19525688]
55. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology.* 2003; 14:451–8. [PubMed: 12843771]
56. Brumback BA, Hernan MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med.* 2004; 23:749–67. [PubMed: 14981673]
57. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf.* 2006; 15:291–303. [PubMed: 16447304]
58. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol.* 2008; 18:637–46. [PubMed: 18652982]
59. Schneeweiss S, Glynn RJ, Tsai EH, Avorn J, Solomon DH. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. *Epidemiology.* 2005; 16:17–24. [PubMed: 15613941]
60. Sturmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med Care.* 2007; 45:S158–65. [PubMed: 17909375]

61. Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration--a simulation study. *Am J Epidemiol.* 2007; 165:1110-8. [PubMed: 17395595]
62. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association.* 1996; 81:444-55.

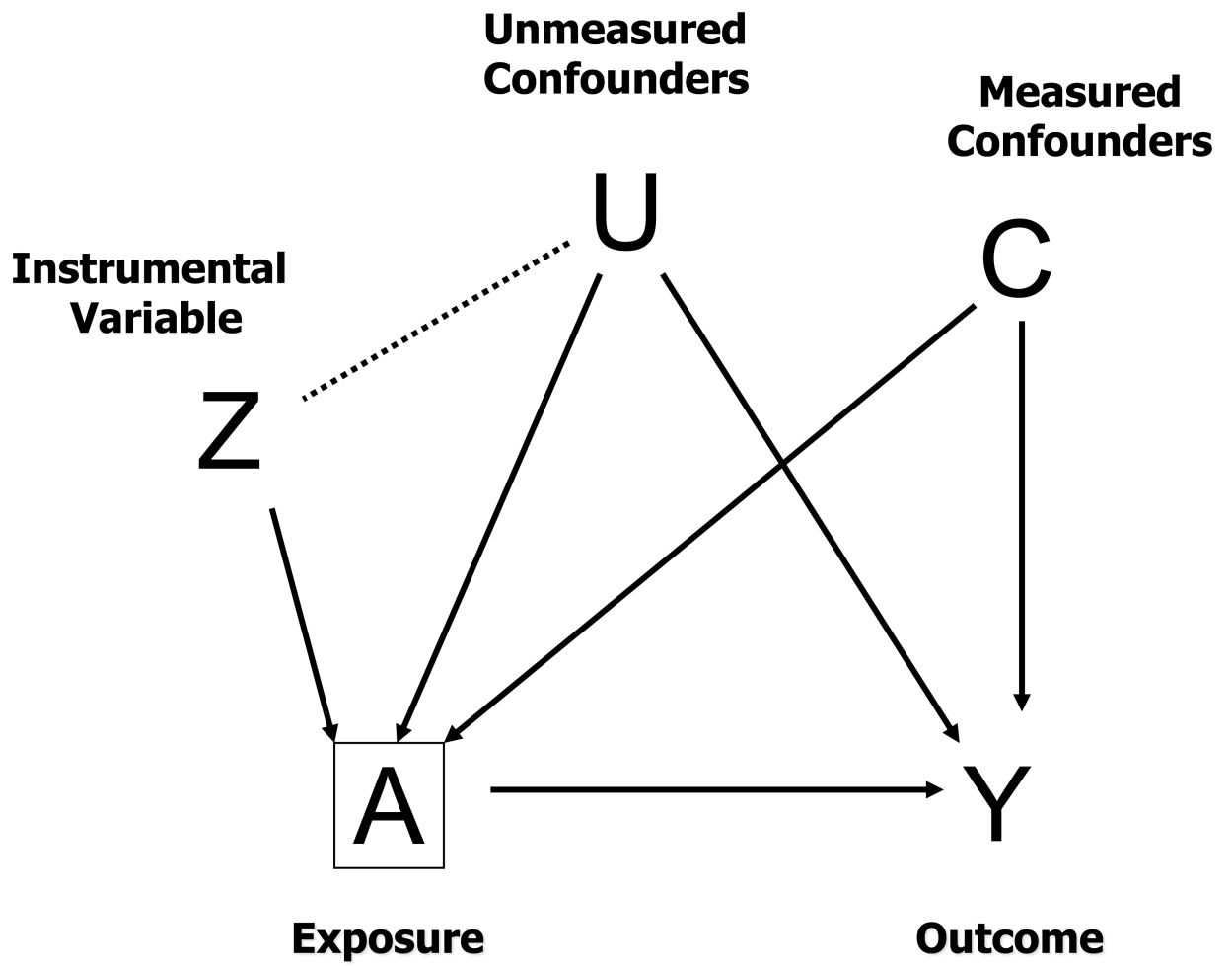


Figure 1.
Z-bias

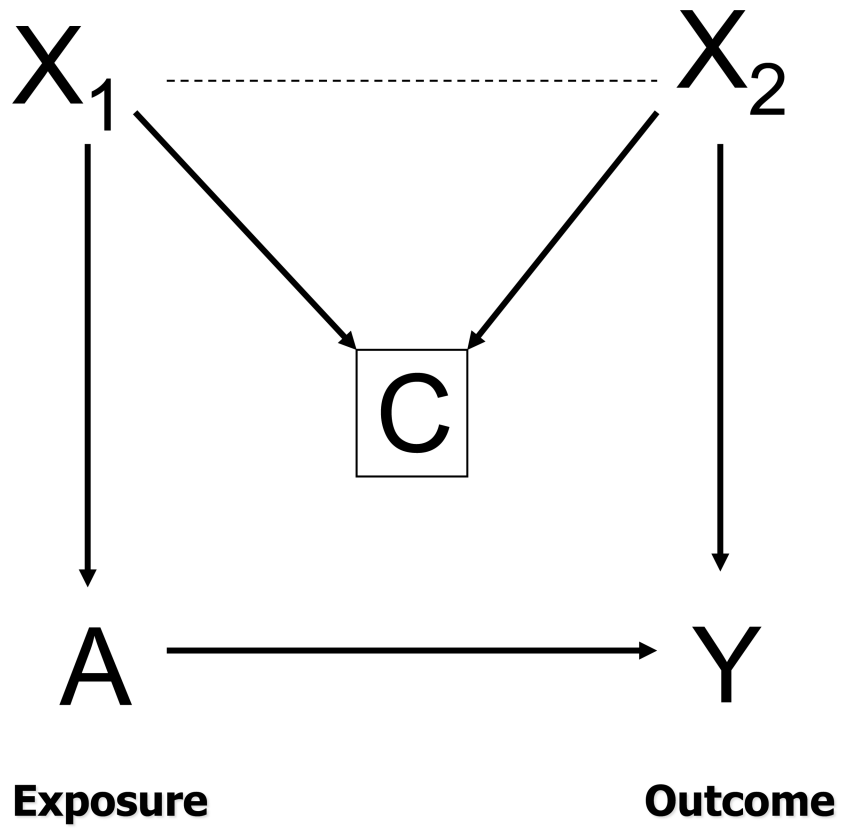


Figure 2.
M-bias