Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2018 September ; 27(9): 1034–1041. doi:10.1002/pds.4615.

Comparison of Privacy-Protecting Analytic and Data-sharing Methods: a Simulation Study

Kazuki Yoshida^{1,2}, Susan Gruber³, Bruce H. Fireman⁴, and Sengwee Toh³

¹Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States

²Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States

³Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, United States

⁴Division of Research, Kaiser Permanente Northern California, Oakland, California, United States

Abstract

PURPOSE—Privacy-protecting analytic and data-sharing methods that minimize the disclosure risk of sensitive information are increasingly important due to the growing interest in utilizing data across multiple sources. We conducted a simulation study to examine how avoiding sharing individual-level data in a distributed data network can affect analytic results.

METHODS—The base scenario had four sites of varying sizes with 5% outcome incidence, 50% treatment prevalence, and seven confounders. We varied treatment prevalence, outcome incidence, treatment effect, site size, number of sites, and covariate distribution. Confounding adjustment was conducted using propensity score or disease risk score. We compared analyses of three types of aggregate-level data requested from sites: risk-set, summary-table, or effect-estimate data (meta-analysis) with benchmark results of analysis of pooled individual-level data. We assessed bias and precision of hazard ratio estimates as well as the accuracy of standard error estimates.

RESULTS—All the aggregate-level data-sharing approaches, regardless of confounding adjustment methods, successfully approximated pooled individual-level data analysis in most simulation scenarios. Meta-analysis showed minor bias when using inverse probability of treatment weights (IPTW) in infrequent exposure (5%), rare outcome (0.01%), and small site (5,000 patients) settings. SE estimates became less accurate for IPTW risk-set approach with less frequent exposure and for PS-matching meta-analysis approach with rare outcomes.

CONCLUSION—Overall, we found that we can avoid sharing individual-level data and obtain valid results in many settings, although care must be taken with meta-analysis approach in infrequent exposure and rare outcome scenarios, particularly when confounding adjustment is performed with IPTW.

MeSH KEY WORDS

distributed databases; privacy-protecting methods; propensity score

INTRODUCTION

Distributed data networks such as the National Patient-Centered Clinical Research Network (PCORnet)¹ and the Sentinel System² are becoming a platform of choice for more rapid generation of real-world evidence. Multi-database analysis typically requires data be shared across sites and combined centrally at a coordinating center. Traditionally, detailed data including individual-level characteristics such as demographics and medical history was shared. However, this need for sharing potentially identifiable information across data partners can hinder multi-database research due to data partners' concerns about patient privacy, data security, and inappropriate or unauthorized use of data^{3,4}.

There have been efforts to create data-sharing platforms or technical infrastructure that offer better patient privacy protection^{5–8}. Multiple analytic and data-sharing methods have also been proposed to address these concerns⁹. Rassen *et al.* compared privacy-protecting methods including propensity score (PS)-based methods and full covariate sharing in two empirical examples – the association of proton pump inhibitors and myocardial infarction¹⁰, and the association of biological antirheumatic drugs and myocardial infarction¹¹. They found similar results from full covariate sharing and individual-level PS sharing. In the setting of survival analysis, an alternative approach to sharing individual-level data is to request risk-set information. Each risk set forms a single dataset record consisting of counts of individuals experiencing the study outcome at each relevant time point during follow-up and individuals remaining at risk at that time point¹². Toh *et al.* examined several methods, including pooled individual-level data analysis, risk-set data analysis, and meta-analysis¹², in empirical examples of the association between bariatric surgery and re-hospitalization³ and the association between anti-hypertensives and angioedema¹³. The latter study¹³ included a simulation sub-study limited to a setting without confounding.

To our knowledge, these more privacy-protecting methods have not been directly compared in robust simulation, and their relative performance remains poorly understood. This motivated us to conduct the current study to assess the relative performance of various approaches to estimating a conditional hazard ratio (HR) in time-to-event data in a simulated distributed data network. Specifically, we aimed to examine how different data-sharing methods in combination with several confounding adjustment methods affect analytic performance across various simulation scenarios compared to sharing individual-level data.

METHODS

We compared the performance of various data-sharing methods on estimation of HRs conditional on data partners (details and examples in eAppendix Section 1), using results from pooled individual-level analysis as the reference.

Data generation

We generated seven independent covariates in the base scenario. The first covariate was a standard normal variable. The other variables were binary Bernoulli variables with prevalence varying from 5% to 50%. We generated treatment assignment probability from a main terms logistic regression model. Variables were given varying levels of influence by assigning different coefficients – the odds ratio of treatment ranged from 0.3 to 3.0. Survival outcome assignment used the exponential model. We varied the incidence of the outcome (0.01%, 0.1%, 1%, and 5% annual incidence), prevalence of the treatment (5%, 10%, 25%, and 50%), size of sites (500, 1,000, 5,000, 10,000, and 20,000), number of sites (2, 4, 8, and 16), treatment effect (HR 0.8, 1.0, and 1.2), and variability of covariate distribution across sites (none, small, moderate, and large) to generate a range of simulation scenarios (eAppendix Section 2).

Confounder summary score estimation

We used PS¹⁴ and disease risk score (DRS)¹⁵ as the confounder summary score. PS was estimated within each site by regressing the treatment variable on all covariates. DRS was estimated among the untreated individuals within each site although in practice estimation using historical data is increasingly recommended¹⁶. Both models were correctly specified. eAppendix (Section 3) provides more details.

Confounding control using summary scores

We performed one-to-one PS and DRS matching of untreated to treated subjects using nearest neighbor method with a caliper¹⁷ set to twenty percent of the sample standard deviation¹⁸. PS matching was on the logit scale and DRS was matched on the log relative hazard scale. Stratification was based on deciles of the scores in the entire cohort. Weights were constructed as the inverse probability of treatment weights (IPTW)^{19,20} and matching weights (MW)^{21,22}. See Table 1 for classification and eAppendix Section 4 for more detailed description of the estimands.

Data reduction, sharing, and analysis

We compared analyses of individual-level data pooled across all participating sites with analyses of three types of summarized data requested from sites: (1) risk-set data, (2) summary-table data, or (3) effect-estimate data with variances (eAppendix Sections 5 and 6). Individual-level data contained the outcome, exposure, and one of the following: match set indicators, stratum indicators, or weights. The risk-set data contained risk set composition at each unique event time within the matched cohort or each stratum. In the case of weighting, the risk sets also carried information on the total weights and the variance of weights. The summary-table data contained a two-by-two table in the matched cohort or within each stratum. The effect-estimate data contained the site-specific point estimate of treatment effect and its variance. With a matched design, these data outputs were generated using site-specific Cox regression. Stratified Cox regression (allowing for stratum-specific baseline hazard function) was used in the within-site stratified analysis. The within-site weighted analyses were performed using weighted Cox regression with the robust sandwich variance estimator²³.

We aggregated data into one combined dataset for the final analysis. Final data analysis differed based on the data-sharing methods but was always stratified by site (eAppendix Section 6). This mimicked the process and analysis typically conducted in an actual distributed data network, where stratification by site may mitigate potential differences in important characteristics, such as baseline outcome rate, across the participating sites.

Performance assessment

The average of point estimates was used for two types of comparisons. The first comparison was against the underlying true estimand in the data generation process, which gave the *bias*, i.e., average deviation from the truth. The second comparison was against the pooled individual-level data analysis, which we called *agreement* for distinction.

Two types of precision measures were obtained: the *simulation-based standard error* (SE; standard deviation of point estimates across iterations) and *average estimated SE*. We took the former as the true variability of the estimator. The average estimated SE was then compared to the simulation-based SE to examine accuracy.

We examined the proportion of simulation iterations that failed to produce interpretable results. A low proportion indicated that the method failed less often, though it might not necessarily mean more accurate results. We give the precise definitions of these assessment metrics in eAppendix Section 7.

RESULTS

Base scenario

Figure 1 shows the distribution of point estimates by confounding adjustment method and data-sharing method. All PS and DRS methods produced average estimates between 0 and 0.05 in this base scenario that had a true log HR of 0 (HR 1.0). They were all substantially less biased than the unadjusted results (unadjusted log HR > 0.75 and HR > 2.1). All data-sharing methods agreed with the pooled individual-level data analysis within each confounding adjustment method. PS performed slightly better than DRS regarding bias. Matching and both PS weighting methods were less biased than stratification on summary score decile.

The ratios of average SEs to the corresponding simulation SE (eFigure 1) were on average within $\pm 5\%$ of 1.0 for all methods, indicating accurate estimation of variability. We did not observe substantial difference across data-sharing methods within each adjustment method. PS IPTW sometimes produced some outliers, including those overestimating by 25% (confidence interval would be too wide) although it was less extreme when combined with the meta-analysis method.

In the subsequent sections, we summarize findings from our analyses for each performance metric.

Bias and agreement

Across three true conditional HRs (1.2, 1.0, and 0.8), both PS and DRS stratification methods were increasingly biased with decreasing exposure prevalence (Figure 2) because the strata were defined based on the distribution of these scores in the entire cohort and lowering the prevalence of the exposure resulted in strata with no exposed individuals, requiring collapsing strata. The agreement between the pooled individual-level data analysis and other data-sharing methods was generally excellent except for PS IPTW, where the meta-analysis approach produced HR estimates that were slightly different (< 0.05 on the HR scale) from the estimates in the pooled individual-level data analysis at exposure prevalence 10%. Upon further investigation, it appeared to be due to small sites contributing biased results with relatively large inverse variance weights, biasing the pooled results.

Varying the incidence of outcome influenced in-cohort estimation of DRS, making the results from DRS analyses more biased as incidence decreased. PS methods were relatively stable regardless of data-sharing method (Figure 3). At an outcome incidence of 0.01%, the meta-analysis method departed from the pooled individual-level data analysis, but less so if done with PS-matched site-specific effect estimates. This departure was also due to the contribution of biased results from smaller sites. PS matching was less affected because smaller sites often had even smaller matched cohorts, which might not have events in both treated and untreated groups to contribute to the meta-analysis.

When we varied site size and number of sites involved, we found good agreement among data-sharing methods except for meta-analysis (Figure 4) in settings with 500 and 1,000 individuals at each site. The DRS methods were more biased, likely due to the lack of the outcome at these site sizes, but the bias reduced with increasing site sizes. The disagreement between meta-analysis and pooled individual-level data analysis was most noticeable, albeit small (< 0.05 on the HR scale), in PS IPTW. This disagreement became less prominent when the individual site size increased. Increasing the number of sites of the same size did little for reducing bias.

Heterogeneity of covariate distributions across sites had little impact on bias (eFigure 2) (< 0.01 on log HR scale). Agreement with the pooled individual-level data analysis was good, except for PS IPTW, which exhibited minor disagreement.

True and estimated variability of point estimates

The precision of estimates generally improved as exposure prevalence increased (eFigure 3). The accuracy of SE estimates was similar across the data-sharing methods except PS IPTW in which minor disagreement across methods was observed, particularly with the experimental weighted risk-set data analysis (eFigure 4).

Across all methods, the precision of HR estimates improved as outcome incidence increased (eFigure 5). Varying the incidence of outcome produced different degrees of accuracy of the SE estimates, most noticeably in PS matching and DRS methods (eFigure 6). In PS matching, lowering the outcome incidence led to overestimation of the SEs, in particular for meta-analyses. The SEs for DRS matching and stratification were underestimated at the

incidence of 0.1%, but became more accurate at 0.01%. Importantly, this result at the incidence of 0.01% was not accompanied by improvement in bias (Figure 3).

When we varied the site size and number of sites, the underlying true variability were similar across data-sharing methods and decreased with increasing site size or number of sites (eFigure 7). The SE estimates demonstrated similar trends when reducing incidence at small site sizes of 500 and 1,000 (eFigure 8).

Varying the covariate distributions across sites had little impact on the simulation SE (eFigure 9). SE estimates were on average accurate and none of the methods underestimated by greater than 5% (eFigure 10).

Proportions of successful simulation iterations

At outcome incidence of 0.01%, model fit failure occurred in a small fraction of iterations for all confounding adjustment methods and data-sharing methods (Figure 3, in-panel numbers). Lack of any events in either the treated or untreated group, which made the dataset non-informative in the Cox regression, caused this issue. Therefore, it was particularly noticeable in the PS matching, which further trimmed the cohort, and in meta-analysis approach that required an effect estimate be available within each site. Although DRS matching had fewer failures, this was because the DRS model performed so poorly that most individuals received similar DRS regardless of their covariate values. This made matching easy, but failed to adjust for confounding adequately. Model fitting failure was not noticeable in other scenarios for any of the methods examined.

DISCUSSION

We examined various privacy-protecting analytic and data-sharing methods through a simulation study to assess whether restricting data sharing could affect the performance of analytic methods compared to the pooled individual-level data analysis. Overall, all the data-sharing methods assessed demonstrated similar performance to the pooled individual-level data analysis in most simulation scenarios. Small but noticeable differences were observed in extreme settings such as rare outcome (0.01%), infrequent exposure (5%), and small sites (500 or 1,000 patients). Aggregating and sharing event counts and person-time data into a summary-level table worked well in this constant-hazard setting with low incidence. Risk-set data analysis was shown to be a useful intermediary between summary-table sharing and individual-level data sharing. The extension of risk-set data analysis to PS-weighted data produced results on bias and precision of HR estimates comparable to pooled individual-level data analysis, although the SE estimates showed a minor difference when used with PS IPTW in infrequent exposure settings.

Data-sharing methods had little impact on bias and precision of log HR estimates within each confounding adjustment method in most simulated scenarios with exposure prevalence > 10%, outcome incidence > 0.1%, and each site contributing > 10,000 patients. This implies that in the setting where each data partner provides a sufficient sample of similar patient populations, a meta-analysis of site-specific effect estimates, the data-sharing method with the least amount of data sharing, may be an attractive option. The additional benefit of

this approach is that the investigator at the coordinating center will always see the site-specific effect estimates, which may help identify treatment effect heterogeneity.

In the presence of treatment effect heterogeneity across sites, the issue about the appropriateness of pooling data from all sites applies to all data-sharing methods studied. When there is treatment effect heterogeneity, presenting site-specific results separately along with respective baseline patient characteristics may be a better option than providing a single pooled effect estimate. Unlike publication-based meta-analysis, investigators can often remove some sources of heterogeneity (e.g., different eligibility criteria, different analytic methods) in a distributed data network. If the assessment of site-specific baseline patient characteristics indicates different compositions of patients, treatment effect modification by baseline characteristics should be investigated. Meta-analysis can be conducted within subgroups defined by potential effect modifiers. Similarly, risk sets can be constructed within levels of potential effect modifiers when conducting risk-set data sharing. However, some differences, for example, varying coding practice across sites, may be difficult to identify in distributed data networks.

The summary-table approach, a two-by-two table analysis of person-time data, can be reasonable in studies with time-constant hazards. The risk-set data and the individual-level data both preserve time component of the time-to-event data, allowing for examination of time-varying treatment effect if necessary. This becomes particularly important when depletion of susceptibles²⁴ is more likely because of high incidence and/or long follow up.

Our study complements the existing literature in several ways. A previous review⁹ provides an overview of confounding adjustment and outcome analysis in distributed data networks. In this paper, we systematically classified various distributed analytic and data-sharing methods using three different domains – the type of summary scores, confounding adjustment methods, and data-sharing methods (Table 1). We considered PS and DRS, two of the most commonly used confounder summary scores in the context of matching, stratification, and weighting. Data-sharing methods, the last axis of classification, was the main interest of the current study. This classification is by no means comprehensive. It nevertheless clarifies how these three axes are orthogonal and can be combined.

We implemented an experimental method for risk-set data combined with PS weighting. This method produced point estimates comparable to the corresponding pooled individual-level data-sharing method. SE estimates were also similar in most scenarios. However, in the scenario with very different treatment arm sizes (10% treated), the IPTW SE estimates were further from the actual variability than individual-level data analysis (eFigure 2) although MW was less affected. Additional assessment of this promising approach is warranted.

To our knowledge, the current study is the most extensive examination of different data-sharing methods in combination with various commonly used confounding adjustment techniques. We introduced some restrictions to the data generation mechanisms to make the comparison of the methods tractable. The number of covariates was small, and they were independent. High-dimensional and collinear covariates may require special handling, such

as penalized regression²⁵ for PS and DRS modeling. We made the true underlying treatment effect of interest homogeneous across data partners to emulate a setting in which data pooling across sites is appropriate. In practice, theoretical and empirical evidence of homogeneity of treatment effects across sites should determine the appropriateness of combining data from different sites. The survival outcome was generated with a constant hazard of events over time. Poisson regression using summary-table data sharing that reduces the full time-to-event data into totals of event counts and person-time for the treated and untreated is comparable to Cox regression if the underlying hazards are time-constant²⁶. When the underlying hazards are time-varying, for example, higher adverse event rates are expected early after medication initiation, summary-table data may require stratification into clinically meaningful time periods. In this case, one may prefer sharing fully time-stratified risk-set data. We applied IPTW without truncation or trimming. This may have resulted in sensitivity to covariate non-overlap and instability compared to MW. We did not examine methods that include PS or DRS as independent variables in outcome regression models because they require other data-sharing techniques that are beyond the scope of our study (e.g., distributed regression)³.

To conclude, data-sharing methods had little impact on bias and precision of log HR estimates in most simulated scenarios. Small differences were observed in extreme settings such as rare outcome, infrequent exposure, and small sites. In practice, meta-analysis of site-specific effect estimates is useful if treatment effect heterogeneity across sites is minimal. In the presence of substantial heterogeneity, investigation of treatment effect modification can be done by sharing site-specific estimates within levels of potential effect modifiers or covariate-stratified risk-set data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING INFORMATION: This study was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1403-11305). 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. The authors have no competing interests to disclose. Dr. Toh is also supported by the National Institute of Biomedical Imaging and Bioengineering (U01EB023683). Dr. Yoshida received funding from Harvard T.H. Chan School of Public Health (partially supported by training grants from Pfizer, Takeda, Bayer, and PhRMA), Japan Student Services Organization, and Honjo International Scholarship Foundation.

The authors would like to thank Mia Gallagher for her project management support.

References

1. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc*. 2014; 21:578–582. DOI: 10.1136/amiajnl-2014-002747 [PubMed: 24821743]
2. Platt R, Carnahan RM, Brown JS, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol Drug Saf*. 2012; 21:1–8. DOI: 10.1002/pds.2343
3. Toh S, Shetterly S, Powers JD, Arterburn D. Privacy-preserving analytic methods for multisite comparative effectiveness and patient-centered outcomes research. *Med Care*. 2014; 52:664–668. DOI: 10.1097/MLR.000000000000147 [PubMed: 24926715]

4. Mazor K, Richards A, Gallagher M, et al. Stakeholders' views on data sharing in multi-center studies. [in press]. *J Comp Eff Res*. 2017
5. Carter KW, Francis RW, Carter KW, et al. ViPAR: a software platform for the Virtual Pooling and Analysis of Research Data. *Int J Epidemiol*. 2015; doi: 10.1093/ije/dyv193
6. Ohno-Machado L, Bafna V, Boxwala AA, et al. iDASH: integrating data for analysis, anonymization, and sharing. *J Am Med Inform Assoc*. 2012; 19:196–201. DOI: 10.1136/amiajnl-2011-000538 [PubMed: 22081224]
7. Gaye A, Marcon Y, Isaeva J, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. *Int J Epidemiol*. 2014; 43:1929–1944. DOI: 10.1093/ije/dyu188 [PubMed: 25261970]
8. Davies M, Erickson K, Wyner Z, Malenfant J, Rosen R, Brown J. Software-Enabled Distributed Network Governance: The PopMedNet Experience. *EGEMS (Wash DC)*. 2016; 4:1213.doi: 10.13063/2327-9214.1213 [PubMed: 27141522]
9. Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med Care*. 2013; 51:S4–10. DOI: 10.1097/MLR.0b013e31829b1bb1
10. 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. DOI: 10.1002/pds.1867 [PubMed: 20162632]
11. Rassen JA, Solomon DH, Curtis JR, Herrinton L, Schneeweiss S. Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologics. *Med Care*. 2010; 48:S83–89. DOI: 10.1097/MLR.0b013e3181d59541 [PubMed: 20473213]
12. Fireman B, Lee J, Lewis N, Bembom O, van der Laan M, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. *Am J Epidemiol*. 2009; 170:650–656. DOI: 10.1093/aje/kwp173 [PubMed: 19625341]
13. 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; 22:1171–1177. DOI: 10.1002/pds.3483 [PubMed: 23878013]
14. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55. DOI: 10.1093/biomet/70.1.41
15. Hansen BB. The prognostic analogue of the propensity score. *Biometrika*. 2008; 95:481–488. DOI: 10.1093/biomet/asn004
16. 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. DOI: 10.1002/pds.3231 [PubMed: 22552989]
17. Sekhon JS. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J Stat Soft*. 2011; 42:1–52.
18. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011; 46:399–424. DOI: 10.1080/00273171.2011.568786 [PubMed: 21818162]
19. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
20. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health*. 2010; 13:273–277. DOI: 10.1111/j.1524-4733.2009.00671.x [PubMed: 19912596]
21. Li L, Greene T. A weighting analogue to pair matching in propensity score analysis. *Int J Biostat*. 2013; 9:215–234. DOI: 10.1515/ijb-2012-0030 [PubMed: 23902694]
22. Yoshida K, Hernandez-Diaz S, Solomon DH, et al. Matching Weights to Simultaneously Compare Three Treatment Groups: Comparison to Three-way Matching. *Epidemiology*. 2017; 28:387–395. DOI: 10.1097/EDE.0000000000000627 [PubMed: 28151746]
23. Lumley T. *Complex Surveys: A Guide to Analysis Using R*. 1. Hoboken, N.J: Wiley; 2010.
24. Hernán MA. The Hazards of Hazard Ratios. *Epidemiology*. 2010; 21:13–15. DOI: 10.1097/EDE.0b013e3181c1ea43 [PubMed: 20010207]

25. Franklin JM, Eddings W, Glynn RJ, Schneeweiss S. Regularized Regression Versus the High-Dimensional Propensity Score for Confounding Adjustment in Secondary Database Analyses. *Am J Epidemiol*. 2015; 182:651–659. DOI: 10.1093/aje/kwv108 [PubMed: 26233956]
26. Holford TR. The analysis of rates and of survivorship using log-linear models. *Biometrics*. 1980; 36:299–305. [PubMed: 7407317]

Figure 1.

Distribution of log hazard ratio point estimates by confounding adjustment method (panels) and data-sharing method (X axis) in the base scenario.

The true underlying log HR was 0 (HR 1.0) in the base scenario.

Abbreviations: HR: hazard ratio; PS: Propensity score; IPTW: Inverse probability of treatment weights; MW: Matching weights; DRS: Disease risk score; meta: Meta-analysis using site-specific effect-estimate data; summary: Summary-table data analysis; risk set: Risk-set data analysis; individual: Pooled individual-level data analysis (reference method).

Figure 2.

Bias assessment when varying exposure prevalence at different treatment effects.

The columns represent different summary score-based confounding adjustment methods.

The rows represent results at different true conditional hazard ratios (0.8, 1.0, and 1.2). In

each panel, lines represent the same data-sharing methods.

Abbreviations: HR: hazard ratio; PS: Propensity score; IPTW: Inverse probability of treatment weights; MW: Matching weights; DRS: Disease risk score; meta: Meta-analysis using site-specific effect-estimate data; summary: Summary-table data analysis; risk set: Risk-set data analysis; individual: Pooled individual-level data analysis (reference method).

Figure 3.

Bias assessment when varying outcome incidence at different treatment effects.

The columns represent different summary score-based confounding adjustment methods.

The rows represent results at different true conditional hazard ratios (0.8, 1.0., and 1.2). In each panel, lines represent the same data-sharing methods. Overlapping lines indicate good agreement among data-sharing methods. The numbers show in the boxes within each panel are % failed iterations. From the top to the bottom, they represent meta-analysis, summary-table data analysis, risk-set data analysis, and pooled individual-level data analysis.

Abbreviations: HR: Hazard ratio; PS: Propensity score; IPTW: Inverse probability of treatment weights; MW: Matching weights; DRS: Disease risk score; meta: Meta-analysis using site-specific effect-estimate data; summary: Summary-table data analysis; risk set: Risk-set data analysis; individual: Pooled individual-level data analysis (reference method).

Figure 4.

Bias assessment when varying the size of sites at different numbers of sites.

The columns represent different summary score-based confounding adjustment methods.

The rows represent results at different numbers of sites in a distributed data network. In each panel, lines represent the same data-sharing methods. Overlapping lines indicate good agreement among data-sharing methods.

Abbreviations: HR: Hazard ratio; PS: Propensity score; IPTW: Inverse probability of treatment weights; MW: Matching weights; DRS: Disease risk score; meta: Meta-analysis using sites-specific effect-estimate data; summary: Summary-table data analysis; risk set: Risk-set data analysis; individual: Pooled individual-level data analysis (reference method).

Table 1

Combination of confounding adjustment methods (both margins) and data-sharing methods (within each cell).

Confounding adjustment method			
Confounder summary scores	Matching	Stratification	Weighting
<i>Propensity score</i>	Individual-level data	Individual-level data	Individual-level data
	Risk-set data	Risk-set data	Risk-set data
	Summary-table data	Summary-table data	-
	Effect-estimate data	Effect-estimate data	Effect-estimate data
<i>Disease risk score</i>	Individual-level data	Individual-level data	-
	Risk-set data	Risk-set data	-
	Summary-table data	Summary-table data	-
	Effect-estimate data	Effect-estimate data	-

-: Methods not currently available or tested in our study.