HHS Public Access

Author manuscript

Epidemiology. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Epidemiology. 2017 March ; 28(2): 249–257. doi:10.1097/EDE.0000000000000595.

A Propensity score based fine stratification approach for confounding adjustment when exposure is infrequent

Rishi J Desai[1], **Kenneth J Rothman**[2,3], **Brian T Bateman**[1,4], **Sonia Hernandez-Diaz**[5], and **Krista F Huybrechts**[1]

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA

[2]Research Triangle Institute, Research Triangle Park, NC, USA

[3]Boston University School of Public Health, Boston, MA, USA

[4]Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[5]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Abstract

Background—When exposure is infrequent, propensity score matching results in reduced precision because it discards a large proportion of unexposed patients. To our knowledge, the relative performance of propensity score stratification in these circumstances has not been examined.

Methods—Using an empirical example of the association of first-trimester statin exposure (prevalence=0.04%) with risk of congenital malformations and 1,000 simulated cohorts (n=20,000) with eight combinations of exposure prevalence (0.5%, 1%, 5%, 10%) and outcome risk (3.5%, 10%), we compared four propensity score based approaches to confounding-adjustment: 1) matching (1:1, 1:5, full), 2) stratification in 10, 50, and 100 strata by entire cohort propensity score distribution, 3) stratification in 10, 50, and 100 strata by exposed group propensity score distribution, 4) standardized mortality ratio (SMR) weighting. Weighted generalized linear models were used to derive effect estimates after weighting unexposed according to the distribution of the exposed in their stratum for the stratification approaches.

Results—In the empirical example, propensity score stratification (cohort) approaches resulted in greater imbalances in covariate distributions between statin-exposed and unexposed compared

Correspondence: Rishi J Desai, MS, PhD, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street, Suite 3030-R, Boston, MA 02120, USA, Phone: 617-278-0932 | Fax: 617-232-8602, rdesai@bwh.harvard.edu.

Data and code availability: Computing codes used in this study will be made available at http://www.drugepi.org/dope-downloads/. The data are not available for replication because of IRB restrictions.

Conflict of interest/Financial disclosures:

Dr. Huybrechts is supported by a career development award from the National Institute of Mental Health (K01 MH099141). Dr. Bateman is supported by a career development award from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the NIH (K08HD075831). Dr. Hernandez-Diaz is supported by the NIH grant R01 MH100216 and has consulted for AstraZeneca (London, UK) for unrelated projects. The other authors declare no other relationships or activities that could appear to have influenced the submitted work.

with propensity score stratification (exposed) and matching. In simulations, propensity score stratification (exposed) resulted in smaller relative bias than the cohort approach with 10 and 50 strata, and greater precision than matching and SMR weighting at 0.5% and 1% exposure prevalence; but similar performance at 5% and 10%.

Conclusion—For exposures with prevalence under 5%, propensity score stratification with fine strata, based on the exposed-group propensity score distribution, produced the best results. For more common exposures, all approaches were equivalent.

Keywords

Propensity score stratification; simulation study; rare exposure

Introduction

Propensity scores, which are balancing scores defined as the predicted probability of receiving a particular exposure given covariate realizations of a study subject,[1] are commonly used to account for a large number of confounders efficiently in pharmacoepidemiology studies, especially when the outcome is infrequent and adjusting for a large number of confounders in conventional multivariable models may result in overfitting and biased estimates.[2] Two of the most widely used propensity score-based approaches to control confounding are matching a comparison group to the exposed group by propensity score and stratifying subjects by a function of the score.[3] Propensity score matching has previously been shown to be superior with respect to bias reduction to stratification using five strata.[4,5] Matching, however, reduces precision because, along with exposed patients for whom no reference patient with comparable propensity is found, it also omits those comparison patients who would be matchable but who are not selected because a pre-specified number of matches were already selected for each exposed patient.[4,5] This loss of subjects is especially relevant when exposure is rare, in which case propensity-score matching excludes a large proportion of comparison subjects and results in considerable loss of information. In contrast, all these potential comparison subjects would be retained in an analysis that stratifies or weights by a function of the propensity score.

It has long been thought that creating five strata based on a continuous variable, such as quintiles of the propensity score, with the stratum boundaries determined by its distribution in the exposed and the comparison group combined, eliminates approximately 90% of measured confounding.[6,7] When exposure is infrequent, however, determining the stratum boundaries by the propensity score values for the combined groups of exposed and comparison patients may result in all the exposed patients being aggregated in one or more extreme strata. This approach is therefore prone to considerable residual confounding. This aggregation of exposed subjects in the extreme strata can be mitigated by increasing the number of strata, or by forming the stratum boundaries based on the values of the PS for the exposed group alone. While both these approaches have been discussed in the literature,[8,9] they have not been compared with other standard confounding control approaches in settings with infrequent exposure, where the aggregation of exposed at one end of the PS distribution is most severe. Therefore, using both a simulation study and a previously published empirical example of the fetal safety of first-trimester statin use during pregnancy,[10] we

assessed confounding control and precision for various PS stratification approaches and for PS matching and weighting.

Methods

Empirical example

For our empirical evaluation, we assessed the prevalence of congenital malformations among live-born infants after first-trimester statin exposure, compared with non-exposure. Statins are contraindicated during pregnancy owing to teratogenicity concerns from animal-studies. Further, the indications for statin use are uncommon in women of reproductive age. Therefore, exposure to statins during pregnancy is rare. We studied a previously assembled cohort of Medicaid-enrolled women aged 12 to 55 years with completed pregnancies resulting in live-born infants, from 46 U.S. states and Washington, DC, for the period of 2000–2007.[11] The use of this de-identified database for research was approved by the Institutional Review Board of Brigham and Women's Hospital. We estimated the date of last menstrual period (LMP) based on the delivery date combined with a validated algorithm based on diagnosis codes for preterm delivery.[12] We required women to have Medicaid eligibility continuously from 3 months before the estimated LMP month through one month postpartum. Similarly, to have complete information on malformations during the outcome measurement period (first 3 months of life), linked infants were also required to have continuous enrollment in Medicaid for at least 3 months following birth, unless they died in which case a shorter eligibility period was allowed. We excluded pregnancies with exposure to a known teratogenic medication (i.e., lithium, antineoplastic agents, retinoids, or thalidomide) during the first trimester and pregnancies in which the infant was diagnosed with a chromosomal abnormality.

For this analysis, women who filled at least two prescriptions for statins during the first trimester were defined as exposed. Statins included in our study were: simvastatin, lovastatin, pravastatin, fluvastatin, atorvastatin, cerivastatin, and rosuvastatin. We only considered statin exposure in the first trimester because it is the etiologically relevant window of exposure for the outcome of congenital malformations. Women not meeting this exposure definition were classified as unexposed. The outcome of interest was any congenital malformation in the infant. These malformations were identified based on the presence of ICD-9 diagnostic codes on two or more separate days in the infant inpatient or outpatient records during the first 3 months of life. The potential confounders considered were demographic characteristics of maternal age, race and region, maternal comorbid conditions including diabetes, hypertension, dyslipidemia, obesity, tobacco use, chronic renal disease, illicit drug and alcohol abuse, maternal co-medication use including other suspected teratogenic medications, oral anti-diabetic medications, hypertension medications, and insulin, and healthcare utilization variables including the number of distinct prescription medications (other than statins) and physician visits. Details on the study design, variable measurements, and substantive results have been previously published.[10]

Simulation study

To augment the findings from the empirical example, we also conducted a simulation study with 1,000 simulated cohorts of 20,000 women. To base our simulation parameters on realistic covariate distributions, we generated ten confounders (c1-c10) using confounder–exposure associations, confounder–outcome associations and confounder prevalences observed in the empirical study. Additionally, we generated two hypothetical non-confounding exposure determinants (c11, c12) and two hypothetical non-confounding outcome determinants (c13, c14). eTable 1 summarizes the parameters that were used to generate simulated data.

A binary exposure variable was generated indicating first-trimester statin use versus non-use. A logistic regression model was used to predict the probability of the exposure using ten confounders and two exposure-determinants (c1-c12). A binary outcome variable was generated indicating the presence or absence of congenital malformations, also using logistic regression models with ten confounders (c1-c10) and two outcome-determinants (c13, c14). A true null exposure effect of statins on the risk of congenital malformations was generated by not including the simulated exposure variable in the logistic regression model predicting the outcome. We used logistic regression models to generate the binary outcome variable and retrieve the true effect estimates as odds ratios (ORs) in our simulations as these models are theoretically more robust to non-convergence compared with log-binomial models[13–15] and when the treatment effect is null, ORs and risk ratios (RRs) coincide.

We generated eight simulation scenarios by varying the exposure prevalence and outcome risks. We varied simulated exposure prevalence (0.5%, 1%, 5%, and 10%) by changing the intercept term in the logistic regression models used to generate the exposure variable, while keeping other coefficients constant. Two outcome risk scenarios were generated for each exposure scenario; the first was based on the risk observed in the empirical example (3.5%) and the second was chosen to represent more common outcomes (10%). The null exposure effect was held constant across the simulation scenarios.

Statistical analysis

We implemented propensity score-based methods in both the empirical example and the simulation studies to account for measured confounders. The propensity score was estimated as the predicted probability of statin exposure during the first trimester using logistic regression models. In the empirical study, all the confounders described above (under Empirical example) were included in the propensity-score model, while in the simulation studies the ten confounders, c1-c10, were included in the propensity-score models. After propensity-score estimation, the following three approaches were used to derive adjusted associations between statin exposure and congenital malformations.

1. Propensity score stratification: Two approaches were used to conduct stratification: 1) creating equally-sized propensity score strata, numbering 10, 50, or 100, after ranking the entire-cohort based on the propensity score (hereafter referred to as the propensity-score strata cohort approach); and 2) creating unequally sized propensity-score strata, numbering 10, 50, or 100, after ranking only the exposed patients based on the propensity score and assigning unexposed

patients to these strata based on their PS (hereafter referred to as the propensity-score strata exposed approach).

Weighted regression models were used to derive an adjusted exposure effect after stratification, in which each exposed patient received a weight of 1 and unexposed patients were weighted in proportion to the distribution of the exposed in the stratum into which they fell. The unexposed group weights were scaled to sum to the number of unique unexposed individuals included in the analysis ($\text{unexposed weights} = \frac{(N_{\text{exposed in strata i}}/N_{\text{total exposed}})}{(N_{\text{unexposed in strata i}}/N_{\text{total unexposed}})}$). Effectively, this weighting creates a pseudo-population in which confounder distribution concordance is achieved between the exposed and unexposed groups, to the extent that it is achieved within each stratum. The exposure effect estimate can then obtained by computing the marginal effect estimates in this weighted population using generalized linear models, with exposure term as the only independent variable. Since the weighting in this approach aims to make the confounder distribution among the unexposed akin to their distribution in the exposed, the marginal estimates computed in the weighted population consistently estimate the average treatment effect among the exposed (also referred to as the average treatment effect among the treated).[3,16] SAS macros used for the propensity score stratification approaches can be downloaded from our website (http://www.drugepi.org/dope-downloads/).

2. Propensity score matching: We used three propensity score-matching approaches. The first two approaches matched unexposed women to each statin-exposed woman using a nearest neighbor approach with a caliper of 0.01 in ratios of 1:1 and 1:5. For 1:5 matching, we employed a variable-ratio matching strategy, allowing for fewer than the target number of matches as long as at least one match is found, as this strategy has been recommended over fixed-ratio matching as the preferred method for achieving greater confounding control.[17,18] The third propensity score-matching approach consisted of a recently proposed full matching strategy, in which exposed and unexposed individuals were matched on the propensity score to form matched sets that contained at least one exposed and at least one unexposed individual using an optimal matching algorithm seeking to minimize the mean within matched-set differences in the propensity score between the exposed and unexposed individuals.[19] Unlike traditional propensity score-matching approaches, full matching seeks to utilize information on the majority of the original patient population by including all matchable exposed and unexposed individuals in matched sets. Full propensity-score matching was implemented using the R package MatchIt (R Version 3.2.3).[16] After matching, generalized linear models were used to derive average treatment effect estimates among the exposed for all three approaches. For 1:5 and full matching, weights induced by matching were incorporated in the regression models. These weights were created in a manner similar to the stratification weights, where unexposed individuals were weighted according to the distribution of the exposed in each matched set. To account for the clustering

of subjects within matched sets, we used a robust variance estimator to compute 95% confidence intervals (CI) for propensity score-matching methods.[19]

3. SMR weighting: We also used the SMR weighting approach based on the propensity score to derive exposure effect estimates. In this approach, the unexposed patients within our cohort received weights equal to the ratio of (PS/(1 - PS)), while the exposed patients received weights of 1.[20] Marginal effect estimates and 95% CI for exposure effects were computed using weighted generalized linear models and robust variance estimators. Since this approach reweights the unexposed population to be similar to the exposed population with respect to confounder distribution, it also results in estimation of the average treatment effect among the exposed.[21]

We excluded the observations from the non-overlapping regions of the propensity-score distributions among exposed and unexposed populations before conducting propensity-score full-matching, stratification and SMR weighting. This step, also referred to as 'trimming', ensures exclusion of patients who will always or never receive therapy because of indications or contraindications and focuses the estimation of treatment effects in a population with clinical equipoise.[22] The ability of propensity score-based approaches to allow researchers to measure treatment effects in a population with clinical equipoise through trimming is one of its great strengths over traditional multivariable outcome regression models.

Evaluation of performance

In the simulation study, we compared performance across 1000 simulations of each method on the following metrics: 1) confidence limit ratios (CLR),[23] which were computed on the OR scale as the average of ratios of upper to lower 95% confidence limit for the exposure effect estimates, as indicators of precision; 2) relative (%) bias estimates as indicators for the extent of confounding control, which were calculated on the OR scale as $\frac{\bar{\beta}-\beta_{True}}{\beta_{True}} \times 100$; where $\bar{\beta}$ is the exponentiated average of the estimated co-efficients and β_{True} is exponentiated true exposure co-efficient, 3) mean squared errors (MSE), which were computed on the log OR scale as the average of the squared differences between estimated and true exposure coefficients, as indicators for the overall accuracy.[24]

In the empirical study, baseline characteristics of the full cohort were reported by statin exposure status. The differences in the baseline characteristics between the exposed and the unexposed women were summarized using absolute standardized differences.[25] For 1:5 and full PS-matching, all stratification approaches, and the SMR weighting approach, absolute standardized differences were computed after weighting the unexposed observations according to the weight calculations described above. We further computed risk differences (RDs) and RRs along with their 95% confidence interval for each of the method using weighted generalized linear models with identity and log links, respectively in the SAS GENMOD procedure (SAS version 9.3, SAS institute, Cary, NC). Since the truth is unknown in the empirical study, we did not compute other measures of performance (relative bias estimates and MSEs).

Results

Empirical study

The full cohort comprised 886,996 pregnancies. Of these, 335 (0.04%) were classified as exposed to statins during the first trimester. Important baseline differences between statin exposed and unexposed women were observed in the full cohort (Table). Statin-exposed women were older, had a higher prevalence of all of the comorbid conditions considered, and higher use of antihypertensive medications, insulin, and oral diabetes medications. 1:1 and 1:5 PS-matching, SMR weighting as well as propensity score stratification (exposed) approaches resulted in excellent covariate balance between exposed and unexposed groups, as demonstrated by low values of absolute standardized differences (Table 1). Notably, propensity score stratification (cohort) approaches demonstrated remaining imbalances in numerous important baseline characteristics, including diabetes and hypertension diagnosis as well as medication use for these conditions (absolute standardized difference>0.1). Full propensity-score matching also resulted in imbalances in a number of characteristics including race and age categories between exposed and unexposed groups.

In the unadjusted analysis, statin exposure was associated with increased risk of malformations compared with non-exposure (RD 0.048, 95% CI 0.019–0.078 and RR 2.4, 95% CI 1.6–3.4) (Figure 1). After 1:1 matching on the propensity score, this large increase in risk associated with statin exposure was greatly diminished (RR: 1.2, 95% CI: 0.69–2.0). 1:5 and full propensity-score matching resulted in similar attenuation of the RR estimates (RR 1.3, 95% CI 0.86–2.0 and RR 1.2, 95% CI 0.76–1.9, respectively). Trimming non-overlapping regions of the propensity-score distribution resulted in exclusion of 315,908 unexposed patients (35.8%) and none of the exposed patients. In the propensity score stratification (cohort) approach, the RR (95% CI) was estimated at 1.7 (1.2–2.4) with 10 strata. This value was reduced to 1.4 (0.97–2.0) with 50 and 1.3 (0.91–1.8) with 100 strata. In the propensity score stratification (exposed) approach, the RR (95% CI) was estimated at 1.2 (0.86–1.7) with 10 strata, 1.2 (0.85–1.7) with 50 strata, and 1.3 (0.88–1.8) with 100 strata. The SMR weighting approach resulted in an RR (95% CI) of 1.1 (0.79–1.6). The patterns in RDs with each method mirrored the patterns observed in RRs (Figure 1).

Simulation study

For almost all of the simulation scenarios, relative bias was higher with the propensity score stratification (cohort) approach using 10 and 50 strata compared with the propensity score stratification (exposed) approach using the same number of strata (Figure 2). Relative bias with the cohort approach decreased substantially with increasing number of strata as well as increasing exposure prevalence or outcome risk. For all propensity score-matching approaches, exposed approaches, and the SMR weighting approach, relative bias was close to zero regardless of the exposure prevalence and outcome risk.

In terms of precision, propensity score-matching approaches resulted in greater confidence limit ratios (lower precision) compared with all the stratification approaches under all exposure–outcome combinations studied (Figure 3). The SMR weighting approach also resulted in lower precision compared with the stratification approaches. Propensity score

stratification cohort and exposed approaches were equivalent in terms of their precision across all the scenarios tested.

Overall, all stratification approaches as well as the SMR weighting approach resulted in lower MSEs compared with propensity score-matching approaches at 0.5% and 1% exposure prevalences (Figure 4). At the lowest exposure prevalence and outcome risk combination (0.5% exposure-3.5% outcome), increasing the number of strata in the stratification approaches appeared to increase the MSEs. At 5% and 10% exposure prevalences, all approaches resulted in MSEs close to zero, indicating equivalent performance of these methods.

Discussion

At very low exposure prevalences (0.5% and 1%), propensity-score stratification approaches with strata formed based on the exposed group distribution resulted in greater confounding control as compared with stratification approaches based on the whole cohort. Propensity-score stratification based on the exposed group also provided greater precision compared with propensity-score matching and SMR weighting approaches. Regardless of the ranking method used, creating finer strata by increasing the number of strata resulted in greater confounding control without any meaningful loss in precision. At higher exposure prevalences (5% and 10%), very little difference in the performance of these methods was noted.

Our finding of improved precision with propensity-score stratification approaches compared with propensity-score matching is in keeping with theoretical expectations and previous reports.[3,4] In general, propensity-score stratification could be expected to produce estimates with greater precision compared with propensity-score matching approaches that discard unmatched unexposed or exposed patients from the analysis. The higher precision of stratification compared with the SMR weighting approach implies that conducting a weighted analysis based on strata-specific weighting of unexposed may be less prone to the influence of extreme weights than SMR weighting. Full matching also resulted in lower precision compared with stratification approaches despite using a nearly equivalent amount of information. However, the confidence intervals for full propensity-score matching were computed based on robust variance estimators, while confidence intervals for stratification were not. One could argue that propensity-score stratification approaches may also result in formation of clustered samples of individuals in each stratum similar to the clustering observed in propensity-score-matching approaches, and therefore stratification approaches should use robust variance estimators. However, since the number of strata in the stratification approach is typically much smaller than the number of matched-sets in full propensity-matching in practice, computation of robust variance estimators for estimates from stratification approaches is challenging.[26]

The gain in precision with propensity-score stratification approaches can be substantial in circumstances with low exposure prevalence. Exposure prevalence may be low in various settings. In our empirical example, the exposure was a treatment used despite contraindications. Other settings where exposure prevalence might be expected to be low are

safety evaluations of newly marketed drugs, and safety evaluations of infrequently used drugs in other vulnerable populations such as elderly or children. At higher exposure prevalences, the gain in precision offered by propensity-score stratification may be modest compared with propensity-score matching.

Our simulations demonstrate the possibility of substantial residual confounding in coarse stratification approaches with ranking based on the whole cohort (the propensity score stratification cohort approach). This problem can be remedied by finer stratification, but the approach of forming strata based on the exposed only (the propensity score stratification exposed approach) was highly effective. In essence, both these solutions address the problem of residual confounding by defining propensity-score strata narrowly so that they resemble propensity score-matched sets in which multiple exposed individuals are matched to multiple unexposed individuals based on their propensity score. Each narrow stratum contains approximately exchangeable exposed and unexposed individuals. Therefore, an analysis accounting for the narrow-stratum membership of each individual results in excellent confounding control. The problem of residual confounding with coarse stratification in the propensity score stratification cohort approach is especially severe when the exposure is infrequent, as stratum boundaries are almost entirely driven by the propensity score distribution in the unexposed. Consequently, the exposed patients cluster in just a few strata at the extreme, leaving the potential for substantial confounding within strata. Such clustering of exposed patients is avoided in the propensity score stratification exposed approach, because the strata are determined by the propensity score distribution in the exposed (See eTable 2a and 2b). In situations with low exposure prevalence, the propensity score stratification exposed approach is a method that is both efficient and effective in controlling confounding.

An alternative way to create fine strata would be to use fixed stratum boundaries defined on the probability scale. With 10 equally wide strata, this approach would place all the subjects with propensity score values between 0 and 0.1 into the first stratum, and similarly for the other nine strata. The results from this approach applied to our empirical example, using 100 strata, were similar to stratification approaches that use distributions of the propensity scores either in the whole cohort or the exposed group (100 strata with the fixed-width approach [OR 1.34, 95% CI 0.91–1.97], compared with the stratification cohort approach [OR 1.33, 95% CI 0.90–1.95] and stratification-exposed approach [OR 1.27 95% CI, 0.86–1.87]).

Our findings also underscore that it may be difficult to make general recommendations about the optimum number of strata that should be used in propensity-score stratification in situations with low exposure prevalence. Full propensity-score matching is another promising method that uses the majority of the individuals in the cohort and selects the number of strata automatically. However, this method produced disappointing results with low exposure prevalence. In our simulations with 0.5% and 1% exposure prevalence, full propensity-score matching resulted in higher MSEs compared with a majority of other approaches. In a previous simulation study, Austin and Stuart have also reported high MSEs with full propensity-score matching when not using any caliper restrictions compared with nearest-neighbor propensity-score-matching.[19] Introducing a caliper restriction to full propensity-score matching may lead to improved performance.

Traditionally, either Mantel-Haenszel pooling or computing a weighted average of subclass specific estimates have been used to derive adjusted exposure effects after propensity-score stratification. We used weighted regression models after stratification, over these traditionally used approaches, because this approach does not rely on the homogeneity assumption (as opposed to Mantel-Haenszel pooling) and provides stable estimates under circumstances of a large number of strata or sparsely populated strata (as opposed to the weighted average method). In a separate set of simulations, we have found the performance of the Mantel-Haenszel pooling approach to be robust with propensity-score stratification.[27] If researchers can reasonably assume homogeneity of the exposure effect across strata, Mantel-Haenszel pooling is a good, simple alternative to the weighting approach used in this study. The precision of the summary exposure effect estimate using the Mantel-Haenszel estimator is optimal as the strata are pooled using inverse-variance weighting.

In conclusion, our findings indicate that performance of propensity score-based fine stratification in confounding control is equivalent to propensity-score matching at higher exposure prevalence and better than propensity-score matching at low exposure prevalence. Therefore, propensity score-based fine stratification should be considered as a strategy for confounding control in routine pharmacoepidemiology practice. Creation of strata should be based on the propensity-score distribution of the exposed group when evaluating outcomes of an infrequent exposure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding:

This study was funded from internal sources of the Division of Pharamcoepidemiology and Pharmacoeconomics.

References

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70(1):41–55.
2. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996; 49(12):1373–9. [PubMed: 8970487]
3. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011; 46(3):399–424. [PubMed: 21818162]
4. Wang Z. Propensity score methods to adjust for confounding in assessing treatment effects: Bias and precision. The Internet Journal of Epidemiology. 2009; 7(2)
5. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. Stat Med. 2006; 25(12):2084–106. [PubMed: 16220490]
6. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association. 1984; 79(387):516–524.
7. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics. 1968; 24(2):295–313. [PubMed: 5683871]
8. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998; 17(19):2265–81. [PubMed: 9802183]

9. Akers, A. Determination of the optimal number of strata for bias reduction in propensity score matching. University of North Texas; 2010.
10. Bateman BT, Hernandez-Diaz S, Fischer MA, Seely EW, Ecker JL, Franklin JM, Desai RJ, Allen-Coleman C, Mogun H, Avorn J. Statins and congenital malformations: cohort study. BMJ. 2015; 350:h1035. [PubMed: 25784688]
11. Palmsten K, Huybrechts KF, Mogun H, Kowal MK, Williams PL, Michels KB, Setoguchi S, Hernandez-Diaz S. Harnessing the Medicaid Analytic eXtract (MAX) to Evaluate Medications in Pregnancy: Design Considerations. PLoS One. 2013; 8(6):e67405. [PubMed: 23840692]
12. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. Pharmacoepidemiol Drug Saf. 2013; 22(1):16–24. [PubMed: 22550030]
13. Williamson T, Eliasziw M, Fick GH. Log-binomial models: exploring failed convergence. Emerg Themes Epidemiol. 2013; 10(1):14. [PubMed: 24330636]
14. Skov T, Deddens J, Petersen MR, Endahl L. Prevalence proportion ratios: estimation and hypothesis testing. Int J Epidemiol. 1998; 27(1):91–5. [PubMed: 9563700]
15. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol. 2003; 157(10):940–3. [PubMed: 12746247]
16. Ho D, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J Stat Softw. 2011; 42(8):28.
17. Ming K, Rosenbaum PR. Substantial gains in bias reduction from matching with a variable number of controls. Biometrics. 2000; 56(1):118–24. [PubMed: 10783785]
18. Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. Pharmacoepidemiol Drug Saf. 2012; 21(Suppl 2):69–80. [PubMed: 22552982]
19. Austin PC, Stuart EA. Estimating the effect of treatment on binary outcomes using full matching on the propensity score. Stat Methods Med Res. 2015
20. Brookhart MA, Wyss R, Layton JB, Sturmer T. Propensity score methods for confounding control in nonexperimental research. Circ Cardiovasc Qual Outcomes. 2013; 6(5):604–11. [PubMed: 24021692]
21. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. Am J Epidemiol. 2006; 163(3):262–70. [PubMed: 16371515]
22. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006; 98(3):253–9. [PubMed: 16611199]
23. Poole C. Low P-values or narrow confidence intervals: which are more durable? Epidemiology. 2001; 12(3):291–4. [PubMed: 11337599]
24. Devore, JL., Berk, KN. Modern mathematical statistics with applications. New York: Springer; 2012.
25. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Communications in Statistics-Simulation and Computation. 2009; 38(6):1228–1234.
26. Arceneaux K, Nickerson DW. Modeling certainty with clustered data: A comparison of methods. Political Analysis. 2009; 17(2):177–190.
27. Desai RJRK, Bateman B, Hernandez-Diaz S, Huybrechts KF. Confounding control using propensity scores when the exposure is infrequent: making the case for a fine stratification approach. Pharmacoepidemiol & Drug Safety. 2015; 24:217.

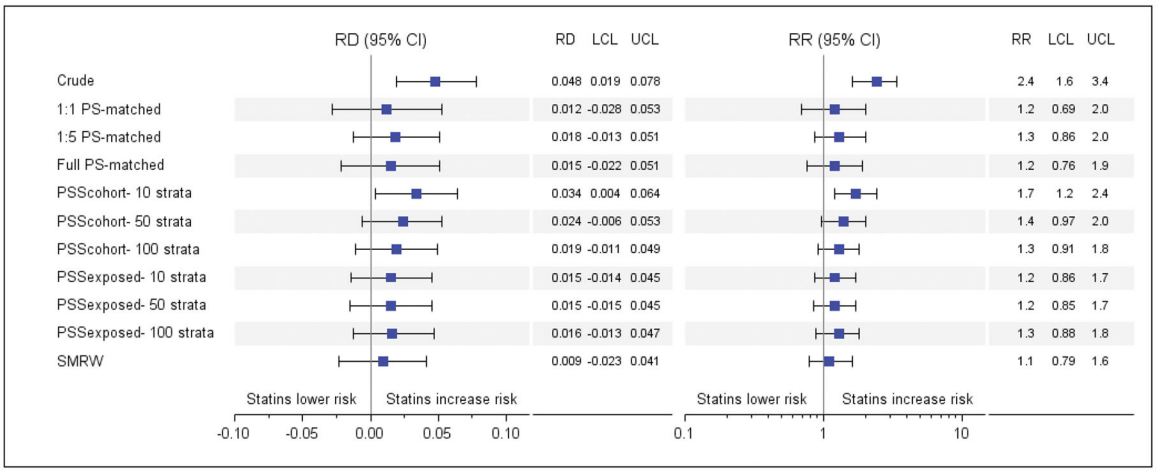

Figure 1.

Measures of association between congenital malformation and first trimester statin exposure versus non-exposure using different analytic approaches, Medicaid Data 2000–2007

Abbreviations: CI- Confidence interval, LCL- Lower confidence limit, PS- Propensity scores, $PSS_{Exposed}$- Propensity score stratification, strata created by ranking only the exposed group, PSS_{Cohort}- Propensity score stratification, strata created by ranking the entire cohort, RD- Risk difference, RR- Risk ratio, SMRW- Standardized mortality ratio weighting, UCL- Upper confidence limit.

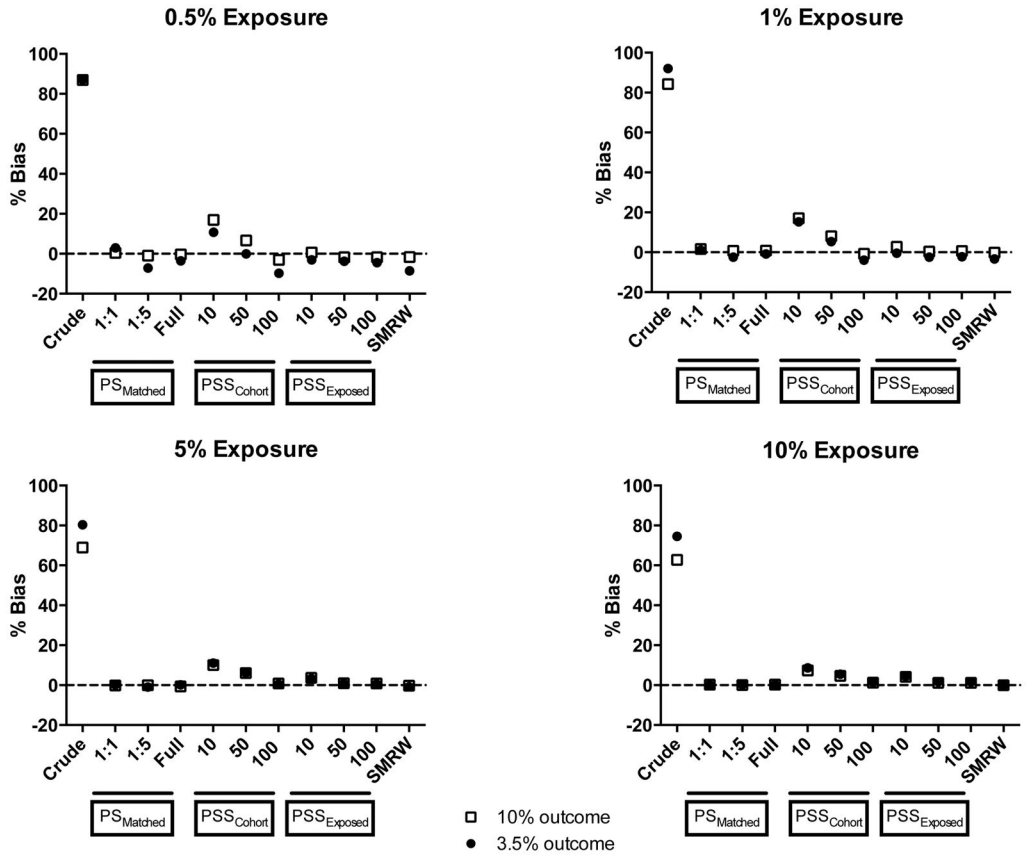

Figure 2.

Relative bias for different analytic approaches over 1,000 simulations

Abbreviations: PS- Propensity scores, $PSS_{Exposed}$- Propensity score stratification, strata created by ranking only the exposed group, PSS_{Cohort}- Propensity score stratification, strata created by ranking the entire cohort, SMRW- Standardized mortality ratio weighting

* The dashed line indicates unbiased estimates (% bias of 0)

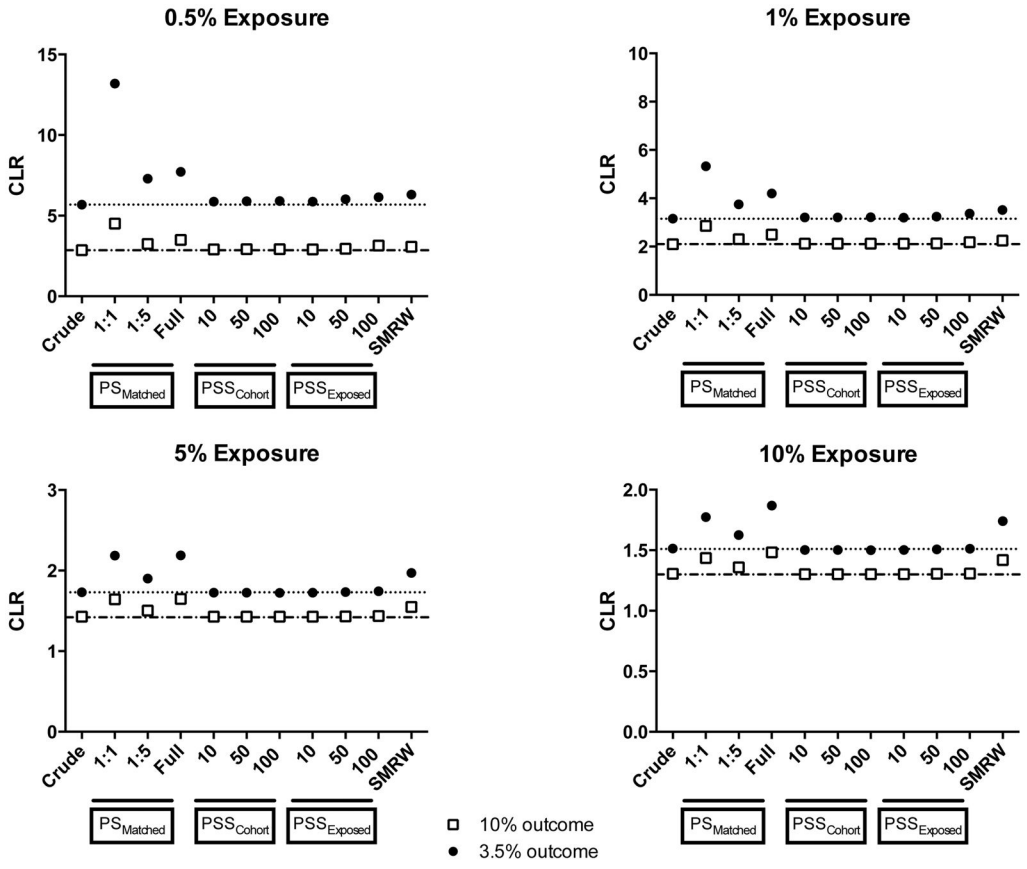

Figure 3.

Confidence limit ratios for different analytic approaches over 1,000 simulations

Abbreviations: CLR- Confidence limit ratio, PS- Propensity scores, $PSS_{Exposed}$- Propensity score stratification, strata created by ranking only the exposed group, PSS_{Cohort}- Propensity score stratification, strata created by ranking the entire cohort, SMRW- Standardized mortality ratio weighting

* The dotted line indicates precision (as measured by the CLR) of the crude estimate at 3.5% outcome risk; the dot-dashed line indicates precision (as measured by the CLR) of the crude estimate at 10% outcome risk.

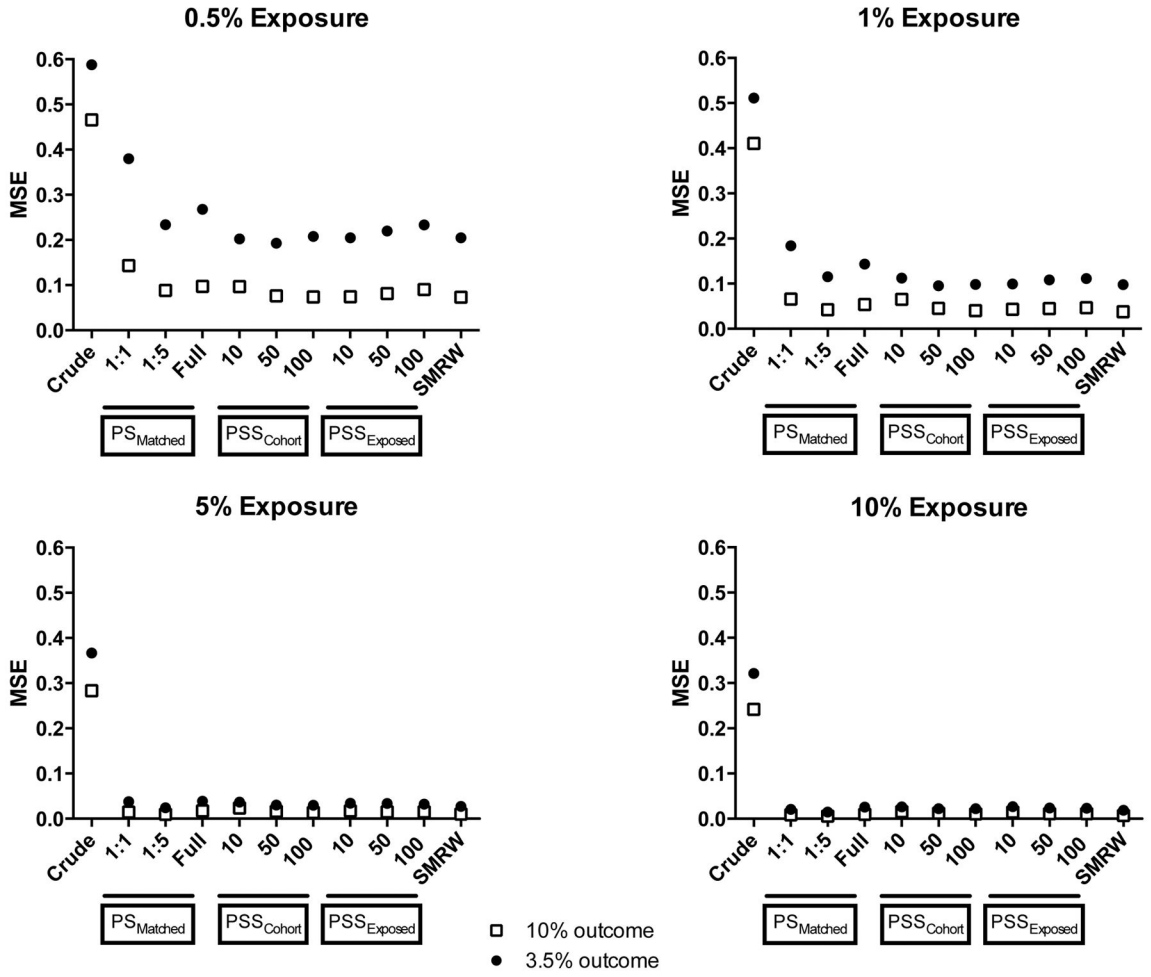

Figure 4.

Mean squared errors for different analytic approaches over 1,000 simulations

Abbreviations: MSE- Mean squared error, PS- Propensity scores, $PSS_{Exposed}$- Propensity score stratification, strata created by ranking only the exposed group, PSS_{Cohort}- Propensity score stratification, strata created by ranking the entire cohort, SMRW- Standardized mortality ratio weighting

Table 1

Distribution of selected baseline characteristics in pregnant women by their statin exposure status in the first trimester and differences in baseline characteristics with different analytic approaches, Medicaid data 2000–2007

Variable	Statin exposed (n=335)	Unexposed (n=886,661)	Absolute standardized differences*										
			Crude	PS-matched			PSS_{Cohort}			$PSS_{Exposed}$			SMRW
	n (%)	n (%)		1:1	1:5	Full	10 strata	50 strata	100 strata	10 strata	50 strata	100 strata	
Age-group													
<19 years	20 (6)	260708 (29)	**0.65**	0.01	0.08	0.02	**0.11**	0.01	0.05	0.04	0.06	0.01	0.01
20–24 years	43 (13)	315473 (36)	**0.55**	0.04	0.04	0.07	**0.2**	0.05	0.02	0.01	0.02	0.06	0.00
25–29 years	58 (17)	180979 (20)	0.08	**0.13**	0.07	**0.12**	**0.12**	0.09	0.08	0.03	0.03	0.02	0.01
30–34 years	104 (31)	84234 (9)	**0.56**	0.02	0.02	0.04	0.07	0.00	0.01	0.01	0.02	0.03	0.01
35–39 years	78 (23)	37142 (4)	**0.58**	0.09	0.01	0.07	**0.19**	0.06	0.03	0.00	0.01	0.02	0.02
>40 years	32 (10)	8125 (1)	**0.40**	0.03	0.04	**0.11**	**0.18**	0.09	0.06	0.02	0.01	0.01	0.03
Race													
Black	80 (24)	302167 (34)	**0.23**	0.04	0.01	**0.43**	0.01	0.02	0.01	0.01	0.00	0.00	0.00
White	163 (49)	360562 (41)	**0.16**	0.08	0.01	**0.31**	0.07	0.02	0.01	0.01	0.00	0.00	0.01
Other	92 (27)	223932 (25)	0.05	0.05	0.02	**0.21**	0.07	0.00	0.00	0.00	0.00	0.01	0.01
Comorbid conditions													
Diabetes	156 (47)	27454 (3)	**1.16**	0.01	0.02	0.09	**0.69**	**0.34**	**0.24**	0.08	0.07	0.07	0.02
Hypertension	155 (46)	44561 (5)	**1.07**	**0.12**	0.03	0.02	**0.47**	**0.20**	**0.12**	0.02	0.01	0.02	0.02
Dyslipidemia	221 (66)	28361 (3)	**1.76**	0.04	0.07	0.03	**0.71**	**0.11**	0.03	0.04	0.06	0.06	0.01
Obesity	66 (20)	47206 (5)	**0.45**	0.04	0.03	0.00	**0.18**	0.05	0.01	0.01	0.02	0.02	0.00
Tobacco use	29 (9)	68922 (8)	0.03	0.01	0.03	0.09	0.03	0.01	0.01	0.01	0.01	0.01	0.04
Comedications													
Oral antidiabetic medications	146 (44)	11552 (1)	**1.18**	0.03	0.04	**0.15**	**0.75**	**0.40**	**0.29**	0.10	0.09	0.09	0.03
Hypertension medications	214 (64)	59287 (7)	**1.49**	0.04	0.01	0.03	**0.43**	**0.23**	**0.12**	0.02	0.00	0.00	0.03
Insulin	107 (32)	11259 (1)	**0.90**	0.01	0.03	0.09	**0.60**	**0.33**	**0.24**	0.08	0.07	0.07	0.01

*ASD>0.1 are in bold, which is suggested by Austin[25] as an indicator for substantial imbalances between the two exposure groups.

Abbreviations: PS- Propensity scores, $PSS_{Exposed}$- Propensity score stratification, strata created by ranking only the exposed group, PSS_{Cohort}- Propensity score stratification, strata created by ranking the entire cohort, SMRW- Standardized mortality ratio weighting