Confounding Adjustment in Comparative Effectiveness Research Conducted Within Distributed Research Networks

Sengwee Toh, ScD,* Joshua J. Gagne, PharmD, ScD,† Jeremy A. Rassen, ScD,‡
Bruce H. Fireman, MA,‡ Martin Kulldorff, PhD,* and Jeffrey S. Brown, PhD*

Background: A distributed research network (DRN) of electronic health care databases, in which data reside behind the firewall of each data partner, can support a wide range of comparative effectiveness research (CER) activities. An essential component of a fully functional DRN is the capability to perform robust statistical analyses to produce valid, actionable evidence without compromising patient privacy, data security, or proprietary interests.

Objectives and Methods: We describe the strengths and limitations of different confounding adjustment approaches that can be considered in observational CER studies conducted within DRNs, and the theoretical and practical issues to consider when selecting among them in various study settings.

Results: Several methods can be used to adjust for multiple confounders simultaneously, either as individual covariates or as confounder summary scores (eg, propensity scores and disease risk scores), including: (1) centralized analysis of patient-level data, (2) case-centered logistic regression of risk set data, (3) stratified or matched analysis of aggregated data, (4) distributed regression analysis, and (5) meta-analysis of site-specific effect estimates. These methods require different granularities of information be shared across sites and afford investigators different levels of analytic flexibility.

Conclusions: DRNs are growing in use and sharing of highly detailed patient-level information is not always feasible in DRNs. Methods that incorporate confounder summary scores allow investigators to adjust for a large number of confounding factors without the need to transfer potentially identifiable information in DRNs. They have the potential to let investigators perform many analyses traditionally conducted through a centralized dataset with detailed patient-level information.

Key Words: comparative effectiveness research, distributed research network, confounding, propensity score, disease risk score, instrumental variable, marginal structural model, pharmacoepidemiology

(*Med Care* 2013;51: S4–S10)

A major goal of comparative effectiveness research (CER) is to provide timely and actionable evidence regarding the relative benefits and risks of various treatments in different patients.^{1–3} Electronic health care databases—such as administrative claims databases and electronic health record databases—have become an important data source for developing evidence on the comparative effectiveness of medical products and delivery of care.^{1,4} These databases chronicle clinical encounters, medical services, and pharmacy prescriptions or dispensings of a large number of individuals. When analyzed appropriately, they can provide useful information to support clinical and regulatory decision making.

Using multiple databases to conduct CER offers a number of advantages. The combined sample size is often large enough to allow evaluations of rare treatments, outcomes, or populations, and can provide more timely evidence. A network of databases with diverse populations allows assessment of treatment heterogeneity and improves generalizability.^{5–7} Although storing all databases in a centralized repository is theoretically appealing, in practice it often creates concerns about privacy, confidentiality, regulations, and proprietary interests. A distributed approach—in which data are kept physically behind data partners' firewalls and under their direct control—is often preferred because it minimizes these concerns and expands the number of data partners willing to contribute information to studies.^{5–9}

A fundamental issue encountered in distributed research network (DRN) studies is the tradeoffs between analytic flexibility and the granularity of information being shared. Many existing DRNs can accommodate simple confounder adjustment using aggregate data (eg, age-stratified and sex-stratified analysis), but investigators generally have to request patient-level analytic datasets to adjust for multiple confounders. Although these analytic datasets typically include little or no Protected Health Information, concerns about privacy and confidentiality still linger due to the difficulty of complete deidentification,¹⁰ as well as

From the *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute; †Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; and ‡Division of Research, Kaiser Permanente Northern California, Oakland, CA.

S.T. and J.S.B. were supported by Agency for Healthcare Research and Quality (AHRQ) grant R01HS019912. J.A.R. is a recipient of a career development award from AHRQ (K01 HS018088).

The authors declare no conflict of interest.

Reprints: Sengwee Toh, ScD, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Ave 6th Floor, Boston, MA 02215. E-mail: darrentoh@post.harvard.edu.

Copyright © 2013 by Lippincott Williams & Wilkins
ISSN: 0025-7079/13/5108-00S4

concerns about loss of control over data security and restrictions on future use.

A fully functional DRN should have the capability to minimize transfer of potentially identifiable and proprietary information while permitting statistical analyses that are unbiased and efficient. However, there is currently no conceptual analytic framework that maps existing methods to CER needs for studies conducted within DRNs. In this paper, we describe: (1) the strengths and limitations of existing approaches that can handle a large number of confounders, either as individual covariates or as confounder summary scores, in CER studies conducted within DRNs; and (2) the theoretical and practical issues to consider when selecting among these approaches in various study settings.

SCOPE

Data harmonization is the most common approach used to perform multisite studies in DRNs. Multisite harmonization through use of a common data model enables investigators to develop and test the analytic code necessary for analysis and allows that program to be executed independently at each site. This paper focuses on studies that are conducted within DRNs with a common data model. We focus solely on observational studies with between-person comparisons because most CER studies compare ≥ 2 groups of patients receiving different treatments. Confounding can sometimes be addressed through within-person comparisons; interested readers are referred to papers that discuss self-controlled designs.^{11,12}

CONFOUNDING ADJUSTMENT IN OBSERVATIONAL CER STUDIES

Confounding may arise when individuals receiving different treatments have unequal underlying risks of developing the outcome of interest. In observational CER studies, confounding is a reflection of real-world clinical practice where patients receive a certain treatment based on, among other things, their clinical condition (indication), disease severity, and prognosis. Within a health plan or delivery system, confounding may also originate at the provider level from differences in treatment preference, care offered, and other ways that affect the outcome. In a network of multiple health plans or delivery systems, confounding may further arise from institutional guidelines or reimbursement policies that prefer one treatment to another based on patients' disease severity or treatment history.

Confounding may sometimes be more problematic in observational studies that examine anticipated treatment effects than those examining unanticipated effects.^{4,13} There may be more confounding when comparing different types of treatments (eg, pharmacotherapy vs. surgery to treat atrial fibrillation) than comparing treatments of the same therapeutic class (eg, captopril vs. enalapril to treat hypertension). Generating valid comparative effectiveness evidence using observational data relies heavily on the ability to identify confounders, the availability and accuracy of confounder information, and the use of appropriate methods to analyze this information.

Individual Confounders Versus Confounder Summary Scores

When the number of confounders is small relative to the number of outcome events, investigators can handle each confounder individually in the analysis. However, in most observational CER studies, adjustment for a large number of confounders is necessary because of the expected imbalances in many outcome risk factors between the treatment groups. Confounder summary scores condense the information of many confounders into a single variable. They obscure potentially identifiable information into nonidentifiable measures and are therefore particularly useful for DRN studies. The exposure propensity score (PS)^{14,15} and the disease risk score (DRS)^{16,17} are the most commonly used confounder summary scores. PSs are the probabilities of having the study exposure given patients' baseline characteristics, whereas DRSs are patients' probabilities or hazards of having the study outcome conditional on their baseline characteristics.

The 2 scores have been shown to provide results comparable to those from individual covariate adjustment.^{18–21} In general, PSs are particularly well suited for CER studies that compare the effects of 2 treatments on multiple outcomes, whereas DRSs are more practical than PSs when there are >2 treatments and a single outcome.¹⁷ When there are 2 treatments and 1 outcome, the choice between PSs and DRSs depends on the prevalence of the exposure and the outcome, the specific need of the study, and the investigator's preference. PSs are more favorable than DRSs when the exposure is common and outcome is infrequent; DRSs are preferred if the study aims include assessment of treatment heterogeneity by baseline outcome risk.

Although there is no standardized threshold for the number of outcome events per confounder over which one would prefer using confounder summary scores to handling each confounder individually, Cepeda et al²² suggested that PSs may perform better than the multivariable outcome logistic regression approach when there were ≤ 7 outcome events per confounder. Peduzzi et al²³ found that at least 10 events per confounder might be needed for the outcome logistic regression model to produce valid estimates.

Estimating Confounder Summary Scores in DRNs

Within DRNs, each database likely includes patients with different characteristics, and each may be created by organizations with distinct practice, coding patterns, or institutional policies. It is generally necessary to estimate confounder summary scores by site. Investigators could have each site run a distributed program that fits a statistical model containing the same covariates. The advantage of this approach is consistency, but by utilizing only variables common to all sites, the approach may not fully utilize the information available at each site.

An alternative would be to have each site fit its own model. This approach reduces residual confounding at sites that can adjust for more confounders; this may be the case if, for instance, certain sites have laboratory results data while others do not. This approach allows each site to provide the maximally adjusted result based on the local data available,

but it is operationally more cumbersome, as each site must build its own model, and some sites may not have the required programming or analytic expertise to do that.

A variant to fitting the site-specific PS model is the high-dimensional PS approach, which allows investigators to prespecify a set of common confounders and use an automated approach to empirically identify additional site-specific confounders.^{24–26} A constraint of this approach is that smaller sites may not have adequate sample size to perform the analysis.²⁷ Another potential weakness is the inclusion of variables that are only predictive of the exposure but not of the outcome except through their associations with the exposure (ie, instrumental variables), although some studies suggested that any bias may be considered trivial relative to the primary source of bias—residual confounding.^{28,29}

Estimating Confounder Summary Scores in CER Studies of Newly Approved Treatments

The issue of confounding is particularly complex in studies that compare a newly approved treatment to existing alternatives.³⁰ As evidence on comparative effectiveness is generally sparse when a new treatment is approved,³¹ some physicians may reserve the treatment for sicker patients or those who fail prior therapies. Patients, physicians, insurers, or delivery systems that adopt a new treatment earlier may be different from those who embrace it later. Treatment choice and characteristics of patients receiving the new treatment will change over time as more is learned of its benefits and risks. Investigators should allow the contribution of individual confounders to vary over time by, for example, estimating the PSs at regular intervals (eg, quarterly), starting from the time the new treatment is available.^{30,32} In contrast, outcome risk factors are generally more stable over time, therefore evolving prescribing dynamics in the early marketing period may have smaller impact on the DRS estimation. For certain outcomes, it may even be possible to fit a DRS model in the period before the introduction of the new treatment, and use the model to estimate exposed and unexposed patients' disease risk after introduction.³³

METHODS FOR CONFOUNDING ADJUSTMENT IN DRNs

Several analytic approaches can handle a large number of confounders, with or without using confounder summary scores, in DRN studies. As discussed below, some allow investigators to conduct an array of prespecified and ad hoc analyses but require more granular information, whereas others limit what investigators can do analytically but provide good protection for patient privacy, data security, and proprietary interests.³⁴

Centralized Analysis of Patient-Level Data

With this approach, the participating sites send the lead team a patient-level analytic dataset with individual covariate information necessary for the analysis, yielding what is essentially a single centralized dataset. Individual confounders can be incorporated into the analysis through restriction, stratification, matching, weighting, or outcome modeling, and the data can be considered all together or

stratified by contributing site.^{35,36} Confounder summary scores can be estimated after centralizing the data. This approach offers the most analytic flexibility at the expense of sharing potentially identifiable patient-level information and participating sites losing operational control over potentially sensitive and proprietary data.³⁷ In principle, most, if not all, Protected Health Information or proprietary information can be removed before leaving participating sites' firewalls. In practice, however, one often cannot completely rule out the possibility of reidentifying distinctive patients, especially at smaller sites,^{10,38} and many data partners may be unwilling to give up operational control over sensitive data, data security, and potential future uses.

Alternatively, each participating site can first estimate confounder summary scores and then send the lead team a patient-level analytic dataset with information on the exposure, outcome, follow-up (for time-to-event analysis), confounder summary scores, and other variables needed for the analysis (eg, age group information if one wishes to perform age-stratified analysis).³⁷ Confounder summary scores can be incorporated into the analysis through restriction, stratification, matching, weighting, or outcome modeling.^{17,39,40} This approach can perform essentially all the prespecified analyses afforded by the approach that shares individual confounder information. Depending on the variables requested, it may or may not be able to accommodate ad hoc analyses. For example, if race is included in PS estimation but race information is not requested separately, investigators will not be able to perform a secondary, race-stratified analysis.

Combining patients with similar values of PS across sites should not be done if PSs were estimated separately within each site. PSs will likely not be comparable across sites as patients' PS values will depend on the prevalence of the exposure in the population in which the PS is estimated. Many factors, including formularies and regional prescribing patterns, will influence the prevalence of exposure. Investigators should account for site in the analysis by either including it as a stratification variable or performing within-site PS matching. In contrast, the influence of risk factors on the outcome is more stable across data sources, even if the outcome incidence varies by site. For example, the relation between male sex and the 1-year risk for acute myocardial infarction, conditional on all other risk factors, should be similar across sites. Therefore, it may be possible to combine DRNs across sites.

Case-centered Logistic Regression of Risk Set Data

This approach was originally developed for vaccine safety research⁴¹ and has since been expanded to studies of other medical products.^{32,42} Sites transfer an aggregated dataset to the lead team that includes 1 record per risk set, with each risk set anchored by a case (ie, patient with the outcome of interest) and comprised of the cases and comparable individuals at risk of the outcome at the time the case occurs (see below). Each record includes a binary variable indicating whether the case is exposed to the treatment of interest and the log odds of the site-specific proportion of exposed patients in the risk set. The lead team fits a logistic regression model with

the indicator variable as the dependent variable and the log odds as the independent variable (specified as an offset). Fireman et al⁴¹ have shown that such a model maximizes the same likelihood as a Cox model fit using patient-level data, and both yield the same parameter estimates.

Confounding adjustment is achieved through the selection of comparable patients into the risk sets. In stratification, the risk set comprises at-risk patients belonging to the same stratum (eg, same age group or PS stratum) as the case. In 1:1 matching, the risk set includes all at-risk patients in the matched cohort. In general, the number of confounders that can be adjusted for may be relatively small if stratifying by or matching on individual confounders. In addition, stratification requires that all stratifying variables be dichotomized or categorized. Investigators may use PSs or DRs to adjust for a larger number of confounders, and handle these scores through stratification or matching.

Stratified or Matched Analysis of Aggregated Data

In a stratified analysis, participating sites send the lead team the total exposed and unexposed persons or person-times, and the number of exposed and unexposed outcomes within each stratum. In the matched analysis, if each site matches in the same fixed ratio, the information needed for the analysis includes only the total exposed and unexposed persons or person-times, and the number of exposed and unexposed outcomes in the matched cohort. As noted above, stratification requires that all stratifying variables be dichotomized or categorized. Therefore the number of confounders that can be adjusted for may be relatively small if stratifying by or matching on individual confounders. Using confounder summary scores allows adjustment of a larger number of confounders, but as discussed above, matching, stratifying, or other treatment of PSs should occur within site rather than across sites.

Distributed Regression Analysis

Distributed regression analysis fits regression models, with or without confounder summary scores, on individual databases within DRNs and produces results identical to those from centralized outcome regression analysis of patient-level data.^{43–46} The approach involves an iterative process, with participating sites transferring only summary statistics to the lead team at each step. The number of iterations depends on the regression model (eg, linear, logistic) and the complexity of the model. For example, fitting a linear regression model is a 2-step process. At step 1, each site executes a distributed program locally and submits intermediate summarized statistical results to the lead team. The lead team combines the intermediate results, and computes the parameter estimates. At step 2, participating sites execute another distributed program and deliver the variance/covariance estimates of the parameter estimates to the lead team to compute the confidence intervals. Although new approach may be developed in the future, this method is currently limited to linear and logistic regression.

Meta-Analysis

In meta-analysis, only site-specific effect estimates and their variances (or other information needed to calculate weight) are sent to the lead team. The site-specific estimates can be obtained from restriction, stratification, matching, weighting, or outcome modeling, with or without using confounder summary scores. These estimates are then pooled through meta-analysis.^{47,48} This approach requires the least amount of potentially identifiable information to leave participating sites' firewalls. It has been shown to produce similar pooled estimates when compared with patient-level data analysis.^{42,49} However, it is very analytically rigid. Every subgroup or sensitivity analysis requires all sites to perform each analysis internally, and then transfer the effect estimates to the lead team. Smaller sites may not be able to perform certain analyses, although sometimes using confounder summary scores may help.

ADDITIONAL CONSIDERATIONS

Unmeasured Confounding

None of the methods discussed thus far is robust against unmeasured confounders. In the presence of unmeasured confounding, instrumental variable analyses may provide valid effect estimates.^{50,51} This approach has been used in observational studies.⁵² In DRNs, a number of potential instrumental variables can be entertained, such as geographic variation in the propensity to use one or another treatment^{53,54} and physician preference.^{55,56} In theory, instrumental variable analyses enable sites to send little amount of patient-level data—only exposure, outcome, and instrumental variable information.

However, instrumental variable analyses are not without limitations. The greatest challenge is to identify a valid instrument. There is no way to verify empirically whether one of the necessary assumptions—the instrument is not associated with the outcome except through the exposure—holds in any observational study.^{57–59} In other words, instrumental variable analyses in observational studies replace the assumption of no unmeasured confounding for the exposure-outcome relation with the assumptions of (i) no unmeasured confounding for the instrument-outcome relation, and (ii) no direct effect of the instrument on the outcome by paths other than through exposure. Investigators have to weigh the plausibility of each set of assumptions when choosing between available methods. In addition, the treatment effect estimated by this approach applies only to the “marginal population” or “complier population,” which is a group of patients that cannot really be identified in practice.⁵¹ To address unmeasured confounding, investigators should always perform sensitivity analysis to examine the robustness of their results.^{60,61}

Time-varying Confounders

Time-dependent treatments are ubiquitous in CER. Sometimes the treatment of interest is a dynamic regimen that depends on patients' responses or prognoses. An example may be “take drug A, if the cholesterol level is still above 240 mg/dL after 2 months, then add drug B.” CER studies of time-dependent treatments must appropriately adjust for time-varying

confounding. Standard approaches, such as matching, stratification, and regression, may introduce bias if the time-varying confounders are also intermediate variables on the causal pathway of the exposure-outcome relation.⁶² To appropriately adjust for such confounders, investigators should instead use methods such as inverse probability weighting,^{63,64} g-estimation,^{65,66} or the g-formula.^{67,68} Although electronic health care databases provide longitudinal information for a large number of patients, the availability and accuracy of time-varying confounder information may be inadequate for certain studies.

Operational Efficiency

For all methods, more operational efficiency can be gained when sites first transform their source data into a common data structure.^{5–7} The lead team can develop and test code that creates the analytic dataset or performs the analysis. Other participating sites can then execute the code, often with minimal or no modification. The methods discussed above require different levels of statistical sophistication at participating sites, those that involve more programming and analytic efforts at participating sites may sometimes be less logistically feasible.³⁴

CER STUDIES WITH PROSPECTIVE DATA COLLECTION

Studies that combine information recorded routinely in electronic health care databases with additional, prospectively collected data are increasingly common.⁶⁹ Such additional information may improve the accuracy of exposure or outcome classification, the ability to study patient-reported outcomes, and the capability to adjust for otherwise unmeasured confounders. It also introduces a number of issues that need to be addressed, many of them may be more challenging in DRNs, as DRNs usually require a concerted effort among sites.

Missing Information in Subset of Study Population

A typical database study can include thousands or even millions of patients. Prospective data collection is often only feasible in a subset of the study population. Even if data are prospectively collected for all patients, nonresponse will lead to missing data. There are many ways to handle missing data. Some, like the complete-case analysis and the missing indicator approach, are easy to implement but are valid only under very strong assumptions.^{70,71} Investigators should consider methods that require weaker assumptions, such as multiple imputation,^{72,73} inverse probability weighting,⁷⁴ and PS calibration.⁷⁵ None of these methods is appropriate for all studies, so investigators should be aware of their strengths and constraints when choosing among them. At the design phase, a 2-stage sampling approach may be considered.⁷⁶

Analyzing Data as They Accrue or After All Data are Collected

Electronic health care databases capture patient experiences longitudinally, so they can be used to conduct prospective studies. If timely comparative effectiveness information is needed, then data should be analyzed weekly, monthly, or quarterly as they arrive using sequential analytic

techniques.⁷⁷ However, the fresher the data, the more likely it is that they may be inaccurate or incomplete because they may not have undergone the usual adjudication process done as part of claims processing or data quality check.⁷⁸ Hence, such sequential analysis should be viewed as an activity that needs follow-up investigations whenever a potential relationship is detected.

CONCLUSIONS

Cross-institutional sharing of detailed patient-level information for CER is not always feasible, which has led to an increase in the use of DRNs. A fully operational DRN should have the capability to perform robust statistical analysis while maintaining patient privacy, data security, and proprietary interests. A range of analytic options are available. Methods that incorporate confounder summary scores have the potential to adjust adequately for a large number of confounders without requiring potentially identifiable information to leave participating sites' firewalls, and allow investigators to perform a wide range of analyses traditionally afforded by a centralized dataset with detailed patient-level information.

REFERENCES

1. Federal Coordinating Council for Comparative Effectiveness Research. *Report to the President and Congress on Comparative Effectiveness Research*. Washington, DC: Department of Health and Human Services; 2009.
2. Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: The National Academies Press; 2009.
3. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *JAMA*. 2012;307:1583–1584.
4. Schneeweiss S. Developments in post-marketing comparative effectiveness research. *Clin Pharmacol Ther*. 2007;82:143–156.
5. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med*. 2009;151:341–344.
6. Brown JS, Holmes JH, Shah K, et al. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care*. 2010;48:S45–51.
7. Toh S, Platt R, Steiner JF, et al. Comparative-effectiveness research in distributed health data networks. *Clin Pharmacol Ther*. 2011;90:883–887.
8. McMurry AJ, Gilbert CA, Reis BY, et al. A self-scaling, distributed information architecture for public health, research, and clinical care. *J Am Med Inform Assoc*. 2007;14:527–533.
9. Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff (Millwood)*. 2009;28:454–466.
10. Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*. 2010;57:1701–1777.
11. Maclure M, Mittleman MA. Should we use a case-crossover design? *Annu Rev Public Health*. 2000;21:193–221.
12. Whitaker HJ, Farrington CP, Spiessens B, et al. Tutorial in biostatistics: the self-controlled case series method. *Stat Med*. 2006;25:1768–1797.
13. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet*. 2004;363:1728–1731.
14. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
15. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984;79:516–524.
16. Miettinen OS. Stratification by a multivariate confounder score. *Am J Epidemiol*. 1976;104:609–620.
17. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res*. 2009;18:67–80.

18. Cook EF, Goldman L. Performance of tests of significance based on stratification by a multivariate confounder score or by a propensity score. *J Clin Epidemiol*. 1989;42:317–324.
19. Arbogast PG, Ray WA. Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. *Am J Epidemiol*. 2011;174:613–620.
20. Sturmer T, Joshi M, Glynn RJ, et al. 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;59:437–447.
21. Cadarette SM, Gagne JJ, Solomon DH, et al. Confounder summary scores when comparing the effects of multiple drug exposures. *Pharmacoepidemiol Drug Saf*. 2010;19:2–9.
22. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol*. 2003;158:280–287.
23. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373–1379.
24. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20:512–522.
25. Toh S, García Rodríguez LA, Hernán MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf*. 2011;20:849–857.
26. Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf*. 2012;21(suppl 1):41–49.
27. Rassen JA, Glynn RJ, Brookhart MA, et al. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol*. 2011;173:1404–1413.
28. Pearl J. On a class of bias-amplifying variables that endanger effect estimates. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010)*. Corvallis, OR: Association for Uncertainty in Artificial Intelligence; 2010:425–432.
29. Myers JA, Rassen JA, Gagne JJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*. 2011;174:1213–1222.
30. Schneeweiss S, Gagne JJ, Glynn RJ, et al. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin Pharmacol Ther*. 2011;90:777–790.
31. Goldberg NH, Schneeweiss S, Kowal MK, et al. Availability of comparative efficacy data at the time of drug approval in the United States. *JAMA*. 2011;305:1786–1789.
32. Fireman B, Toh S, Butler MG, et al. A protocol for active surveillance of acute myocardial infarction in association with the use of a new antidiabetic pharmaceutical agent. *Pharmacoepidemiol Drug Saf*. 2012; 21(suppl 1):282–290.
33. Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies. *Pharmacoepidemiol Drug Saf*. 2012;21(suppl 2):138–147.
34. Rassen JA, Moran J, Toh D, et al. Evaluating strategies for data sharing and analyses in distributed data settings, 2013. Available at: http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Evaluating-Strategies-for-Data-Sharing-and-Analyses.pdf. Accessed March 28, 2013.
35. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
36. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–586.
37. Rassen JA, Solomon DH, Curtis JR, et al. Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologics. *Med Care*. 2010;48:S83–89.
38. Kushida CA, Nichols DA, Jadnick R, et al. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. *Med Care*. 2012;50(suppl):S82–101.
39. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*. 2006;163:262–270.
40. Sturmer T, Schneeweiss S, Brookhart MA, et al. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *Am J Epidemiol*. 2005;161:891–898.
41. Fireman B, Lee J, Lewis N, et al. Influenza vaccination and mortality: differentiating vaccine effects from bias. *Am J Epidemiol*. 2009;170:650–656.
42. Toh S, Reichman ME, Houstoun M, et al. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. *Arch Intern Med*. 2012;172:1582–1589.
43. Karr AF, Lin X, Sanil AP, et al. Secure regression on distributed databases. *J Comput Graph Stat*. 2005;14:263–279.
44. Fienberg SE, Fulp WJ, Slavkovic AB, et al. “Secure” log-linear and logistic regression analysis of distributed databases. *Lect Notes Comput Sci*. 2006;2006:277–290.
45. Lin X, Karr AF. Privacy-preserving maximum likelihood estimation for distributed data. *J Privacy Confidentiality*. 2009;1:213–222.
46. Wu Y, Jiang X, Kim J, et al. Grid Binary LOGistic REGression (GLORE): building shared models without sharing data. *J Am Med Inform Assoc*. 2012;19:758–764.
47. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available at www.cochrane-handbook.org. Accessed February 14, 2013.
48. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–188.
49. Rassen JA, Avorn J, Schneeweiss S. Multivariate-adjusted pharmacoepidemiologic analyses of confidential information pooled from multiple health care utilization databases. *Pharmacoepidemiol Drug Saf*. 2010;19:848–857.
50. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 1996;91:444–455.
51. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health*. 1998;19:17–34.
52. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf*. 2010;19:537–554.
53. Stukel TA, Fisher ES, Wennberg DE, et al. 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;297:278–285.
54. Brooks JM, Chrischilles EA, Scott SD, et al. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. *Health Serv Res*. 2003;38:1385–1402.
55. Korn EL, Baumrind S. Clinician preference and the estimation of causal treatment effects. *Stat Sci*. 1998;13:209–235.
56. Brookhart MA, Wang PS, Solomon DH, et al. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology*. 2006;17:268–275.
57. D’Agostino RB Jr, D’Agostino RB Sr. Estimating treatment effects using observational data. *JAMA*. 2007;297:314–316.
58. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist’s dream? *Epidemiology*. 2006;17:360–372.
59. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29:722–729.
60. Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc*. 1999;47:749–754.
61. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25:1107–1116.
62. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
63. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
64. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat*. 2008;4. Article 22.

65. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the health worker survivor effect. *Math Model.* 1986;7:1393–1512.
66. Wittenman JC, D'Agostino RB, Stijnen T, et al. G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *Am J Epidemiol.* 1998;148:390–401.
67. Young JG, Cain LE, Robins JM, et al. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Stat Biosci.* 2011;2011:119–143.
68. Taubman SL, Robins JM, Mittleman MA, et al. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int J Epidemiol.* 2009;38:1599–1611.
69. EDM Forum. Project profiles, 2011. Available at: <http://www.edm-forum.org/publicgrant/About/projectprofiles/>. Accessed February 14, 2012.
70. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol.* 1995;142:1255–1264.
71. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat.* 2007;61:79–90.
72. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York, NY: John Wiley & Sons; 1987.
73. Raghunathan TE, Lepkowski JM, Van Hoewyk J, et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol.* 2001;27:85–95.
74. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc.* 1994;89:846–866.
75. Sturmer T, Schneeweiss S, Avorn J, et al. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol.* 2005;162:279–289.
76. Collet JP, Schaubel D, Hanley J, et al. Controlling confounding when studying large pharmacoepidemiologic databases: a case study of the two-stage sampling design. *Epidemiology.* 1998;9:309–315.
77. Nelson JC, Cook AJ, Yu O, et al. Methods for observational post-licensure medical product safety surveillance. *Stat Methods Med Res.* 2011. DOI: 10.1177/0962280211413452.
78. Greene SK, Kulldorff M, Yin R, et al. Near real-time vaccine safety surveillance with partially accrued data. *Pharmacoepidemiol Drug Saf.* 2011;20:583–590.