Comparison of Two Methods to Correct for Non-differential Misclassification in Meta-analysis

by

Kyle Johnson

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2019

DEDICATION

This thesis is dedicated to my family, who, through their love and support, made everything possible.

Table of Contents

List of Tables

List of Figures

Abstract

Like all data, administrative health data are susceptible to bias. In this project I focus on bias due to misclassification of disease outcome in the context of a meta-analysis. I propose a novel approach to random effects estimation in the presence of misclassification based on a method proposed in the literature for fixed effects estimation. Both these approaches to meta-analysis in the presence of misclassification adjust the study-specific variance in log odds ratio for the presence of between-study variance in misclassification rates. Monte Carlo simulation is used to compare these variance correction approaches to naïve (non-variance) correction approaches.

The simulation demonstrates that, in fact, the naïve correction procedure yields effect estimates that are less biased than those yielded by the variance correction procedure, and its coverage probability is closer to its nominal value. High false negative rates are observed for all homogeneity statistics, while their false positive rates remain low.

List of Symbols Used

θ_i^{se}	True logit sensitivity in study i
θ_i^{sp}	True logit specificity in study i
θ^{se}	Overall mean logit sensitivity
θ^{sp}	Overall mean logit specificity
T	Between study variance in logit sensitivity *and* logit specificity
se_i	True sensitivity in study i
sp_i	True specificity in study i
se_{ovr}	Sensitivity used in matrix method correction procedure
sp_{ovr}	Specificity used in matrix method correction procedure
$\boldsymbol{V}$	Covariance matrix of corrected log odds ratio variances
μ	True overall log odds ratio
μ_i	True log odds ratio in study i
μ_i^U	Uncorrupted log odds ratio estimate in study i
μ_i^M	Misclassification corrupted log odds ratio estimate in study i
μ_i^{MM}	Matrix method corrected log odds ratio estimate in study i
μ_F^U	Uncorrupted pooled fixed effects log odds ratio estimate
μ_R^U	Uncorrupted pooled random effects log odds ratio estimate
μ_F^M	Pooled misclassification corrupted fixed effects log odds ratio estimate
μ_R^M	Pooled misclassification corrupted random effects log odds ratio estimate
μ_F^{NC}	Naively corrected pooled fixed log odds ratio estimate
μ_R^{NC}	Naively corrected pooled random log odds ratio estimate
μ_F^{VC}	Variance corrected pooled fixed effects log odds ratio estimate
μ_R^{VC}	Variance corrected pooled random effects log odds ratio estimate
v_i^U	Variance of uncorrupted log odds ratio estimate in study i
v_i^M	Variance of misclassification corrupted log odds ratio estimate in study i
v_i^{NC}	Naively corrected variance of matrix method corrected log odds ratio estimate in study i
v_i^{VC}	Variance corrected variance of matrix method corrected log odds ratio estimate in study i

v_F^U	Variance of uncorrupted pooled fixed effects log odds ratio estimate
v_R^M	Variance of uncorrupted pooled random effects log odds ratio estimate
v_F^M	Variance of misclassification corrupted pooled fixed effects log odds ratio estimate
v_R^M	Variance of misclassification corrupted pooled random effects log odds ratio estimate
v_F^{NC}	Variance of naively corrected pooled fixed effects log odds ratio estimate
v_R^{NC}	Variance of naively corrected pooled random effects log odds ratio estimate
v_F^{VC}	Variance of variance corrected pooled fixed effects log odds ratio estimate
v_R^{VC}	Variance of variance corrected pooled random effects log odds ratio estimate
Q^U	Uncorrupted χ^2 homogeneity test statistic
Q^M	Misclassification corrupted χ^2 homogeneity test statistic
Q^{NC}	Naively corrected χ^2 homogeneity test statistic
Q_h	Proposed generalized between study variance χ^2 homogeneity test statistic
X_h^2	Greenland's corrected χ^2 homogeneity test statistic
$EmpSE$	Empirical standard error
n_{sim}	Number of iterations in Monte Carlo simulation

Acknowledgements

This thesis would not have been possible without the help and support of a large number of people. I would first like to thank my committee, who provided invaluable guidance and support. Special thanks goes to my supervisor, Dr. Levy, who first gave me the opportunity to pursue this Master's degree, and whose mentorship and guidance have opened up so many doors.

I would also like to thank the Maritime SPOR Support Unit for their financial assistance in the form of their student award. I would not have had the freedom to focus on this thesis as exclusivey as I did.

Finally, I would like to thank my parents for their unwavering support and encouragement. Without them, none of this would be possible. And Cat. I couldn't have asked for a better partner throughout this journey.

Comparison of Two Methods to Correct for Non-differential Misclassification in Meta-analysis

1 Introduction

The evidence needed to generate an accurate risk-benefit profile for a drug is complex and multi-faceted. It ranges from pre-clinical investigations of potential toxicity *in vitro* or *in vivo,* to population level studies of the risk of adverse drug reactions after the drug has been approved for use in the general population. A life-cycle approach to prescription drug safety places special emphasis on post-market surveillance as a means of addressing the limitations of small, short-term pre-market controlled trials (1).

Many unintended effects of prescription medication can only be observed in large populations followed for long durations. Linked administrative health data sets are well suited to this task. In Canada, unique personal identifiers allow for the linkage of administrative health databases within provinces, which is an advantage any other countries lack (2). The advantages of administrative health data sets for research – their large sample size and long term follow-up – would be maximized if all provincial administrative data were collected in a single repository, but privacy and legal concerns have prevented this from happening (3) (an exception is the Canadian Institute for Health Information's discharge abstract database, which covers nearly the entire population of Canada outside of Quebec (4)).

The Canadian Network for Observational Drug Effect Studies (CNODES) is a distributed network of researchers that spans multiple Canadian provinces (3). It coordinates and directs observational pharmacoepidemiological studies carried out within participating provinces, where each study follows a common protocol and uses administrative health data. The results of the province-specific studies are then synthesized in a meta-analysis.

Like all research, this type of observational pharmacoepidemiology is fraught with potential biases, some of which are more common in studies that use administrative health data. In this project, we focus on a particular source of bias – misclassification of disease outcome – and compare two methods of correcting its effects in the context of a meta-analysis. While misclassification has long interested epidemiologists and biostatisticians, it is understudied in this context.

The two methods for correcting misclassification-induced bias investigated in this project are distinguished according to whether or not the study-specific variance in effect size is adjusted to reflect between-study variance in misclassification rates. The procedure to adjust study-specific variances is based on that articulated in (5) for fixed effects estimation. By proposing a novel Cochran between-study variance statistic (6,7) we show how an adjusted estimate of between-study variance in effect size may be derived. Monte Carlo simulation is used to compare the variance corrected meta-analytic effect estimate to the naively (non-variance) corrected estimate on key performance measures. As the choice of a random or fixed effects model is an essential

part of meta-analysis, we also compare various homogeneity test statistics for their false negative and false positive rates.

It is our hope that by determining which is the superior approach to misclassification-induced bias correction in meta-analysis, this project will encourage more widespread use of quantitative bias analysis in meta-analyses. We feel that this is especially important for studies that use administrative health data, and that seek to inform policy decisions.

2 Background

2.1 Administrative Health Data: Identifying Disease Outcomes

Every encounter with the healthcare system generates data used for administrative purposes. Such data includes physician billing claims, hospital discharge abstracts, and pharmacy claims (8). In Canada, a unique identifier associated with each patient and each physician allows for linkage across databases within provinces or territories (9). Date of admission and discharge for hospital abstracts, and date of service for physician billing and pharmacy claims allow for the definition of time-windows for exposure and disease (10). Health outcomes are represented using diagnostic codes, with Canada adopting the International Classification of Diseases, 10th Revision (ICD-10) in a six year implementation process between 2001 and 2006 (11). ICD-10 provides a standardized system for representing "the universe of diseases, disorders, and other related health conditions" by assigning each such condition a distinct alphanumeric code (12). Standardizing diagnostic codes allows for easier comparison of health information between geographic regions and time-periods.

The use of a coding system like ICD-10 allows for the development of consistent case-identification algorithms that can be applied to administrative health databases in different geographical settings. Standardized diagnostic codes may increase an investigator's confidence that the same construct is being measured in every application of the algorithm. Case-identification algorithms often include additional criteria besides diagnostic and procedure codes, such as prescription records and time constraints. The

Canadian Chronic Disease Surveillance System, for example, uses an algorithm for identifying diabetes in administrative health that defines cases by the presence of one hospitalization or two physician billing claims within two years that contain a diagnostic code for diabetes (13) (for a range of other possible such algorithms, see Lix et al. (14)). However, just as an epidemiological study that does primary data collection must describe its selection process in order to demonstrate its internal and external validity, a study using administrative health data must describe why the database was created, how data is entered and by who, and how these factors may influence the accuracy and completeness of the database (15).

Despite how valuable they are for research, administrative health data are not collected for research purposes. This can affect the accuracy of the diagnostic codes they contain, and hence the performance of a case-identification algorithm. Clerical errors in transcription can lead to inaccurate coding, and a given diagnostic code may represent a rule-out diagnosis rather than a final diagnosis (16,17). However, coding accuracy has been shown to depend on both the data source and the target condition (18), which hints that there are more subtle and often more important sources of coding error than can be captured by notions of simple clerical or transcription error (19). The choice of an appropriate case-identification algorithm must take into consideration the care-seeking behaviour of patients with the target condition, the diagnostic procedures and patterns of treatment associated with that condition, and intrinsic characteristics of the databases employed (20).

The clinical characteristics of the target condition – its acuity, severity, and symptomatology – will influence the care seeking behaviour of patients with that condition. The more acute or severe their symptoms, the more likely a patient will seek care. For this reason, administrative health data are biased toward the more severe cases of the target condition. This poses particular problems for the study of chronic diseases, which progress gradually and may go undiagnosed for some time. One can overcome this limitation by restricting oneself to the most catastrophic cases, increasing the accuracy of case-identification, but doing so may misrepresent the true morbidity of the target condition (20).

The process by which a physician decides on a diagnosis and course of treatment determines, in part, whether the disease status of a given individual is accurately represented in an administrative health database (20,21). An accurate diagnostic code depends on an accurate diagnosis, which in turn depends upon a clear, reciprocal exchange of information between patient and physician that allows the physician to decide on the most appropriate diagnostic tools or procedures. However, this information exchange can be hindered if the patient has poor health literacy, or if the physician has poor communication skills (21). The focus of the practice or institution in which the encounter occurs will also play a part in shaping the expectations of the patient, and hence the information they are likely to provide, as well as the procedures and practices most familiar to the physician. These factors and more are encompassed by the notion of medical practice variation (22,23). Investigators should think critically

about how medical practice variation may affect the performance their chosen case-identification algorithm.

Reimbursement policies can also affect the reliability of administrative health data. Internal auditing of diagnostic codes associated with inpatient hospital stays for insurance purposes, for instance, may render these records more reliable than those associated with outpatient care (20). Formulary restrictions can also lead to certain pharmaceutical exposures being underrepresented in administrative health data (24–26)

All factors influencing the performance of a case-identification algorithm discussed in this section are known to vary by geographic region. Care seeking behaviour varies according to demographic and sociocultural factors that vary between regions (27–29). Geographic variation in medical practice is well documented (22,23,30,31). In Canada, formulary restrictions vary by province (32,33). This makes the analysis of potential variability in the performance of a case-identification algorithm an important consideration when synthesizing information from geographically dispersed administrative health databases.

2.2 Meta-Analysis

Meta-analysis is a statistical technique for the systematic synthesis of distinct research studies with the aim of achieving a more precise understanding of the distribution of a given treatment effect (34). If all studies are considered to be

estimating a common effect, a meta-analysis will in general provide a more precise estimate of the magnitude of that effect than any of the individual included studies (35). If the studies are not estimating a common effect, meta-analysis allows for the quantitative assessment of the heterogeneity between included studies. Investigation of heterogeneity is indeed one of the key motivations for conducting a meta-analysis (34).

The former case is referred to as *fixed effects* meta-analysis, referencing the fact that the effect size being estimated is identical in all included studies. The latter case is referred to as a *random effects* meta-analysis to indicate that there is random variation in that effect size between included studies. To each of these models of the underlying distribution of true effect sizes among the included studies there exists a corresponding estimation procedure, respectively called *fixed effects estimation* and *random effects estimation*. In the random effects case, we assume that the true effect sizes in the included studies, μ_i, are normally distributed about an overall mean μ with variance τ^2.

Since the true underlying model is unknown, a key step in any meta-analysis is the performance of a χ^2 test for homogeneity, which, if statistically significant, is strong evidence for the presence of heterogeneity, and may be taken as justification for the use of random effects estimation (36,37). In case the χ^2 test is not statistically significant, then fixed effects estimation is taken as appropriate. However, these χ^2 tests are known to have poor performance, especially when the number of studies is small, and additional qualitative exploration of heterogeneity is usually necessary (38).

Table 1 lays out the notation used in this manuscript for the various data associated with any given meta-analysis.

	Estimate	Estimand (Target Value)
Effect size in study i	$\widehat{\mu_i}$	μ_i
Variance in study i	$\widehat{v_i}$	v_i
Overall effect size	μ_F, μ_R	μ
Fixed effects estimate	μ_F	μ
Random effects estimate	μ_R	μ
Between study variance	$\widehat{\tau^2}$	τ^2
χ^2 Homogeneity Statistic	Q	NA

Table 1: Notation used for data associated with a meta-analysis. Note that Q does not have a "true value", since it is a statistic calculated from the data

2.2.1 Prospectively Planned Meta-Analysis

CNODES conducts what may be referred to as “prospectively planned” meta-analyses, which pose several unique methodological problems (39). They are prospectively planned because the location of the individual studies is known in advance. An argument can therefore be made that these studies do not comprise a random sample from some larger set of studies. If this is the case the use of a random effects model is theoretically questionable.

Regardless of the validity of the above argument, heterogeneity must still be dealt with, and CNODES studies do employ random effects estimation. As shown in Table 2, four of the last six studies conducted by CNODES chose to use random effects

meta-analysis. Thus, questions concerning the performance of random effects estimation in meta-analysis of administrative health data are still relevant for CNODES (40–45).

	Year Published	Random Effects Model Used?	χ^2 Homogeneity Test Significant?
Azoulay L, Filion KB, Platt RW, Dahl M, Dormuth CR, Clemens KK, et al.	2016	Yes	No
Renoux C, Dell'Aniello S, Khairy P, Marras C, Bugden S, Turin TC, et al.	2016	Yes	No
Henry D, Dormuth C, Winquist B, Carney G, Bugden S, Teare G, et al.	2016	No	N/A
Renoux C, Lix LM, Patenaude V, Bresee LC, Paterson JM, Lafrance J-P, et al.	2015	Yes	Yes
Gomes T, Paterson JM, Mukati M, Henry D, investigators for CNODES.	2015	No	N/A
Dormuth CR, Filion KB, Paterson JM, James MT, Teare GF, Raymond CB, et al.	2014	Yes	Yes

Table 2: Six most recently published CNODES studies, their use of random effects models, and the statistical significance of the χ^2 homogeneity test

2.3 Misclassification

Misclassification is a type of information bias to which all epidemiological studies are susceptible. Studies using administrative health data are especially susceptible, if only because such data is not originally collected for research purposes (other, more systematic reasons for the misrepresentation of health state information in administrative health databases are touched on above. See Terris et al. (21) for a useful conceptual framework). Early work investigating the effect of misclassification led to the development of the *matrix method* of correction, an algebraic technique that can

recover the original data assuming the misclassification rates are known (46). If the misclassification rates are not known, they can be estimated using validation studies (47). This section will first describe the validation study method of arriving at the key measures of misclassification – sensitivity and specificity – and then describe how the matrix method can be used to correct an odds ratio. We will then review the difference between differential and non-differential misclassification, and, finally, describe the magnitude and direction of the bias induced by non-differential misclassification.

2.3.1 Validation Studies

A validation study compares the performance of fallible classification criteria to a "gold-standard", which is difficult and expensive to administer, and assumed to be infallible (47). The fallible classification criteria we are interested in are those that comprise an algorithmic means of identifying disease in administrative health databases, but identical considerations apply to the identification of other outcomes, such as exposure status, for example. Applying both the fallible and infallible criteria to the subjects comprising the validation sample yields data in the form of Table 3, from which four measures of misclassification can be derived: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). While PPV and NPV may be the most intuitive measures of performance, representing the probability that a positive and negative classification, respectively, are correct, their values depend on the prevalence of the outcome in question, meaning they cannot be generalized easily outside the population in which the validation was conducted (15). Therefore, sensitivity and specificity, the probabilities that a person is recorded as having the outcome given

that they truly do, and as not having the outcome given that they truly do not, respectively, are the most common way of quantifying the performance of a classification algorithm.

		Infallible Classifier	
		Positive	Negative
Fallible Classifier	Positive	TP	FP
	Negative	FN	TN

Table 3: Example output of validation Study

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

2.3.2 Matrix-method Correction

Assume that sensitivity and specificity are known or have been estimated in a validation study. Let sensitivity be denoted se and specificity be denoted sp. The corrupting action of misclassification can be represented by the following corruption matrix B:

$$B = \begin{bmatrix} se & 1 - se \\ 1 - sp & sp \end{bmatrix}$$

If $U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the true contingency table, prior to corruption by misclassification, then the corrupted contingency table $C = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$ is given by

$$C = \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} se & 1-se \\ 1-sp & sp \end{bmatrix}$$

Clearly then, given values for se and sp, we can recover the original data U by multiplying C by B^{-1}, where

$$B^{-1} = \left(\frac{1}{se + sp - 1}\right) * \begin{bmatrix} sp & se-1 \\ sp-1 & se \end{bmatrix}$$

2.3.3 Differential versus Non-differential Misclassification

In the above, we assume that sensitivity and specificity are the same in each exposure category. This situation is referred to as non-differential misclassification. If sensitivity and specificity depend on exposure category, we refer to it as differential misclassification.

This project focuses on non-differential misclassification for three reasons. First, data on differential misclassification is more difficult to come by. Depending on the exposure, it may be unlikely that a published validation study reports sensitivity and specificity values separately for the particular exposure categories under investigation. Second, assuming non-differential misclassification allows for a more straightforward model of misclassification and therefore simplifies the simulation. Finally, as argued below, misclassification in our motivating example study is most likely non-differential.

2.3.4 Magnitude and Direction of Misclassification-Induced Bias

Assuming the classifier performs better than chance (i.e. both sensitivity and specificity are greater than 0.5), and that misclassification is non-differential, ignoring misclassification of dichotomous outcomes across a dichotomous exposure will, on average, bias effect estimates toward the null (48). In that paper, Neuhaus derives a bias factor that allows us to approximate the magnitude of this attenuation relative to the true effect size. The amount of attenuation depends on the false positive rate ($\gamma_0 = 1 - sp$), false negative rate ($\gamma_1 = 1 - se$), and prevalence of the outcome in question. Figure 1, reproduced from (48), shows the relation between γ_0, γ_1, and the ratio of observed and true effect sizes, assuming a prevalence of 0.5. We see, for example, that if the false negative and false positive rates are both 0.2, the uncorrected effect is approximately 60% of its true size.

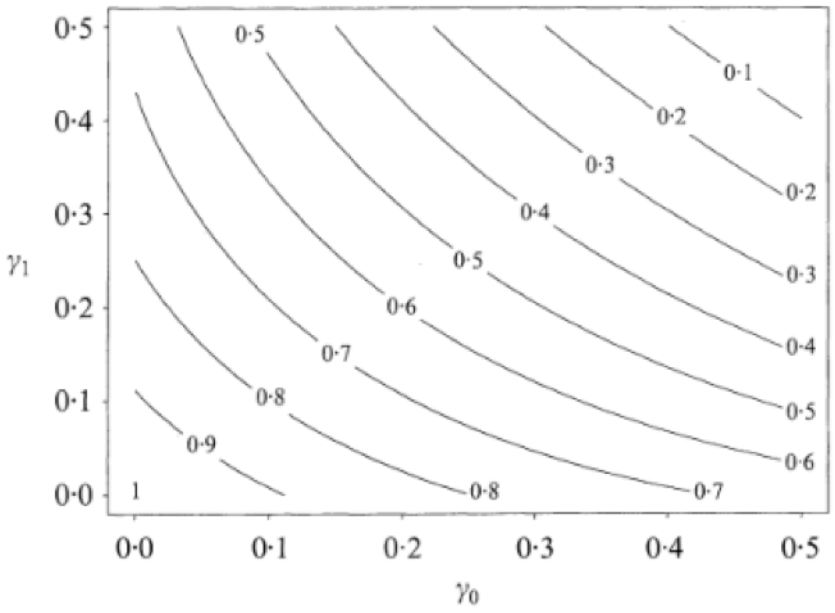


Figure 1: Contour plot of bias in log odds ratio due to misclassified outcomes. Reproduced from Neuhaus, 1999

2.4 Systematic Error Reporting in Observational Studies

Observational studies lack the benefits of randomization that (if properly done) ensure the unbiased allocation of treatment and the applicability of statistical techniques that assume randomization (49)**.** Observational studies therefore have "built-in bias" (50) that must be carefully addressed in the design and execution of the study.

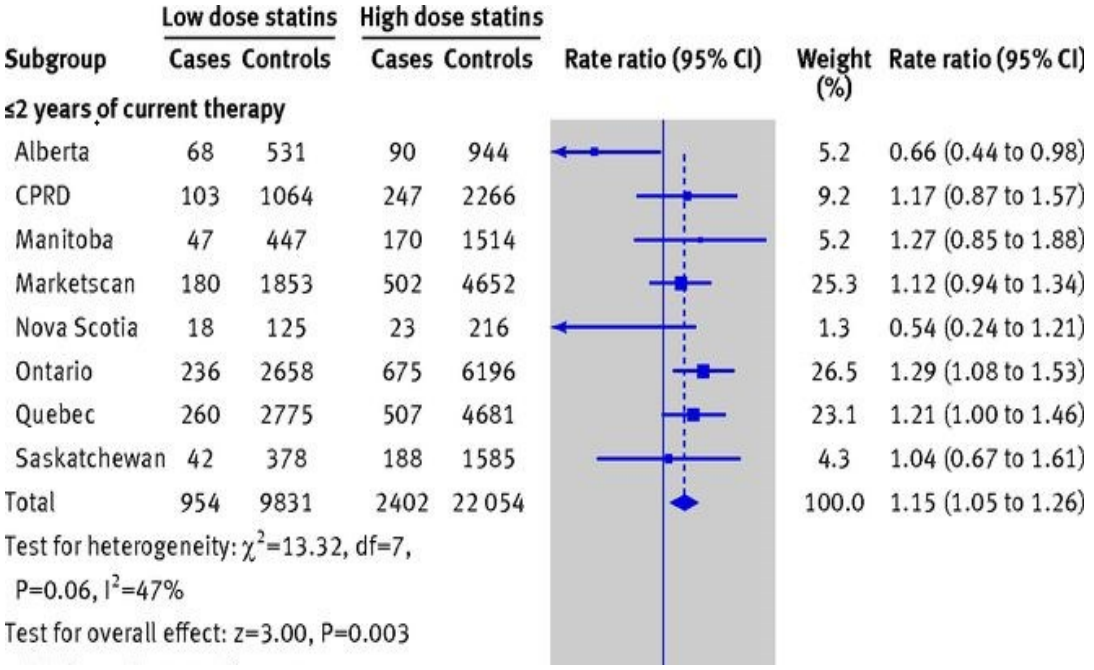
Epidemiological journals require that authors report measures of random error, such as confidence intervals and p-values, but do not require quantitative analyses of potential sources of bias (51). The vast majority of observational studies only report quantitative measures of random error, due to randomization or sampling variance, for example (52). Yet confidence intervals and standard errors give an inadequate representation of uncertainty in the presence of bias, and their interpretation is problematic when no randomization or random sampling has occurred. Potential sources of bias tend to be addressed qualitatively, rather than quantitatively, but techniques for quantitative bias analysis exist and ought to be more widespread (53). Quantitative bias analysis in observational studies allows those who use the research – policy makers, regulators, etc. – to better understand sources of systematic uncertainty (52). This is relevant for the type of pharmacoepidemiological studies undertaken by organizations like CNODES which (i) use administrative databases that cover the entire population rather than a random sample, and (ii) are motivated primarily by the need to provide regulators with timely answers to drug safety queries (3)

2.5 Misclassification in Meta-Analysis: A Motivating Example

Figure 2 presents the results of a 2014 study conducted as part of CNODES that used administrative health data from six Canadian provinces and two international databases to investigate the risk of new diabetes associated with higher potency versus lower potency statin use in patients receiving statins for the secondary prevention of cardiovascular events (45). It was the first study of this association that focused on real-world use, was restricted to secondary prevention, and was specifically designed to evaluate diabetes endpoints, making it an important contribution to the literature. The criteria used in their case-identification algorithm for identifying diabetes from administrative health data were (a) a record of hospitalization with a principal or secondary diagnosis code for diabetes, or (b) a prescription for insulin or oral antidiabetic. As we argue above, the clinical characteristics of diabetes, the patterns of care associated with its diagnosis and treatment, and characteristics of the databases employed will all affect the likelihood that diabetes status is accurately represented in administrative databases (20).

Diabetes is an example of a chronic disease that may be difficult to identify in administrative databases. It is clinically heterogeneous and can remain asymptomatic for long periods (54). Further, there is pronounced variation in diabetes treatment across geographic regions (55). Finally, systematic differences in the health-seeking behaviour of a subpopulation of diabetes patients, due to low health literacy (56–58), for example, could lead to its underrepresentation in administrative health data. If any of these systematic factors impacting the representation of diabetes status in the

databases used for this study vary across the populations covered by these databases, misclassification rates in these populations will also vary.



Subgroup	Low dose statins Cases	Low dose statins Controls	High dose statins Cases	High dose statins Controls	Rate ratio (95% CI)	Weight (%)	Rate ratio (95% CI)
≤2 years of current therapy							
Alberta	68	531	90	944		5.2	0.66 (0.44 to 0.98)
CPRD	103	1064	247	2266		9.2	1.17 (0.87 to 1.57)
Manitoba	47	447	170	1514		5.2	1.27 (0.85 to 1.88)
Marketscan	180	1853	502	4652		25.3	1.12 (0.94 to 1.34)
Nova Scotia	18	125	23	216		1.3	0.54 (0.24 to 1.21)
Ontario	236	2658	675	6196		26.5	1.29 (1.08 to 1.53)
Quebec	260	2775	507	4681		23.1	1.21 (1.00 to 1.46)
Saskatchewan	42	378	188	1585		4.3	1.04 (0.67 to 1.61)
Total	954	9831	2402	22 054		100.0	1.15 (1.05 to 1.26)

Test for heterogeneity: χ^2=13.32, df=7, P=0.06, I^2=47%

Test for overall effect: z=3.00, P=0.003

Figure 2: Rate ratios for new onset diabetes within two years of starting higher potency or lower potency statins after a major cardiovascular event or procedure (Dormuth et al. 2014)

If there is misclassification in this study, it is most likely non-differential. If sensitivity and specificity were to vary between the higher and lower potency statin groups within an individual site, then the choice of diabetes classification criteria would imply that either (i) the risk of hospitalization (and diagnosis of diabetes while in hospital), or (ii) the likelihood of receiving a prescription for insulin or an oral anti-diabetic medication must vary between the exposure groups. The latter is unlikely as no evidence of a clinically meaningful drug-drug interaction between statins and oral anti-diabetics or insulin exists (59) that would contraindicate their prescription based on the potency of the statin. It is also reasonable to assume that cases of diabetes in both

statin groups are clinically similar due to the focus on identifying incident cases. Hence, the patterns of care associated with their diagnosis and treatment will also be similar (20). While it is possible that differences in baseline characteristics make one statin group more likely to be hospitalized in general, inspection of Table 1 in Dormuth et al. (reproduced in Appendix 1) indicates that the number of hospitalizations is similar between exposure groups (45). However, note that non-differential misclassification refers to identical sensitivity and specificity between exposure groups *within a single site*, and not across sites. Misclassification rates could vary between sites and it could still be non-differential.

Dormuth et al. provide a single estimate of sensitivity and specificity – 90% and 92%, respectively – for their chosen case-identification algorithm. However, no bias-correction was performed, and the choice of validation study from which these estimates were derived should raise some reservations. First, the authors state that "[u]sing hospital discharge data to capture diabetes was previously evaluated to have 90% sensitivity and 92% specificity in the Canadian province of Ontario", referring to the validation study of Hux et al. (60). Recall, however, that this was not the case-definition used (one would expect that the inclusion of prescription data would have an impact on sensitivity and specificity). Further, Hux et al. in fact validated an algorithm in which diabetes was defined as the presence of any hospital discharge abstract *or physician service claim* with the presence of a diagnostic code for diabetes. In the Dormuth et al. study, physician services claims were excluded due to concerns with immortal time bias (45).

Independently of the accuracy of sensitivity and specificity estimates, bias-correction for misclassification in meta-analysis poses methodological problems. In meta-analyses of published studies, estimates of misclassification rates in each study will not be available. In a meta-analysis such as Dormuth et al., which is prospectively planned, validation studies could theoretically be conducted at every site, but in practice this is impossible due to resource constraints. In practice, and if possible, investigators conducting a meta-analysis should attempt to find a meta-analysis of validation studies from which to derive the sensitivity and specificity estimates for their chosen classification algorithm.

2.6 Model of Between-Study Variation in Misclassification Rates

Reitsma et al. (61) suggest that meta-analyses of validation studies should model the distribution of logit sensitivity and logit specificity as bivariately normally distributed among the included studies. As opposed to other methods of meta-analyzing validation studies, this method allows for the estimation of between study variance in logit sensitivity and specificity separately and calculation of a confidence ellipse around the pooled mean vector of logit sensitivity and specificity. Further, it allows for the introduction of covariates that have separate effects on sensitivity and specificity (61).

In order to model the distribution of misclassification rates among studies included in a meta-analysis, we chose to follow this bivariate model. Thus, the logit sensitivity and logit specificity at the i^{th} site, θ_i^{se} and θ_i^{sp}, are bivariately normally distributed about an overall mean vector $(\theta^{se}, \theta^{sp})$:

$$\left(\theta_i^{se}, \theta_i^{sp}\right) \sim N\left((\theta^{se}, \theta^{sp}), T\right),$$

where T is a 2x2 covariance matrix.

This model will allow us to easily vary the between-study variance in logit sensitivity and specificity, and provides the ability to accommodate correlation between them (61).

Like the early work on misclassification, which focused on the situation in which the true misclassification rates were known, this project focuses on the situation in which the true overall sensitivity and specificity are known. The overall sensitivity and specificity are assumed to be 90% and 92%, respectively, as reported in Dormuth et al.

2.7 Matrix Method Correction in Meta-Analysis

In a meta-analysis with varying misclassification rates, if sensitivity and specificity estimates could be derived from an internal validation sample at each site, the matrix method could be applied to the study-specific effect estimates separately, before pooling. More realistically, we must use an estimate of the overall misclassification rates (ideally pooled from a meta-analysis of validation studies) and apply the same correction matrix to every site.

2.7.1 Derivation of Common Correction Matrix

In Section 2.2.2, the correction matrix B^{-1} was shown to be of the form

$$B^{-1} = \left(\frac{1}{se + sp - 1}\right) * \begin{bmatrix} sp & se - 1 \\ sp - 1 & se \end{bmatrix}$$

for known sensitivity and specificity se and sp. However, in the context of a meta-analysis in which the true logit misclassification rates follow the distribution specified in Section 2.5, se and sp cannot simply be taken to the inverse-logit transform of the overall means θ^{se} and θ^{sp}. In the case of the study-specific sensitivities se_i, this is because

$$E[se_i] \neq \frac{e^{\theta^{se}}}{1+e^{\theta^{se}}} = se$$

Rather, a Taylor series expansion (62) must be used to approximate $E[se_i]$**.** Let

$$g(u) = logit^{-1}(u) = \frac{e^u}{1+e^u}$$

Then, with θ_i^{se} the logit sensitivity at the i^{th} site,

$$E[g(\theta_i^{se})] = g(\theta^{se}) + \frac{1}{2}g''(\theta^{se}) * \sigma_{se}^2$$
$$E[se_i] = se(1 + \frac{\sigma_{se}^2}{2}(se-1)(2*se-1))$$

Similar considerations hold for specificity.

With these approximations of the true values of $E[se_i]$ and $E[sp_i]$, the correction matrix becomes

$$B^{-1} = \left(\frac{1}{E[se_i]+E[sp_i]-1}\right) * \begin{bmatrix} E[sp_i] & E[se_i]-1 \\ E[sp_i]-1 & E[se_i] \end{bmatrix}$$

2.7.2 Variance Correction

The study-specific variances of the corrected log odds ratio estimates may be calculated directly from the matrix method corrected contingency table using the familiar formula of Woolf (63), a procedure we refer to as “naïve correction”. On the other hand, the study-specific variances may be updated to reflect the variance in misclassification rates, a procedure we refer to as “variance correction”.

Our variance correction procedure will follow that articulated by Greenland in (5) for non-differential misclassification. However, whereas variation in misclassification rates in that paper was assumed to represent the *sampling variance* associated with the validation study from which the misclassification rates were derived, we are interested in the case in which variance in misclassification rates takes the form of *between study variance*. In this section, we describe the variance correction procedure proposed by Greenland and how it can be adapted for our purposes.

For a given study i, let f_1 and f_0 respectively denote the corrected proportion of exposed and unexposed non-cases, and let e_1 and e_0 denote the corrected proportion of exposed and unexposed cases. If we let $*$ in the superscript indicate that the parameter refers instead to the uncorrected proportion, then Greenland provides the following formula for the corrected variance of the log odds ratio at study i:

$$V_i = (Var[sens] * \left(\frac{1}{f_1} - \frac{1}{f_0}\right)^2 + Var[spec] * \left(\frac{1}{e_1} - \frac{1}{e_0}\right)^2 + \sum_t C_t)/D^2$$

$$C_t = e_t^* f_t^* / (M_t e_t^2 f_t^2), \quad t = 0,1$$

$$D = sens + spec - 1$$

where M_1 is the total number of subjects in the exposed group, M_0 is the total number of subjects in the unexposed, and $sens$ and $spec$ are the estimated misclassification rates. Then $Var[sens]$ and $Var[spec]$ can be derived from the between-study variance in logit sensitivity, σ_{se}^2, and logit specificity σ_{sp}^2, by an application of the delta method. See Appendix 3.

The corrected weights for pooling the study-specific estimates into a variance corrected fixed effects estimate μ_F^{VC} of an assumed common log odds ratio μ are given by first forming the covariance matrix $\boldsymbol{V}$ whose ji^{th} $(j \neq i)$ element is given by

$$V_{ji} = \sum_{h,t} \frac{(-1)^{h+t}\left(\frac{Var[sens]}{f_{hj}f_{ti}} + \frac{Var[spec]}{e_{hj}e_{ti}}\right)}{D^2}$$

and whose diagonal elements are the corrected study-specific variances:

$$V_{ii} = V_i$$

Then, letting V^{ji} denote the ji^{th} element of $\boldsymbol{V^{-1}}$, $W_i = \sum_j V^{ji}$, and $W_+ = \sum_i W_i$, the estimate of the common log odds ratio, μ_F^{VC}, is given by:

$$\mu_F^{VC} = \frac{\sum_i W_i \mu_i^{MM}}{W_+}$$

where μ_i^{MM} is the matrix method corrected log odds ratio estimate from the i^{th} study. A consistent variance estimator for μ_F^{VC} is given by $1/W_+$.

2.7.3 Why Matrix Method?

There are many ways of correcting for misclassification-induced bias. The following section provides a rationale for focusing on the matrix method.

The matrix method has the advantage over the inverse matrix method of utilizing sensitivity and specificity rather than predictive values. The latter depend on the prevalence of the outcome in question, meaning estimates provided in the literature have limited generalizability, as the prevalence in the validation sample from which they are derived is likely different from that in the sample to they are to be applied. Additionally, in the case of a prospective meta-analysis in the style of CNODES, in which an internal validation sample is theoretically possible, such a sample taken from one included study's (site's) population would face the same generalizability problems if the included study populations are heterogeneous. Finally, the inverse matrix method is most naturally applicable in the case of differential misclassification (64), which is not being considered here.

The matrix method is also significantly less computationally intensive than the modified maximum likelihood approach. Because this project is looking to make a preliminary contribution to the literature on quantitative bias analysis in meta-analysis, we choose to simplify the simulation by focusing on the matrix-method. Further, given that a central concern of the present study is to argue for greater use of quantitative bias analysis in epidemiological studies, it makes sense to focus on the simplest, most

methodologically accessible approach. Future work ought to investigate the modified maximum likelihood method, however, which may provide better performance.

Iterative approaches to bias analysis, such as probabilistic bias analysis and Bayesian bias analysis (52,65,66) depend on one of the above methods of correction. These more sophisticated approaches choose one method of bias correction, and repeatedly apply it to adjust a corrupted effect estimate, contingent on the values of relevant bias parameters (e.g. misclassification rates) that are iteratively drawn from pre-specified probability distributions on that parameter. The performance of iterative approaches to bias analysis therefore depends on the performance of the chosen bias correction method. By evaluating the matrix method of bias correction, we provide a basis upon which the validity of probabilistic and Bayesian bias analysis depends.

2.8 Misclassification and Between-Study Variation in Effect Size

The biasing effect of non-differential misclassification, because it draws the k study-specific estimates toward the null, will tend to shrink the differences between those estimates. The conventional $k-1$ degree of freedom χ^2 statistic

$$Q = \sum w_i(\hat{\mu}_i - \bar{\theta}_F)^2$$

would then be biased toward the null and would suffer reduced power, decreasing the probability that heterogeneity would be detected. However, the study-specific variances v_i are also shrunk due to the biasing effect of misclassification, increasing the weights

w_i, and hence biasing the Q statistic away from the null and increasing the probability a positive Q test..

A reduction in the power of the homogeneity test, which is already underpowered (67), can lead to the inappropriate use of a fixed-effects model. Fixed-effects models generate tighter confidence intervals than random-effects (68) and would misrepresent the true extent of uncertainty about our pooled estimate. In the presence of misclassification, then, we risk overstating the precision of an inaccurate estimate. The direction of bias due to non-differential misclassification means the true effect of a drug will be underestimated. Unless misclassification is adequately addressed in the planning and interpretation of a study, this could mean that a harmful drug remains in use, or that a beneficial drug does not make it to market.

2.8.1 Corrected χ^2 Test

In addition to providing formulas for corrected study-specific log odds ratio variances and pooled fixed-effects overall log odds ratio estimate, (5) also provides a corrected $k-1$ degree of freedom χ^2 statistic, X_h^2, for testing for the homogeneity of the study-specific effect estimates:

$$X_h^2 = \sum_{ji} V^{ji} \mu_j^{MM} \mu_i^{MM} - W_{+(}\mu_F^{VC})^2$$

However, this statistic is of a different form that the conventional Q statistic. The Q statistic presented above is a kind of "generalized Cochran between-study variance statistic" (6,7), which are χ^2 statistics taking the general form

$$Q_a = \sum_i a_i (y_i - \bar{\theta}_F)^2$$

for any positive weights a_i, observations y_i, and weighted mean $\bar{\theta}_F = \frac{\sum_i a_i y_i}{\sum_i a_i}$. By equating Q_a to its expected value, one can derive a generalized method of moments (GMM) estimator of the between study variance:

$$\tau^2_{GMM} = \max\left(0, \frac{Q_a - \left(\sum_i a_i v_i - \frac{\sum_i a_i^2 v_i}{\sum_i a_i}\right)}{\sum_i a_i - \frac{\sum_i a_i^2}{\sum_i a_i}}\right)$$

When the a_i are inverse variance weights, $a_i = v_i^{-1} = w_i$, τ^2_{GMM} reduces to the familiar Dersimonian and Laird estimator.

Because the corrected χ^2 statistic, X^2_h is not a generalized Cochran between-study variance statistic, it is not obvious how to derive a corresponding τ^2 estimate. In an attempt to address this limitation of X^2_h, we propose to use a generalized Cochran between-study variance statistic

$$Q_h = \sum_i W_i (\mu_i^{MM} - \mu_F^{VC})^2$$

where the a_i are the corrected study weights W_i defined in Section 2.7.2, the μ_i^{MM} refer to the matrix method corrected effect estimate from site i and μ_F^{VC} refers to the corrected pooled fixed effects estimate. A GMM estimator τ_h^2 can then be derived from Q_h, allowing for the calculation of new study weights W_i^* as follows. Form a new covariance matrix $\boldsymbol{V}^*$ from $\boldsymbol{V}$ by adding τ_h^2 to the diagonal entries representing the corrected study-specific variances:

$$V_{ii}^* = V_{ii} + \tau_h^2$$
$$V_{ji}^* = V_{ji}, \qquad j \neq i$$

Then, if V^{ji} denotes the ji^{th} element of $\boldsymbol{V}^{*-\mathbf{1}}$, let

$$W_i^* = \sum_j V^{ji}$$
$$W_+^* = \sum_i W_i^*$$

The corrected pooled random effects estimator μ_R^{VC} can be calculated as

$$\mu_R^{VC} = \frac{\sum_i W_i^* \mu_i^{MM}}{W_+^*}$$

With this we have in hand two variance corrected effect estimates, μ_F^{VC} and μ_R^{VC}, corresponding to fixed and random effects estimation, respectively. We are now ready to compare their performance to the corresponding naively (non-variance) corrected estimates μ_F^{NC} and μ_R^{NC}, and determine which is more suitable for use in quantitative bias analysis.

3 Objectives

The objectives of this study are, using a Monte Carlo simulation modeling a meta-analysis of administrative health data,

(1) To compare the performance of the variance corrected random effects estimator μ_R^{VC} to that of the naively corrected random effects estimator μ_R^{NC} in situations in which there is between study heterogeneity in both effect size and misclassification rate

(2) To compare the false negative rates and false positive rates of the χ^2 statistics (i) Q, corresponding to the non-variance corrected estimate, (ii) X_h^2, Greenland's variance corrected estimate, and (iii) Q_h, our proposed generalized Cochran between study variance statistic

(3) To compare the performance of the variance corrected fixed effects estimator (due to Greenland) μ_F^{VC} to the naively corrected fixed effects estimator μ_F^{NC}

(4) To determine whether the incorrect specification of the meta-analytic model (i.e. choosing a fixed effects estimator when in fact there is between study variance in effect size) affects the naively corrected estimator more than the variance corrected estimator

4 Methods

This section will provide a description of the Monte Carlo simulation to be conducted. The simulation is coded using R. We begin by defining key simulation parameters. Following this, we specify our chosen performance measures, and trace how these measures are calculated, starting from site level effect size determination, through meta-analysis level pooling and hypothesis testing, and finally to what we call the context level, which is the level at which the performance measures are determined.

4.1 Parameters of Simulation

We call those parameters whose effect on the outcomes we wish to control the "variable" parameters, referring to the fact that their range is pre-specified, and their value changes at the direction of the user. The variable parameters in this simulation are the overall log odds ratio μ, data generating model (DGM) $Model$, and between study variance in misclassification rates T. The former two parameters parameterize the distribution of study-specific estimands in the simulated meta-analyses: if $Model$ is $Fixed$ then the study-specific estimands μ_i are identical to μ. If $Model$ is $Random$, $\mu_i \sim N(\mu, \tau^2_{eff})$, where τ^2_{eff} is derived from χ^2 statistic reported in Dormuth et al. T parameterizes the distribution of the study-specific logit misclassification rates:

$$T = \begin{bmatrix} \sigma^2_{se} & \sigma_{sesp} \\ \sigma_{sesp} & \sigma^2_{se} \end{bmatrix},$$

where σ^2_{se} and σ^2_{se} are the variances of logit sensitivity and specificity, respectively, and σ_{sesp} is their covariance. In order to ease the computational burden, we will consider

only those T for which $\sigma_{sesp} = 0$ and $\sigma_{se}^2 = \sigma_{sp}^2$. We therefore use T to refer to both σ_{se}^2 and σ_{sp}^2 in what follows.

Different combinations of variable parameters correspond to different contexts in which a meta-analysis can be conducted. We define a *data generating context* (DGC) as a triple $(\mu, Model, T)$. The range of μ differs according to the value of $Model$ to reflect the fact that Dormuth et al. observed a fixed effects estimated odds ratio of 1.15, and a random-effects estimated odds ratio of 1.11. The range of T is independent of the other two variable parameters, taking on 10 equally spaced values in the range [0, 1]. Table 2 lays out the possible combinations of variable parameters that define the DGCs considered in the simulation.

	Possible Values	
	Fixed Effects Model	*Random Effects Model*
Overall Log Odds Ratio	log(1), log(1.15), log(2.3)	log(1), log(1.11), log(2.22)
T	$[0, 0.1, 0.2, \dots, 1]$	$[0, 0.1, 0.2, \dots, 1]$

Table 4: Values of the Overall Odds Ratio and T for different values of Model

The between study variance in effect τ_{eff}^2 is an example of a fixed parameter. Fixed parameters are those that are constant across all iterations in all DGCs. Other key examples of fixed parameters are the overall sensitivity and specificity values used by Dormuth et al., which are 90% and 92%, respectively. See Appendix 2, Table 1 for a full list of fixed parameters and their values.

The stochastic variables are those that follow a probability distribution parameterized by the fixed and variable parameters just described. Appendix 2, Table 2

lists the stochastic parameters, the fixed and variable parameters determining their distributions, and the distributions themselves.

4.2 Simulation Outcomes

In each DGC, 10000 meta-analyses will be simulated. The number 10000 was settled on because of the possibility that the variance of the variance corrected effect estimates will turn out to be negative, and this happens in a non-negligible number of iterations. It is it difficult to determine the number of iterations necessary to achieve a given level of precision when the probability of a negative variance is unknown. 10000 was judged to be a sufficient number of iterations to achieve a high level of precision (low Monte Carlo standard error) while allowing for the negative variance observations. The issue of negative variances will be discussed shortly.

Table 4 lays out the key estimands, estimates, and statistics that will be calculated at the study and meta-analysis levels of the simulation. The following section describes how the parameters are determined at each level.

Level	Estimands / Simulated parameters	Estimates	Statistics
Site	True overall log odds ratio μ_i True sensitivity and specificity values se_i and sp_i	Uncorrupted log odds ratio μ_i^U and variance v_i^U Misclassified log odds ratio μ_i^M and variance v_i^M Matrix method corrected log odds ratio μ_i^{MM} Naively corrected variance v_i^{NC} Variance corrected variance v_i^{VC}	N/A
Meta-analysis	Overall odds ratio μ Data generating model $Model$ Between study variance in logit misclassification rates T	Misclassified pooled fixed and random effects estimates μ_F^M and μ_R^M and their variance v_F^M and v_R^M Uncorrupted pooled fixed and random effects estimates μ_F^U and μ_R^U and their variance v_F^U and v_R^U Naively corrected pooled fixed and random effects estimates μ_F^{NC} and μ_R^{NC} and their variance v_F^{NC} and v_R^{NC} Variance corrected pooled fixed and random effects estimates μ_F^{VC} and μ_R^{VC} and their variance v_F^{VC} and v_R^{VC}	Homogeneity statistics: Uncorrupted Q^U Misclassified Q^M Naively corrected Q^{NC} Proposed Generalized Cochran Q_h Greenland's corrected X_h^2

Table 5: Site-level and Meta-analysis level simulation parameters

4.3 Simulation Procedure

4.3.1 Site-Level

Every simulated meta-analysis consists of 8 sites. Let the number of simulated subjects in the k^{th} treatment group of the i^{th} site be denoted J_{ik}, where this value is taken from the Dormuth et al. study (see Appendix 2, Table 1). The disease status of the j^{th} individual in this treatment group, S_{jik}, is determined by a draw from a binomial distribution

$$S_{jik} \sim Bin(1, p_{ik}),$$

where p_{ik} is the probability of disease in the k^{th} treatment group of the i^{th} site. The probability of disease in the lower potency statin group, p_{i0}, is a fixed parameter taken from Dormuth et al. To derive the probability of disease in the higher potency statin group, p_{i1}, from the known parameters p_{i0} and study-specific log odds ratio μ_i, first note that the odds ratio $\varphi = e^{\mu_i}$ is given by

$$\varphi = \frac{p_{i1}(1 - p_{i0})}{p_{i0}(1 - p_{i1})}$$

Multiplying by p_{i0} lets us begin with the following simplified term:

$$\varphi * p_{i0} = \frac{p_{i1}(1 - p_{i0})}{(1 - p_{i1})}$$

Our goal is to determine an expression γ such that $(1 - p_{i1}) * \gamma = (1 - p_{i0})$. To do this, see that

$$(1 - p_{i0}) = (1 - p_{i0}) * (1 - p_{i1} + p_{i1})$$

$$= (1 - p_{i0}) * (1 - p_{i1}) + (1 - p_{i0}) * p_{i1}$$

Then γ is given by:

$$(1 - p_{i1}) * \gamma = (1 - p_{i0}) * (1 - p_{i1}) + (1 - p_{i0}) * p_{i1}$$

$$\gamma = (1 - p_{i0}) + \frac{(1 - p_{i0}) * p_{i1}}{1 - p_{i1}} = (1 - p_{i0}) + p_{i0} * \varphi$$

So, we have the following expression for p_{i1}:

$$p_{i1} = p_{i0} * \frac{e^{\mu_i}}{1 - p_{i0} + p_{i0} * e^{\mu_i}},$$

where the study-specific log odds-ratio $\mu_i = \mu$ if $Model = Fixed,$ or $\mu_i \sim N(\mu, \tau_{eff}^2)$ if $Model = Random.$ Choosing p_{i0} as a fixed parameter is necessary, as μ_i does not uniquely determine the probabilities of disease in both groups.

The simulated subjects are then partitioned into a contingency table U_i according to their disease status S_{jik} and their treatment group k. U_i therefore represents the true cross-classification of disease and exposure, from which the uncorrupted log odds ratio and its variance can be calculated:

$$U_i = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\mu_i^U = \log\left(\frac{ad}{bc}\right)$$

$$v_i^U = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

To corrupt U_i by misclassification and generate the misclassified contingency table M_i, first let $\theta^{se} = logit(0.9)$ and $\theta^{sp} = logit(0.92)$ denote the overall logit sensitivity and specificity. Then the study-specific logit sensitivity and logit sensitivity θ_i^{se} and θ_i^{sp} are stochastically determined:

$$\left(\theta_i^{se}, \theta_i^{sp}\right) \sim N\left((\theta^{se}, \theta^{sp}), T\right)$$

If se_i and sp_i are the value of the misclassifications rates corresponding to θ_i^{se} and θ_i^{sp}, then

$$M_i = \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} sens_i & 1 - sens_i \\ 1 - spec_i & spec_i \end{bmatrix}$$

This table yields the misclassified log odds ratio μ_i^M and its variance v_i^M.

Then, to yield the matrix method corrected log odds ratio μ_i^{MM}, we correct the corrupted table M_i using the expected study-specific sensitivity and specificity values se_{ovr} and sp_{ovr} determined from θ^{se} and θ^{sp} according to the procedure outlined in Section 2.7.1 to yield the corrected table MM_i:

$$MM_i = \left(\frac{1}{se_{ovr} + sp_{ovr} - 1}\right) * M_i * \begin{bmatrix} sp_{ovr} & se_{ovr} - 1 \\ sp_{ovr} - 1 & se_{ovr} \end{bmatrix}$$

The naively corrected variance v^{NC} is calculated using the standard formula $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$ using the corrected cell counts in MM_i.

Finally, the corrected variance v_i^{VC} is determined using the formula in Section 2.6.

Matrix method correction in this way can possibly introduce negative cell counts in the corrected contingency table. If this happens, new study-specific misclassification rates are drawn and the corruption/correction procedure is redone. This occurs until all cell counts are non-negative. The total number of iterations that generate negative cells counts will be reported in Appendix 4.

4.3.2 Meta-analysis Level

Once the study-specific data has been generated, the meta-analytic outcomes may be calculated. We use the usual formulas for calculating the fixed effects estimate for the uncorrupted, misclassified and naively corrected observations

$$\mu_F^+ = \sum_i w_i^+ \mu_i^+ / \sum_i w_i^+$$
$$w_i^+ = 1/v_i^+$$

where the " $+$ " indicates that this formula is used for these three observations.

Finally, the variance corrected fixed effects estimate μ_F^{VC} is calculated according to Greenland's procedure as described in Section 2.6.

The various homogeneity statistics are calculated as described in Section 2.7.1. Each are compared to a χ^2 distribution with 7 degrees of freedom.

An estimate of between study heterogeneity τ^2 is derived using the GMM procedure with weights corresponding to the type of observations used. For the

uncorrupted, misclassified, and naively corrected data, the random effects estimator μ_R^+ is calculated as usual

$$\mu_R^+ = \sum_i z_i^+ \mu_i^+ / \sum_i z_i^+$$
$$z_i = 1/(v_i^+ + \tau^2)$$

while the variance corrected random effects estimator is calculated as described in Section 2.7.1.

4.4 Performance Measures

Following Morris et al. (69), we use bias, coverage probability and empirical standard error to quantify the performance of an estimation procedure, and false negative and false positive rates to quantify the performance of the homogeneity test. For each of these measures, we calculate a corresponding Monte Carlo standard error, which represents the uncertainty about that measure due to the fact that only a finite number of iterations are run (69). Table 5 lays out the definition of these performance measures, their estimates, and the Monte Carlo standard error of these estimates. We suppress superscripts and subscripts, and use the "hat" notation to represent an estimated value, but recall that performance measures are calculated for all types of observations calculated at the meta-analysis level of the simulation.

Measure	Definition	Estimate	Monte Carlo SE of estimate
Bias	$E[\hat{\mu}] - \mu$	$\frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} \hat{\mu}_j - \mu$	$\sqrt{\frac{1}{n_{sim}(n_{sim}-1)} \sum_{j=1}^{n_{sim}} (\hat{\mu}_j - \bar{\mu})^2}$
Coverage	$\mathrm{P}(\hat{\mu}_{low} < \mu < \hat{\mu}_{high})$	$\frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} 1(\hat{\mu}_{low} < \mu < \hat{\mu}_{high})$	$\sqrt{\frac{\widehat{Cover}(1-\widehat{Cover})}{n_{sim}}}$
Empirical SE	$\sqrt{Var(\hat{\mu})}$	$\sqrt{\frac{1}{n_{sim}-1} \sum_{j=1}^{n_{sim}} (\hat{\mu}_j - \bar{\mu})^2}$	$\frac{\widehat{EmpSE}}{\sqrt{2(n_{sim}-1)}}$
False Negative Rate	$\Pr(\chi^2 < \alpha \mid Model = Random)$	$\frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} \mathbf{1}(\chi^2 < \alpha, Random)$	$\sqrt{\frac{\widehat{FNR}(1-\widehat{FNR})}{n_{sim}}}$
False Positive Rate	$\Pr(\chi^2 > \alpha \mid Model = Fixed)$	$\frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} \mathbf{1}(\chi^2 > \alpha, Fixed)$	$\sqrt{\frac{\widehat{FPR}(1-\widehat{FPR})}{n_{sim}}}$

Table 6: Performance measures used to evaluate and compare the naively corrected and variance corrected estimators. Notation: n_{sim} = number of iterations; $\hat{\mu}_{low}$, $\hat{\mu}_{high}$ = bounds of 95% CI; α = significance level of χ^2 homogeneity test.

4.5 Summary

In order to address objective 1, to investigate the problem of between study variance in misclassification rates in the presence of between study variance in effect size, we implement a random effects estimation procedure that corrects the study specific log odds ratio variances for variance in misclassification rates. The procedure is modeled on that proposed for fixed effects estimation presented in (5). We compare this variance correction procedure to a naïve procedure in which the study-specific

variances are calculated directly from their respective contingency tables, after matrix method correction, using Woolf's formula (63). The variance correction procedure relies upon a novel generalized Cochran between study variance statistic, and to meet objective 2, the performance of this novel statistic is compared to the corrected χ^2 homogeneity test statistic proposed in (5). To address objective 3, we compare the fixed effects estimation procedure proposed in (5) to a corresponding naïve correction procedure that does not perform any variance correction. Finally, The impact of incorrectly specifying the true distribution of underlying effect sizes is determined for both variance correction procedures, thereby meeting objective 4.

5 Results

5.1 Comparison of Variance Corrected and Naively Corrected Random Effects Estimators in Random Contexts

Figure 3 compares the bias, coverage probability, and empirical standard error, along with corresponding Monte Carlo standard errors, of the naively corrected, misclassified, and variance corrected random effects estimates in *Random* contexts. It is clear from this figure that our proposed variance correction procedure yields random effects estimates that are more biased than the naively corrected estimates. The difference in bias between the two estimator types grows as the overall odds ratio gets larger. When the overall OR is 1.11, the bias of the variance correted estimator in fact approaches that of the uncorrected estimator.

The relative performance in terms of coverage probability of the two estimation methods depends on the overall log odds ratio. When the overall OR is 1 or 1.11, the variance corrected estimator has a higher coverage probability, while the situation is reversed when the overall OR is 2.22 However, it is important to note that a higher coverage probability is not necessarily desirable. Ideally, the coverage probability would be equal to its nominal value of 95%, but the average coverage probability for the variance corrected estimate exceeds 95% when the overall OR is 1 and 1.11.

Somewhat surprisingly, the variance corrected random effects estimates tend to have lower empirical standard error than the naively corrected estimates. In the long run, then, the variance corrected estimates have greater precision. However, the

variance correction procedure produces much less accurate *estimates* of its own empirical standard error than does naïve correction. Table 7 shows the percentage by which the estimated standard error (derived from the meta-analytic variance estimator) differs from the empirical standard error for the random effects estimators in *Random* contexts. The variance correction procedure consistently overestimates the true standard error by between 50% and 735%, depending on the level of between study variance in misclassification rates but with no apparent relation to the overall odds ratio. In contrast, the naive correction procedure consistently underestimates its true standard error (in DGCs where misclassification variance $T \neq 0$), and the range of the magnitude of the underestimation for the naïve procedure depends on the overall odds ratio.

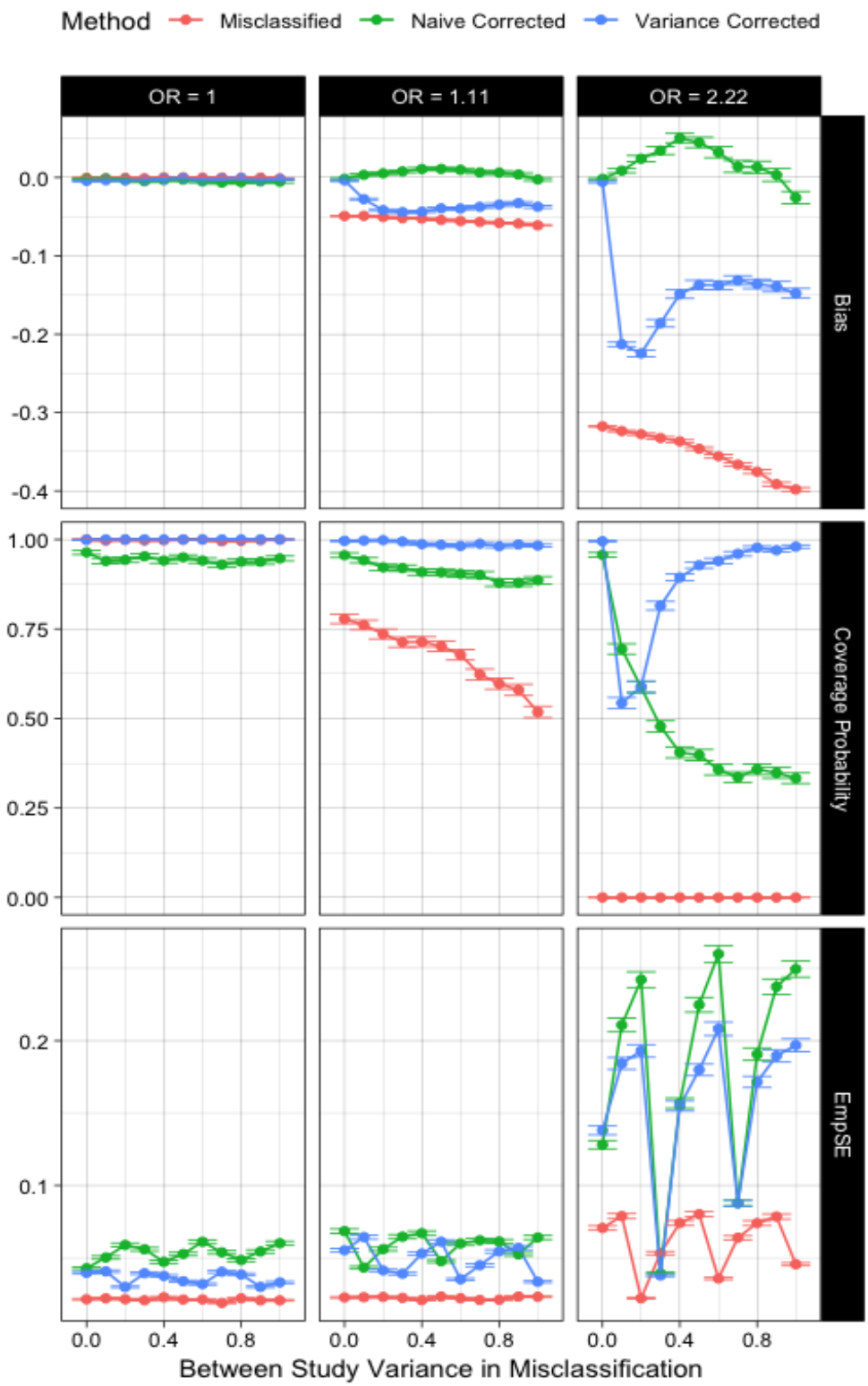


Figure 3: Bias, coverage probability, and empirical standard error for the misclassified, naively corrected, and variance corrected methods of random effects estimation in Random *contexts. Error bars represent Monte Carlo standard errors.*

T	OR = 1		OR = 1.11		OR = 2.22	
	Naively Corrected	Variance Corrected	Naively Corrected	Variance Corrected	Naively Corrected	Variance Corrected
0	7.39%	56.29%	5.46%	50.69%	8.22%	51.94%
0.1	-0.08%	96.91%	-1.9%	98.62%	-45.89%	87.52%
0.2	-0.7%	133.33%	-8.72%	148.84%	-60.31%	88.25%
0.3	-1.02%	204.13%	-11.47%	230.53%	-66.29%	196.95%
0.4	-5.14%	237.65%	-15.94%	308.49%	-71.35%	266.01%
0.5	-4.95%	288.22%	-19.02%	456.29%	-73.36%	367.24%
0.6	-10.8%	436.92%	-19.35%	544.62%	-75.13%	393.64%
0.7	-15.24%	439.94%	-17.22%	594.34%	-76.14%	489.04%
0.8	-14.52%	498.42%	-23.77%	727.77%	-76.6%	537.62%
0.9	-9.33%	561.79%	-24.33%	679.72%	-77.73%	614.37%
1	-6.3%	630.52%	-18.04%	734.93%	-77.28%	635.84%

Table 7: Percentage Error in estimated standard error relative to empirical standard error for the random effects estimators in Random contexts

5.2 Comparison of Variance Corrected and Naively Corrected Fixed Effects Estimators in Fixed Contexts

Similar results are observed for the variance corrected fixed effects estimator, which is equivalent to that proposed in (5). It is more biased and has greater standard error than the naively corrected fixed effects estimator. The superior performer with respect to coverage probability depends on the value of *T* and the overall OR. Figure 4 plots the performance measures for the naively corrected, misclassified, and variance corrected fixed effects estimates in *Fixed* contexts, and Figure 6 presents bias and empirical standard error in contexts with overall OR=1 and OR=1.15.

Table 8 shows the percentage error in estimated standard error relative to empirical standard error for the fixed effects estimators in *Fixed* contexts. Like the above results for *Random* contexts, the naïve correction procedure consistently underestimates true standard error, and, when the overall odds ratio is 1 or 1.15, the variance correction procedure overestimates true standard error. The error in the variance corrected standard error estimates is not as high as in *Random* contexts, however, and the true standard error is in fact underestimated when $T \geq 0.3$ and the overall odds ratio is 2.3.

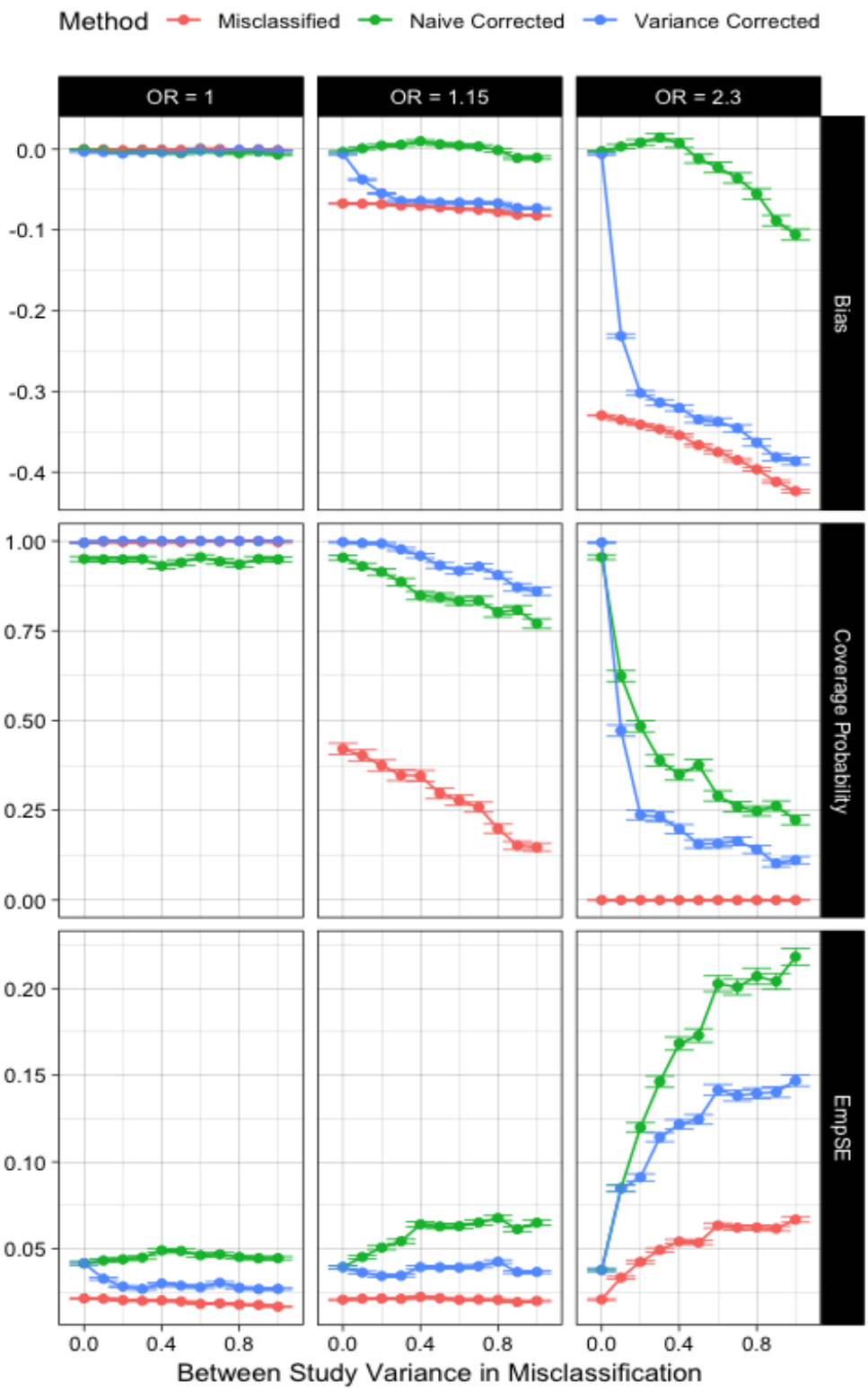


Figure 4: Bias, coverage probability, and empirical standard error of the misclassified, naively corrected, and variance corrected methods of fixed effects estimation in Fixed contexts. Error bars represent Monte Carlo standard error

T	OR = 1		OR = 1.15		OR = 2.3	
	Naively Corrected	Variance Corrected	Naively Corrected	Variance Corrected	Naively Corrected	Variance Corrected
0	-1.42%	52.46%	2.33%	57.04%	0.73%	51.28%
0.1	-3.55%	103.47%	-9.4%	102.28%	-54.71%	46.35%
0.2	-3.62%	144.71%	-18.09%	127.27%	-67.42%	29.41%
0.3	-5.22%	161.68%	-22.51%	133.23%	-72.7%	-3.55%
0.4	-11.7%	142.03%	-33.02%	106.62%	-76.06%	-15.47%
0.5	-10.91%	155.58%	-32.26%	102.26%	-76.75%	-21.65%
0.6	-6.39%	164.29%	-32.5%	102.8%	-79.96%	-34.68%
0.7	-9.21%	136.07%	-34.59%	95.68%	-79.95%	-33.38%
0.8	-7.31%	158.14%	-38.03%	82.05%	-80.64%	-36.14%
0.9	-6.51%	166.58%	-32.98%	102.73%	-80.82%	-38.85%
1	-8.03%	160.76%	-36.94%	106.16%	-82.12%	-40.67%

Table 8: Percentage error in estimated standard error relative to empirical standard error for the fixed effects estimators in Fixed contexts

5.3 Performance of Homogeneity Test Statistics

All homogeneity test statistics calculated in the simulation exhibit a high false negative rate, consistent with previous Monte Carlo simulation studies (70). The naively corrected, and generalized Cochran χ^2 statistics, as well as that proposed by in (5) are so high as to render them useless. On the other hand, all homogeneity test statistics other than the misclassified have a low false positive rate. Table 9 presents the average false positive and false negative rates of the various homogeneity test statistics over all values of *T*.

	Uncorrupted Q^U	Misclassified Q^M	Naïve Corrected Q^{NC}	Variance Corrected X_h^2	Generalized Cochran Q_h
FNR (mean)					
OR = 1	89.31%	65.47%	99.95%	100%	99.99%
OR = 1.11	89.47%	57.76%	99.96%	100%	99.96%
OR = 2.22	89.21%	11.54%	56.93%	96.75%	99.89%
FPR (mean)					
OR = 1	9.55%	33.12%	0.036%	0.0%	0.0%
OR = 1.15	9.83%	48.89%	0.0275	0.0%	0.073%
OR = 2.3	9.66%	89.93%	45.71%	3.63%	0.082%

Table 9: Mean false negative rate (FNR) and false positive rate (FPR) of homogeneity test statistics across levels of T *at a significance level of 0.1*

5.4 Impact of incorrect Model Choice

The consequence of an incorrect model choice stemming from an erroneous homogeneity test is the use a fixed effects approach in *Random* contexts, or vice-versa. We call this incorrect usage "off-model" estimation in contrast to "on-model" estimation, where the estimation approach (fixed or random effects) matches the underlying distribution of study-specific effect sizes.

Because qualitatively similar results are observed in all DGCs, Figures 5 and 6 compare the bias, coverage probability, and empirical standard error only for the *Fixed* and *Random* DGCs mimicking those of Dormuth et al., which have overall ORs of 1.15 and 1.11, respectively. Table 10 shows the relative percentage difference in

performance between the on-model and off-model estimators for all DGCs.

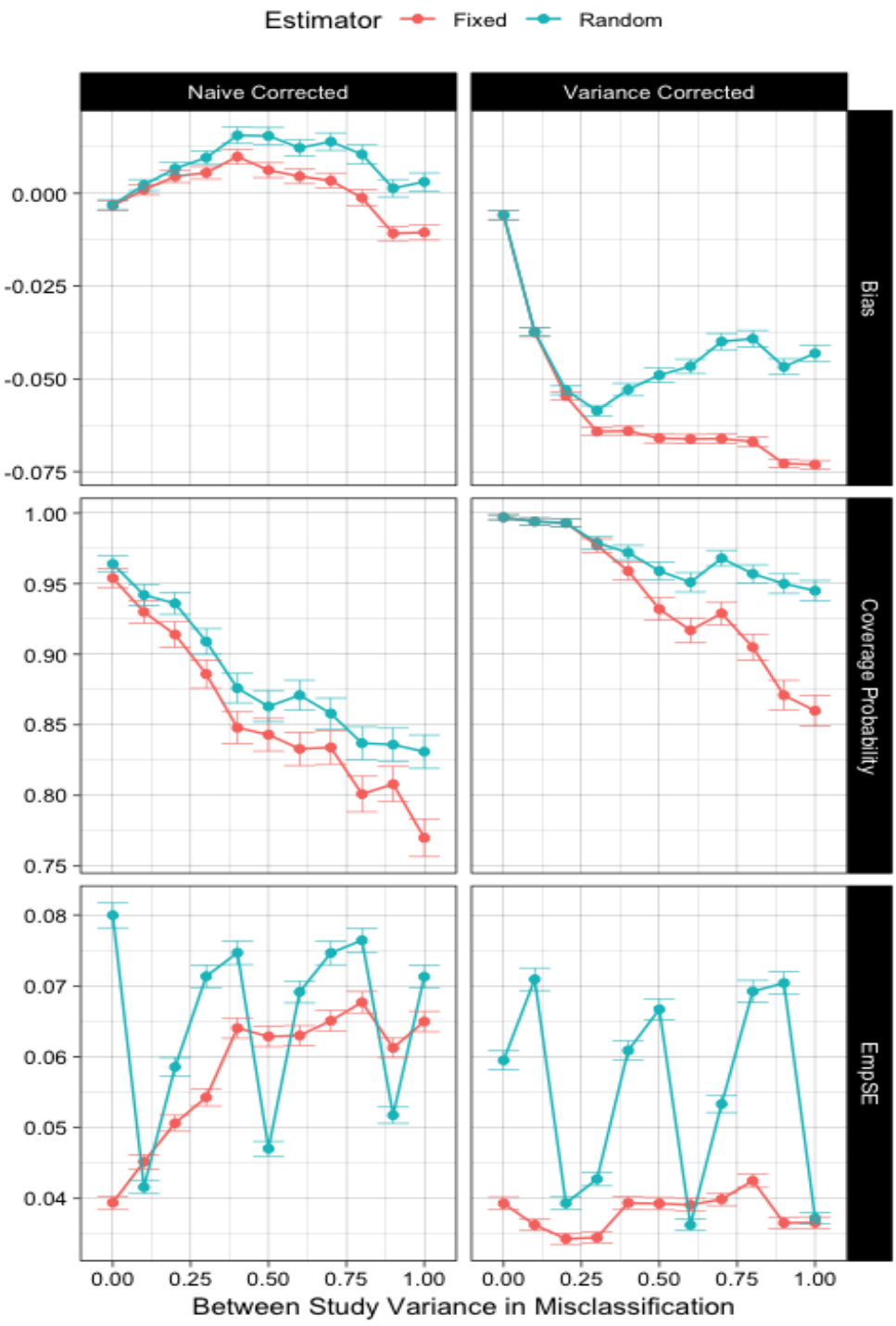


Figure 5: Bias, coverage probability, and empirical standard error for the fixed and random effects estimators corresponding to the naively corrected and variance corrected procedures in the Fixed *context with overall OR=1.15*

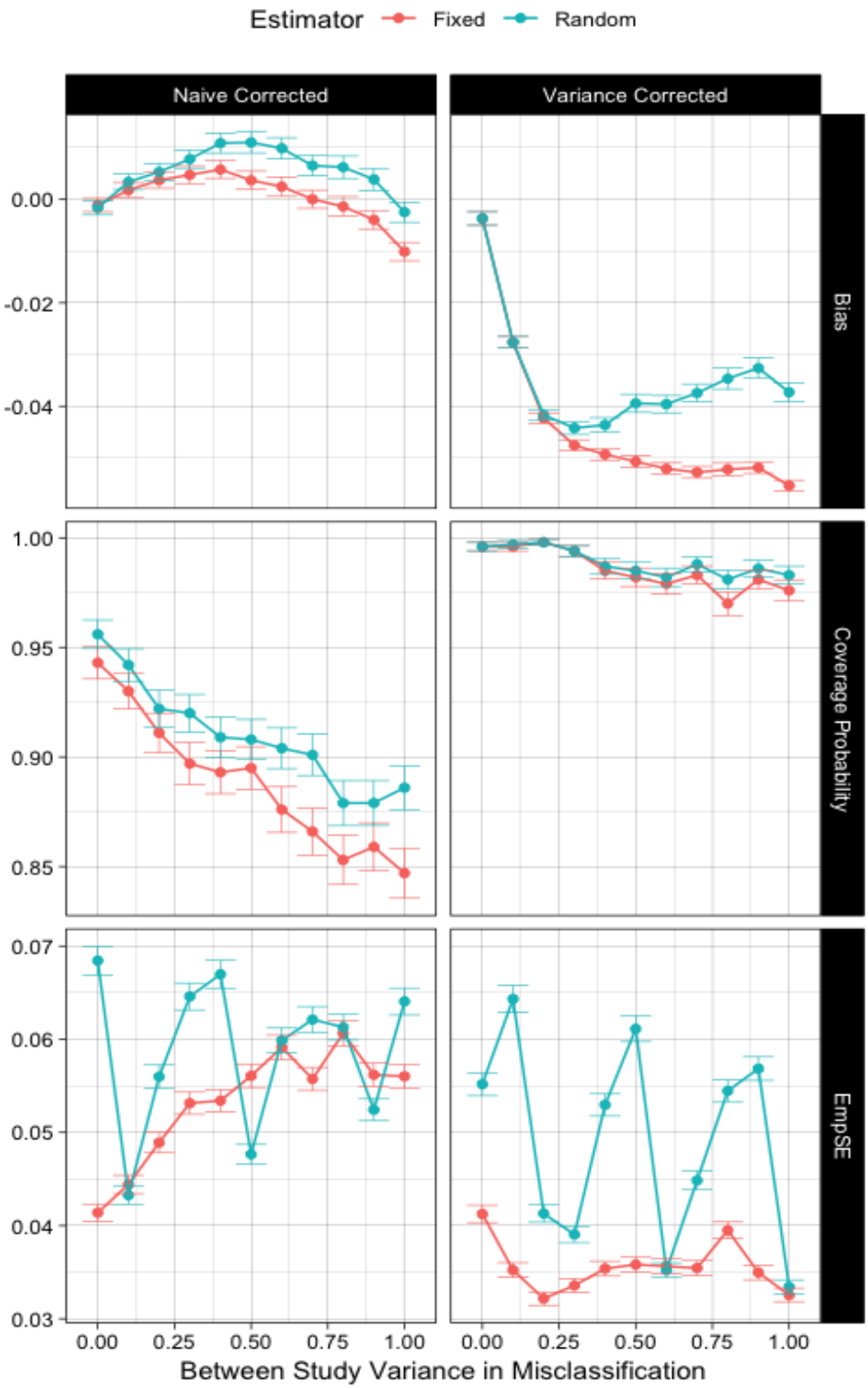


Figure 6: Bias, coverage probability, and empirical standard error for the fixed and random effects estimators corresponding to the naive correction and variance correction methods in the Random *context with overall OR=1.11*

	Fixed Contexts			Random Contexts		
	OR=1	OR=1.15	OR=2.3	OR=1	OR=1.11	OR=2.22
Bias						
VC	9.54%	-22.05%	-43.61%	-10.14%	25.52%	101.26%
NC	14.58%	140.5%	99.17%	-8.76%	18.95%	255.08%
Coverage						
VC	0.0%	3.37%	406.24%	0.0%	-0.34%	-65.0%
NC	0.48%	3.30%	19.08%	-0.82%	-2.38%	-17.11%
Emp SE						
VC	14.62%	44.57%	35.99%	-12.27%	-24.17%	-26.35%
NC	15.79%	11.53%	16.18%	-12.21%	-9.18%	-12.68%

Table 10: Relative percentage difference in performance measures between on-model and off-model estimators in Fixed *and* Random *effects contexts. Difference calculated as (off-model value)-(on-model value)/(on model value). VC = Variance Corrected, NC = Naively Corrected.*

6 Discussion

6.1 Summary of Results

In terms of estimation accuracy and precision, the naively corrected random effects estimator performs better than our proposed variance corrected random effects estimator. It has markedly lower bias and empirical standard error, and suffers none of the statistical issues involved with negative variance estimates. Further, the impact of an incorrect model specification is less severe with naïve correction, regardless of the true underlying model (*Fixed* or *Random*). However, under certain conditions, the coverage probability is higher for the variance corrected estimator than the naively corrected estimator. This is due to pathologically large variance estimates associated with the variance corrected estimator. Similar results are observed for the variance corrected fixed effects estimator. It too is more biased and less precise than its naïvely corrected counterpart.

The poor performance of the variance corrected estimators can be partially attributed to the multiple delta-method/Taylor series approximations involved in the variance correction procedure. The value of the Taylor series approximation of the expected study-specific misclassification rates, described in Section 2.7.1, is used as an input to the delta-method approximation of the covariance matrix of the corrected study-specific log odds ratios. Error therefore compounds upon error. Additionally, Taylor series approximations, being linear approximations, are known to be poor when the function being approximated is highly non-linear, and the logit transform of

sensitivity and specificity is highly non-linear when their values are close to 1. Because the overall sensitivity and specificity in the simulation are set at 0.9 and 0.92, respectively, a poor approximation may be unavoidable.

Further contributing to this poor performance is the fact that the corrected variance estimates depend on the corrected and uncorrected probabilities of disease in both exposure groups (see Section 2.7.2) . Correlation between study-specific effect estimates and study weights is known to cause bias in the pooled meta-analytic effect estimate (38,71,72). Appendix 5 contains an analysis of this correlation in our results. Linear correlation exists when there is no between study variance in misclassification rates. The introduction of misclassification variance removes the linear correlation, but a clear relationship remains. The more different from 0 the effect estimate, the less likely it is to receive a large weight.

Finally, generalized method of moments estimators for between study variance in effect size are often criticized for their reliance on the assumption that the study-specific variances are known exactly, an assumption never met in practice (73–75). Error in study weights will bias the estimate of between study variance in effect size. By using a delta-method approximation to correct study-specific variances, an extra source of error is introduced into the estimation of the true variance of the study-specific effect sizes, exacerbating this issue.

Other methods of estimating between study variance in effect size exist, and can outperform the Dersimonian and Laird estimator. See (6,72,76,77) for overviews and

comparisons of different between study variance estimators. Iterative approaches may be an especially fruitful area of future inquiry.

A different approach to modeling between study variance in misclassification rates could model the distribution of misclassification rates using a beta-distribution. Beta-distributions are well known to be the conjugate prior of binomial distributions (1). Using a beta-distribution would remove the need for a delta-method approximation of the mean and variance of sensitivity and specificity, as these can be derived exactly from the parameters of the beta-distribution. This could potentially render the corrected study-specific variance estimate more precise. Constructions of bivariate beta distributions exist that can model possible correlation between misclassification rates (2,3). The bivariate normal distribution of Reitsma and colleagues (4) was chosen in this thesis because it agrees with the most familiar meta-analytic assumptions – i.e. that data are normally distributed. It was my thinking that meta-analysis of validation studies would be most likely to use the techniques corresponding to the assumption of normality. However, techniques for meta-analysis of beta-distributed data do exist (5).

All homogeneity test statistics investigated in this project have very high false negative rates, making the detection of heterogeneity in effect size unlikely. Caution should therefore be used before using the outcome of a homogeneity test to inform a decision about whether to use a fixed or random effects estimator. The best course of action is likely to present the results of both the random and fixed effects estimation

regardless of the outcome of the homogeneity test and explore potentially sources of heterogeneity qualitatively (78,79).

While it has arguably become a commonplace to compare generalized Cochran Q statistics to a χ^2 distribution, this is based on a misunderstanding (38). A more accurate test would compare a Q statistic to a gamma distribution, as suggested in (80) when the effect measure is a log odds ratio. Beyond theoretical issues with the use of χ^2 homogeneity statistics, the extra error introduced into the estimated study-specific variances by the delta-method correction procedure just described will also adversely affect the ability of these statistics to detect heterogeneity. In fact, this is the fundamental reason the estimate of between study variance in effect size is biased when study-specific variances are not known exactly. Finally, our choice of between study variance in effect size, which is derived from Dormuth et al. is small – approximately 0.017. This likely also contributes to the high false negative rate of the homogeneity statistics.

The I^2 statistic represents a way of quantifying the amount of heterogeneity between studies included in a meta-analysis (6,7). It represents the *amount* (percentage) of variation between studies that is due to heterogeneity rather than chance. This is in contrast to the *Q* statistic, which, when used in a statistical test for heterogeneity, is meant to reveal only whether or not *there is* heterogeneity. The *Q* statistic therefore does not give a measure of the *extent* of heterogeneity. However, because $I^2 = 100\% * (Q - df)/Q$, where Q is Cochran's between study variance

statistic, and df is its degrees of freedom, it will also be affected by bias in the Q statistic induced by misclassification.

6.2 Contextualization

By showing that the naïve correction procedure – leaving study specific variances untouched – outperforms the variance correction procedure, we lay the groundwork for future quantitative analysis of misclassification-induced bias in the context of a meta-analysis. Investigators wishing to conduct a quantitative bias analysis in their meta-analysis can safely use the naïve procedure, which is simpler and more straightforward to implement. However, our results depend on a number of simplifying assumptions, and future work ought to consider more complex situations. The modified maximum likelihood approach to bias correction may yield better performance than the matrix method approach and warrants further investigation. Differential misclassification is common. Novel approaches to homogeneity testing may allow for better detection of heterogeneity in the presence of misclassification.

It is our hope that this project will encourage more widespread use of quantitative bias analysis in meta-analysis. Despite focusing on meta-analyses of administrative health data, our results are generalizable to other forms of meta-analysis that are susceptible to misclassification. A fruitful direction for future work is the investigation of methods to correct other forms of bias affecting epidemiological effect estimates, such as selection biases and other forms of information bias than misclassification.

6.3 Strengths

The main strength of this study is that it is an empirical investigation of sophisticated theoretical algebraic techniques for variance correction that shows their insufficiency under realistic conditions. While the numerous approximations used in the variance correction procedure may be theoretically justified, they fail upon empirical examination. We have thus demonstrated the importance of Monte Carlo simulation in the evaluation of proposed statistical techniques.

6.4 Limitations

A key limitation of this study is that we restricted ourselves to a fairly small set of controlled parameters – namely, only three overall odds ratios were used for each underlying data generating mechanism (e.g. *Fixed* or *Random*). Our choice of values for the misclassification variance was limited in three ways: we assumed that sensitivity and specificity had identical variances, that there was no correlation between sensitivity and specificity, and, finally, we chose equally spaced values for the variance between 0 and 1. Our motivation for the latter decision was to determine whether the estimation procedures investigated were robust to *increasing* misclassification variance. However, this may be at the expense of more realistic variances occurring near the lower end of this range.

All simulation studies must make certain concessions by limiting themselves to a small subset of the whole parameter space. A quote attributed to Patrick Royston in (69)

sums up the issue nicely: “Simulation studies reveal points of light on a landscape, but can not illuminate the entire landscape.”

6.5 Directions for Future Research

Future research should focus on determining a realistic prior distribution on the distribution of misclassification rates among epidemiological studies. Meta-analyses of validation studies can be consulted for this purpose. These may give a better representation of the range of between study variance in misclassification rates that could be used in a sensitivity analysis of misclassification in meta-analysis. This should include investigations of the case where sensitivity and specificity have different variances, and cases in which there is correlation between the two.

Another direction would involve the evaluation of a different effect estimator – the arcsine difference – for its performance in a meta-analysis with between study variance in misclassification rates. The arcsine difference has been proposed for use in meta-analysis to accommodate studies with cell counts of 0 (81). However, its main interest to us would be its variance properties. In particular, the study-specific variance of the arcsine difference is given by

$$\frac{1}{4 * (a + b)} + \frac{1}{4 * (c + d)}$$

where a, b, c, d are entries in a contingency table.

Notice that, unlike the variance of the log odds ratio, the variance of the arcsine difference does not depend on the individual cell counts, but only on the row margins $(a + b)$ and $(c + d)$. Because misclassification of disease outcome leaves the row margins constant (assuming there is no additional misclassification of exposure status), the variance of the arcsine difference will be unaffected by misclassification. This may greatly simplify any variance correction procedures.

7 Conclusion

Misclassification has the potential to severely bias an epidemiological effect estimate and render a study useless, or worse, misleading. A qualitative assessment of the appropriateness of a chosen case-identification algorithm should be complemented with a quantitative sensitivity analysis of bias due to misclassification when administrative health data is used. For organizations that prospectively plan meta-analyses of administrative health data with the explicit aim of informing regulatory decision-making, such a sensitivity analysis is crucial. When conducting a sensitivity analysis of a meta-analysis of administrative health data in which disease misclassification is expected, variance correction should not be performed. The simpler, naïve correction method performs much better, and should be sufficient for such purposes.

References

1. Psaty BM, Burke SP. Protecting the Health of the Public — Institute of Medicine Recommendations on Drug Safety. N Engl J Med. 2006;355(17):1753–5.

2. Quan H, Smith M, Bartlett-Esquilant G, Johansen H, Tu K, Lix L. Mining Administrative Health Databases to Advance Medical Science: Geographical Considerations and Untapped Potential in Canada. Can J Cardiol. 2012;28(2):152–4.

3. Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. Open Med [Internet]. 2012;6(4):e134. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23687528

4. Discharge Abstract Database metadata [Internet]. [cited 2019 Jul 10]. Available from: https://www.cihi.ca/en/discharge-abstract-database-metadata

5. Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. Stat Med. 1988;7(7):745–57.

6. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to Estimate the Between-Study Variance and Its Uncertainty in Meta-Analysis. Res Synth Methods. 2016;7(1):55–79.

7. Dersimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. Contemp Clin Trials. 2007;28(2):105–14.

8. Lucyk K, Tang K, Quan H. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. BMC Health Serv Res. 2017;17(1):1–10.

9. Lucyk K, Lu M, Sajobi T, Quan H. Administrative health data in Canada: Lessons from history. BMC Med Inform Decis Mak [Internet]. 2015;15(1):1–6. Available from: http://dx.doi.org/10.1186/s12911-015-0196-9

10. Cadarette SM, Wong L. An Introduction to Health Care Administrative Data. Can J Hosp Pharm [Internet]. 2015;68(3):232–7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4485511/pdf/cjhp-68-232.pdf

11. Walker RL, Hennessy DA, Johansen H, Sambell C, Lix L, Quan H. Implementation of ICD-10 in Canada: how has it impacted coded hospital discharge data? BMC Health Serv Res [Internet]. 2012 Jun 10;12:149. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22682405

12. International Classification of Diseases (ICD) Information Sheet.

13. Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes

cases from administrative data: a population-based validation study. BMC Health Serv Res. 2018;18(1):1–8.

14. Lix L, Yogendran M, Burchill C, Metge C, Mckeen N, Moore D, et al. Defining and validating chronic diseases: an administrative data approach. Vol. 2006. 2006. 197 p.

15. Van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol [Internet]. 2012;65(2):126–31. Available from: http://dx.doi.org/10.1016/j.jclinepi.2011.08.002

16. Curtis JR, Martin C, Saag KG, Patkar NM, Kramer J, Shatin D, et al. Confirmation of administrative claims-identified opportunistic infections and other serious potential adverse events associated with tumor necrosis factor α antagonists and disease-modifying antirheumatic drugs. Arthritis Care Res (Hoboken). 2007;57(2):343–6.

17. Claster S, Termuhlen A, Schrager SM, Wolfson JA, Iverson E. Pitfalls of using administrative data sets to describe clinical outcomes in sickle cell disease. Pediatr Blood Cancer [Internet]. 2013/08/26. 2013 Dec;60(12):1936–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24039121

18. Johnson EK, Nelson CP. Values and pitfalls of the use of administrative databases for outcomes assessment. J Urol [Internet]. 2013/04/20. 2013 Jul;190(1):17–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23608038

19. Savitz DA, Wellenius GA. Interpreting Epidemiologic Evidence: Connecting Research to Applications. Oxford University Press; 2016.

20. Lanes S, Brown JS, Haynes K, Pollack MF, Walker AM. Identifying health outcomes in healthcare databases. Pharmacoepidemiol Drug Saf [Internet]. 2015;24(10):1009–16. Available from: http://dx.doi.org/10.1002/pds.3856

21. Terris DD, Litaker DG, Koroukian SM. Health state information derived from secondary databases is affected by multiple sources of bias. J Clin Epidemiol. 2007;60(7):734–41.

22. Corallo AN, Croxford R, Goodman DC, Bryan EL, Srivastava D, Stukel TA. A systematic review of medical practice variation in OECD countries. Health Policy (New York). 2014;114(1):5–14.

23. Mercuri M, Gafni A. Medical practice variations: what the literature tells us (or does not) about what are warranted and unwarranted variations. J Eval Clin Pract [Internet]. 2011 Aug 1;17(4):671–7. Available from: https://doi.org/10.1111/j.1365-2753.2011.01689.x

24. Gagne JJ. Restrictive Reimbursement Policies: Bias Implications for Claims-Based Drug Safety Studies. Drug Saf [Internet]. 2014;37(10):771–6. Available from:

https://doi.org/10.1007/s40264-014-0220-5

25. Gamble J-M, Johnson JA, Majumdar SR, McAlister FA, Simpson SH, Eurich DT. Evaluating the introduction of a computerized prior-authorization system on the completeness of drug exposure data. Pharmacoepidemiol Drug Saf [Internet]. 2013 May 1;22(5):551–5. Available from: https://doi.org/10.1002/pds.3427

26. Filion KB, Eberg M, Ernst P. Confounding by drug formulary restriction in pharmacoepidemiologic research. Pharmacoepidemiol Drug Saf [Internet]. 2015/12/09. 2016 Mar;25(3):278–86. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26648236

27. Alberts JF, Sanderman R, Gerstenbluth I, van den Heuvel WJA. Sociocultural variations in help-seeking behavior for everyday symptoms and chronic disorders. Health Policy (New York) [Internet]. 1998;44(1):57–72. Available from: http://www.sciencedirect.com/science/article/pii/S0168851098000153

28. Chomi EN, Mujinja PGM, Enemark U, Hansen K, Kiwara AD. Health care seeking behaviour and utilisation in a multiple health insurance system: does insurance affiliation matter? Int J Equity Health. 2014;13(1):25.

29. Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. BMC Fam Pract. 2016;17(1):38.

30. Birkmeyer JD, Sharp SM, Finlayson SRG, Fisher ES, Wennberg JE. Variation profiles of common surgical procedures. Surgery. 1998;124(5):917–23.

31. Wennberg J, Gittelsohn A. Small Area Variations in Health Care Delivery. Science (80-) [Internet]. 1973 Dec 14;182(4117):1102 LP – 1108. Available from: http://science.sciencemag.org/content/182/4117/1102.abstract

32. Stephens MJ, Finnell JT, Simonaitis L, Overhage JM. Variability in drug formularies and implications in decision support. AMIA . Annu Symp proceedings AMIA Symp [Internet]. 2011/10/22. 2011;2011:1327–36. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22195194

33. Morgan S, Hanley G, Raymond C, Blais R. Breadth, Depth and Agreement among Provincial Formularies in Canada. Healthc Policy [Internet]. 2009 May;4(4):e162–84. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20436800

34. Haidich AB. Meta-analysis in medical research. Hippokratia [Internet]. 2010 Dec;14(Suppl 1):29–37. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21487488

35. Jackson D, Turner R. Power analysis for random-effects meta-analysis. Res Synth Methods. 2017;8(3):290–302.

36. Barili F, Parolari A, Kappetein PA, Freemantle N. Statistical Primer: heterogeneity,

random- or fixed-effects model analyses?†. Interact Cardiovasc Thorac Surg [Internet]. 2018 Jun 2;27(3):317–21. Available from: https://doi.org/10.1093/icvts/ivy163

37. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002 Jun;21(11):1539–58.

38. Hoaglin DC. Misunderstandings about Q and 'Cochran's Q test'' in meta-analysis.' Stat Med [Internet]. 2016 Feb 20;35(4):485–95. Available from: https://doi.org/10.1002/sim.6632

39. Platt RW, Chateau D, Dormuth CR, Filion K. Observational Studies of Drug Safety in Multi-Database Studies: Methodological Challenges and Opportunities. eGEMs (Generating Evid Methods to Improv patient outcomes) [Internet]. 2016 May;4(1):1221. Available from: http://repository.edm-forum.org/egems/vol4/iss1/9

40. Azoulay L, Filion KB, Platt RW, Dahl M, Dormuth CR, Clemens KK, et al. Association Between Incretin-Based Drugs and the Risk of Acute PancreatitisIncretin-Based Drugs and the Risk of Acute PancreatitisIncretin-Based Drugs and the Risk of Acute Pancreatitis. JAMA Intern Med [Internet]. 2016 Oct 1;176(10):1464–73. Available from: https://doi.org/10.1001/jamainternmed.2016.1522

41. Renoux C, Dell'Aniello S, Khairy P, Marras C, Bugden S, Turin TC, et al. Ventricular tachyarrhythmia and sudden cardiac death with domperidone use in Parkinson's disease. Br J Clin Pharmacol [Internet]. 2016/05/19. 2016 Aug;82(2):461–72. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27062307

42. Henry D, Dormuth C, Winquist B, Carney G, Bugden S, Teare G, et al. Occurrence of pregnancy and pregnancy outcomes during isotretinoin therapy. CMAJ [Internet]. 2016/04/25. 2016 Jul 12;188(10):723–30. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27114489

43. Renoux C, Lix LM, Patenaude V, Bresee LC, Paterson JM, Lafrance J-P, et al. Serotonin-Norepinephrine Reuptake Inhibitors and the Risk of AKI: A Cohort Study of Eight Administrative Databases and Meta-Analysis. Clin J Am Soc Nephrol [Internet]. 2015/07/31. 2015 Oct 7;10(10):1716–22. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26231193

44. Gomes T, Paterson JM, Mukati M, Henry D, investigators CN for ODES. Retrospective analysis of trends in dispensing long-acting non-tamper-resistant oxycodone near the Canada-United States border. C open [Internet]. 2015 Apr 2;3(2):E231–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26393233

45. Dormuth CR, Filion KB, Paterson JM, James MT, Teare GF, Raymond CB, et al. Higher potency statins and the risk of new diabetes: multicentre, observational study of administrative databases. Vol. 348, The BMJ. 2014.

46. Barron BA. The Effects of Misclassification on the Estimation of Relative Risk. Biometrics. 1977;33(2):414–8.

47. Tenenbein A. A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications. J Am Stat Assoc. 1970;65(331):1350–61.

48. Neuhaus JM. Bias and efficiency loss due to misclassified responses in binary regression. Biometrika. 1999;86(4):843–55.

49. Vandenbroucke JP. When are observational studies as credible as randomised trials? Lancet. 2004;363(9422):1728–31.

50. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002;359(9302):248–52.

51. Lash Timothy L, Fink Aliza K. Semi-Automated Sensitivity Analysis to Assess Systematic Errors in Observational Data. Epidemiology. 2003;14(4):451–8.

52. Greenland S. Multiple-Bias Modelling for Analysis of Observational Data. J R Stat Soc Ser A (Statistics Soc. 2005;168(2):267–306.

53. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. Springer Science & Business Media; 2011.

54. Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L. The many faces of diabetes: A disease with increasing heterogeneity. Lancet [Internet]. 2014;383(9922):1084–94. Available from: http://dx.doi.org/10.1016/S0140-6736(13)62219-9

55. Gamble J-M, Butalia S. Medical practice variations in diabetes mellitus. Med Pract Var. 2016;1–40.

56. Powell CK, Hill EG, Clancy DE. The Relationship Between Health Literacy and Diabetes Knowledge and Readiness to Take Health Actions. Diabetes Educ. 2007;33(1):144–51.

57. Al Sayah F, Williams B. An integrated model of health literacy using diabetes as an exemplar. Can J Diabetes. 2012;36(1):27–31.

58. Gazmararian JA, Williams M V, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Educ Couns. 2003;51(3):267–75.

59. May M, Schindler C. Clinically and pharmacologically relevant interactions of antidiabetic drugs. Ther Adv Endocrinol Metab. 2016/03/31. 2016 Apr;7(2):69–83.

60. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care. 2002;25(3):512.

61. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH.

Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol [Internet]. 2005;58(10):982–90. Available from: http://www.sciencedirect.com/science/article/pii/S0895435605001629

62. Hendeby G, Gustafsson F. On nonlinear transformations of Gaussian distributions. Tech Rep from Autom Control Link? pings Univ. 2007;

63. Woolf B. ON ESTIMATING THE RELATION BETWEEN BLOOD GROUP AND DISEASE. Ann Hum Genet. 1955;19(4):251–3.

64. Morrissey MJ, Spiegelman D. Matrix Methods for Estimating Odds Ratios with Misclassified Exposure Data: Extensions and Comparisons. Biometrics. 1999;55(2):338–44.

65. Jonsson Funk M, Landi S. Misclassification in Administrative Claims Data: Quantifying the Impact on Treatment Effect Estimates. Curr Epidemiol Reports. 2014;1(4):175–85.

66. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. Int J Epidemiol. 2014;43(6):1969–85.

67. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ [Internet]. 2003 Sep;327(7414). Available from: http://bmj.com/content/327/7414/557.full.pdf

68. Borenstein M, Hedges L V, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97–111.

69. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med [Internet]. 2019 May 20;38(11):2074–102. Available from: https://doi.org/10.1002/sim.8086

70. Sánchez-Meca J, Marín-Martínez F. Homogeneity tests in meta-analysis: a Monte Carlo comparison of statistical power and Type I error. Int J Methodol. 1997;31(4):385–99.

71. Shuster JJ. Empirical vs natural weighting in random effects meta-analysis. Stat Med [Internet]. 2010 May 30;29(12):1259–65. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19475538

72. van Aert RCM, Jackson D. Multistep estimators of the between-study variance: The relationship with the Paule-Mandel estimator. Stat Med [Internet]. 2018/04/26. 2018 Jul 30;37(17):2616–29. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29700839

73. Böhning D, Sarol Jr J. Estimating risk difference in multicenter studies under baseline-risk heterogeneity. Biometrics. 2000;56(1):304–8.

74. Böhning D, Malzahn U, Dietz E, Schlattmann P, Viwatwongkasem C, Biggeri A. Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. Biostatistics. 2002;3(4):445–57.

75. Novianti PW, Roes KCB, van der Tweel I. Estimation of between-trial variance in sequential meta-analyses: A simulation study. Contemp Clin Trials [Internet]. 2014;37(1):129–38. Available from: http://www.sciencedirect.com/science/article/pii/S1551714413001948

76. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Res Synth Methods [Internet]. 2019 Mar 1;10(1):83–98. Available from: https://doi.org/10.1002/jrsm.1316

77. Petropoulou M, Mavridis D. A comparison of 20 heterogeneity variance estimators in statistical synthesis of results from studies: a simulation study. Stat Med [Internet]. 2017 Nov 30;36(27):4266–80. Available from: https://doi.org/10.1002/sim.7431

78. Song F, Sheldon TA, Sutton AJ, Abrams KR, Jones DR. Methods for Exploring Heterogeneity in Meta-Analysis. Eval Health Prof [Internet]. 2001 Jun 1;24(2):126–51. Available from: https://doi.org/10.1177/016327870102400203

79. Petitti DB. Approaches to heterogeneity in meta-analysis. Stat Med. 2001;20(23):3625–33.

80. Kulinskaya E, Dollinger MB. An accurate test for homogeneity of odds ratios based on Cochran's Q-statistic. BMC Med Res Methodol. 2015;15(1):49.

81. Rücker G, Schwarzer G, Carpenter J, Olkin I. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. Stat Med. 2009;28(5):721–38.

Appendices

Appendix 1: Baseline Characteristics of study population from Dormuth et al.

Baseline characteristics of the overall study population (patients newly prescribed higher potency statins and those prescribed lower potency statins during 1997-2011), matched and unmatched on propensity score. Values are numbers (percentages) of patients

	Matched cohort*		Full cohort	
Characteristic	Higher potency statins (n=30 843)	Lower potency statins (n=30 843)	Higher potency statins (n=89 077)	Lower potency statins (n=47 889)
Age (years):				
40-49	2 449 (7.9)	2 451 (7.9)	7 455 (8.4)	3 410 (7.1)
50-59	5 834 (18.9)	5 837 (18.9)	16 057 (18.0)	7 783 (16.3)
60-64	3 852 (12.5)	3 846 (12.5)	9 382 (10.5)	5 087 (10.6)
65-69	3 415 (11.1)	3 406 (11.0)	11 527 (12.9)	6 386 (13.3)
70-74	4 350 (14.1)	4 373 (14.2)	13 305 (14.9)	7 832 (16.4)
75-79	4 369 (14.2)	4 343 (14.1)	12 745 (14.3)	7 411 (15.5)
≥80	6 574 (21.3)	6 587 (21.4)	18 606 (20.9)	9 980 (20.8)
Sex:				
Women	11 926 (38.7)	11 926 (38.7)	32 504 (36.5)	18 598 (38.8)
Men	18 917 (61.3)	18 917 (61.3)	56 573 (63.5)	29 291 (61.2)
Cohort defining events and procedures:				
Myocardial infarction	18 628 (60.4)	18 629 (60.4)	60 751 (68.2)	29 155 (60.9)
Stroke	7 653 (24.8)	7 646 (24.8)	15 668 (17.6)	11 124 (23.2)
CABG	3 032 (9.8)	3 042 (9.9)	7 702 (8.6)	4 970 (10.4)
PCTA	3 564 (11.6)	3 564 (11.6)	15 237 (17.1)	5 859 (12.2)
Diagnoses:				
Hypertensive disease	17 636 (57.2)	17 328 (56.2)	49 010 (55.0)	25 991 (54.3)
Hypercholesterolemia	9 741 (31.6)	9 700 (31.4)	25 646 (28.8)	15 614 (32.6)
Peripheral vascular disease	1 111 (3.6)	1 125 (3.6)	2 633 (3.0)	1 794 (3.7)
Congestive heart failure	5 472 (17.7)	5 477 (17.8)	14 523 (16.3)	8 400 (17.5)
Injury or poisoning	12 977 (42.1)	12 968 (42.0)	37 601 (42.2)	17 691 (36.9)
No of hospitalisations:				
0	24 865 (80.6)	24 962 (80.9)	70 142 (78.7)	37 656 (78.6)
1	4 095 (13.3)	4 080 (13.2)	12 329 (13.8)	6 962 (14.5)
2	1 259 (4.1)	1 193 (3.9)	4 422 (5.0)	2 125 (4.4)
≥3	624 (2.0)	608 (2.0)	2 184 (2.5)	1 146 (2.4)
Drugs:				
Prescription acetaminophen	9 604 (31.1)	9 504 (30.8)	23 413 (26.3)	13 577 (28.4)
Prescription NSAID	7 289 (23.6)	7 187 (23.3)	19 802 (22.2)	11 610 (24.2)
ACE inhibitor	18 467 (59.9)	18 377 (59.6)	55 084 (61.8)	27 632 (57.7)
Angiotensin II receptor blocker	3 523 (11.4)	3 330 (10.8)	10 850 (12.2)	4 631 (9.7)
Thiazide diuretics	5 688 (18.4)	5 486 (17.8)	15 096 (16.9)	8 134 (17.0)
Loop diuretics	5 690 (18.4)	5 615 (18.2)	14 593 (16.4)	8 714 (18.2)
Potassium sparing diuretics	2 212 (7.2)	2 153 (7.0)	5 747 (6.5)	3 453 (7.2)
Other diuretics	1 096 (3.6)	1 060 (3.4)	2 928 (3.3)	1 775 (3.7)
β blockers	21 658 (70.2)	21 509 (69.7)	63 559 (71.4)	33 531 (70.0)
Calcium channel blockers	8 526 (27.6)	8 295 (26.9)	22 386 (25.1)	13 701 (28.6)
Antibiotics	13 248 (43.0)	13 034 (42.3)	36 303 (40.8)	19 835 (41.4)

CABG= coronary artery bypass graft, PCTA=percutaneous coronary intervention, NSAID=non-steroidal anti-inflammatory drug, ACE=angiotensin converting enzyme.

*Matching was to allow comparison of patients conditional on propensity score.

Table 11: Baseline characteristics of study population in Dormuth et al. 2014 (reproduced from Dormuth et al. 2014)

Appendix 2: Fixed Parameters

Parameter Name	Level of Simulation	Value(s)		
Probability of disease in unexposed / lower potency statin group	Study-specific	Alberta		0.114
		CPRD:		0.088
		Manitoba:		0.095
		Marketscan:		0.088
		Nova Scotia:		0.126
		Ontario:		0.082
		Quebec:		0.086
		Saskatchewan:		0.100
Row Margins	Study Specific	Alberta	Exposed:	1034
			Unexposed:	599
		CPRD	Exposed:	2513
			Unexposed:	1167
		Manitoba	Exposed:	1684
			Unexposed:	494
		Marketscan	Exposed:	5154
			Unexposed:	2033
		Nova Scotia	Exposed:	239
			Unexposed:	143

Parameter Name	Level of Simulation	Value(s)		
Row Margins	Study Specific	Ontario	Exposed:	6871
			Unexposed:	2894
		Quebec	Exposed:	5188
			Unexposed:	3035
		Saskatchewan	Exposed:	1773
			Unexposed:	420
Number of Sites	Meta-analysis Level	8		
Between study variance in effect size	Meta-analysis Level	$\chi^2 = 13.32 \Rightarrow \tau^2_{eff} = .016$		
Overall sensitivity and specificity	Meta-analysis Level	Sensitivity		90%
		Specificity		92%

Table 12: Description of fixed parameters

Appendix 3: Delta Method Approximation of Variance of Sensitivity and Specificity from Between Study Variance in Logit Sensitivity and Logit Specificity

Assume that the study-specific logit sensitivity and logit specificity, θ_i^{se} and θ_i^{sp}, are bivariately normally distributed about an overall mean vector $(\theta^{se}, \theta^{sp})$:

$$\left(\theta_i^{se}, \theta_i^{sp}\right) \sim N\left((\theta^{se}, \theta^{sp}), T\right),$$

where $T = \begin{bmatrix} \sigma_{se}^2 & 0 \\ 0 & \sigma_{sp}^2 \end{bmatrix}$ – i.e. we are assuming no correlation between θ_i^{se} and θ_i^{sp}.

We assumed no correlation between θ_i^{se} and θ_i^{sp} so, $Var[sens]$ can be derived from the between-study variance in logit sensitivity, σ_{se}^2, by an application of the delta method as follows:

$$\theta^{se} = \text{logit}(\text{sens}) \rightarrow sens = g(\theta^{se}) = \frac{e^{\theta^{se}}}{1 + e^{\theta^{se}}}$$

$$Var[sens] = \sigma_{se}^2 * [g'(\theta^{se})]^2$$
$$Var[sens] = \sigma_{se}^2 \left(se(1 - se)\right)^2$$

A similar procedure can be performed for $Var[spec]$.

Appendix 4: Number of Negative Cell Counts

	Number of Iterations with Negative Cell Counts

T	OR = 1	OR = 1.11	OR = 2.22
0	0	0	0
0.1	0	0	0
0.2	4	3	7
0.3	33	28	28
0.4	102	106	94
0.5	306	233	221
0.6	533	456	459
0.7	796	765	751
0.8	1168	1042	1063
0.9	1621	1547	1374
1	2025	1886	1788

Table 13: Number of iterations in each Random *DGC for which negative cell counts were observed in at least one site after matrix method correction.*

T	Number of Iterations with Negative Cell Counts		
	OR = 1	OR = 1.15	OR = 2.3
0	0	0	0
0.1	0	0	0
0.2	5	4	3
0.3	36	37	32
0.4	129	104	105
0.5	277	254	254
0.6	516	414	446
0.7	809	729	712
0.8	1287	1019	1053
0.9	1610	1369	1397
1	2121	1781	1833

Table 14: Number of iterations in each Fixed *DGC for which negative cell counts were observed in at least one site after matrix method correction*

Appendix 5: Correlation Between Study Weights and Corrected Site specific Effect Estimates

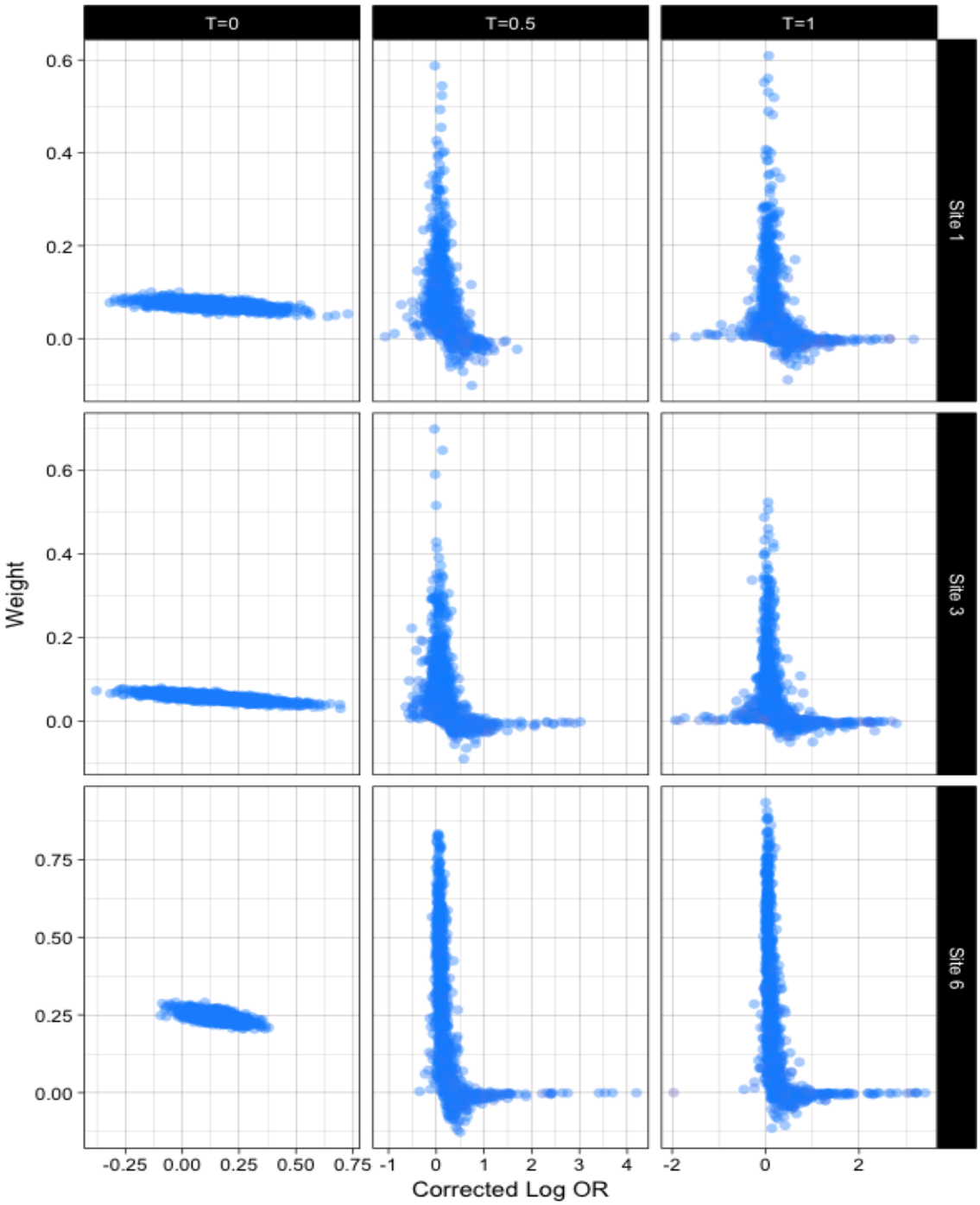


Figure 7: Correlation between fixed effects weights and corrected fixed effects estimates in three studies at three different levels of between study variance in misclassification rates T, when the context is Fixed*. Note the linear correlation when T=0. When $T \neq 0$, corrected log ORs closer to 0 are more likely to receive high weight than those further away.*

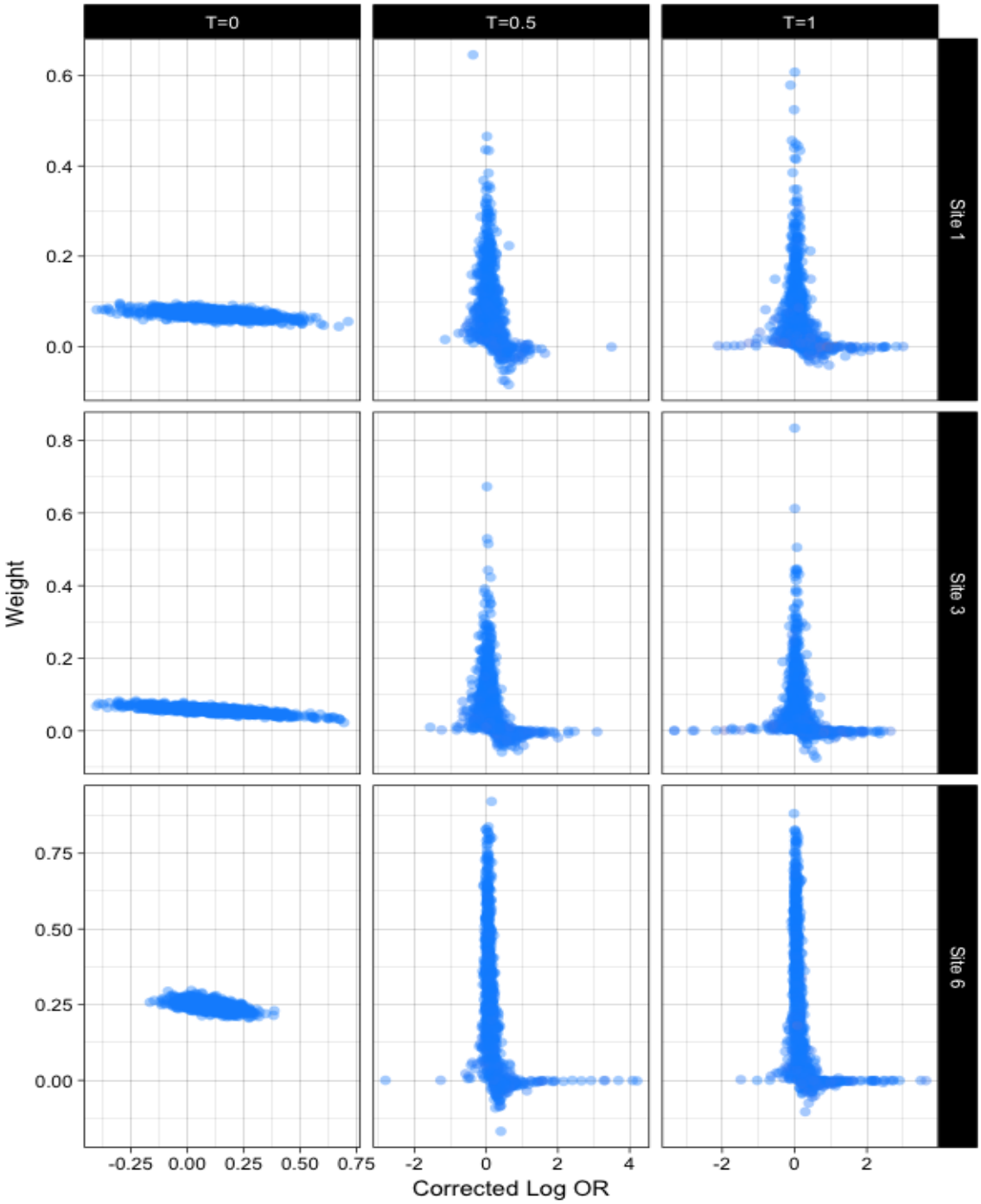


Figure 8: Correlation between random effects weights and corrected study-specific effect estimates in three studies at three different levels of between study variance in misclassification rates T, when the context is Random. *Note the linear correlation when T=0. When $T \neq 0$, corrected log ORs closer to 0 are more likely to receive high weight than those further away.*

Appendix 6: Simulation Code

```
library(dplyr)
library(ggplot2)

do_fixed = TRUE
do_random = TRUE

####
####
#### ---------------------------------- GENERATE FIXED PARAMETER VALUES AND DATA STRUCTURES ------------------------------
####
####

#---------------------- Create Set of Controlled Variable Parameters
trueLogORs_fixed = log(c(1,1.15,2.3))
trueLogORs_random = log(c(1, 1.11, 2.22))

miscBetweenStudyVariances = seq(from=0, to=1, by=0.1)

variableParameters_fixed = expand.grid(x=trueLogORs_fixed, y=miscBetweenStudyVariances)
variableParameters_random = expand.grid(x=trueLogORs_random, y=miscBetweenStudyVariances)
names(variableParameters_fixed) <- c("logOR", "T")
names(variableParameters_random) <- c("logOR", "T")


#----------------- Specify Fixed Parameters
#All taken from Dormuth's forest plot
casesExposed = c(90, 247, 170, 502, 23, 675, 507, 188)
casesUnexposed = c(68, 103, 47, 180, 18, 236, 260, 42)
numUnexposed = c(599, 1167, 494, 2033, 143, 2894, 3035, 420)
numExposed = c(1034, 2513, 1684, 5154, 239, 6871, 5188, 1773)
baselineProb = casesUnexposed/numUnexposed

#calculate the study-specific variance of logORs. Use the confidence intervals and estimates provided in Dormuth et
al's forest plot
lbs = c(.44, .87, .85, .94, .24, 1.08, 1, .67)
ubs = c(.98, 1.57, 1.88, 1.34, 1.21, 1.53, 1.46, 1.61)
est = log(c(.66, 1.17, 1.27, 1.12, .54, 1.29, 1.21, 1.04))

variances = ((est - log(lbs))/1.96)^2
w = 1/variances
#Dormuth et al report a Q of 13.32, and has 7 df
tauSqr = (13.32 - 7)/(sum(w) - (sum(w^2)/sum(w)))

estimatedSensitivity = 0.9
estimatedSpecificity = 0.92

overallSensitivity = 0.9
overallSpecificity = 0.92

overallLogitSensitivity = log(overallSensitivity/(1-overallSensitivity))
overallLogitSpecificity = log(overallSpecificity/(1-overallSpecificity))

numMAs = 1000                  #Needs to be integer. controls how many instances are simulated

numSites = rep(8, numMAs)          #Here, constant for every MA instance
```

```
numValidationSites = rep(1, numMAs)   #Here, constant for every MA instance
siteNames = c("Alberta", "CPRD", "Manitoba", "MarketScan","NS", "Ontario","Quebec","Saskatchewan")

QSigLevel = 0.1
significanceLevel = 0.05


#------------------- Generate empty data structures
hom.df       = data.frame(corrupted=numeric(numMAs),
                     corrected=numeric(numMAs),
                     uncorrupted=numeric(numMAs),
                     gCorrected=numeric(numMAs),
                     genCochQ=numeric(numMAs))

df           = data.frame(corrupted=numeric(numMAs),
                    corrected=numeric(numMAs),
                    uncorrupted=numeric(numMAs),
                    gCorrected=numeric(numMAs))

siteData.df = data.frame(matrix(0, ncol=length(siteNames), nrow=numMAs)); names(siteData.df) = siteNames

siteLogORData = list(corrupted = siteData.df,
                corrected = siteData.df,
                uncorrupted = siteData.df)

weightData    = list(corrupted=siteData.df,
                corrected=siteData.df,
                uncorrupted=siteData.df,
                gCorrected=siteData.df)

contextStruct =   list(FP_Q              = hom.df,
                  FN_Q              = hom.df,
                  homStats          = hom.df,
                  coveredTrueLogORFE  = df,
                  coveredTrueLogORRE  = df,
                  falseNegativesRE    = df,
                  falseNegativesFE    = df,
                  varFE             = df,
                  varRE             = df,
                  FE                = df,
                  RE                = df,
                  t2                = df,
                  numNegCellInstances = df,
                  FEWeights         = weightData,
                  REWeights         = weightData,
                  siteLogORs        = siteLogORData,
                  siteSens          = siteData.df,
                  siteSpec          = siteData.df
                  )

contextData_fixed = rep(list(contextStruct), nrow(variableParameters_fixed))
contextData_random = rep(list(contextStruct), nrow(variableParameters_random))

unlogit = function(x) {return(exp(x)/(1+exp(x)))}

####
####
####---------------------------- FUNCTIONS FOR GENERATING SITE-LEVEL DATA -------------------------
```

```
####
####

createSite = function(siteId, numUnexposed, numExposed, overallLogitSensitivity, overallLogitSpecificity, miscT,
                trueLogOR, baselineProb = NA, overallSens, overallSpec)
{
 pDiseaseUnexposed = baselineProb


 pDiseaseExposed = pDiseaseUnexposed * (exp(trueLogOR) / (1 - pDiseaseUnexposed + pDiseaseUnexposed *
exp(trueLogOR)))


 #Generate the vector of true observed diseased statuses in each group
 trueStatusExp         = rbinom(numExposed, 1, pDiseaseExposed)
 trueStatusUnexp       = rbinom(numUnexposed, 1, pDiseaseUnexposed)

 site = list()

 a = sum(trueStatusExp)
 b = numExposed - a
 c = sum(trueStatusUnexp)
 d = numUnexposed - c

 site$uncorruptedTable = matrix(c(a,b,c,d), nrow=2, ncol=2, byrow=TRUE)

 #The expected value of the site-specific sensitivity and specificity is not simply the inverse logit of the overall
 #logit misclassification rates. E[sens_i] != unlogit(trueOverallLogitSensitivity) - i.e. E[sens_i] != .9, and similarly
 #for E[spec_i]. Therefore, we use as a correcting sensitivity a taylor approximation to the mean:
 #E[sens_i] = unlogit(trueOverallLogitSensitivity) + .5*unlogit''(trueOverallSensitivity)*Var[trueOverallLogitSensitivity]
 #where unlogit'' is the second derivative of the inverse logit transform. Similar for E[spec_i]
 correctingSens = overallSens + .5*overallSens*(overallSens-1)*(2*overallSens-1)*miscT
 correctingSpec = overallSpec + .5*overallSpec*(overallSpec-1)*(2*overallSpec-1)*miscT

 negCellCountFlag = TRUE #negative cell counts for naively corrected

 numNegCellInstances = 0
 while(negCellCountFlag)
 {
  siteSpecificSens = unlogit(rnorm(1, overallLogitSensitivity, sqrt(miscT)))
  siteSpecificSpec = unlogit(rnorm(1, overallLogitSpecificity, sqrt(miscT)))


  site$corruptingM = matrix(c(siteSpecificSens, 1-siteSpecificSens,
                      1-siteSpecificSpec, siteSpecificSpec),
                    nrow=2, ncol=2, byrow=TRUE)

  site$corruptedTable = site$uncorruptedTable %*% site$corruptingM

  #### Matrix-method correction matrix formed from estimated sensitivity and specificity
  site$correctingM = (1/(correctingSens+correctingSpec-1)) * matrix(c(correctingSpec, correctingSens-1,
                                          correctingSpec-1, correctingSens),
                                        nrow=2, ncol=2, byrow=TRUE)

  #### Get and round the corrected table
  site$correctedTable = round(site$corruptedTable %*% site$correctingM)
  site$corruptedTable = round(site$corruptedTable)
```

```
   if(!any(site$correctedTable <= 0))
   {
    negCellCountFlag = FALSE
   }
   else
   {
    numNegCellInstances = numNegCellInstances+1
   }
  }



  #### The logOR actually observed - i.e. from the Corrupted Table
  site$corruptedLogOR = with(site,
log((corruptedTable[1,1]*corruptedTable[2,2])/(corruptedTable[1,2]*corruptedTable[2,1])))
  site$corruptedVariance = sum(1/site$corruptedTable)

  #### The logOR that would have been observed absent misclassification - i.e. from the Uncorrupted Table
  site$uncorruptedLogOR = with(site,
log((uncorruptedTable[1,1]*uncorruptedTable[2,2])/(uncorruptedTable[1,2]*uncorruptedTable[2,1])))
  site$uncorruptedVariance = sum(1/site$uncorruptedTable)

  #### The logOR observed after matrix-method correction - i.e. from the Corrected Table
  site$mmCorrectedLogOR = with(site,
log((correctedTable[1,1]*correctedTable[2,2])/(correctedTable[1,2]*correctedTable[2,1])))
  site$NCVariance = sum(1/site$correctedTable)

  #### Get the corrected site-specific variance
  site$gCorrectedVariance = getGreenlandCorrectedVariance(site$corruptedTable, site$correctedTable, overallSens,
overallSpec, miscT)

  site$sensitivity = siteSpecificSens
  site$specificity = siteSpecificSpec
  site$correctingSens = correctingSens
  site$correctingSpec = correctingSpec

  site$numNegCellInstances = numNegCellInstances
  return(site)
}

getGreenlandCorrectedVariance = function(corruptedTable, correctedTable, overallSens, overallSpec, miscT)
{
  #### Implements Equation 2 of Greenland 1988

  correctingSens = overallSens + .5*overallSens*(overallSens-1)*(2*overallSens-1)*miscT
  correctingSpec = overallSpec + .5*overallSpec*(overallSpec-1)*(2*overallSpec-1)*miscT
  Dsqr = (correctingSens+correctingSpec-1)^2

  #### Delta-method approximation to variance of sensitivity and specificity
  vSens = miscT*(overallSens*(1-overallSens))^2 + (miscT^2)*(overallSens*(overallSens-1)*(2*overallSens-1))^2
  vSpec = miscT*(overallSpec*(1-overallSpec))^2 + (miscT^2)*(overallSpec*(overallSpec-1)*(2*overallSpec-1))^2

  m1 = corruptedTable[1,1] + corruptedTable[1,2]
  m0 = corruptedTable[2,1] + corruptedTable[2,2]

  f1 = correctedTable[1,2]/m1 # #non-cases classified as exposed after correction
  f0 = correctedTable[2,2]/m0 # #non-cases classified as unexposed
```

```
 e1 = correctedTable[1,1]/m1 # #cases classified as exposed
 e0 = correctedTable[2,1]/m0 # #cases classified as unexposed

 f1_star = corruptedTable[1,2]/m1
 f0_star = corruptedTable[2,2]/m0

 e1_star = corruptedTable[1,1]/m1
 e0_star = corruptedTable[2,1]/m0



 c1 = (e1_star*f1_star)/(m1*e1^2*f1^2)
 c0 = (e0_star*f0_star)/(m0*e0^2*f0^2)
 gCorrectedVariance = (vSens*(1/f1-1/f0)^2 + vSpec*(1/e1-1/e0)^2 + c1 + c0)/Dsqr

 return(gCorrectedVariance)
}

####
####
####----------------------------------------FUNCTIONS FOR GENERATING META-ANALYSIS LEVEL DATA ----------------------
####
####

getMACovMat = function(sites, vSens, vSpec, Dsqr)
{
 #Creates covariance matrix of corrected study-specific log odds ratios. Implements equation (4) of Greenland, 1988
 gCovMat = matrix(0, nrow=length(sites), ncol=length(sites))

 for(j in 1:length(sites))
 {
  m1j = sites[[j]]$corruptedTable[1,1] + sites[[j]]$corruptedTable[1,2]
  m0j = sites[[j]]$corruptedTable[2,1] + sites[[j]]$corruptedTable[2,2]

  f1j = sites[[j]]$correctedTable[1,2]/m1j # #exposed classified as non-case after correction
  f0j = sites[[j]]$correctedTable[2,2]/m0j # #unexposed classified non-case

  e1j = sites[[j]]$correctedTable[1,1]/m1j # #exposed classified as cases
  e0j = sites[[j]]$correctedTable[2,1]/m0j # #unexposed classified as cases
  for(k in 1:length(sites))
  {
   if(k==j)
   {
    gCovMat[j,k] = sites[[j]]$gCorrectedVariance
   }
   else
   {
    m1k = sites[[k]]$corruptedTable[1,1] + sites[[k]]$corruptedTable[1,2]
    m0k = sites[[k]]$corruptedTable[2,1] + sites[[k]]$corruptedTable[2,2]

    f1k = sites[[k]]$correctedTable[1,2]/m1k # #exposed classified as non-case after correction
    f0k = sites[[k]]$correctedTable[2,2]/m0k # #unexposed classified non-case

    e1k = sites[[k]]$correctedTable[1,1]/m1k # #exposed classified as cases
    e0k = sites[[k]]$correctedTable[2,1]/m0k # #unexposed classified as cases

    gCovMat[j,k] = (vSens/(f0j*f0k)+vSpec/(e0j*e0k))/Dsqr - (vSens/(f0j*f1k)+vSpec/(e0j*e1k))/Dsqr -
     (vSens/(f1j*f0k)+vSpec/(e1j*e0k))/Dsqr + (vSens/(f1j*f1k)+vSpec/(e1j*e1k))/Dsqr
```

```
    }
   }
  }

  return(gCovMat)
}

getGCorrectedQ = function(gCovMat_inv, logORs, FE, w_plus)
{
  #Get corrected chi2 statistic as described by Greenland 1988
  Q = 0
  for(j in 1:nrow(gCovMat_inv))
  {
    for(k in 1:ncol(gCovMat_inv))
    {
      Q = Q + gCovMat_inv[j,k]*logORs[j]*logORs[k]
    }
  }
  Q = Q - w_plus*FE^2

  return(Q)
}

getGreenlandCorrectedPooledEstimates = function(sites, logORs, miscT, overallSens, overallSpec)
{
  #Gets the pooled fixed and random effects estimates that use the corrected study specific variances

  corrSens = overallSens + .5*overallSens*(overallSens-1)*(2*overallSens-1)*miscT
  corrSpec = overallSpec + .5*overallSpec*(overallSpec-1)*(2*overallSpec-1)*miscT
  Dsqr = (corrSens+corrSpec-1)^2
  ret = vector("list")


  #### Delta-method approximation to variance of sensitivity and specificity
  vSens = miscT*(overallSens*(1-overallSens))^2 + (miscT^2)*(overallSens*(overallSens-1)*(2*overallSens-1))^2
  vSpec = miscT*(overallSpec*(1-overallSpec))^2 + (miscT^2)*(overallSpec*(overallSpec-1)*(2*overallSpec-1))^2

  gCovMat = getMACovMat(sites, vSens, vSpec, Dsqr)
  gCovMat_inv = solve(gCovMat)

  w = colSums(gCovMat_inv)
  w_plus = sum(w)

  ret$FE = sum(w*logORs)/w_plus
  ret$var_FE = 1/w_plus

  ret$gCorrQ = getGCorrectedQ(gCovMat_inv, logORs, ret$FE, w_plus)
  ret$genCochQ = sum(w*(logORs-ret$FE)^2)
  ret$FEWeights = w

  v = diag(gCovMat)
  num = ret$genCochQ - sum(w*v) + sum(w^2*v)/sum(w)
  denom = sum(w) - sum(w^2)/sum(w)

  ret$t2 = num/denom
  if(ret$t2<0)
    ret$t2 = 0
```

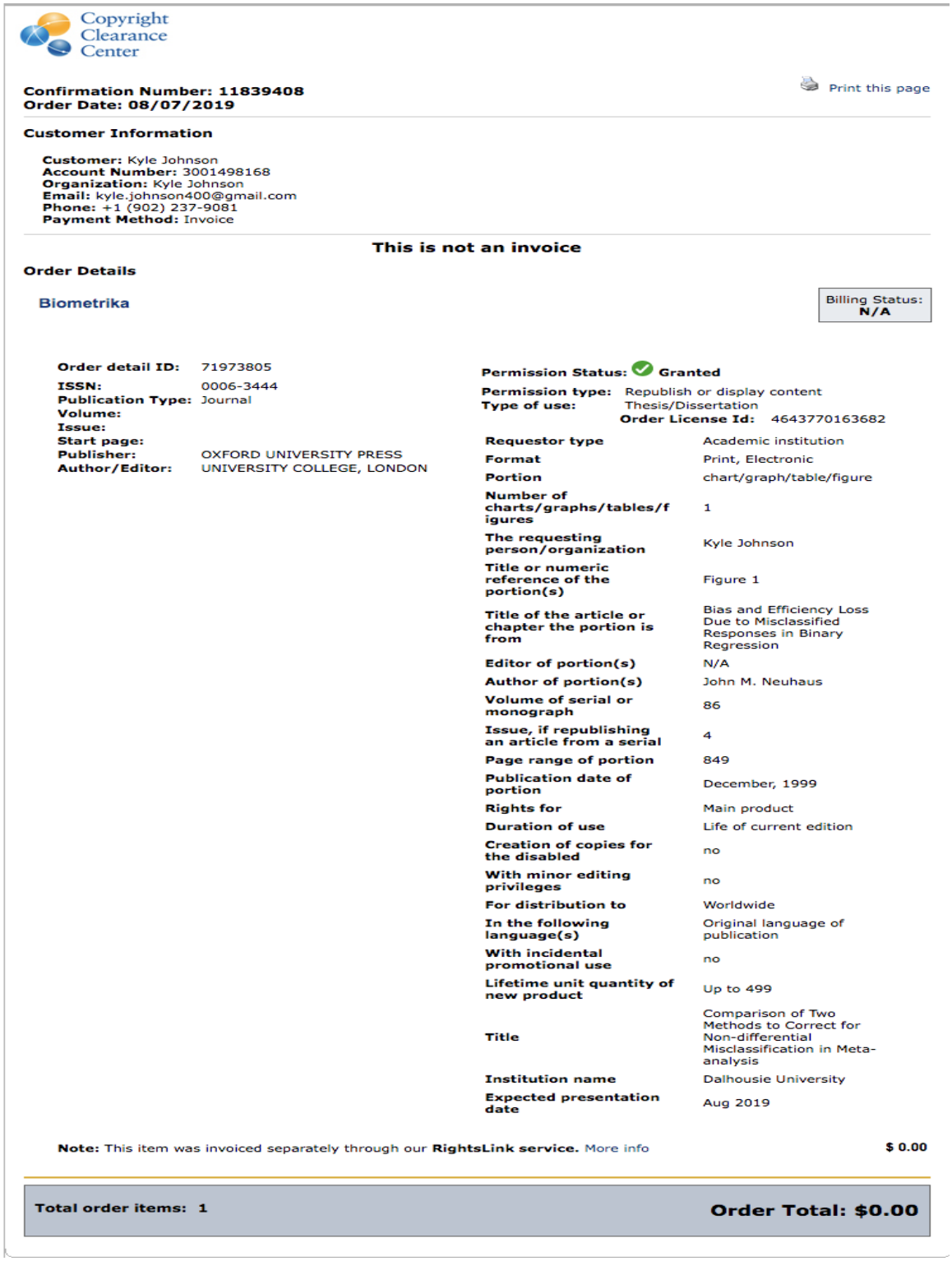
```
 diag(gCovMat) = diag(gCovMat) + ret$t2
 gCovMat_inv = solve(gCovMat)
 w_star = colSums(gCovMat_inv)
 ret$RE = sum(w_star*logORs)/sum(w_star)
 ret$var_RE = 1/sum(w_star)
 ret$REWeights = w_star

 return(ret)
}

getMetaAnalyticOutcomes = function(sites,  miscT, overallSens, overallSpec)
{
 #Gets the outcomes for a given instance of a meta-analysis
 variances = data.frame(corrupted=sapply(sites, function(x){return(x$corruptedVariance)}),
                  corrected=sapply(sites, function(x){return(x$NCVariance)}),
                  uncorrupted=sapply(sites, function(x){return(x$uncorruptedVariance)}))

 logORs = data.frame(corrupted=sapply(sites, function(x){return(x$corruptedLogOR)}),
                corrected=sapply(sites, function(x){return(x$mmCorrectedLogOR)}),
                uncorrupted=sapply(sites, function(x){return(x$uncorruptedLogOR)}))

 numNegCellInstances = sum(sapply(sites, function(x){return(x$numNegCellInstances)}))


 w = 1/variances
 FE = colSums(w * logORs)/colSums(w)
 var_FE = 1/colSums(w)

 s = cbind(logORs[,1]-FE[1], logORs[,2]-FE[2], logORs[,3]-FE[3])
 Q = colSums(w * s^2)

 C = colSums(w) - (colSums(w^2)/colSums(w))
 t2 = (Q - length(sites)+1)/C


 if(any(t2[!is.na(t2)]<0))
 {
  t2[which(t2<0)] = 0
 }

 w_star = 1/(variances+t2)
 RE = colSums(w_star * logORs)/colSums(w_star)
 var_RE = 1/colSums(w_star)

 gCorrectedEstimates = getGreenlandCorrectedPooledEstimates(sites, logORs$corrected, miscT, overallSens,
overallSpec)

 FE['gCorrected']    = gCorrectedEstimates$FE
 RE['gCorrected']    = gCorrectedEstimates$RE
 var_FE['gCorrected'] = gCorrectedEstimates$var_FE
 var_RE['gCorrected'] = gCorrectedEstimates$var_RE
 Q['gCorrected']     = gCorrectedEstimates$gCorrQ
 Q['genCochQ']       = gCorrectedEstimates$genCochQ
 t2['gCorrected']    = gCorrectedEstimates$t2


 logORCI_FE = data.frame(ub=numeric(4), lb=numeric(4), Type=numeric(4))
```

```
 logORCI_FE$Type = c("Corrupted", "Corrected", "Uncorrupted", "gCorrected")

 logORCI_FE$lb = FE - qnorm(1-significanceLevel/2)*sqrt(var_FE)
 logORCI_FE$ub = FE + qnorm(1-significanceLevel/2)*sqrt(var_FE)

 logORCI_RE = data.frame(ub=numeric(4), lb=numeric(4), Type=numeric(4))
 logORCI_RE$Type = c("Corrupted", "Corrected", "Uncorrupted", "gCorrected")

 logORCI_RE$lb = RE - qnorm(1-significanceLevel/2)*sqrt(var_RE)
 logORCI_RE$ub = RE + qnorm(1-significanceLevel/2)*sqrt(var_RE)

 FEWeights = list(corrupted=w$corrupted,
             corrected=w$corrected,
             uncorrupted=w$uncorrupted,
             gCorrected=gCorrectedEstimates$FEWeights)

 REWeights = list(corrupted=w_star$corrupted,
             corrected=w_star$corrected,
             uncorrupted=w_star$uncorrupted,
             gCorrected=gCorrectedEstimates$REWeights)


 sens = sapply(sites, function(x){return(x$sensitivity)})
 spec = sapply(sites, function(x){return(x$specificity)})

 return(list(FE, RE, Q, var_FE, var_RE, t2, logORCI_FE, logORCI_RE, numNegCellInstances, FEWeights, REWeights,
logORs, sens, spec))
}

####
####
####---------------------------------- FUNCTION FOR GENERATING CONTEXT-LEVEL OUTCOMES, E.G. PERFORMANCE
MEASURES----------------
####
####
getContextOutcomes = function(params, contextData, maData, maNum, QSigLevel, fixed=FALSE)
{
 #### Essentially just a function for taking the MA data that was just generated and putting into one of the
 #### contextData structs
 logOR = params$logOR

 negCellInstances = 0
 coversLogOR = function(logOR, CI) {
  return(CI$lb <= logOR & logOR <= CI$ub)
 }

 coversNull = function(CI) {
  return(CI$lb <= 0 & CI$ub >= 0)
 }

 contextData = within(contextData, {
  p        = parent.env(environment())
  logOR    = p$logOR
  maNum    = p$maNum
  maData   = p$maData

  coveredTrueLogORFE[maNum,] = coversLogOR(logOR, maData$logORCI_FE)
  coveredTrueLogORRE[maNum,] = coversLogOR(logOR, maData$logORCI_RE)
```

```
  falseNegativesFE[maNum,]   = coversNull(maData$logORCI_FE) & (logOR != 0)
  falseNegativesRE[maNum,]   = coversNull(maData$logORCI_RE) & (logOR != 0)

  t2[maNum,] = maData$t2
  FE[maNum,] = maData$FE
  RE[maNum,] = maData$RE

  varFE[maNum,] = maData$var_FE
  varRE[maNum,] = maData$var_RE

  FEWeights$corrupted[maNum,] = maData$FEWeights$corrupted
  FEWeights$corrected[maNum,] = maData$FEWeights$corrected
  FEWeights$uncorrupted[maNum,] = maData$FEWeights$uncorrupted
  FEWeights$gCorrected[maNum,] = maData$FEWeights$gCorrected

  REWeights$corrupted[maNum,] = maData$REWeights$corrupted
  REWeights$corrected[maNum,] = maData$REWeights$corrected
  REWeights$uncorrupted[maNum,] = maData$REWeights$uncorrupted
  REWeights$gCorrected[maNum,] = maData$REWeights$gCorrected

  siteLogORs$corrupted[maNum,] = maData$SiteLogORs$corrupted
  siteLogORs$corrected[maNum,] = maData$SiteLogORs$corrected
  siteLogORs$uncorrupted[maNum,] = maData$SiteLogORs$uncorrupted

  siteSens[maNum,] = maData$SiteSens
  siteSpec[maNum,] = maData$SiteSpec

  homStats[maNum,] = maData$Q
  negCellInstances = negCellInstances + maData$numNegCellInstances

  sig = maData$Q > qchisq(1-p$QSigLevel, 7)
  if(p$fixed)
    FP_Q[maNum,] = sig
  else
    FN_Q[maNum,] = 1-sig
 })
}


####
####
####------------------------------ MAIN SIMULATION LOOP------------------------------
####
####

if(do_fixed)
{
 for(curr_context in 1:nrow(variableParameters_fixed))
 {
  for(i in 1:numMAs)
  {
    #---------------Specify MA-specific values------------------
    siteLogOR = variableParameters_fixed$logOR[curr_context]

    #---------------Set Site Specific Values-----------------------------
    sites = vector("list", numSites[i])
    for(j in 1:numSites[i])
```

```
    {
      sites[[j]] = createSite(j, numUnexposed[j], numExposed[j], overallLogitSensitivity, overallLogitSpecificity,
                        variableParameters_fixed$T[curr_context], siteLogOR, baselineProb[j],
                        overallSensitivity, overallSpecificity)
    }
    names(sites) <- siteNames

    maData = getMetaAnalyticOutcomes(sites, variableParameters_fixed$T[curr_context], overallSensitivity,
overallSpecificity)
    names(maData) <- c("FE", "RE", "Q", "var_FE", "var_RE", "t2", "logORCI_FE", "logORCI_RE",
"numNegCellInstances",
                  "FEWeights", "REWeights","SiteLogORs", "SiteSens", "SiteSpec")

    contextData_fixed[[curr_context]] = getContextOutcomes(variableParameters_fixed[curr_context,],
                                          contextData_fixed[[curr_context]], maData,
                                          i, QSigLevel, TRUE)

    print(paste(paste("MA:", i), paste("Fixed Context:", curr_context)))
   }
  }
}

if(do_random)
{
 for(curr_context in 1:nrow(variableParameters_random))
 {
   for(i in 1:numMAs)
   {
    #---------------------- Set MA-specificy Values -----------------
    siteLogOR = rnorm(numSites[i], variableParameters_random$logOR[curr_context], tauSqr)


    #----------------------Specify Site-Specific Values ------------------------
    sites = vector("list", numSites[i])
    for(j in 1:numSites[i])
    {
      sites[[j]] = createSite(j, numUnexposed[j], numExposed[j], overallLogitSensitivity, overallLogitSpecificity,
                        variableParameters_random$T[curr_context], siteLogOR[j], baselineProb[j], .9, .92)
    }
    names(sites) <- siteNames

    maData = getMetaAnalyticOutcomes(sites, variableParameters_random$T[curr_context], overallSensitivity,
overallSpecificity)
    names(maData) <- c("FE", "RE", "Q", "var_FE", "var_RE", "t2", "logORCI_FE", "logORCI_RE",
"numNegCellInstances",
                  "FEWeights", "REWeights","SiteLogORs", "SiteSens", "SiteSpec")

    contextData_random[[curr_context]] = getContextOutcomes(variableParameters_random[curr_context,],
                                  contextData_random[[curr_context]], maData,
                                  i, QSigLevel, FALSE)
    print(paste(paste("MA:", i), paste(", Random Context:", curr_context)))
   }
  }
}
```

Appendix 7: Copyright Permissions



Copyright Clearance Center

Confirmation Number: 11839408
Order Date: 08/07/2019

Print this page

Customer Information

Customer: Kyle Johnson
Account Number: 3001498168
Organization: Kyle Johnson
Email: kyle.johnson400@gmail.com
Phone: +1 (902) 237-9081
Payment Method: Invoice

This is not an invoice

Order Details

Biometrika

Billing Status: **N/A**

Order detail ID:	71973805
ISSN:	0006-3444
Publication Type:	Journal
Volume:	
Issue:	
Start page:	
Publisher:	OXFORD UNIVERSITY PRESS
Author/Editor:	UNIVERSITY COLLEGE, LONDON

Permission Status: Granted
Permission type: Republish or display content
Type of use: Thesis/Dissertation
Order License Id: 4643770163682

Requestor type	Academic institution
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/figures	1
The requesting person/organization	Kyle Johnson
Title or numeric reference of the portion(s)	Figure 1
Title of the article or chapter the portion is from	Bias and Efficiency Loss Due to Misclassified Responses in Binary Regression
Editor of portion(s)	N/A
Author of portion(s)	John M. Neuhaus
Volume of serial or monograph	86
Issue, if republishing an article from a serial	4
Page range of portion	849
Publication date of portion	December, 1999
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	Comparison of Two Methods to Correct for Non-differential Misclassification in Meta-analysis
Institution name	Dalhousie University
Expected presentation date	Aug 2019

Note: This item was invoiced separately through our **RightsLink service.** More info **$ 0.00**

Total order items: 1 **Order Total: $0.00**

Permission for Figure 1 in this thesis

License Details

This Agreement between Kyle A Johnson ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

Print Copy

License Number	4652590476323
License date	Aug 19, 2019
Licensed Content Publisher	BMJ Publishing Group Ltd.
Licensed Content Publication	The BMJ
Licensed Content Title	Higher potency statins and the risk of new diabetes: multicentre, observational study of administrative databases
Licensed Content Author	Colin R Dormuth, Kristian B Filion, J Michael Paterson, Matthew T James, Gary F Teare, Colette B Raymond, Elham Rahme, Hala Tamim, Lorraine Lipscombe
Licensed Content Date	May 29, 2014
Licensed Content Volume	348
Type of Use	Dissertation/Thesis
Requestor Type	Individual Account
Format	Print and electronic
Portion Used	Figure/table/extract
Number of figure/table/extracts	2
Description of figure/table/extracts	Table 1 Figure 1
Will you be translating?	No
Circulation/distribution	1
Title of your thesis / dissertation	Comparison of Two Methods to Correct for Non-differential Misclassification in Meta-analysis
Expected completion date	Aug 2019
Estimated size(pages)	1
Requestor Location	Kyle A Johnson 6147 Charles Street Halifax, NS B3K 1L2 Canada Attn:
Publisher Tax ID	GB674738491
Total	**0.00 CAD**

Permission for Figure 2 and Table 11 in this thesis