

Postmyocardial Infarction Statin Exposure and the Risk of Stroke with Weighting for Outcome Misclassification

Christopher A. Gravel,^{a,b,c} Kristian B. Filion,^{a,b,d} Pauline M. Reynier,^b and Robert W. Platt^{a,b,c}

Background: Observational healthcare data can be used for drug safety and effectiveness research. The use of inverse probability of treatment weights (IPW) reduces measured confounding under the assumption of accurate measurement of the outcome variable; however, many datasets suffer from systematic outcome misclassification. **Methods:** We introduced a modification to IPW to correct for the presence of outcome misclassification. To demonstrate the utility of these modified weights in realistic settings, we investigated postmyocardial infarction statin use and the 1-year risk of stroke in the Clinical Practice Research Datalink.

Results: We computed an IPW-adjusted odds ratio (OR = 0.67; 95% confidence interval (CI) = 0.48, 0.93). We employed a technique to modify IPW for the presence of outcome misclassification using linked hospital records for outcome validation (modified IPW adjusted OR = 0.77; 95% CI = 0.52, 1.15) and compared the results with a meta-analysis of randomized controlled trials (RCTs) (pooled OR = 0.80; 95% CI = 0.74, 0.87). Finally, we present simulation studies to investigate the impact of model selection on bias reduction and variability.

Conclusion: Ignoring outcome misclassification yielded biased estimates whereas the use of the modified IPW approach produced encouraging results when compared with the meta-analytic RCT findings.

Keywords: binary outcomes, inverse probability of treatment weighting, myocardial infarction, outcome misclassification, statins, stroke

(*Epidemiology* 2020;31: 880–888)

The practice of collecting and warehousing patient healthcare information has been steadily increasing.^{1,2} Currently, academic researchers,³ government regulatory agencies,⁴ and private interests have begun using this information to conduct and support research initiatives throughout the regulatory approval process of prescription drugs. These real-world data sources are routinely collected in a manner that reflects dimensions of standard healthcare practice, rather than an experimental setting, such as in randomized controlled trials (RCTs).

Computerized administrative healthcare datasets and electronic medical records are common sources of information used to conduct pharmacoepidemiologic research examining the utilization, effectiveness, and safety of prescription drugs in real-world settings. An observational data source that is well established for such studies is the Clinical Practice Research Datalink (CPRD), which contains the medical records of 674 general practitioner practices in the United Kingdom, constituting approximately 6.9%⁵ of the United Kingdom's population.

Estimation of the causal effect of a drug on health outcomes requires accurate recording and classification of the variables of interest. For example, Herrett et al.⁶ studied the completeness and diagnostic validity of acute myocardial infarctions (MIs) recorded in the CPRD. They linked primary care records from the CPRD to inpatient hospital admission records from the Hospital Episode Statistics (HES)⁷ and a national registry for acute coronary syndromes, the Myocardial Ischaemia National Audit Project (MINAP). The presence of diagnoses of MI in all the three datasets within a 30-day interval was defined as an accurate record of an MI.

Table 1⁶ shows the concordance in the recording of these events across datasets. Of the 17,964 patients with at least one nonfatal MI, only 50.1% of patients had concordant

Submitted August 6, 2019; accepted July 29, 2020.

From the ^aDepartment of Epidemiology Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada; ^bCentre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, Quebec, Canada; ^cDepartment of Pediatrics, McGill University, Montreal, Quebec, Canada; ^dDepartment of Medicine, McGill University, Montreal, Quebec, Canada; and ^eSchool of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada

This study was conducted using data acquired and the authors thank the Canadian Network for Observational Drug Effect Studies (CNODES), a collaborating center of the Drug Safety and Effectiveness Network (DSEN), funded by the Canadian Institutes of Health Research (Grant Number DSE-146021). C.A.G. holds a Postdoctoral Mitacs Accelerate Internship with McGill University and Risks Sciences International (www.risksciences.com). K.F. holds a salary support award from the *Fonds de recherche du Québec – santé* (Quebec Foundation for Health Research) and a William Dawson Scholar award from McGill University. R.W.P. holds the Albert Boehringer I Chair in Pharmacoepidemiology at McGill University.

R.W.P. has received consulting fees from Amgen, Analysis Group, Biogen, Depuy, Eli Lilly, and Merck, outside the scope of this work. The other authors have no conflicts of interest to declare.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Christopher A. Gravel, Department of Epidemiology, Biostatistics and Occupational Health, 1020 Pine Avenue West, McGill University, Montreal, Quebec, H3A 1A2, Canada. E-mail: christopher.gravel2@mail.mcgill.ca.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/20/3106-0880

DOI: 10.1097/EDE.0000000000001253

TABLE 1. Number and Percentage of Records of Nonfatal Myocardial Infarctions in the CPRD, HES, and MINAP⁶

	Record in MINAP		No Record in MINAP		HES/CPRD Only		Total
	Record in CPRD	No Record in CPRD	Record in CPRD	No Record in CPRD	Record in CPRD	No Record in CPRD	
Record in HES	5,561 (31%)	1,290 (7%)	3,532 (20%)	1,806 (10%)	9,093 (51%)	3,096 (17%)	12,189 (68%)
No Record in HES	1,099 (6%)	1,488 (8%)	3,188 (18%)	-	4,287 (24%)	1,488 (8%)	5,775 (32%)
				Total	13,380 (75%)	4,584 (25%)	17,964 (100%)

This table describes the rate of concordance in recording of acute nonfatal myocardial infarctions between the Clinical Practice Research Datalink (CPRD), the Hospital Episode Statistics (HES) and the Myocardial Ischaemia National Audit Project (MINAP). Among the 17,964 patients with at least one record of a nonfatal MI in one of these sources, concordance was defined as the presence of records for similar diagnoses in all the three datasets within a 30-day interval. This study was conducted on patient records between January 2003 and March 2009 and was part of the Cardiovascular Disease Research using Linked Bespoke studies and Electronic Health Records Programme.⁶ To describe the rates of concordance between the HES and CPRD data exclusively, the marginal counts are also provided.

records between the CPRD and HES implying the presence of high proportions of outcome misclassification in these data. In addition, 1,488 patients (8%) with a record for a nonfatal MI in MINAP were missing from both the CPRD and HES further, adding to the uncertainty surrounding the accuracy of the recording of events. The authors concluded that failure to use linked records for outcome ascertainment may lead to biased estimates of MI incidence. However, the choice to restrict studies to only individuals with concordant records may introduce bias if the process underlying the missingness of the diagnosis in each dataset is informative.

It is typically assumed that nondifferential outcome misclassification (i.e., the error is independent of exposure) biases expected values towards the null hypothesis of no effect. However, the directionality of the bias is not guaranteed when the validity of the assumption is only approximate,⁸ which might motivate the use of quantitative misclassification bias correction methods. Both the internal validation of a subset of records and information drawn from an external source can be used for the estimation of the outcome error rates,^{9,10} although external validation requires similarity in the characteristics of both data sources, which is often difficult to justify. Hence, internal validation is preferable. In addition, the lack of randomization of exposure assignment in observational studies may confound estimation of causal effects. Inverse probability of treatment (propensity score; the probability of exposure given covariates) weights (IPW) can be used to create a simulated pseudo-population in which the marginal causal effect of exposure is identifiable and can be estimated consistently.¹¹

Gravel and Platt¹² introduced a novel set of weights that modify IPW and, assuming accurately measured exposure, can be used to produce consistent estimation in confounded binary outcome data that are subject to misclassification. To study the utility of these modified weights in a realistic setting, we applied this approach to CPRD data in a retrospective cohort study of post-MI statin use and the 1-year risk of stroke, using records from the HES data, deterministically linked to the CPRD data to validate the defined stroke outcomes.¹³ A comparison between a crude analysis, an IPW analysis ignoring misclassification, an analysis using the proposed approach, and results from a

published meta-analysis of RCTs considering stroke incidence in high-risk populations¹⁴ was conducted. Finally, we conducted simulation studies to explore the performance of the proposed weights under incorrect misclassification model specifications, smaller validation sample sizes, and with a known target parameter to study the consistency of the method.

METHODS

Modified IPW for Outcome Misclassification and Confounding

IPW can rebalance covariate distributions across treatment groups to simulate a pseudo-population from which consistent estimation of the effect of interest is possible. Individuals are assigned a weight that is proportional to the estimated probability of receiving their actual treatment, creating simulated independence of the exposure variable and the measured confounders. This approach allows for the identification and consistent estimation of the marginal causal effect of treatment under certain assumptions (exchangeability, positivity, and consistency).¹⁵ Outcome misclassification creates a violation of the consistency assumption or the notion that the observed outcome is the potential outcome under the observed treatment. In the presence of such misclassification, the resulting simulated pseudo-population will fail at balancing the covariate distributions, resulting in potential residual bias.

Gravel and Platt¹² modified the traditional IPW approach to account for outcome misclassification. The modification is a function of the rates of misclassification and the bias-corrected predicted values from the outcome model. The resulting quantity is used in conjunction with the inverse propensity scores to weight the observed, possibly misclassified outcomes restoring covariate balance in the pseudo-population. These weights are presented in equation 1:

$$W_a = \begin{cases} \left[PS_a \left(SE_a + \{1 - SP_a\} \frac{1 - \pi_a}{\pi_a} \right) \right]^{-1}, & Y^* = 1 \\ \left[PS_a \left(\{1 - SE_a\} \frac{\pi_a}{1 - \pi_a} + SP_a \right) \right]^{-1}, & Y^* = 0 \end{cases} \quad (1)$$

where Y^* denotes the original, possibly misclassified binary outcome, PS_a denotes the propensity score, SE_a and SP_a denote the sensitivity and specificity of outcome ascertainment, respectively, and π_a denotes the probability of having the true outcome, all for the a^{th} treatment, $a = 0, 1$. Note that, for simplicity, all the quantities in equation 1 are written as a function of the exposure variable, A , but can also be modeled as a function of a vector of covariates.

These quantities can be estimated in several ways.^{16–18} In the following example, we followed the approach described by Lyles et al.¹⁰ to estimate the misclassification-corrected quantities and standard logistic regression to estimate the propensity scores. Briefly, the authors employ a likelihood-based approach, under the assumption of the availability of “gold-standard” internal validation data, to produce consistent estimates of these model parameters in differentially misclassified binary outcome data (please see eAppendix 1; <http://links.lww.com/EDE/B719> for a technical summary of this methodology). Gravel and Platt¹² showed that the standard sandwich variance estimator will fail and proposed a modified bootstrap estimator (please see eAppendix 2; <http://links.lww.com/EDE/B719> for more details) in which the validation of the resampled binary outcomes is generated by a parametric Monte Carlo approach using the misclassification rates observed in the validation data. The recommendation for a partially parametric bootstrap is based on the likely scenario of low misclassification rates, leading to even lower rates of misclassified observations in the validation data, which can cause the nonparametric bootstrap to fail (particularly in a rare-event setting).

EXAMPLE

Postmyocardial Infarction Statin Use and the 1-Year Risk of Stroke

To demonstrate how the modified IPW approach can be implemented, we conducted a retrospective cohort study examining post-MI statin use and the 1-year risk of stroke using the CPRD, acknowledging the potential for outcome misclassification.⁶ To explore the utility of the modified IPW estimator in this context, we used a dataset, made up of deterministically linked records from the CPRD and HES data.¹³ The HES is a collection of hospital admission records that provides additional details on a subset of the population and acts as a rich source of information to validate stroke diagnoses in the CPRD.

The study population included patients, aged 18 years or older, with a recorded diagnostic code (based on the Read coding system) for MI in the CPRD between April 1, 1998 and March 31, 2012. Inclusion was restricted to patients that were registered with English practices that consented to be linked to the HES (58% of all UK general practices; 76% of English practices),⁶ the internal validation source for the modified IPW approach. The CPRD and its linkage with other

National Health Service data holdings have been described in detail elsewhere.^{6,13} We excluded patients with less than a year of history in the CPRD at the time of cohort entry (defined below) as well as those with a history of stroke, acute coronary syndrome, or MI. We employed a new user design in which inclusion was restricted to individuals with no recorded statin use in the previous year. Patients entered the cohort 30 days after the date of their first recorded MI in the CPRD, denoted as t_0 , and statin use was defined in this 30-day window as a binary exposure (i.e., using an approach analogous to an intent-to-treat analysis). Patients were excluded if they had a stroke within this interval. Individuals were followed until a recorded stroke or censoring due to end of follow-up (1 year), departure from the CPRD, death, or end of the study period. The final study sample size was $n = 20,957$. This study was approved by the Independent Scientific Advisory Committee for MHRA database research (protocol number 14018A3; the study protocol was made available to journal reviewers) and the Research Ethics Board of the Jewish General Hospital in Montreal, Canada.

To link CPRD- and HES-defined strokes, we considered records to be concordant if they occurred in a ± 30 day window of each other. In other words, if an individual had a record of an eligible stroke in the CPRD, we searched for a concordant entry in HES ± 30 days around the original event. Because the lower bound of this interval can extend into the exposure window (i.e., before t_0), we ended up excluding 12 additional patients who had a stroke recorded in HES during the 30 days post-MI but not in the CPRD. Consequently, our sample size reduced to $n = 20,945$. Similarly, given that the upper bound for validation could exceed the maximum 1-year of follow-up, we searched the HES for validation information 30 days past the end of follow-up (i.e., records can be validated up to $t_0 + 395$ days). Finally, for individuals with no record of a stroke in the CPRD, we defined validation by an absence of stroke in HES at any time during follow-up ± 30 days.

Table 2 presents the baseline characteristics for the 20,945 patients, measured at, or within the year before, the date of the first record of an MI in the CPRD, cross-classified by exposure.

We conducted a crude and an IPW-adjusted analysis on the CPRD data, ignoring misclassification, and estimated the propensity score using logistic regression with a set of potential confounders measured at the cohort entry-defining MI. Next, we implemented the modified IPW approach using the linked HES data as internal validation information. The cross-classification of statin exposure, CPRD-defined strokes, and HES-validated strokes, using the previously described linkage procedure is presented in Table 3. These data will be used to produce the plug-in estimates needed in equation 1.

To generate estimates of the misclassification parameters we modeled SE_a and SP_a conditional on statin exposure, to avoid constraining the analyses without justification.

TABLE 2. Baseline Characteristics of Subjects by Statin Exposure Measured at, or Within a Year of, the Myocardial Infarction Used to Define Cohort Entry

	Statin		No Statin	
Characteristics	n = 13,232	%	n = 7,713	%
Male	9,309	70	4,577	59
Age				
0–39	244	2	119	2
40–49	1,475	11	470	6
50–59	3,113	24	1,074	14
60–69	3,479	26	1,548	20
≥70	4,921	37	4,502	58
Smoking				
Yes	6,780	51	3,105	40
No	3,117	24	1,899	25
Missing ^a	3,335	25	2,709	35
Alcohol	162	1	86	1
Obesity				
Yes	1,936	15	908	12
No	5,513	42	3,216	42
Missing ^a	5,783	43	3,589	46
Total Cholesterol (mmol/L)				
<5.2	1,351	10	755	10
5.2–6.2	1,442	11	558	7
>6.2	1,186	9	410	5
Missing ^a	9,253	70	5,990	78
LDL Cholesterol (mmol/L)				
<1.8	35	0	34	0
1.8–2.6	215	2	126	2
2.6–3.3	470	4	219	3
3.4–4.1	645	5	222	3
4.1–4.9	338	3	99	1
>4.9	167	1	53	1
Missing ^a	11,362	85	6,960	90
Year of Cohort Entry				
1998	268	2	546	7
1999	502	4	828	11
2000	723	6	828	11
2001	864	7	889	12
2002	1,113	8	760	10
2003	1,247	9	698	9
2004	1,223	9	599	8
2005	1,062	8	489	6
2006	1,102	8	409	5
2007	1,045	8	395	5
2008	972	7	337	4
2009	960	7	325	4
2010	984	8	277	4
2011	913	7	261	3
2012	254	2	72	1
Comorbidities				
Atrial fibrillation	305	2	369	5
Cerebrovascular disease	69	1	81	1

(Continued)

TABLE 2. (Continued)

	Statin		No Statin	
Chronic obstructive pulmonary disease	465	4	396	5
Congestive heart failure	367	3	556	7
Coronary artery disease	435	3	525	7
Diabetes mellitus	785	6	754	10
Hypercholesterolemia	293	2	103	1
Hypertension	788	6	565	7
Peripheral vascular disease	145	1	138	2
Previous coronary revascularization	1,254	10	225	3
Medication Use				
ACE inhibitors	2,199	17	1,344	17
ARBs	567	4	324	4
Aspirin	2,347	18	1,866	24
Beta-blockers	2,331	18	1,360	18
Calcium-channel blockers	2,009	15	1,427	19
Diuretics	2,533	19	2,474	32
Fibrates	48	0	32	0
NSAIDs	3,037	23	1,715	22
Prescription Count				
0–4	7,186	54	3,343	43
5–7	2,711	21	1,572	20
8–11	1,959	15	1,416	18
≥12	1,376	10	1,382	18
Number of Hospitalizations				
0–1	12,368	94	7,186	93
≥2	864	6	527	7

^aMissing data were imputed using multiple imputation. Due to the high amount of missing total and LDL cholesterol data, the main analyses excluded these covariates from the propensity score model. A sensitivity analysis was conducted with imputed cholesterol data (see the Limitations section for more detail).

ACE indicates angiotensin-converting enzyme; ARB, angiotensin receptor blocker; NSAID, nonsteroidal anti-inflammatory; MI, myocardial infarction.

We then followed the approach described by Lyles et al.¹⁰ to estimate the misclassification bias corrected outcome classification probabilities, $\pi_a(l_i)$, $a = 0, 1$, where l_i denotes the covariate vector for the i^{th} individual. The π -parameters were estimated using the same covariates as in the propensity score models. These values, along with the propensity score predicted probabilities, were plugged into the weights in equation 1, and a logistic regression was run on the weighted outcome variable and the statin exposure variable producing the marginal results in Figure 1. Given that the outcome model is saturated, we used nonstabilized weights for the IPW and modified IPW analyses. Variance estimation used a modified bootstrap (eAppendix 2; <http://links.lww.com/EDE/B719>).¹²

The results for the crude analysis (OR = 0.41; 95% CI = 0.30, 0.56), the IPW analysis ignoring misclassification (OR = 0.67; 95% CI = 0.48, 0.93), and the modified IPW analysis using 100% of the HES information to validate the CPRD-identified strokes (OR = 0.77; 95% CI = 0.52, 1.15) are

TABLE 3. Cross-Classification of CPRD-Defined Strokes, Statin Exposure, and HES-Validation Information

	Statin		No Statin		Total
	CPRD Recorded Stroke	No Stroke in CPRD	CPRD Recorded Stroke	No Stroke in CPRD	
HES Validated ^a Stroke	42	38	49	59	188
HES Validated ^b Negative	28	13,124	50	7,555	20,757
Total	70	13,162	99	7,614	20,945

^aValidation was conducted by searching patient HES records for a stroke within a 30-day interval around the CPRD-recorded stroke.

^bValidated negatives are defined as the absence of any recording of a stroke in the HES within follow-up.

CPRD indicates Clinical Practice Research Datalink; HES, Hospital Episode Statistics.

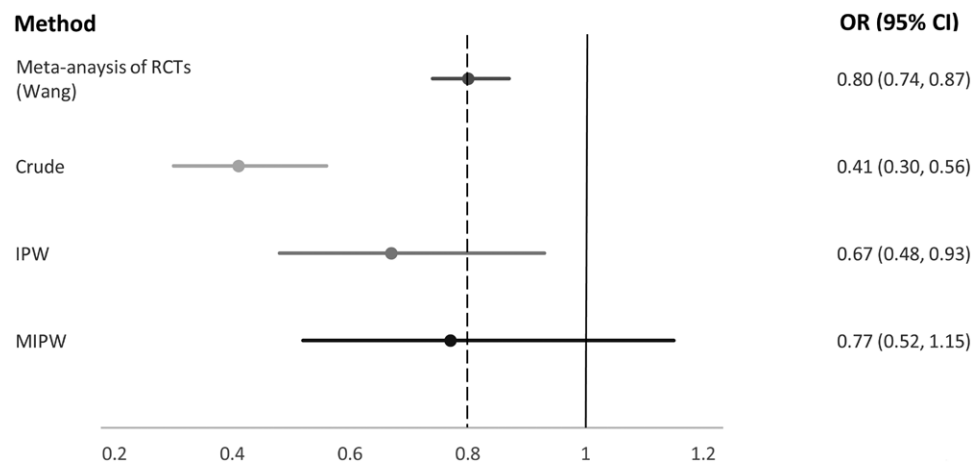


FIGURE 1. Forest plot for the results of the meta-analysis, the crude analysis, the IPW analysis and the MIPW analysis with 100% validated observations. This forest plot presents 95% confidence intervals and odds ratio estimates for four analyses. The first denotes the pooled meta-analytic odds ratio estimate of overall stroke incidence and statin exposure in those individuals at high-risk of stroke due to prevalent conditions based on 18 RCTs (10; plotted as a dashed vertical line). The second denotes the results of the crude analysis based on the individuals with records taken only from the CPRD (ignoring misclassification). The third denotes the results based on the IPW analysis using the CPRD data only with the set of potential confounders in Table 2 (excluding the total and LDL cholesterol due to the high amount of missing data) used to estimate the propensity scores. The fourth denotes the modified IPW analysis validating 100% of the outcomes with linked HES records and using the same covariate set to estimate the propensity scores. CI indicates confidence interval; IPW, inverse probability of treatment weighting; MIPW, modified inverse probability of treatment weighting; OR, odds-ratio; RCT, randomized controlled trial.

displayed in Figure 1. We validated outcomes for all patients for demonstrative purposes and explored lesser validation sample sizes through simulation studies; in practice, this information is often costly or difficult to obtain. Note that the modified IPW estimate is closer to the null value of 1 than the IPW or crude estimates. The IPW and modified IPW estimates are quite different from the crude, which could imply the presence of a notable amount of confounding. Finally, the modified IPW estimate is more variable than the IPW estimate, which is to be expected given that more quantities require estimation.

It is difficult to evaluate the performance of the modified IPW approach in this context, given that the “true” effect of statins on the 1-year risk of stroke in a high-risk population is unknown. In addition, the real-world setting of this study may introduce several additional uncertainties; for example, adherence

to treatment is expected to be lower than that observed in RCTs, which would produce an effect estimate that is closer to the null than that observed in RCTs. Wang et al.¹⁴ conducted a meta-analysis of randomized, placebo-controlled statin trials examining the incidence of stroke in patients defined as high-risk due to prevalent conditions (coronary heart disease, diabetes mellitus, hypertension, myocardial ischemia, and hypercholesterolemia). They included 18 trials and computed a pooled odds ratio (OR) estimate of 0.80 (95% confidence interval [CI] = 0.74, 0.87). In Figure 1, we also plot the results for comparison. The black vertical line denotes the meta-analytic pooled estimate of 0.80, and it seems that the modified IPW estimate is closer than the IPW estimate ignoring misclassification; however, the modified IPW interval is wider.

These results are promising for the use of the modified IPW OR estimator; however, this study has potential

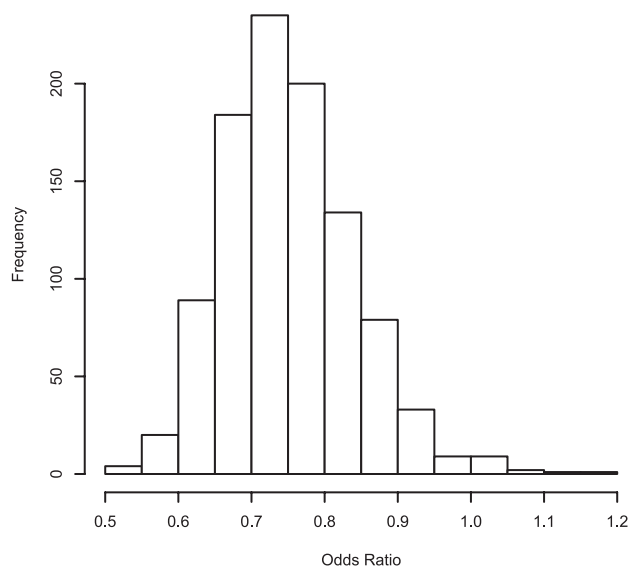


FIGURE 2. Histogram of 1,000 estimates of the odds ratio computed using the modified inverse probability weighting (IPW) approach with 50% validation sample. This histogram presents 1,000 estimates of the odds ratio for the association between postmyocardial infarction statin use and the 1-year risk of stroke derived using the modified IPW approach with validation samples comprised of 10,472 observations (50% validation sample proportion).

limitations that merit further investigation. In the following section, we describe these limitations and present simulation studies designed to evaluate the potential impact of these issues on the performance of the modified IPW approach.

LIMITATIONS

The first potential limitation was the use of CPRD-defined MI to define cohort entry without considering its potential for misclassification. While this is a potential limitation, we posit that its impact is likely minimal. A recorded MI is likely to have high positive predictive value, and the high error rates discussed in Herrett et al.⁶ are related to false negatives. Hence, the cohort entry MI was used to increase the likelihood of drawing individuals from the target population of interest, which is likely unaffected by high false-negative rates. Namely, individuals with a heightened risk for a cardiovascular event and a medical history supporting the use of statins for secondary prevention.

Next, the method of validating outcomes has two limitations. First, using 100% of the HES information for validation is unrealistic. In fact, these results can be thought of as the hypothetical “target” of an analysis in which a smaller subset of the observations is validated. To study this setting, we recreated the previous study using a 50% validation sample (chosen due to the rarity of strokes in the data) and employed a Monte Carlo approach to observe whether a smaller validation sample size consistently estimates the value produced

with a 100% sample. For each of 1,000 iterations, we randomly selected 10,472 observations from the CPRD data for validation using HES. We then computed the weights in the same manner as before but using the 50% subsample. Figure 2 displays a histogram of the 1,000 estimates of the OR. The average value is 0.75, compared with the estimate using a 100% validation sample of 0.77. However, this 50% sample estimate is accompanied by a wide range, which is expected given the decreased validation sample size and the increased variability of the modified IPW estimator.

A second limitation is that the HES data are not a gold standard and likely have some systematic outcome misclassification themselves. For example, fatal strokes that do not present to the hospital may not be recorded in HES, while the individual’s general practitioner would be informed of these events. Assessing the extent of the residual bias incurred from an imperfect source of validation is not straightforward; although the use of such a source has been studied under a Bayesian framework.¹⁹

In building the propensity score models, we considered total and LDL cholesterol levels; however, as noted in Table 2, large amounts of data were missing (72; 8% for total cholesterol and 87.5% for LDL cholesterol). Given the potential for violations of the requisite missing at random assumption for multiple imputations, we excluded the cholesterol data from the main analyses and conducted a sensitivity analysis in which these data were included using multiple imputations. This sensitivity analysis produced identical results (after rounding).

Finally, we modeled the misclassification rates differentially with respect to statin exposure alone, ignoring any dependency on other covariates. This modeling choice is likely incorrect and as such residual bias may influence the results as shown in Figure 1. The inaccuracy of the pooled estimate complicates our ability to assess the impact of model misspecification on the precision of these estimates.

To study the robustness of the modified IPW estimator to misclassification model misspecification and a small validation sample size, we conducted simulation studies with a data generation process similar to that described in the study by Gravel and Platt¹² and Setoguchi et al.²⁰ For each of 500 datasets, we generated 10 covariates, both discrete and continuous, a binary exposure, dependent on a subset of the covariates, and a binary outcome, dependent on the generated exposure and other covariates. Four of the covariates were generated to be confounders, three to be predictors of the outcome, and three to be predictors of exposure. The target marginal causal OR was set at $\exp(-0.4) = 0.6703$ (chosen to be similar to the study by Gravel and Platt),¹² and the studies were conducted under a rare event setting to reflect the CPRD study.

Next, we misclassified the generated outcomes differentially across 11 sets of studies using the following models:

$$SE_a(l_i) = \text{expit}(\alpha_1 + \alpha_2 a + \alpha_3 l_{ki}) \quad (2)$$

$$SP_a(l_i) = 1 - \text{expit}(\beta_1 + \beta_2 a + \beta_3 l_{ki})$$

where l_{ki} denotes the value of a single covariate for the i^{th} individual, $i = 1, \dots, n$, indexed by k , to which the misclassification rates depend, and a denotes the exposure variable, $a = 0, 1$. This covariate was generated to be a predictor of exposure only (i.e., an instrument), a predictor of the outcome only, or a confounder of the outcome-exposure relationship (please see Table 4). For each study, we generated from nine sets of coefficients characterizing the misclassification parameters starting with high error rates and decreasing across the nine simulated parameter sets to observe the impact of subtle changes in the amount of misclassification. Note that the rest of the parameters in the models are held fixed across all simulations using the values specified in the study by Gravel and Platt.¹²

For each set of misclassification parameters, we display 95% CIs for the IPW OR ignoring misclassification, the modified IPW OR modeled correctly (as generated), and misspecified either nondifferentially or dependent only on exposure for studies 3–11 (i.e., ignoring the additional covariate). We computed estimates of the variance using the bootstrap approach described in the study by Gravel and Platt.¹² We used a sample size of $n = 1,000$ and ran a validation sample size of 20% (see Figure 3).

The black horizontal line represents the target log OR and the dotted line represents the null value of 0.0. For each simulation, we generated a sample size of $n = 1,000$ individuals with a validation sample size of 20%. We computed point estimates using the R built-in function `svyglm()` which was also used to compute the variance estimates for the IPW estimator ignoring outcome misclassification. For the modified IPW estimator, a bootstrap approach described in the study by Gravel and Platt¹² was used with 200 bootstrap samples which was justified as sufficiently large for variance estimation in the study by Efron and Tibshirani.²¹

For all 11 sets of simulation studies, similar trends emerged. The correctly modeled modified IPW analysis appeared to have a minimal bias as expected; however, the intervals were wider, which is consistent with the inclusion of additional parameters in the model. The misspecified modified IPW estimators were biased but less so than the IPW approach across all studies. Misspecification of the misclassification models by excluding the additional covariate biased the results in different directions depending on the type of covariate generated; however, modeling a dependency on exposure consistently produced less bias than the studies in which we assumed nondifferential error.

DISCUSSION

When estimating causal effects in the presence of confounding, residual bias may remain after adjustment due to outcome misclassification. Using the approach of Gravel and Platt¹² and linked hospital records as a source of diagnostic validation, we produced estimates of the association between post-MI statin use and the 1-year risk of stroke that appeared closer to estimates produced by a meta-analysis of RCTs¹⁴ than those computed ignoring misclassification.

In this study, we chose to model diagnostic error as depending only on exposure status; however, this may not be realistic. For example, these error rates may also depend on aspects of clinical practice that can be observed through a regional covariate or the size or type of hospital. We demonstrated through simulation that the choice to model misclassification differentially only on exposure when the data-generating model depends on a single additional covariate, produced some residual bias, although much less than when incorrectly assuming nondifferential misclassification. In addition, the magnitude of the residual confounding depended on the amount of misclassification bias and was relatively low for the studies with lower rates of misclassification. Alternatively, the choice to erroneously assume accurate measurement of the

TABLE 4. Parameter Sets Chosen to Misclassify Simulated Outcomes Specified by Equation 2

Simulation ID	Studies 1–2				Studies 3–11					
	α_1	α_2	β_1	β_2	α_1	α_2	α_3	β_1	β_2	β_3
1	1	0.5	–1	–0.2	1	0.5	–1	–1	–0.2	–0.5
2	1.2	0.5	–1.2	–0.2	1.2	0.5	–1	–1.2	–0.2	–0.5
3	1.4	0.5	–1.4	–0.2	1.4	0.5	–1	–1.4	–0.2	–0.5
4	1.6	0.5	–1.6	–0.2	1.6	0.5	–1	–1.6	–0.2	–0.5
5	1.8	0.5	–1.8	–0.2	1.8	0.5	–1	–1.8	–0.2	–0.5
6	2	0.5	–2	–0.2	2	0.5	–1	–2	–0.2	–0.5
7	2.2	0.5	–2.2	–0.2	2.2	0.5	–1	–2.2	–0.2	–0.5
8	2.4	0.5	–2.4	–0.2	2.4	0.5	–1	–2.4	–0.2	–0.5
9	2.6	0.5	–2.6	–0.2	2.6	0.5	–1	–2.6	–0.2	–0.5

The covariate in equation 2, l_{ki} , was generated to be a confounder for studies 3–5, a predictor of the exposure only for studies 6–8, and a predictor of the outcome only for studies 9–11. For studies 1 and 2, this covariate was not included in the misclassification model. These parameter values were chosen to gradually decrease the magnitude of the outcome misclassification across simulation studies.

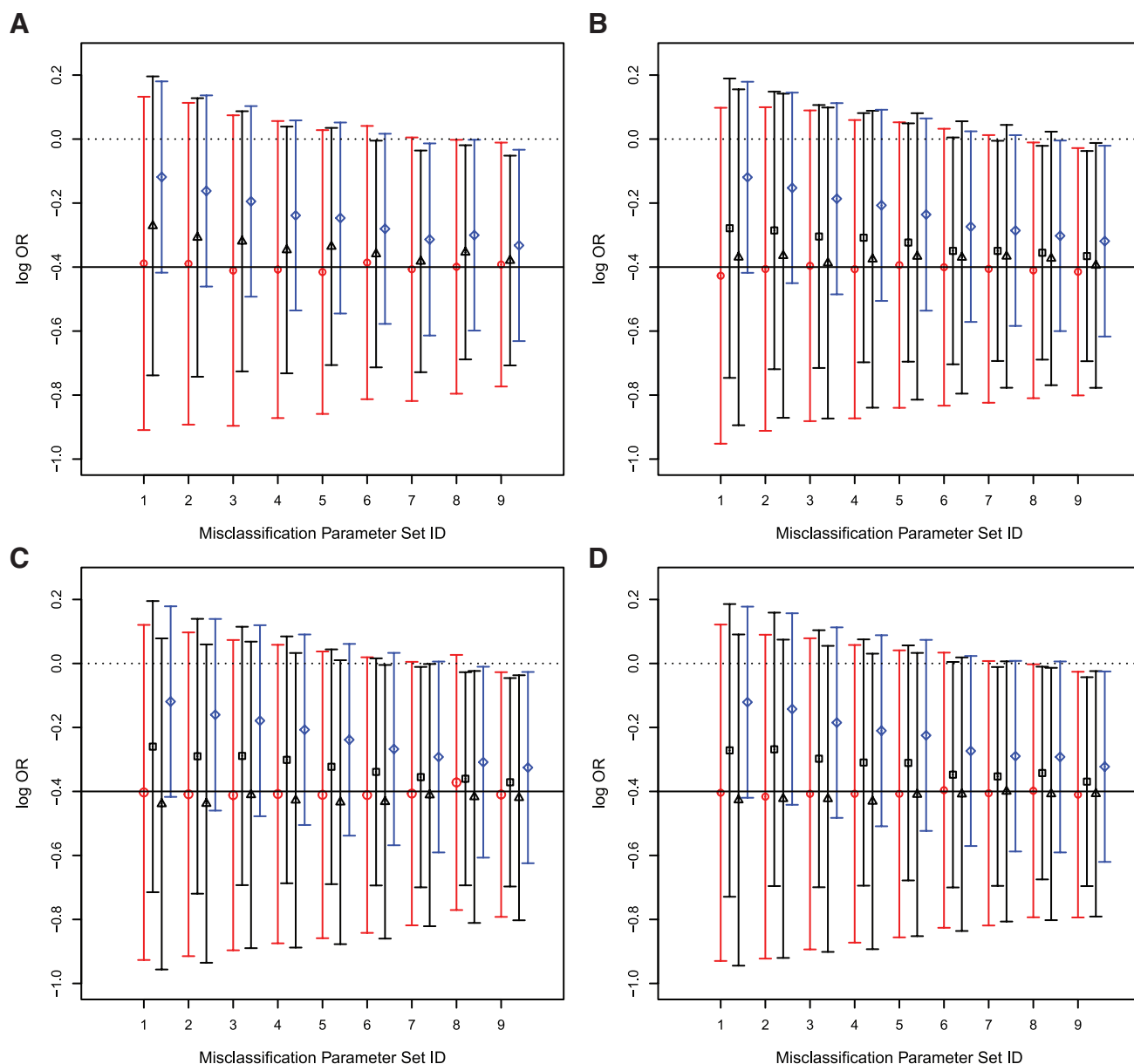


FIGURE 3. Log odds ratio estimates and 95% confidence intervals based on the IPW and modified IPW approach across 11 simulation studies in which differential outcome misclassification is generated based on Equation 2 and Table 4.

- + The MIPW approach correctly modeled
- The MIPW approach modeled nondifferentially
- ▲ MIPW approach modeled dependent on exposure only (possibly ignoring the additional covariate)
- ◆ The IPW approach

Symbols for CI's are as follows: circle denotes the results for the MIPW approach correctly modeled; square denotes the results for the MIPW approach modeled non-differentially; triangle denotes the results for the MIPW approach modeled dependent on exposure only; diamond denotes the results for the IPW approach. These plots present the results of 11 simulation studies to investigate the robustness of the modified IPW estimator to misclassification model misspecification and a small validation sample size. Each of 500 simulated datasets was generated using a process similar to that described in Gravel and Platt¹² and Setoguchi et al.²⁰ In addition, four covariates were generated as confounders, three as predictors of the outcome, and three as predictors of exposure. Outcomes were misclassified differentially across the 11 studies based on nine sets of coefficients characterizing the misclassification parameters starting with high error rates and decreasing across the nine parameter sets as described in Table 4. For each set of misclassification parameters, the log odds ratio estimates and 95% confidence intervals are displayed along with a black horizontal line denoting the log of the target marginal causal odds ratio, which is set at -0.4 . IPW indicates inverse probability of treatment weighting; MIPW, modified inverse probability of treatment weighting.

outcome and use the standard IPW approach produced a large amount of residual bias. As expected, correct model specification produced consistent estimation regardless of the magnitude of the error generated; however, the more parameters included in the models, the higher the variability. In addition, this added variability partly depended on the chosen validation sample size, as demonstrated in the study by Gravel and Platt.¹² However, the minimization of validation sample size may result in additional bias if the expected number of validated observations in any subgroup is small.

The modified IPW approach of Gravel and Platt¹² produces less bias than the IPW approach ignoring misclassification but at the cost of increased variability. However, failing to employ these correction methods creates uncertainty in the point estimate itself. Variance reduction for the modified IPW approach might be accomplished in several ways. Increasing the number of validated observations will produce more precise estimates as noted in the study by Gravel and Platt.¹² Gravel, Farrell, and Krewski²² pointed out that conditionally sampling on the original outcome status (case or control) produced more efficient estimation in a (multiple) contingency table setting. Another approach might be to employ a Bayesian technique to derive point estimates for the quantities in equation 1 (see Paulino et al.).¹⁷ Finally, in settings for which either the sensitivity or specificity is extremely high, the respective misclassification parameter(s) can be fixed at 1, lessening the number of estimated quantities in the modified IPW approach. This may incur some additional bias when incorrect; however, for some adverse health outcomes (e.g., mortality), diagnostic validity and accuracy in healthcare databases may be unquestioned.

The increasing global interest in the use of real-world evidence for drug safety and effectiveness research creates a need for methodologic rigor with respect to addressing potential biases. The presence of outcome misclassification in an observational study that has conducted substantial confounder adjustment can continue to produce important bias, leading to incorrect decisions. Record linkage is increasingly being used to put data sources and their potential for bias in perspective when designing research questions. When the accuracy of outcome ascertainment is thought to be questionable, quantitative bias correction methods should be considered. The modified IPW approach creates a structure to characterize this bias, which gives investigators a tool that can be used to assess the robustness of their findings through sensitivity analyses or to attempt to correct the bias using internal validation sampling.

ACKNOWLEDGMENTS

The data used in this article were obtained from the Clinical Practice Research Datalink. A copy of the study protocol and the original data acquisition request has been

included with the submission. The authors acknowledge Mr. Josselin Cabaussel for his programming support.

REFERENCES

1. Benchimol EI, Smeeth L, Guttman A, et al; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12:e1001885.
2. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018;363:k3532.
3. Suissa S, Henry D, Caetano P, et al; Canadian Network for Observational Drug Effect Studies (CNODES). CNODES: the Canadian network for observational drug effect studies. *Open Med*. 2012;6:e134–e140.
4. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA*. 2018;320:867–868.
5. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44:827–836.
6. Herrett E, Shah AD, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ*. 2013;346:f2350.
7. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data resource profile: Hospital Episode Statistics admitted patient care (HES APC). *Int J Epidemiol*. 2017;46:1093–1093i.
8. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol*. 2005;34:680–687.
9. Greenland S. Maximum-likelihood and closed-form estimators of epidemiologic measures under misclassification. *J Stat Planning Inference*. 2008;138:528–538.
10. Lyles RH, Tang L, Superak HM, et al. Validation data-based adjustments for outcome misclassification in logistic regression: an illustration. *Epidemiology*. 2011;22:589–597.
11. Rubin D. Estimating causal effects of treatments in randomized and non-randomized studies. *J Educ Psychol*. 1974;66:688–701.
12. Gravel CA, Platt RW. Weighted estimation for confounded binary outcomes subject to misclassification. *Stat Med*. 2018;37:425–436.
13. Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol*. 2019;34:91–99.
14. Wang W, Zhang B. Statins for the prevention of stroke: a meta-analysis of randomized controlled trials. *PLoS One*. 2014;9:e92388.
15. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009;20:3–5.
16. Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. *Am J Epidemiol*. 1997;146:195–203.
17. Paulino CD, Soares P, Neuhaus J. Binomial regression with misclassification. *Biometrics*. 2003;59:670–675.
18. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for misclassified outcomes in binary regression models using multiple imputation with internal validation data. *Am J Epidemiol*. 2013;177:904–912.
19. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*. 1995;141:263–272.
20. Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiol Drug Saf*. 2008;17:546–555.
21. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.
22. Gravel CA, Farrell PJ, Krewski D. Conditional validation sampling for consistent risk estimation with binary outcome data subject to misclassification. *Pharmacoepidemiol Drug Saf*. 2019;28:227–233.
23. R Core Team. *R: a Language and Environment for Statistical Computing*. R Foundation FOR Statistical Computing. Vienna, Austria; 2017. <https://www.R-project.org/>
24. SAS Institute, Inc. *SAS/STAT 9.4 User's Guide*. Cary, NC: SAS Institute, Inc; 2013.