Published in final edited form as:
Am J Epidemiol. 2019 April 01; 188(4): 709–723. doi:10.1093/aje/kwy265.

Privacy-Protecting Analytical Methods Using Only Aggregate-Level Information to Conduct Multivariable-Adjusted Analysis in Distributed Data Networks

Xiaojuan Li, **Bruce H. Fireman**, **Jeffrey R. Curtis**, **David E. Arterburn**, **David P. Fisher**, **Érick Moyneur**, **Mia Gallagher**, **Marsha A. Raebel**, **W. Benjamin Nowell**, **Lindsay Lagreid**, and **Sengwee Toh**

Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts (Xiaojuan Li, Mia Gallagher, and Sengwee Toh); Division of Research, Kaiser Permanente Northern California, Oakland, California (Bruce H. Fireman); University of Alabama at Birmingham, Birmingham, Alabama (Jeffrey R. Curtis); Kaiser Permanente Washington Health Research Institute, Seattle, Washington (David E. Arterburn); The Permanente Medical Group, Kaiser Permanente Northern California, Oakland, California (David P. Fisher); StatLog Econometrics Inc., Montreal, Quebec, Canada (Érick Moyneur); Institute for Health Research, Kaiser Permanente Colorado, Denver, Colorado (Marsha A. Raebel); Global Healthy Living Foundation, CreakyJoints, Upper Nyack, New York (W. Benjamin Nowell); and Limeade®, Bellevue, Washington (Lindsay Lagreid).

Abstract

Distributed data networks enable large-scale epidemiologic studies but protecting privacy while adequately adjusting for a large number of covariates continues to pose methodological challenges. Using two empirical examples within a three-site distributed data network, we tested combinations of three aggregate-level data-sharing approaches (risk-set, summary-table, effect-estimate), four confounding adjustment methods (matching, stratification, inverse probability weighting, matching weighting), and two summary scores (propensity score, disease risk score) for binary and time-to-event outcomes. We assessed the performance of these data-sharing and adjustment method combinations by comparing their results against the results from the corresponding pooled individual-level data analysis (reference). For both outcome types, the method combinations examined yielded identical or comparable results to the reference in most scenarios. Within each data-sharing approach, comparability between aggregate- and individual-level data analysis depended on adjustment method, e.g., risk-set data sharing with matched or stratified analysis of summary scores produced identical results, while weighted analysis showed some discrepancies. Across adjustment methods examined, risk-set data sharing generally performed better while summary-table and effect-estimate data sharing more often produced discrepancies in settings of rare outcome and small sample size. Valid multivariable-adjusted analysis can be performed in distributed data networks without sharing individual-level data.

Correspondence to Dr. Xiaojuan Li, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 401 Park Drive, Suite 401 East, Boston, MA 02215 (xiaojuan_li@harvardpilgrim.org).

Conflict of interest: none declared.

Keywords

confounding control; data sharing; disease risk score; multi-center studies; distributed data networks; meta-analysis; privacy-protecting; propensity score

Multi-center distributed data networks support rapid evidence generation in large and diverse populations, assessment of treatment effect heterogeneity, and evaluation of rare exposures or outcomes (1–3). Existing large-scale networks include the Sentinel System (4, 5), the Health Care Systems Research Collaboratory (6), and the National Patient-Centered Clinical Research Network (7). However, efficient and privacy-protecting data sharing remains a challenge in distributed data network studies. To maximize analytical validity, researchers have traditionally requested detailed individual-level data to control for confounding and other biases. However, sharing detailed data about patients raises concerns about privacy. Even when participating organizations are open to sharing individual-level data, the required legal and contractual agreements and ethical reviews are often labor-intensive and time-consuming, making a study less efficient or even unachievable.

Privacy-protecting analytical methods can help address this challenge (8–11). The theoretical properties of these methods have been previously explored (12). Using only aggregate-level information, these methods can produce results consistent with those from the pooled individual-level data analysis, but evidence supporting their validity is limited in epidemiologic research. Prior empirical examinations showed that propensity score (PS)-stratified analysis of risk-set data and meta-analysis of site-specific effect-estimate data can achieve similar levels of statistical sophistication as their corresponding pooled individual-level analyses (13, 14); but simulation studies also suggested that these methods could produce different results with sparse data (14). Using two empirical examples from a distributed data network, we assessed the performance of different combinations of data-sharing approaches and confounding adjustment methods across a range of scenarios that researchers could encounter in real-world studies.

METHODS

This study focused on the statistical performance of various combinations of data-sharing approaches and confounding adjustment methods for binary and time-to-event outcomes as evaluated by the concordance between their results and those from the corresponding pooled individual-level data analyses, which served as the reference of our assessment (Table 1). The two empirical examples were comparative effectiveness and safety research topics on obesity and rheumatoid arthritis. The clinical contexts of these examples have been explored elsewhere (15–19). Both examples drew data from three integrated health care delivery systems, organized as a three-site distributed data network: Kaiser Permanente Colorado, Kaiser Permanente Northern California, and Kaiser Permanente Washington. These systems have previously transformed their electronic health data into research-ready datasets with a common data structure (20). The Institutional Review Board at Harvard Pilgrim Health Care approved this study; the three participating delivery systems ceded their Institutional Review Board oversight to Harvard Pilgrim Health Care.

Empirical examples

Example 1.—The first example assessed the comparative effectiveness and safety of adjustable gastric banding and Roux-en-Y gastric bypass. We identified a retrospective cohort of patients ≥18 years who underwent one of these procedures between 1/1/2005 and 9/30/2015. Eligible patients had continuous health plan enrollment with medical and pharmacy benefits, at least one recorded body mass index measurement ≥35 kg/m^2, and no exposure to any major gastrointestinal procedures during the 365-day period preceding the initial bariatric procedure.

The effectiveness outcomes of interest were achievement of clinically meaningful changes in body mass index from baseline (e.g., ≥10%) within the first post-procedure year. The safety outcomes included re-intervention and all-cause hospitalization within the first post-procedure year (15, 21). We analyzed both effectiveness and safety outcomes as binary and time-to-event outcomes. We defined binary safety and effectiveness outcomes as occurrence of outcomes of interest closet to the end of the first post-procedure year, and time-to-event outcomes as time to the first occurrence of outcomes of interest within the same follow-up period. Follow-up began on the day after the discharge date of the index procedure hospitalization and ended at the earliest occurrence of an outcome event, 365 days of follow-up, death, end of health plan enrollment, or 9/30/2015. We identified potential confounders (Web Table 1) during the 365-day period preceding the index procedure based on subject-matter knowledge and prior studies (15, 16, 21).

Example 2.—The second example compared the effectiveness and safety of tumor necrosis factor-alpha inhibitor biologics (adalimumab, certolizumab pegol, etanercept, golimumab, or infliximab) and non-tumor necrosis factor-alpha inhibitor biologics (abatacept, rituximab, or tocilizumab) for rheumatoid arthritis. We identified a retrospective cohort of patients ≥18 years with rheumatoid arthritis who had a first dispensing of a study drug between 1/1/2001 and 9/30/2015. Eligible patients had continuous health plan enrollment with medical and pharmacy benefits, no exposure to any study drugs during the 365-day period preceding initial dispensing. We excluded patients who had an outcome event of interest, cancer (excluding non-melanoma skin cancer), human immunodeficiency virus infection or acquired immune deficiency syndrome, or organ transplantation during the 365-day baseline period.

The effectiveness outcome was an adapted version of a validated claims-based clinical effectiveness measure operationalized for use with health plan data (22). The safety outcomes included bacterial infections requiring hospitalization and hypersensitivity reaction, identified using previously validated algorithms (19, 23), in the year following the index dispensing. We analyzed both effectiveness and safety outcomes as binary and time-to-event outcomes. We defined binary outcomes as occurrence of outcomes of interest closet to the end of the first year following the index dispensing, and time-to-event outcomes as time to the first occurrence of outcomes of interest within the same follow-up period, except for the time-to-event effectiveness outcome, which was defined as time to the first occurrence of switching to another biologic anti-rheumatic drug to which the patient had no prior exposure (a component of the validated claims-based clinical effectiveness measure).

Follow-up began on the date of index dispensing and ended on the earliest occurrence of an outcome event of interest, 365 days of follow-up, death, end of health plan enrollment, cessation of initial biologic treatment, initiation of another biologic treatment, or 9/30/2015. We identified pre-specified potential confounders during the 365-day baseline period preceding the index dispensing (Web Table 1).

Data-sharing approaches examined

We tested three aggregate-level data-sharing approaches that require varying levels of information to be shared by data-contributing sites. The appendix of Mazor et al. (24) and an introductory video (25), both freely available, provide examples of analytical datasets typically shared by a site using these approaches. We used pooled individual-level data from the three sites in the reference analysis.

Risk-set data—This approach aggregated individual-level data into a dataset that included one record per risk-set, with each risk-set anchored by a unique outcome event time. A risk-set comprised patients who experienced the outcome and patients who were still at risk of developing the outcome at that time point. Each record of the shared risk-set data included the unique event time, number of exposed events, number of unexposed events, size of the exposed risk-set, and size of the unexposed risk-set. With different confounding adjustment methods, as discussed below, the base for at-risk patients varied. For example, when confounding was adjusted through PS matching, the risk-set included all at-risk patients in the PS-matched cohort within the same site.

Summary-table data—This approach further reduced the data into an aggregated dataset that resembled two-by-two summary tables. Depending on the outcome type, this aggregated dataset contained the total number of persons (for binary outcomes) or total person-times (for time-to-event outcomes), as well as the number of outcome events in each exposure group. As with risk-set data sharing, the number of two-by-two summary tables depended on the confounding adjustment method. For example, when confounding was adjusted through PS matching, only a single two-by-two summary table was necessary for the PS-matched cohort within each site.

Effect-estimate data—This approach shared the least amount of data—an aggregated dataset that only contained the site-specific effect estimate and the corresponding variance, obtained by analyzing the individual-level data within each site using the same confounding adjustment method used for the corresponding reference analysis. For example, when PS-matching was used for confounding adjustment, the site-specific effect estimates were obtained by analyzing individual-level data at each site using PS-matching.

Confounder summary scores examined

To adjust for the large number of pre-specified confounders, we used two confounder summary scores—PS and disease risk score (DRS)—to condense the information contained in individual confounders into a single variable. PSs are the probabilities of having the study exposure given patients' baseline characteristics (26), while DRSs are patients' probabilities or hazards of having the study outcome conditional on their baseline characteristics (27).

Confounding adjustment methods examined

We performed within-site confounding adjustment by incorporating the two confounder summary scores into the analysis via matching, stratification, or weighting (except for DRS for which weighted analysis has not been established for single- or multi-database settings). We evaluated two types of PS weights—inverse probability treatment weights (28) and matching weights (29, 30). When estimated correctly, these summary scores provide results comparable to those from individual covariate adjustment (27, 31).

Statistical analysis

Analysis of individual-level data (reference analysis)—We analyzed the pooled individual-level data across three sites and used the results as the reference to evaluate the performance of other approaches that analyzed aggregate-level datasets. We used site-stratified logistic regression to obtain odds ratios (ORs) and 95% confidence intervals (CIs) for binary outcomes, and site-stratified Cox proportional hazards regression to estimate hazard ratios (HRs) and 95% CIs for time-to-event outcomes.

Analysis of risk-set data—For time-to-event outcomes, we analyzed the risk-set data by fitting a logistic regression model with the proportion of exposed outcome events among all events as the dependent variable and the log odds of having the study exposure in the risk-set as the independent variable (specified as an offset). This approach has been shown to be mathematically equivalent to a stratified Cox regression with individual-level data (9). For binary outcomes, we used logistic regression with count data.

Analysis of summary-table data—For binary outcomes, we fit a site-stratified logistic regression model for grouped data, with the number of outcomes/total number of persons as the dependent variable and the exposure variable as the independent variable. For time-to-event outcomes, we fit a site-stratified conditional Poisson regression model with the natural log of person-time as the offset. When confounding adjustment was done through stratification, we also included the quintile indicator of the confounder summary score as another stratification variable. In situations where the regression-based analysis was not feasible, we used the Mantel-Haenszel method (32) to compute a weighted estimate for the desired effect measure across strata. Weighted analysis has not been established to analyze summary-table data.

Analysis of effect-estimate data—With the site-specific effect-estimate data, we performed an inverse variance-weighted meta-analysis using the DerSimonian and Laird's fixed-effect and random-effects models to obtain the overall effect estimate and 95% CI (33).

Assessment of treatment effect heterogeneity across sites—The goal of the study was to assess the performance of various data-sharing and analytical method combinations when the decision to pool data across sites had been made. However, we used Cochran's Q test to examine treatment effect heterogeneity across sites for illustrative purposes (34).

Assessment of statistical performance—To assess the statistical performance of different combinations of data-sharing approaches and confounding adjustment methods, we compared their results against their corresponding pooled individual-level data analyses. We did not compare the results across methods (e.g., PS matching versus PS stratification) because they estimated different treatment effects in different target populations.

RESULTS

Example 1: comparative effectiveness and safety of bariatric procedures

We identified 584 eligible adjustable gastric banding patients and 8,777 eligible Roux-en-Y gastric bypass patients. Web Table 2 summarizes their baseline characteristics.

PS-based analyses

Binary outcomes.: All aggregate-level data-sharing approaches generated results similar to their references for all confounding adjustment methods examined (Table 2). In fact, the results from risk-set and summary-table data sharing were identical to the reference. Both fixed-effect and random-effects meta-analyses of effect-estimate data produced comparable results for effectiveness outcomes, with the random-effects model showing slightly more variation. For safety outcomes, the two meta-analyses of effect-estimate data produced somewhat different results, with greater discrepancy observed in inverse probability weighted analyses.

Time-to-event outcomes.: Risk-set data sharing produced results identical to the reference in all confounding adjustment methods assessed (Table 3). Summary-table data sharing generated numerically different but qualitatively similar results in matched and stratified analysis of summary scores. Fixed-effect meta-analysis of effect-estimate data produced results compatible to the reference, while random-effects meta-analysis produced slightly different results that did not materially change the overall inference for effectiveness outcomes. For safety outcomes, the effect-estimate data-sharing approach showed discrepant results, with greater divergence seen in inverse probability weighted analyses.

DRS-based analyses—As with PS-based analyses, we observed similar performance for various data-sharing and adjustment method combinations when used with DRS for both binary and time-to-event outcomes (Table 4). Analyses of risk-set data produced results identical to the reference. Summary-table data sharing generated identical results for binary outcomes but slightly different results for time-to-event outcomes when compared with the reference. The two meta-analyses of effect-estimate data produced slightly different results for both outcome types. When compared across the same confounding adjustment method (i.e., stratification or matching) for any specific outcome, DRS-based analyses generally produced results consistent with those from PS-based analyses.

Treatment effect heterogeneity across sites—The Q-statistic suggested potential treatment effect heterogeneity across the three data-contributing sites for most outcomes examined (Tables 2–4).

Example 2: comparative effectiveness and safety of biologic disease-modifying anti-rheumatic drugs

We identified 7,419 patients who initiated a tumor necrosis factor-alpha inhibitor and 407 patients who initiated a non-tumor necrosis factor-alpha inhibitor biologic. Web Table 3 summarizes their baseline characteristics. Due to the low outcome occurrences as well as the limited sample size, we present results for switching for the effectiveness outcome and serious infections for the safety outcome, the only outcomes for which we could obtain reliable estimates.

PS-based analyses

Binary outcomes.: Similar to the bariatric procedure example, all three data-sharing approaches generated results similar to the reference (Table 5). The results from risk-set and summary-table data sharing were identical to the reference when confounding was adjusted through stratification or matching. The two meta-analyses of effect-estimate data also produced comparable results. When using inverse probability weighting for confounding adjustment, divergence from the reference was observed, especially for the serious infections outcome, which had lower incidence compared to treatment switching.

Time-to-event outcomes.: Sharing of risk-set data produced results identical to the reference except when confounding was adjusted through weighting—divergence from the reference was observed in the 95% CIs, especially for the serious infections outcome (Table 6). Both meta-analyses of effect-estimate data produced results compatible to the reference, with the random-effects showing slightly more variation. However, different from the bariatric procedure example, summary-table data sharing generated results concordant with the reference.

DRS-based analyses—We observed similar findings for both binary and time-to-event outcomes when comparing results from the aggregate-level analytical methods with the reference (Table 7). Risk-set and summary-table data sharing generated identical results for both binary and time-to-event outcomes. Meta-analyses of effect-estimate data produced slightly different results for both outcome types but did not change the overall inference. For any specific outcome, DRS-based analyses generated results consistent with those from PS-based analyses when using the same confounding adjustment method (i.e., stratification or matching).

Treatment effect heterogeneity across sites—The Q-statistic suggested no treatment effect heterogeneity across the three data-contributing sites for most outcomes examined (Tables 5–7).

DISCUSSION

Using two empirical examples within a three-site distributed data network, we tested combinations of three aggregate-level data-sharing approaches, four confounding adjustment methods, and two confounder summary scores and assessed their performance in multivariable-adjusted analysis of binary and time-to-event outcomes. The empirical examples included a range of exposure prevalences and outcome incidences, allowing for assessment under various real-world settings. For both outcome types, these aggregate-level data-sharing approaches yielded results identical or comparable to those from their corresponding pooled individual-level data analyses in most scenarios examined.

Summary of findings

For a given data-sharing approach, the comparability between aggregate- and individual-level data analysis depended on the confounding adjustment method. For example, with risk-set data sharing, matched or stratified analysis of confounder summary scores returned identical results, while weighted analysis showed some variation. This was true for both PS- and DRS-based analysis. Our finding on the equivalence between PS-stratified analysis of risk-set data and the pooled individual-level data analysis was consistent with a previous empirical examination (13). Our study also confirmed the high comparability between inverse probability weighted analysis of risk-set data and the corresponding reference analysis in most scenarios, which was previously demonstrated in a simulation study (35).

Sharing of summary-table data only requires aggregated information by exposure group, but analyses using this approach were sensitive to outcome type, outcome incidence, and sample size. Across confounding adjustment methods, this data-sharing approach yielded results identical to the reference for binary outcomes but discrepant results for time-to-event outcomes in some scenarios. For example, the HR estimate for <5% change in body mass index from the PS-based analysis was 3.48 (95% CI: 3.11, 3.89) with summary-table data sharing while the reference was 2.20 (1.97, 2.46) (Table 3). This discordance was not surprising because summary-table data sharing for time-to-event outcomes was, in essence, performing a Poisson regression analysis, which assumes constant hazards. In the situation of time-varying hazards, this approach would generate results different from the Cox proportional hazards regression used in the pooled individual-level data analysis. This difference indicates that analysis of summary-table data may not be appropriate for certain time-to-event outcomes, especially when the hazards of outcome under study are not constant.

Meta-analysis of effect-estimate data requires the least amount of information be shared across sites, but our empirical examples suggest that this approach was sensitive to outcome incidence and sample size. The discordance between results from this approach and the reference was evident for the ≥30% change in body mass index outcome and the safety outcomes in the bariatric procedure example, and the serious infections outcome in the biologic anti-rheumatic drugs example. These outcomes incidences were ≤3.5% at some sites, much lower compared to the other outcomes. In addition, some outcomes only occurred in one exposure group at some sites, making the data uninformative in meta-analyses of effect-estimate data. Conversely, other data-sharing approaches can utilize data from sites with outcome occurring in only one of the exposure groups. When the outcome under study was common across exposure groups and across sites, effect-estimate data sharing, using both fixed-effect and random-effects modeling, produced estimates similar to the reference. This finding was consistent with the results from previous simulation studies (14, 35).

Synthesis of evidence on the performance of methods examined

Results from this empirical study confirmed and complemented those from a simulation study that examined the performance of these methods in a wider range of scenarios with varying treatment prevalence, outcome incidence, treatment effect, site size, number of sites, and covariate distributions (35). Simulation and empirical studies showed that these method combinations produced highly comparable results to those from their corresponding pooled individual-level analysis when the exposure prevalence was high, the outcome incidence was high, and the site size was adequate. The performance of these method combinations varied in scenarios with low exposure prevalence, low outcome incidence, and small site size. Web Table 4 summarizes the strengths and limitations of these methods examined in both studies and how their performance may be influenced by key parameters in a given multi-center study. This table can serve as a guide for researchers interested in applying these methods. In general, risk-set data sharing is the method of choice in matched or stratified analysis of confounder summary scores because of its mathematical equivalence to its corresponding pooled individual-level data analysis. We demonstrated this equivalence in simulation and empirical studies. Meta-analysis of effect-estimate data is a valid alternative if all data-contributing sites are able to produce an effect estimate. Summary-table data sharing can also be considered when the hazards of study outcome are constant.

Additional considerations

We evaluated the performance of these methods in a distributed data network that had a common data model and reliable data quality. However, we do not expect their relative performance to differ in settings with less standardized data infrastructure, because the pooled individual-level data analysis would be equally susceptible to the same data issues. In practice, it may be more challenging to apply certain privacy-protecting methods in settings with less standardized data infrastructure. The use of these methods may also require more programming resources at each site and more coordination across sites. These operational challenges, though important, were beyond the scope of this study, which focused on the statistical performance of the methods.

It is not uncommon to have richer data at certain sites in a multi-center study. Researchers can estimate confounder summary scores using a common set of covariates or site-specific covariates. Both approaches have unique strengths and limitations that may vary by setting (12). Again, this issue applies to all data-sharing approaches, including approaches that share individual-level data. Using a common model to estimate summary scores ensures consistency across sites, but this approach may not fully utilize the information available at each site. Estimating site-specific summary scores theoretically allows better confounding adjustment at each site but may require more programming resources when using aggregate-level data sharing. Some semi-automated modeling techniques, such as the high-dimensional PS approach (36), may help improve the feasibility of estimating site-specific summary scores. In practice, it is generally worthwhile to estimate summary scores in multiple ways to examine the robustness of the results.

Strengths

To our knowledge, this is the first study to systematically and comprehensively assess these newer privacy-protecting analytical and data-sharing methods for distributed data network studies. We used the results from pooled individual-level data analysis as the benchmark to evaluate the results from these more privacy-protecting methods. Although the referent pooled individual-level data analysis might not necessarily yield the true treatment effect, it represents the best possible analysis one could perform in multi-center studies; a more privacy-protecting method is a reasonable alternative if it produces identical or comparable results. It is also reassuring that our empirical studies produced results consistent with findings from prior methodological (8, 12–14, 35, 37, 38) and clinical studies (15–19). Data from the three integrated delivery systems allowed us to assess the performance of these methods in settings that researchers may encounter in real-world studies with different outcome incidences and exposure prevalences. We also produced empirical evidence to support the use of DRS in combination with aggregate-level data-sharing approaches, which has not been previously evaluated.

Limitations

Due to small sample sizes and rare outcomes in some scenarios, certain analyses were not feasible or produced unreliable estimates. However, our study offers a realistic scenario involving sparse data at participating sites, a setting that necessitates multi-center studies. Our distributed data network comprised only three sites whose data had been converted into standardized formats. Future studies need to assess the validity of these methods in networks with more data-contributing sites, larger sample sizes, and more diverse databases.

The combinations of data-sharing approaches and confounding adjustment methods evaluated were by no means exhaustive. We did not consider distributed regression (10, 11, 39–41), which could be used in combination with confounder summary scores (42). We tested for treatment effect heterogeneity across sites but did not address it in our analyses other than accounting for it in the random-effects meta-analysis. In the presence of treatment effect heterogeneity by site, issues around the appropriateness of combining data across sites apply to all data-sharing approaches, including approaches that share individual-level data. All methods we examined can accommodate assessment of treatment effect heterogeneity, either by site or by specific patient characteristics, if researchers specify potential effect modifiers in advance and request data accordingly. It is worth noting that the performance of the various method combinations examined was similar in both empirical examples, one of which showed substantial treatment effect heterogeneity and the other did not.

Conclusion

When used in conjunction with confounder summary scores, several combinations of data-sharing approaches and confounding adjustment methods allow researchers to perform multivariable-adjusted analysis using only aggregate-level information from participating sites and produce results identical or comparable to those from pooled individual-level data analysis. These more privacy-protecting analytical methods can be viable alternatives when sharing of individual-level data is not feasible or preferred in multi-center studies. Generally, risk-set data sharing is the method of choice in matched or stratified analysis of confounder summary scores. Meta-analysis of effect-estimate data is a valid alternative if all data-contributing sites can produce an effect estimate. Summary-table data sharing can also be considered when the hazards of study outcome are constant. Researchers should carefully evaluate exposure prevalence and outcome incidence when choosing among available data-sharing approaches and confounding adjustment methods in multi-center studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This work was supported by the Patient Centered Outcomes Research Institute (PCORI) [contract numbers ME-1403–11305 and PPRN-1306–04811]. All statements in this article, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI or PCORI's Board of Governors or Methodology Committee. Dr. Toh was also supported in part by the National Institutes of Health [grant number U01EB023683] and Agency for Healthcare Research and Quality [grant number R01HS026214].

The authors would like to thank project managers, programmers, and staff from the three data-contributing sites in this study.

This work was funded by the Patient Centered Outcomes Research Institute awards ME-1403–11305 and PPRN-1306–04811.

References

1. 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(6 Suppl):S45–S51. [PubMed: 20473204]
2. Maro JC, Platt R, Holmes JH, et al. Design of a National Distributed Health Data Network. Ann Intern Med. 2009;151(5):341–344. [PubMed: 19638403]
3. Toh S, Platt R, Steiner JF, et al. Comparative-Effectiveness Research in Distributed Health Data Networks. Clin Pharmacol Ther. 2011;90(6):883–887. [PubMed: 22030567]
4. Behrman RE, Benner JS, Brown JS, et al. Developing the Sentinel System—a National Resource for Evidence Development. N Engl J Med. 2011;364:498–499. [PubMed: 21226658]
5. Sentinel Initiative. The Sentinel System, https://www.sentinelinitiative.org/; 2018 [accessed 6 January 2018].
6. NIH Health Care Systems Research Collaboratory, https://www.nihcollaboratory.org/Pages/default.aspx; 2018 [accessed 6 January 2018].
7. Patient-Centered Outcomes Research Institute. PCORnet: The National Patient-Centered Clinical Research Network, http://www.pcori.org/funding-opportunities/pcornet-national-patient-centered-clinical-research-network/; 2018 [accessed 6 January 2018].
8. 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(8):848–857. [PubMed: 20162632]
9. Fireman B, Lee J, Lewis N, et al. Influenza Vaccination and Mortality: Differentiating Vaccine Effects from Bias. Am J Epidemiol. 2009;170(5):650–656. [PubMed: 19625341]
10. Karr AF, Lin X, Sanil AP, et al. Secure Regression on Distributed Databases. J Comput Graph Stat. 2005;14:263–279.
11. 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(5):758–764. [PubMed: 22511014]
12. Toh S, Gagne JJ, Rassen JA, et al. Confounding Adjustment in Comparative Effectiveness Research Conducted within Distributed Research Networks. Med Care. 2013;51:S4–10.

13. Toh S, Shetterly S, Powers JD, et al. Privacy-Preserving Analytic Methods for Multisite Comparative Effectiveness and Patient-Centered Outcomes Research. Med Care. 2014;52(7):664–668. [PubMed: 24926715]

14. Toh S, Reichman ME, Houstoun M, et al. Multivariable Confounding Adjustment in Distributed Data Networks without Sharing of Patient-Level Data. Pharmacoepidemiol Drug Saf. 2013;22(11): 1171–1177. [PubMed: 23878013]

15. Arterburn D, Powers JD, Toh S, et al. Comparative Effectiveness of Laparoscopic Adjustable Gastric Banding vs Laparoscopic Gastric Bypass. JAMA Surg. 2014;149(12):1279–1287. [PubMed: 25353723]

16. Maciejewski ML, Arterburn DE, Van Scoyoc L, et al. Bariatric Surgery and Long-Term Durability of Weight Loss. JAMA Surg. 2016;151(11):1046–1055. [PubMed: 27579793]

17. Toh S, Li L, Harrold LR, et al. Comparative Safety of Infliximab and Etanercept on the Risk of Serious Infections: Does the Association Vary by Patient Characteristics? Pharmacoepidemiol Drug Saf. 2012;21(5):524–534. [PubMed: 22411435]

18. Grijalva CG, Chen L, Delzell E, et al. Initiation of Tumor Necrosis Factor-Alpha Antagonists and the Risk of Hospitalization for Infection in Patients with Autoimmune Diseases. JAMA. 2011;306(21):2331–2339. [PubMed: 22056398]

19. Curtis JR, Patkar N, Xie A, et al. Risk of Serious Bacterial Infections among Rheumatoid Arthritis Patients Exposed to Tumor Necrosis Factor Alpha Antagonists. Arthritis Rheum. 2007;56(4): 1125–1133. [PubMed: 17393394]

20. Curtis LH, Brown J, Platt R. Four Health Data Networks Illustrate the Potential for a Shared National Multipurpose Big-Data Network. Health Aff. 2014;33(7):1178–1186.

21. Flum DR, Belle SH, King WC, et al. Perioperative Safety in the Longitudinal Assessment of Bariatric Surgery. N Engl J Med. 2009;361(5):445–454. [PubMed: 19641201]

22. Curtis JR, Baddley JW, Yang S, et al. Derivation and Preliminary Validation of an Administrative Claims-Based Algorithm for the Effectiveness of Medications for Rheumatoid Arthritis. Arthritis Res Ther. 2011;13(5):R155. [PubMed: 21933396]

23. Walsh KE, Cutrona SL, Foy S, et al. Validation of Anaphylaxis in the Food and Drug Administration’s Mini-Sentinel. Pharmacoepidemiol Drug Saf. 2013;22(11):1205–1213. [PubMed: 24038742]

24. Mazor KM, Richards A, Gallagher M, et al. Stakeholders’ Views on Data Sharing in Multicenter Studies. J Comp Eff Res. 2017;6(6):537–547. [PubMed: 28805448]

25. Privacy-Protecting Analytic and Data-Sharing Methods for Multi-Database Studies. Privacy-Protecting Methods. https://www.distributedanalysis.org/educational-materials; 2018 [Assessed June 27, 2018].

26. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 1983;70(1):41–55.

27. 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(5):613–620. [PubMed: 21749976]

28. Xu S, Ross C, Raebel MA, et al. Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence Intervals. Value Health. 2010;13(2):273–277. [PubMed: 19912596]

29. Li L, Greene T. A Weighting Analogue to Pair Matching in Propensity Score Analysis. Int J Biostat. 2013;9(2):215–234. [PubMed: 23902694]

30. Yoshida K, Hernandez-Diaz S, Solomon DH, et al. Matching Weights to Simultaneously Compare Three Treatment Groups: Comparison to Three-Way Matching. Epidemiology. 2017;28(3):387–395. [PubMed: 28151746]

31. 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(4):317–324. [PubMed: 2723692]

32. Mantel N, Haenszel W. Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. J Natl Cancer Inst. 1959;22(4):719–748. [PubMed: 13655060]

33. DerSimonian R, Laird N. Meta-Analysis in Clinical Trials. Control Clin Trials. 1986;7(3):177–188. [PubMed: 3802833]

34. Cochran WG. The Combination of Estimates from Different Experiments. Biometrics. 1954;10(1): 101–129.

35. Yoshida K, Gruber S, Fireman BH, et al. Comparison of Privacy-Protecting Analytic and Data-Sharing Methods: a Simulation Study. Pharmacoepidemiol Drug Saf. 2018;27(9):1034–1041. [PubMed: 30022561]

36. 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(4):512–522. [PubMed: 19487948]

37. 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(20):1582–1589. [PubMed: 23147456]

38. 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. [PubMed: 22262618]

39. Fienberg SE, Fulp WJ, Slavkovic AB, et al. “Secure” Log-Linear and Logistic Regression Analysis of Distributed Databases. Lect Notes Comput Sci. 2006;4302:277–290.

40. Lin X, Karr AF. Privacy-Preserving Maximum Likelihood Estimation for Distributed Data. J Priv Confid. 2009;1:213–222.

41. Her QL, Malenfant JM, Malek S, et al. A Query Workflow Design to Perform Automatable Distributed Regression Analysis in Large Distributed Data Networks. eGEMs. 2018;6(1):11. [PubMed: 30094283]

42. Toh S, Wellman R, Coley RY, et al. Combining Distributed Regression and Propensity Scores: A Doubly Privacy-Protecting Analytic Method for Multi-Center Studies. Clin Epidemiol. 2018; 10:1773–1786. [PubMed: 30568510]

Table 1.

Combinations of Confounder Summary Score, Confounding Adjustment Method, Data-Sharing Approach, and Outcome Type Evaluated

Confounding adjustment method & data-sharing approach	Statistical analysis performed at the analysis center	
	Binary outcome[a]	Time-to-event outcome[b]
	Propensity score	
Stratification		
Pooled individual-level	PS- and site-stratified (Reference)	PS- and site-stratified (Reference)
Risk-set	PS- and site-stratified	Case-centered logistic regression[c]
Summary-table[d]	PS- and site-stratified	PS- and site-stratified conditional Poisson regression
Effect-estimate	Inverse variance-weighted meta-analysis	Inverse variance-weighted meta-analysis
Matching		
Pooled individual-level	PS-matched, site-stratified (Reference)	PS-matched, site-stratified (Reference)
Risk-set	PS-matched, site-stratified	Case-centered logistic regression
Summary-table	PS-matched, site-stratified	PS-matched, site-stratified conditional Poisson regression
Effect-estimate	Inverse variance-weighted meta-analysis	Inverse variance-weighted meta-analysis
Inverse probability weighting		
Pooled individual-level	Inverse probability weighted, site-stratified (Reference)	Inverse probability weighted, site-stratified (Reference)
Risk-set	Inverse probability weighted, site-stratified	Inverse probability weighted, site-stratified
Summary-table	Not established	Not established
Effect-estimate	Inverse variance-weighted meta-analysis	Inverse variance-weighted meta-analysis
Matching weighting		
Pooled individual-level	Matching weighted, site-stratified (Reference)	Matching weighted, site-stratified (Reference)
Risk-set	Matching weighted, site-stratified	Matching weighted, site-stratified
Summary-table	Not established	Not established
Effect-estimate	Inverse variance-weighted meta-analysis	Inverse variance-weighted meta-analysis
	Disease risk score	
Stratification		
Pooled individual-level	DRS- and site-stratified (Reference)	DRS- and site-stratified (Reference)
Risk-set	DRS- and site-stratified	Case-centered logistic regression
Summary-table	DRS- and site-stratified	DRS- and site-stratified conditional Poisson regression
Effect-estimate	Inverse variance-weighted meta-analysis	Inverse variance-weighted meta-analysis

Confounding adjustment method & data-sharing approach	Statistical analysis performed at the analysis center	
	Binary outcome[a]	Time-to-event outcome[b]
Matching		
Pooled individual-level	DRS-matched, site-stratified (Reference)	DRS-matched, site-stratified (Reference)
Risk-set	DRS-matched, site-stratified	Case-centered logistic regression
Summary-table	DRS-matched, site-stratified	DRS-matched, site-stratified conditional Poisson regression
Effect-estimate	Inverse variance-weighted meta-analysis	Inverse variance-weighted meta-analysis
Inverse probability weighting		
Pooled individual-level	Not established	Not established
Risk-set	Not established	Not established
Summary-table	Not established	Not established
Effect-estimate	Not established	Not established
Matching weighting		
Pooled individual-level	Not established	Not established
Risk-set	Not established	Not established
Summary-table	Not established	Not established
Effect-estimate	Not established	Not established

Note: DRS= disease risk score; PS= propensity score

[a] Unless otherwise specified, logistic regression was used to obtain estimates of odds ratios and their 95% confidence intervals for binary outcomes.

[b] Unless otherwise specified, Cox proportional hazards regression was used to obtain estimates of hazard ratios and their 95% confidence intervals for time-to-event outcomes.

[c] Case-centered logistic regression is a logistic regression model with the proportion of exposed outcome events among all events as the dependent variable and the log odds of having the study exposure in the risk-set as the independent variable, specified as an offset (9). Each risk-set, anchored by a unique outcome event time, comprises patients who experienced the outcome and patients who were still at risk of developing the outcome at that time point. When combined with confounder summary scores, the risk-set is created within a matched cohort or stratum defined by the confounder summary score within a site. In this particular analysis, each risk-set comprised the patient or patients who developed the outcome plus all other at-risk patients belonging to the same propensity score stratum at the time of the event within each site.

[d] In situations where the regression-based analysis was not feasible for the summary-table data-sharing approach, we used the Mantel-Haenszel method to compute a weighted estimate for the desired effect estimate.

Table 2.

Empirical Example 1: Results for Binary Outcomes from Propensity Score-Adjusted Analyses using Different Combinations of Confounding Adjustment Method and Data-Sharing Approach, AGB vs RYGB[a]

Confounding adjustment method & data-sharing approach	Effectiveness outcome, change in body mass index[b]										Safety outcome[c]			
	<5%		≥5%		≥10%		≥20%		≥30%		Re-hospitalization		Re-intervention	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Stratification														
Pooled individual-level	3.74	3.08, 4.54	0.13	0.08, 0.22	0.12	0.10, 0.16	0.09	0.07, 0.11	0.08	0.05, 0.11	0.85	0.62, 1.15	0.82	0.59, 1.11
Risk-set	3.74	3.08, 4.54	0.13	0.08, 0.22	0.12	0.10, 0.16	0.09	0.07, 0.11	0.08	0.05, 0.11	0.85	0.62, 1.15	0.82	0.59, 1.11
Summary-table	3.74	3.08, 4.54	0.13	0.08, 0.22	0.12	0.10, 0.16	0.09	0.07, 0.11	0.08	0.05, 0.11	0.85	0.62, 1.15	0.82	0.59, 1.11
Effect-estimate, fixed-effect	3.73	3.08, 4.52	0.13	0.08, 0.20	0.12	0.09, 0.16	0.09	0.07, 0.11	0.09	0.06, 0.12	0.94	0.69, 1.28	0.85	0.62, 1.15
Effect-estimate, random-effects	3.59	1.52, 8.48	0.15	0.06, 0.33	0.12	0.06, 0.23	0.07	0.04, 0.12	0.05	0.01, 0.16	0.85	0.29, 2.46	0.74	0.41, 1.32
Measure of heterogeneity, Q[d]	22.59	<.0001	2.41	0.2983	4.67	0.0962	7.23	0.0269	7.79	0.0203	12.79	0.0017	4.31	0.1156
Matching														
Pooled individual-level	3.66	2.85, 4.73	0.15	0.05, 0.37	0.17	0.11, 0.26	0.10	0.07, 0.13	0.08	0.06, 0.12	0.87	0.58, 1.30	0.78	0.52, 1.17
Risk-set	3.66	2.85, 4.73	0.15	0.05, 0.37	0.17	0.11, 0.26	0.10	0.07, 0.13	0.08	0.06, 0.12	0.87	0.58, 1.30	0.78	0.52, 1.17
Summary-table	3.66	2.85, 4.73	0.15	0.05, 0.37	0.17	0.11, 0.26	0.10	0.07, 0.13	0.08	0.06, 0.12	0.87	0.58, 1.30	0.78	0.52, 1.17
Effect-estimate, fixed-effect	3.63	2.84, 4.66	0.16	0.07, 0.40	0.19	0.12, 0.29	0.10	0.07, 0.13	0.09	0.06, 0.13	0.93	0.62, 1.38	0.81	0.55, 1.21
Effect-estimate, random-effects	2.93	1.08, 7.94	0.16	0.07, 0.40	0.19	0.12, 0.29	0.10	0.07, 0.13	0.06	0.02, 0.20	0.73	0.25, 2.11	0.64	0.24, 1.72
Measure of heterogeneity, Q[d]	16.27	0.0003	1.72	0.4217	1.64	0.4385	1.71	0.4234	7.05	0.0294	7.91	0.0191	7.71	0.0211
Inverse probability weighting														
Pooled individual-level	3.16	2.67, 3.75	0.14	0.09, 0.20	0.11	0.08, 0.13	0.10	0.09, 0.12	0.09	0.06, 0.11	0.85	0.65, 1.12	0.86	0.65, 1.14
Risk-set	3.16	2.67, 3.75	0.14	0.10, 0.21	0.11	0.09, 0.13	0.10	0.09, 0.12	0.09	0.06, 0.12	0.85	0.65, 1.12	0.86	0.65, 1.14
Effect-estimate, fixed-effect	3.15	2.65, 3.74	0.11	0.08, 0.17	0.11	0.09, 0.13	0.10	0.09, 0.12	0.12	0.09, 0.17	1.16	0.86, 1.55	1.09	0.81, 1.45
Effect-estimate, random-effects	3.11	2.55, 3.80	0.14	0.03, 0.60	0.09	0.05, 0.16	0.10	0.04, 0.29	0.02	0.00, 0.37	0.78	0.16, 3.71	0.43	0.09, 1.93
Measure of heterogeneity, Q[d]	2.27	0.3208	4.10	0.1281	6.42	0.0403	41.94	<.0001	12.10	0.0024	31.41	<.0001	16.87	0.0002
Matching weighting														
Pooled individual-level	3.74	2.92, 4.80	0.15	0.06, 0.36	0.12	0.07, 0.19	0.09	0.07, 0.12	0.08	0.06, 0.12	0.80	0.54, 1.17	0.82	0.56, 1.22
Risk-set	3.74	2.92, 4.80	0.15	0.06, 0.35	0.12	0.07, 0.19	0.09	0.07, 0.12	0.08	0.06, 0.12	0.80	0.54, 1.17	0.82	0.56, 1.22
Effect-estimate, fixed-effect	3.73	2.90, 4.80	0.15	0.06, 0.37	0.12	0.08, 0.20	0.09	0.07, 0.12	0.09	0.06, 0.13	0.82	0.55, 1.23	0.83	0.56, 1.24

Confounding adjustment method & data-sharing approach	Effectiveness outcome, change in body mass index[b]										Safety outcome[c]			
	<5%		≥5%		≥10%		≥20%		≥30%		Re-hospitalization		Re-intervention	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Effect-estimate, random-effects	3.53	1.42, 8.77	0.15	0.06, 0.37	0.12	0.08, 0.20	0.08	0.05, 0.13	0.05	0.01, 0.17	0.78	0.24, 2.50	0.79	0.49, 1.29
Measure of heterogeneity, Q[d]	11.59	0.0030	0.38	0.8233	0.88	0.6426	2.89	0.2352	6.52	0.0384	8.50	0.0142	2.38	0.3040

Note: AGB= Adjusted gastric banding; RYGB= Roux-en-Y gastric bypass; OR= odds ratio; CI= confidence interval

[a] There were 584 (6.2%) patients who underwent AGB and 8,777 (93.8%) patients who underwent RYGB.

[b] The incidences for <5%, ≥5%, ≥10%, ≥20%, and ≥30% change in body mass index were 68.0%, 93.7%, 76.5%, 31.2%, and 6.8%, respectively, for the AGB users; 32.9%, 99.1%, 96.8%, 83.8%, and 47.7%, respectively, for the RYGB users. These effectiveness outcomes were defined as the occurrence of the outcomes of interest closet to the end of the first post-procedure year so the incidences for <5% change in body mass index and ≥5% change in body mass index do not sum up to 100%.

[c] The incidences for re-hospitalization and re-intervention were 9.9% and 9.3%, respectively, for the AGB users, and 11.7% and 11.6%, respectively, for the RYGB users.

[d] Q is a measure of heterogeneity among the three data-contributing sites. The summary statistic and p-value from Cochran’s Q test are shown here.

Table 3.

Empirical Example 1: Results for Time-to-Event Outcomes from Propensity Score-Adjusted Analyses using Different Combinations of Confounding Adjustment Method and Data-Sharing Approach, AGB vs RYGB[a]

Confounding adjustment method & data-sharing approach	Effectiveness outcome, change in body mass index[b]										Safety outcome[c]			
	<5%		≥5%		≥10%		≥20%		≥30%		Re-hospitalization		Re-intervention	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Stratification														
Pooled individual-level	2.20	1.97, 2.46	0.53	0.49, 0.58	0.36	0.32, 0.40	0.17	0.15, 0.20	0.10	0.07, 0.13	0.84	0.64, 1.12	0.82	0.61, 1.09
Risk-set	2.20	1.97, 2.46	0.53	0.49, 0.58	0.36	0.32, 0.40	0.17	0.15, 0.20	0.10	0.07, 0.13	0.84	0.64, 1.12	0.82	0.61, 1.09
Summary-table	3.48	3.11, 3.89	0.42	0.39, 0.46[d]	0.37	0.34, 0.41[d]	0.22	0.19, 0.26[d]	0.12	0.08, 0.17	0.84	0.62, 1.11	0.81	0.60, 1.09
Effect-estimate, fixed-effect	2.26	2.02, 2.52	0.54	0.49, 0.59	0.36	0.33, 0.40	0.17	0.15, 0.20	0.11	0.08, 0.15	0.97	0.73, 1.28	0.85	0.64, 1.14
Effect-estimate, random-effects	2.35	1.21, 4.56	0.58	0.41, 0.81	0.36	0.27, 0.47	0.15	0.10, 0.22	0.06	0.02, 0.19	0.85	0.32, 2.22	0.76	0.45, 1.28
Measure of heterogeneity, Q[e]	25.29	<.0001	15.08	0.0005	7.80	0.0202	6.55	0.0377	6.66	0.0357	13.50	0.0012	3.97	0.1371
Matching														
Pooled individual-level	2.25	1.91, 2.66	0.50	0.44, 0.57	0.35	0.31, 0.40	0.17	0.15, 0.21	0.10	0.07, 0.14	0.85	0.60, 1.22	0.78	0.54, 1.11
Risk-set	2.25	1.91, 2.66	0.50	0.44, 0.57	0.35	0.31, 0.40	0.17	0.15, 0.21	0.10	0.07, 0.14	0.85	0.60, 1.22	0.78	0.54, 1.11
Summary-table	3.49	2.95, 4.13	0.40	0.36, 0.45	0.38	0.33, 0.43	0.23	0.19, 0.27	0.12	0.09, 0.17	0.85	0.58, 1.23	0.77	0.52, 1.12
Effect-estimate, fixed-effect	2.23	1.89, 2.63	0.50	0.45, 0.57	0.35	0.31, 0.40	0.17	0.15, 0.21	0.11	0.08, 0.15	0.91	0.63, 1.32	0.81	0.56, 1.18
Effect-estimate, random-effects	1.95	0.92, 4.16	0.56	0.32, 0.99	0.36	0.27, 0.47	0.17	0.12, 0.23	0.07	0.02, 0.23	0.73	0.29, 1.87	0.64	0.24, 1.72
Measure of heterogeneity, Q[e]	15.58	0.0004	22.17	<.0001	4.79	0.0911	3.24	0.1976	5.58	0.0612	7.65	0.0218	7.71	0.0211
Inverse probability weighting														
Pooled individual-level	1.92	1.73, 2.13	0.49	0.45, 0.54	0.32	0.29, 0.35	0.19	0.17, 0.22	0.10	0.08, 0.14	0.84	0.65, 1.08	0.85	0.65, 1.12
Risk-set	1.93	1.51, 2.47	0.49	0.40, 0.59	0.32	0.26, 0.38	0.19	0.14, 0.27	0.10	0.06, 0.17	0.84	0.46, 1.53	0.80	0.60, 1.07
Effect-estimate, fixed-effect	1.92	1.73, 2.13	0.50	0.46, 0.54	0.32	0.29, 0.35	0.20	0.18, 0.23	0.15	0.11, 0.20	1.16	0.89, 1.51	1.07	0.81, 1.40
Effect-estimate, random-effects	1.85	1.55, 2.21	0.54	0.43, 0.68	0.32	0.25, 0.41	0.19	0.11, 0.33	0.02	0.00, 0.42	0.74	0.22, 2.51	0.45	0.11, 1.80
Measure of heterogeneity, Q[e]	3.31	0.1905	9.67	0.0079	9.24	0.0098	22.16	<.0001	10.49	0.0053	27.68	<.0001	15.27	0.0005
Matching weighting														
Pooled individual-level	2.29	1.94, 2.71	0.53	0.46, 0.59	0.34	0.30, 0.39	0.17	0.14, 0.20	0.10	0.07, 0.14	0.80	0.56, 1.15	0.83	0.57, 1.20
Risk-set	2.33	2.08, 2.62	0.52	0.47, 0.57	0.34	0.31, 0.38	0.17	0.14, 0.20	0.10	0.07, 0.13	0.80	0.60, 1.07	0.85	0.57, 1.28
Effect-estimate, fixed-effect	2.26	1.91, 2.67	0.52	0.46, 0.59	0.34	0.30, 0.39	0.17	0.14, 0.20	0.11	0.08, 0.15	0.83	0.57, 1.20	0.83	0.57, 1.22

Confounding adjustment method & data-sharing approach	Effectiveness outcome, change in body mass index[b]										Safety outcome[c]			
	<5%		≥5%		≥10%		≥20%		≥30%		Re-hospitalization		Re-intervention	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Effect-estimate, random-effects	2.23	1.11, 4.49	0.56	0.40, 0.78	0.34	0.26, 0.44	0.14	0.09, 0.23	0.06	0.02, 0.20	0.77	0.27, 2.24	0.81	0.53, 1.25
Measure of heterogeneity, Q[e]	10.84	0.0044	7.90	0.0192	4.05	0.1316	6.28	<.0431	6.20	0.0450	8.35	0.0154	2.22	0.3285

Note: AGB= Adjusted gastric banding; RYGB= Roux-en-Y gastric bypass; HR= hazard ratio; CI= confidence interval

[a] There were 584 (6.2%) patients who underwent AGB and 8,777 (93.8%) patients who underwent RYGB.

[b] The incidences for <5%, ≥5%, ≥10%, ≥20%, and ≥30% change in body mass index were 68.0%, 93.7%, 76.5%, 31.2%, and 6.8%, respectively, for the AGB users; 32.9%, 99.1%, 96.8%, 83.8%, and 47.7%, respectively, for the RYGB users. These effectiveness outcomes were defined as the occurrence of the outcomes of interest closet to the end of the first post-procedure year so the incidences for <5% change in body mass index and ≥5% change in body mass index do not sum up to 100%.

[c] The incidences for re-hospitalization and re-intervention were 9.9% and 9.3%, respectively, for the AGB users, and 11.7% and 11.6%, respectively, for the RYGB users.

[d] These were calculated using the Mantel-Haenszel approach, because the exact confidence intervals for the regression-based analysis could not be obtained.

[e] Q is a measure of heterogeneity among the three data-contributing sites. The summary statistic and p-value from Cochran’s Q test are shown here.

Table 4.

Empirical Example 1: Results from Disease Risk Score[a]-Adjusted Analyses using Different Combinations of Confounding Adjustment Method and Data-Sharing Approach, AGB vs RYGB[b]

Confounding adjustment method & data-sharing approach	Effectiveness outcome, change in body mass index[c]										Safety outcome[d]			
	<5%		≥5%		≥10%		≥20%		≥30%		Re-hospitalization		Re-intervention	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Stratification														
Pooled individual-level	3.95	3.26, 4.79	0.15	0.10, 0.23	0.12	0.09, 0.15	0.08	0.07, 0.10	0.07	0.05, 0.10	0.86	0.63, 1.15	0.81	0.58, 1.09
Risk-set	3.95	3.26, 4.79	0.15	0.10, 0.23	0.12	0.09, 0.15	0.08	0.07, 0.10	0.07	0.05, 0.10	0.86	0.63, 1.15	0.81	0.58, 1.09
Summary-table	3.95	3.26, 4.79	0.15	0.10, 0.23	0.12	0.09, 0.15	0.08	0.07, 0.10	0.07	0.05, 0.10	0.86	0.63, 1.15	0.81	0.58, 1.09
Effect-estimate, fixed-effect	3.93	3.25, 4.76	0.14	0.09, 0.21	0.11	0.09, 0.14	0.08	0.07, 0.10	0.09	0.06, 0.12	0.95	0.70, 1.28	0.83	0.62, 1.12
Effect-estimate, random-effects	3.73	1.52, 9.12	0.16	0.08, 0.34	0.11	0.04, 0.28	0.07	0.03, 0.12	0.04	0.01, 0.16	0.85	0.30, 2.36	0.75	0.46, 1.21
Measure of heterogeneity, Q[e]	25.49	<.0001	2.46	0.2922	11.25	0.0036	10.50	0.0052	8.27	0.0159	12.04	0.0024	3.29	0.1930
Matching														
Pooled individual-level	3.99	3.10, 5.15	0.08	0.02, 0.24	0.11	0.06, 0.18	0.08	0.06, 0.10	0.08	0.06, 0.12	0.86	0.57, 1.30	0.93	0.61, 1.42
Risk-set	3.99	3.10, 5.15	0.08	0.02, 0.24	0.11	0.06, 0.18	0.08	0.06, 0.10	0.08	0.06, 0.12	0.86	0.57, 1.30	0.93	0.61, 1.42
Summary-table	3.99	3.10, 5.15	0.08	0.02, 0.24	0.11	0.06, 0.18	0.08	0.06, 0.10	0.08	0.06, 0.12	0.86	0.57, 1.30	0.93	0.61, 1.42
Effect-estimate, fixed-effect	3.93	3.06, 5.05	0.08	0.02, 0.26	0.11	0.06, 0.18	0.08	0.06, 0.10	0.09	0.06, 0.13	0.89	0.61, 1.27	0.95	0.63, 1.44
Effect-estimate, random-effects	3.42	1.06, 10.98	0.08	0.02, 0.26	0.11	0.06, 0.18	0.06	0.04, 0.11	0.05	0.01, 0.16	0.79	0.38, 1.67	0.79	0.34, 1.83
Measure of heterogeneity, Q[e]	20.91	<.0001	0.00	0.9985	0.21	0.8981	3.32	0.1898	7.47	0.0239	4.11	0.1275	5.51	0.0633
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Stratification														
Pooled individual-level	2.23	2.01, 2.48	0.51	0.47, 0.56	0.34	0.31, 0.38	0.17	0.14, 0.19	0.09	0.07, 0.13	0.83	0.63, 1.08	0.79	0.59, 1.05
Risk-set	2.23	2.00, 2.47	0.51	0.47, 0.56	0.34	0.31, 0.38	0.17	0.14, 0.19	0.09	0.07, 0.13	0.82	0.63, 1.08	0.79	0.59, 1.05
Summary-table	3.35	2.98, 3.79	0.41	0.38, 0.46	0.36	0.33, 0.39	0.22	0.19, 0.25	0.12	0.08, 0.17	0.81	0.61, 1.07	0.78	0.56, 1.03
Effect-estimate, fixed-effect	2.31	2.08, 2.57	0.52	0.47, 0.57	0.34	0.31, 0.38	0.17	0.15, 0.20	0.11	0.08, 0.14	0.94	0.72, 1.23	0.81	0.61, 1.08
Effect-estimate, random-effects	2.28	1.17, 4.42	0.54	0.38, 0.75	0.34	0.26, 0.45	0.14	0.09, 0.22	0.06	0.02, 0.18	0.81	0.34, 1.96	0.75	0.49, 1.14
Measure of heterogeneity, Q[e]	28.81	<.0001	14.65	0.0007	8.62	0.0134	7.96	0.0186	7.06	0.0292	12.00	0.0025	2.99	0.2232
Matching														

Confounding adjustment method & data-sharing approach	Effectiveness outcome, change in body mass index[c]										Safety outcome[d]			
	<5%		≥5%		≥10%		≥20%		≥30%		Re-hospitalization		Re-intervention	
Pooled individual-level	2.41	2.03, 2.85	0.49	0.43, 0.55	0.34	0.30, 0.39	0.16	0.14, 0.19	0.10	0.07, 0.13	0.85	0.59, 1.22	0.90	0.61, 1.32
Risk-set	2.41	2.03, 2.85	0.49	0.43, 0.56	0.34	0.30, 0.39	0.16	0.14, 0.19	0.10	0.07, 0.13	0.85	0.59, 1.22	0.90	0.61, 1.32
Summary-table	3.75	3.16, 4.46	0.39	0.35, 0.44	0.37	0.32, 0.42	0.22	0.18, 0.26	0.12	0.08, 0.17	0.85	0.58, 1.24	0.89	0.59, 1.32
Effect-estimate, fixed-effect	2.36	1.99, 2.80	0.49	0.43, 0.55	0.34	0.30, 0.39	0.16	0.14, 0.20	0.11	0.08, 0.15	0.88	0.61, 1.27	0.92	0.62, 1.36
Effect-estimate, random-effects	2.22	0.88, 5.59	0.52	0.36, 0.74	0.35	0.24, 0.50	0.13	0.08, 0.21	0.05	0.01, 0.20	0.79	0.40, 1.55	0.79	0.37, 1.67
Measure of heterogeneity, Q[e]	21.66	<.0001	9.16	0.0102	7.95	0.0187	6.67	0.0356	6.90	0.0317	4.17	0.1242	5.07	0.0792

Note: AGB= Adjusted gastric banding; RYGB= Roux-en-Y gastric bypass; OR= odds ratio; CI= confidence interval; HR= hazard ratio

[a] The disease risk score was estimated using Cox proportional hazards regression on patients receiving Roux-en-Y gastric bypass procedure.

[b] There were 584 (6.2%) patients who underwent AGB and 8,777 (93.8%) patients who underwent RYGB.

[c] The incidences for <5%, ≥5%, ≥10%, ≥20%, and ≥30% change in body mass index were 68.0%, 93.7%, 76.5%, 31.2%, and 6.8%, respectively, for the AGB users; 32.9%, 99.1%, 96.8%, 83.8%, and 47.7%, respectively, for the RYGB users. These effectiveness outcomes were defined as the occurrence of the outcomes of interest closet to the end of the first post-procedure year so the incidences for <5% change in body mass index and ≥5% change in body mass index do not sum up to 100%.

[d] The incidences for re-hospitalization and re-intervention were 9.9% and 9.3%, respectively, for the AGB users, and 11.7% and 11.6%, respectively, for the RYGB users.

[e] Q is a measure of heterogeneity among the three data-contributing sites. The summary statistic and p-value from Cochran’s Q test are shown here.

Table 5.

Empirical Example 2: Results for Binary Outcomes from Propensity Score-Adjusted Analyses using Different Combinations of Confounding Adjustment Method and Data-Sharing Approach, Non-TNFi vs TNFi[a]

Confounding adjustment method & data-sharing approach	Effectiveness outcome		Safety outcome	
	Treatment switching[b]		Serious infections[c]	
	OR	95% CI	OR	95% CI
Stratification				
Pooled individual-level	0.52	0.34, 0.76	0.97	0.46, 1.86
Risk-set	0.52	0.34, 0.76	0.97	0.46, 1.86
Summary-table	0.52	0.34, 0.76	0.97	0.46, 1.86
Effect-estimate, fixed-effect	0.54	0.36, 0.79	1.06	0.56, 2.03
Effect-estimate, random-effects	0.53	0.32, 0.86	1.03	0.56, 2.03
Measure of heterogeneity, Q[d]	2.23	0.3268	0.78	0.6741
Matching				
Pooled individual-level	0.47	0.30, 0.73	1.07	0.46, 2.37
Risk-set	0.47	0.30, 0.73	1.07	0.46, 2.37
Summary-table	0.47	0.30, 0.73	1.07	0.46, 2.37
Effect-estimate, fixed-effect	0.47	0.31, 0.72	1.23	0.58, 2.64
Effect-estimate, random-effects	0.47	0.31, 0.72	1.23	0.58, 2.64
Measure of heterogeneity, Q[d]	0.55	0.7589	0.03	0.9820
Inverse probability weighting				
Pooled individual-level	0.36	0.23, 0.57	2.88	1.97, 4.21
Risk-set	0.36	0.23, 0.57	3.11	2.12, 4.55
Effect-estimate, fixed-effect	0.39	0.25, 0.61	3.17	2.16, 4.65
Effect-estimate, random-effects	0.45	0.18, 1.09	3.17	2.16, 4.65
Measure of heterogeneity, Q[d]	4.62	0.0992	0.85	0.6531
Matching weighting				
Pooled individual-level	0.51	0.32, 0.81	0.89	0.39, 2.04
Risk-set	0.51	0.32, 0.81	0.95	0.41, 2.19
Effect-estimate, fixed-effect	0.51	0.32, 0.82	0.93	0.40, 2.17
Effect-estimate, random-effects	0.51	0.32, 0.82	0.93	0.40, 2.17
Measure of heterogeneity, Q[d]	1.30	0.5213	0.42	0.8105

Note: TNFi=tumor necrosis factor-alpha inhibitor; OR= odds ratio; CI= confidence interval

[a] There were 407 (5.2%) patients who initiated non-TNFi, and 7,419 (94.8%) patients who initiated TNFi.

[b] The incidences for treatment switching were 7.6% in new users of non-TNFi and 11.2% in new users of TNFi.

[c] The incidences for serious infections were 2.9% in new users of non-TNFi and 3.1% in new users of TNFi.

[d] Q is a measure of heterogeneity among the three data-contributing sites. The summary statistic and p-value from Cochran's Q test are shown here.

Table 6.

Empirical Example 2: Results for Time-to-Event Outcomes from Propensity Score-Adjusted Analyses using Different Combinations of Confounding Adjustment Method and Data-Sharing Approach, Non-TNFi vs TNFi[a]

Confounding adjustment method & data-sharing approach	Effectiveness outcome		Safety outcome	
	Treatment switching[b]		Serious infections[c]	
	HR	95% CI	HR	95% CI
Stratification				
Pooled individual-level	0.59	0.41, 0.86	0.94	0.50, 1.77
Risk-set	0.59	0.41, 0.86	0.94	0.50, 1.77
Summary-table	0.59	0.39, 0.85	0.94	0.45, 1.78
Effect-estimate, fixed-effect	0.64	0.44, 0.93	1.04	0.55, 1.96
Effect-estimate, random-effects	0.63	0.26, 1.50	1.04	0.55, 1.96
Measure of heterogeneity, Q[d]	4.21	0.1215	0.83	0.6599
Matching				
Pooled individual-level	0.55	0.37, 0.83	1.07	0.51, 2.22
Risk-set	0.55	0.37, 0.83	1.07	0.51, 2.22
Summary-table	0.55	0.35, 0.82	1.07	0.47, 2.34
Effect-estimate, fixed-effect	0.57	0.38, 0.84	1.23	0.58, 2.61
Effect-estimate, random-effects	0.58	0.34, 1.02	1.23	0.58, 2.61
Measure of heterogeneity, Q[d]	2.37	0.3053	0.03	0.9701
Inverse probability weighting				
Pooled individual-level	0.60	0.38, 0.93	2.67	1.86, 3.84
Risk-set	0.60	0.35, 1.01	2.67	0.53, 13.52
Effect-estimate, fixed-effect	0.69	0.44, 1.07	2.93	2.03, 4.22
Effect-estimate, random-effects	0.79	0.26, 2.42	2.93	2.03, 4.22
Measure of heterogeneity, Q[d]	7.61	0.0222	0.81	0.6656
Matching weighting				
Pooled individual-level	0.60	0.39, 0.94	0.85	0.38, 1.92
Risk-set	0.60	0.41, 0.88	0.85	0.45, 1.63
Effect-estimate, fixed-effect	0.62	0.40, 0.96	0.89	0.39, 2.05
Effect-estimate, random-effects	0.60	0.30, 1.23	0.89	0.39, 2.05
Measure of heterogeneity, Q[d]	2.62	0.2689	0.37	0.8304

Note: TNFi=tumor necrosis factor-alpha inhibitor; HR= hazard ratio; CI= confidence interval

[a] There were 407 (5.2%) patients who initiated non-TNFi, and 7,419 (94.8%) patients who initiated TNFi.

[b] The incidences for treatment switching were 7.6% in new users of non-TNFi and 11.2% in new users of TNFi.

[c] The incidences for serious infections were 2.9% in new users of non-TNFi and 3.1% in new users of TNFi.

[d] Q is a measure of heterogeneity among the three data-contributing sites. The summary statistic and p-value from Cochran’s Q test are shown here.

Table 7.

Empirical Example 2: Results from Disease Risk Score[a]-Adjusted Analyses using Different Combinations of Confounding Adjustment Method and Data-Sharing Approach, Non-TNFi vs TNFi[b]

Confounding adjustment method & data-sharing approach	Effectiveness outcome		Safety outcome	
	Treatment switching[c]		Serious infections[d]	
	OR	**95% CI**	**OR**	**95% CI**
Stratification				
Pooled individual-level	0.53	0.35, 0.78	0.88	0.42, 1.64
Risk-set	0.53	0.36, 0.78	0.88	0.42, 1.64
Summary-table	0.53	0.35, 0.78	0.88	0.42, 1.64
Effect-estimate, fixed-effect	0.56	0.38, 0.82	0.97	0.52, 1.82
Effect-estimate, random-effects	0.51	0.28, 0.96	0.97	0.52, 1.82
Measure of heterogeneity, Q[e]	2.67	0.2624	0.24	0.8832
Matching				
Pooled individual-level	0.41	0.26, 0.63	0.86	0.37, 1.93
Risk-set	0.41	0.26, 0.63	0.86	0.37, 1.93
Summary-table	0.41	0.26, 0.63	0.86	0.37, 1.93
Effect-estimate, fixed-effect	0.41	0.27, 0.63	0.89	0.40, 1.98
Effect-estimate, random-effects	0.41	0.27, 0.63	0.89	0.40, 1.98
Measure of heterogeneity, Q[e]	1.77	0.4115	0.00	0.9961
	HR	**95% CI**	**HR**	**95% CI**
Stratification				
Pooled individual-level	0.59	0.41, 0.84	0.86	0.47, 1.57
Risk-set	0.59	0.41, 0.84	0.86	0.47, 1.57
Summary-table	0.58	0.39, 0.83	0.86	0.42, 1.57
Effect-estimate, fixed-effect	0.63	0.44, 0.91	0.95	0.52, 1.75
Effect-estimate, random-effects	0.59	0.26, 1.32	0.95	0.52, 1.75
Measure of heterogeneity, Q[e]	3.83	0.1466	0.21	0.8989
Matching				
Pooled individual-level	0.46	0.31, 0.68	0.89	0.42, 1.86
Risk-set	0.46	0.31, 0.68	0.89	0.42, 1.86
Summary-table	0.45	0.30, 0.68	0.88	0.38, 1.95
Effect-estimate, fixed-effect	0.47	0.32, 0.70	0.91	0.42, 2.00
Effect-estimate, random-effects	0.52	0.22, 1.25	0.91	0.42, 2.00
Measure of heterogeneity, Q[e]	4.01	0.1345	0.00	1.0000

Note: TNFi= tumor necrosis factor-alpha inhibitor; OR= odds ratio; CI= confidence interval; HR= hazard ratio

[a] The disease risk score was estimated using Cox proportional hazards regression on patients receiving tumor necrosis factor-alpha inhibitor biologics.

[b] There were 407 (5.2%) patients who initiated non-TNFi, and 7,419 (94.8%) patients who initiated TNFi.

[c] The incidences for treatment switching were 7.6% in new users of non-TNFi and 11.2% in new users of TNFi.

[d] The incidences for serious infections were 2.9% in new users of non-TNFi and 3.1% in new users of TNFi.

[e] Q is a measure of heterogeneity among the three data-contributing sites. The summary statistic and p-value from Cochran's Q test are shown here.