Version: 1

Table of Contents

# 1 List of Abbreviations

|  |  |
| --- | --- |
| CI | Confidence interval |
| COVID-19 | Coronavirus disease 2019 |
| EHR | Electronic health record |
| HR | Hazard ratio |
| NPV | Negative predictive value |
| OR | Odds ratio |
| PPV | Positive predictive value |
| QBA | Quantitative bias analysis |
| RCT | Randomized-controlled clinical trial |
| RR | Relative risk |
| RWD | Real-world data |
| RWE | Real-world evidence |

# 2 Investigators

|  |  |
| --- | --- |
| Investigator | Contact |
| James Weaver 1,2,3 \* | [james.weaver@ndorms.ox.ac.uk](mailto:james.weaver@ndorms.ox.ac.uk) |
| Patrick B. Ryan 2,3,4 | [ryan@ohdsi.org](mailto:ryan@ohdsi.org) |
| Victoria Strauss 1,2 | [victoria.strauss@csm.ox.ac.uk](mailto:victoria.strauss@csm.ox.ac.uk) |
| Marc A. Suchard 3,5 | [msuchard@ucla.edu](mailto:msuchard@ucla.edu) |
| Daniel Prieto-Alhambra 1,3,6 | [daniel.prietoalhambra@ndorms.ox.ac.uk](mailto:daniel.prietoalhambra@ndorms.ox.ac.uk) |

## 2.1 Disclosures

This study is undertaken within Observational Health Data Sciences and Informatics (OHDSI), an open science collaboration. **JW** and **PR** are employees of shareholders of Janssen R&D (a Johnson & Johnson company). **MAS** receives grant support from the US National Institutes of Health, US Food & Drug Administration and US Department of Veterans Affairs and contracts from Janssen R&D. **VS**... **DPA**’s research group has received grant support from Amgen, Chesi-Taylor, Novartis, and UCB Biopharma. His department has received advisory, consultancy fees from Amgen, Astellas, AstraZeneca, Johnson & Johnson, and UCB Biopharma and fees for speaker services from Amgen and UCB Biopharma. Janssen, on behalf of IMI-funded EHDEN and EMIF consortiums, and Synapse Management Partners have supported training programmes organised by DPA’s department and open for external participants organized by his department outside submitted work.

## 2.2 Collaborators

Eli Harriss of Oxford Bodleian Libraries

# 3 Review information

## 3.1 Anticipated or actual start date

2021-11-01

## 3.2 Anticipated completion date

2022-06-01

## 3.3 Stage of review

|  |  |  |
| --- | --- | --- |
| Review stage | Started | Completed |
| Preliminary searches | Yes | No |
| Piloting of the study selection process | No | No |
| Formal screening of search results against eligibility criteria | No | No |
| Data extraction | No | No |
| Risk of bias (quality) assessment | No | No |
| Data analysis | No | No |

# 4 Abstract

**Background**

Outcome misclassification is acknowledged but rarely corrected for in causal inference studies using observational data. Quantitative bias analysis (QBA) is a method for outcome misclassification correction, but it is unclear how often it is used and the extent to which it minimizes bias in effect estimates.

The existing literature generally reports study results from one exposure comparison, for one outcome, in one database, using one QBA type. Existing evaluations of QBA performance are subject to variability from different research questions, study designs, analytic implementations, and contexts where there is no a priori expectation about the true result. This review will inform whether novel QBA methodological research should be subsequently designed and executed using strictly consistent study design, analytic implementation, and applied to research questions where the estimates are expected to be null (i.e., negative control exposure-outcome pairs).

**Objectives**

* Characterize use of QBA for outcome misclassification correction in pharmacoepidemiologic literature
* Assess the impact QBA for outcome misclassification correction on effect measure estimates
* Explore QBA methodological variables that may explain impact

**Search methods**

We searched X electronic databases from 2006-01-01 to YYYY-MM-DD. Sentence on validated search strategy here. We restricted to English language studies.

**Selection criteria**

We queried and examined clinical and methodological studies that used validation study results as QBA input to generate effect estimates corrected for outcome misclassification.

**Data collection and analysis**

The following outcome measures will be collected from included studies: relative risks or rate ratios (RR), odds ratios (OR), and hazard ratios (HR) point estimates and associated standard errors. The systematic review outcome measures are bias, mean squared error, and precision difference.

# 5 Background

Randomized-controlled clinical trials (RCT) are the gold-standard for making causal inferences between exposures and outcomes but are resource intensive, slow to execute, lack generalizability, and are under powered for detecting many adverse events. Causal inferences made from the analysis of real-world data[[1]](#footnote-34) (RWD) are of increasing importance for medical decision-making and can be the basis for formulating clinical, policy, and regulatory strategies if properly conducted [[1](#ref-Sherman2016-yn)]. Healthcare utilization databases, such as administrative claims and electronic health records (EHR), are a main source of observational data used in pharmacoepidemiologic research. Appropriate analysis of observational data offers the potential to rapidly produce generalizable, precise, real-world evidence (RWE) that accurately reflects the heterogeneity of patient experience in routine clinical care outside the confines of strictly defined experimental conditions[[2](#ref-Stuart2015-gg)]. However, RWE validity relies on the extent to which study design and analysis can address the limitations inherent to observational data. Major limitations of observational data include the presence of selective treatment assignment and measurement errors which can manifest as confounding and information biases in results.

Extensive literature is available on methods to correct for observed confounding from selective treatment assignment using exposure propensity scores, many of which have been widely studied, endorsed, and routinely used in observational research using large healthcare databases [[3](#ref-Rosenbaum1983-hi)–[7](#ref-Tian2018-zy)]. However, and despite similar availability and endorsement, quantitative assessment of the direction, magnitude, and uncertainty of bias from measurement errors is rarely included clinical studies in epidemiology [[8](#ref-Hunnicutt2016-gn),[9](#ref-Petersen2021-yk)].

Quantitative bias analysis (QBA) is a set of methods for measuring and correcting selection, confounding, and information biases in observational studies. A prior systematic review covering 2006-2019 [[9](#ref-Petersen2021-yk)] found that from over 400,000 publications returned by searching “epidemiology,” there were 697 candidate studies from which 238 were reviewed. The authors conclude that QBA use is rare but increasing, often did not alter results interpretation, the methodology reporting was often incomplete, and the degree of bias correction varied across studies and methods. An earlier, more specific review in pharmacoepidemiology literature found 15 studies[[8](#ref-Hunnicutt2016-gn)]. The authors reviewed use of probabilistic QBA and reported that use is rare and methodological approaches and reporting are suboptimal.

The scope of this systematic review is to identify and evaluate observational, pharmacoepidemiology studies intended to estimate the causal effect of a drug exposure on a clinical outcome that used QBA to correct for outcome misclassification. Because this review focuses specifically on QBA for outcome misclassification correction, all subsequent references to QBA will refer to this particular use.

## 5.1 Description of the problem

Measurement errors typically occur when empirical information is incorrectly recorded in a data source or when there is discordance between a true clinical/biological definition and its operationalization in research. For example, an improperly calibrated continuous glucose monitor may record inaccurate glucose measurements. Further, a clinical definition of type 2 diabetes mellitus as two recordings of fasting blood glucose of ≥126 mg/dL from separate tests is unlikely to be accurately captured in a study that defines patients with diabetes as those for whom the E11.9 ICD-10-CM code is observed. A contemporary example is the ascertainment of coronavirus disease 2019 (COVID-19) cases in administrative claims databases with linkages to laboratory testing information. In many instances, patients with a COVID-19 diagnosis are also observed to have varying patterns of positive and negative COVID-19 testing results temporally proximal to their diagnosis date, which complicates the process of defining case definitions.

Misclassification can affect exposure, confounder, or outcome measurement which can render study results subject to information bias. In the relatively novel context of distributed observational database network studies, a single clinical outcome definition is generally applied across multiple, disparate data sources where data sources vary widely in terms of intent (e.g., reimbursement, clinical care), data capture process, population coverage, temporal coverage, and content. Given database variability, it follows that measurement error for clinical outcome definitions will vary accordingly and mostly unpredictably. Therefore, the scope of this review will cover methods for correcting for outcome misclassification.

## 5.2 Description of the method

This review will investigate a subset of methods for addressing outcome misclassification in pharmacoepidemiologic studies. The focus is on 4 methods introduced by Lash et al[[10](#ref-Lash2009-ch)] and all methods require bias parameter information (i.e. sensitivity and specificity or positive predictive value (PPV) and negative predictive value (NPV) as input. The bias parameters can be derived from internal or external validation studies, literature sources, or failing those, expert opinion. Figure 5.1 displays a classification tree of QBA methods.

**Simple bias analysis** rearranges the observed distribution of a categorical contingency table of exposure and outcome patient counts to reflect the true distribution were there no misclassification. Algebraic equations that relate effect estimate calculation to the outcome definition bias parameters are used. Standard errors are calculated using the method by Chu[[11](#ref-Chu2006-zo)].

**Multidimensional bias analysis** is an extension of simple bias analysis where a range of plausible bias parameters are used as input to produce a corresponding range of misclassification corrected estimates.

In **probabilistic bias modeling**, the bias parameters are selected by drawing from a bias parameter distribution (e.g., uniform, normal, triangular, etc.). The drawn bias parameters are used is input to a multidimensional bias analysis that produced a corresponding distribution of misclassification corrected estimates, where the estimate distribution median could be considered the final, corrected estimate.

**Multiple bias modeling** incorporates correlated selection, confounder, and measurement bias parameter distributions and is infrequently used in practice[2]. Care must be taken to ensure that implausible bias parameter combinations are not selected to avoid calculating invalid misclassification-corrected estimates.

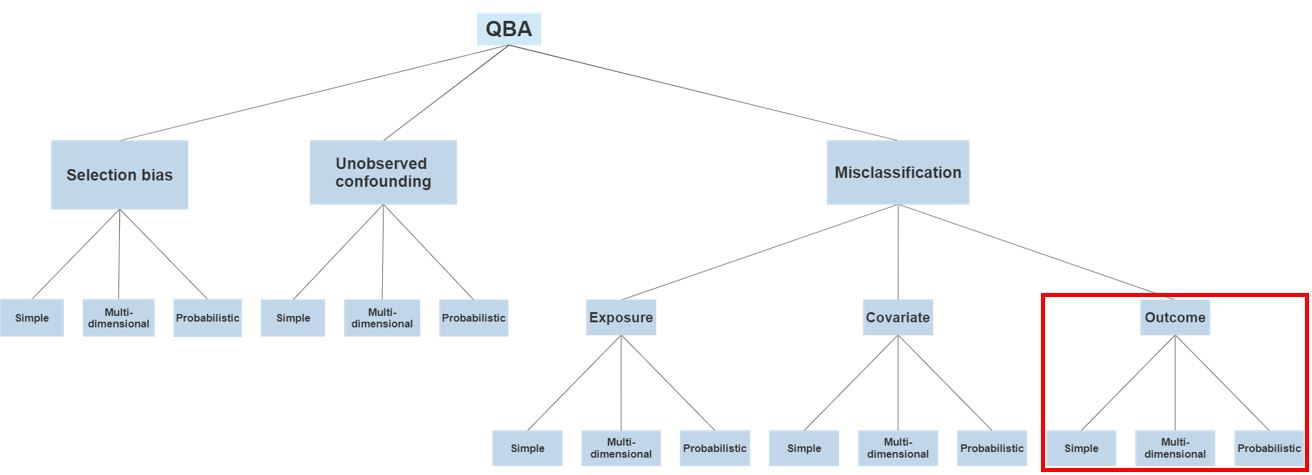


Figure 5.1: Schematic of QBA methods classification. This review focuses on simple, multidimensional, and probabilistic methods for outcome misclassification correction highlighted in the red.

## 5.3 Intent of the method under review

Applying QBA can quantify the direction, magnitude, and uncertainty of systematic error that results from outcome misclassification and correct for it in an estimate of the causal effect of an exposure on an outcome. A popular heuristic rule states that outcome misclassification that is non-differential with respect to exposure status will bias estimation results towards the null, although in practice this is an unreliable assumption[[12](#ref-Funk2014-bk),[13](#ref-Lash2014-jd)]. Rather than relying on heuristics, QBA for outcome misclassification correction attempts to replace the measured outcome occurrence with the true outcome occurrence in a causal analysis (Figure 5.2). The direction of bias owing to outcome misclassification results from its dependence and differentially with respect to the exposure[[14](#ref-VanderWeele2012-ua)].

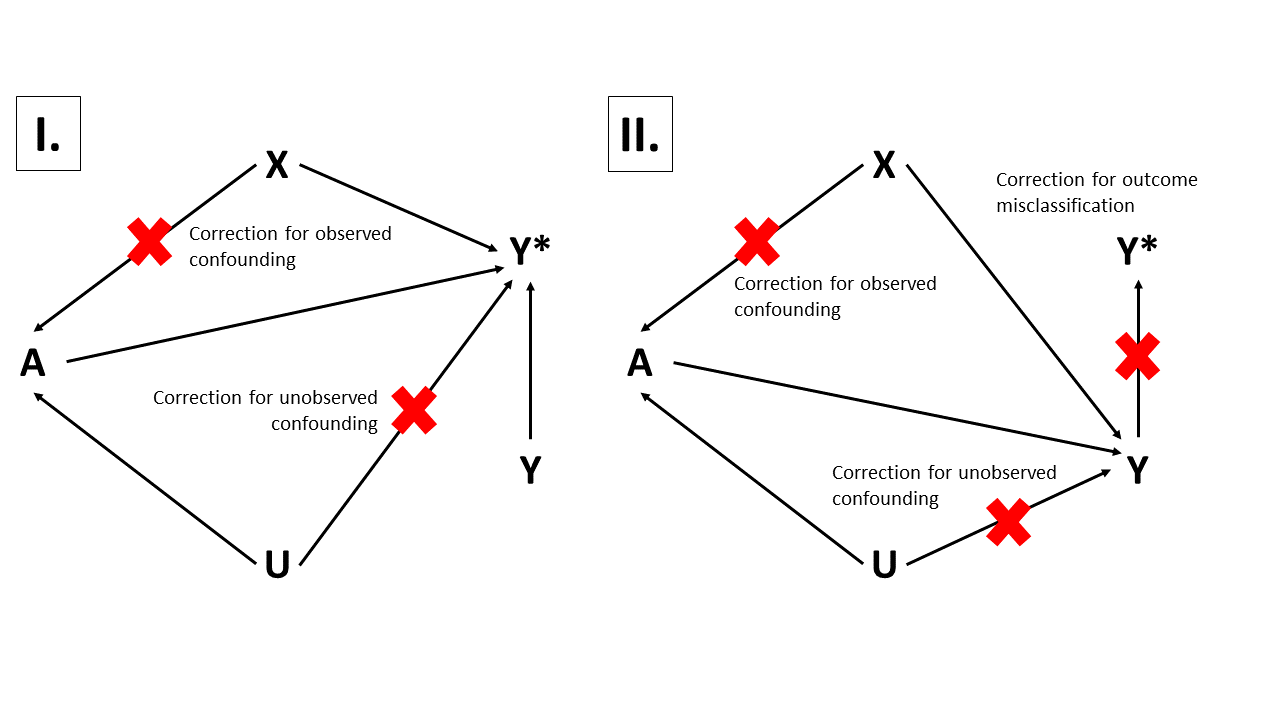


Figure 5.2: Directed acyclic graph depicting a simplified causal structure for assessing the effect of A on Y with correction for observed confounding, unobserved confounding, and outcome misclassification. A: exposure, Y: true outcome, Y\*: measured outcome, X: observed confounders, U: unobserved confounders, A → Y\*: target association of inference subject to information bias, A → Y: target association of inference, A ← X → Y: observed confounding, A ← U → Y = unobserved confounding, Y → Y\*: outcome misclassification. This scenario assumes exposure and observed confounders are accurately measured and that outcome misclassification is independent and non-differential with respect to exposure status. Panel I depicts the causal structure before outcome misclassification correction. Panel II depicts the causal structure after outcome misclassification correction.

## 5.4 Rationale for this review

The existing empirical literature generally reports results from one exposure comparison, for one outcome, in one database, using one QBA type. Evaluating QBA performance is subject to variability from different research questions, study designs, analytic implementations, in contexts where there is no *a priori* expectation about the true result. This review will determine whether novel QBA methodological research should be designed using strictly consistent study design, and analytic implementation, and applied to research questions where the estimates that are expected to be null (i.e., negative control exposure-outcome pairs).

# 6 Review question

What is the performance of quantitative bias analysis (QBA) for outcome misclassification correction in observational, causal effect estimation studies pharmacoepidemiology?

## 6.1 Objectives

* Characterize use of QBA for outcome misclassification correction in pharmacoepidemiological literature. We will report the number of studies that include QBA methods by year. We will report the proportion of QBA studies by study characteristics [Section XX.X](#study-characteristics) and by QBA characteristics [Section XX.X](#qba-characteristics).
* Assess the impact QBA for outcome misclassification correction on effect measure estimates
* Explore study characteristics and QBA characteristics that may explain impact

# 7 Methods

## 7.1 Sources to identify studies

* Systematic review registries
* Cochrane Database of Systematic Reviews
* Cochrane Methodology Register
* PROSPERO register
* Bibliographic databases
  + MEDLINE via PubMed
  + EMBASE
  + Web of Science/Web of Social Science
* Reference lists of review articles
* Articles that cite seminal QBA literature

## 7.2 Search dates

2021-11-01 to YYYY-MM-DD

## 7.3 Restrictions

* Human studies
* English language
* Publication period: 2006-01-01 to YYYY-MM-DD

## 7.4 Scientific domain

Observational, causal effect estimation studies pharmacoepidemiology. Studies intended to estimate the causal effect of a therapeutic exposure on an outcome.

## 7.5 Study inclusion criteria

We will query the literature for clinical and methodological studies that used validation study results as QBA input to generate effect estimates corrected for outcome misclassification. We will not restrict studies to any specific therapeutic area, study design, data type, or QBA approach. We will only require that the clinical studies were intended to estimate the causal effect of a therapeutic exposure on an outcome using observational data and include effect estimates with and without QBA-based correction. Methodological studies must be intended to assess a method for outcome misclassification correction in studies with this same intent. Methodological studies could be based on empirical or simulated data.

### 7.5.1 Participants/population

No restrictions on patient population or therapeutic area.

### 7.5.2 Intervention

4 types QBA for outcome misclassification correction from Lash (2009)[[10](#ref-Lash2009-ch)]

* Simple
* Multidimensional
* Probabilistic
* Multiple bias modeling (correlated biases)

### 7.5.3 Comparator

No QBA for outcome misclassification correction

### 7.5.4 Types of studies to be included

* Cohort
* Case-control
* Self-controlled
* Methodological (empirical or simulated)

## 7.6 URL to search strategy

Search strategy for each source here. Searches will be rerun prior to final analysis.

## 7.7 Main outcomes

* **Bias difference**, defined as the difference between the log of the QBA-corrected effect estimate and the log of the uncorrected estimate)
  + log(nonQbaEstimate) – log(qbaEstimate)
* **Relative bias**, defined as the percent difference between the QBA corrected effect estimate and the uncorrected estimate)
  + [nonQbaEstimate – qbaEstimate] / nonQbaEstimate \* 100

## 7.8 Additional outcomes

* **Squared error**, defined as a directionless quantity representing only the magnitude of bias correction
  + [log(nonQbaEstimate) – log(qbaEstimate)]2
* **Precision difference**, defined as the difference in the precision of the QBA-corrected point estimate and the uncorrected estimate)
  + 1/(SE(log(nonQbaEstimate))2 – 1/(SE(log(qbaEstimate))2
* **Coverage**, defined as whether the confidence intervals (CI) of the QBA-corrected and uncorrected estimates include the null

## 7.9 Study selection

Two reviewers (JW and ??) will independently apply eligibility criteria to the de-duplicated set of studies identified by the search strategies and will be blinded to decisions of the other reviewer. Included study discrepancies will be adjudicated by a third reviewer (DPA).

## 7.10 Data extraction

Two reviewers (JW and ??) with independently extract data from included studies. Data extraction discrepancies will be adjudicated by a third reviewer (DPA). Data will be collected in a data collection tool. Missing data will be allowed. Study authors will be contacted if data are missing for main or additional outcomes.

The following data will be extracted from each included study. A single study can contain ≥1 effect estimates so there can be ≥1 row of data collected per study. The unit of analysis of this review is each effect estimate, grouped at a second level by study.

### 7.10.1 Study indentification

* First author (text)
* Year published (numeric)
* PMID, if available (numeric)

### 7.10.2 Outcomes

Extracted effect estimates will be those adjusted for any other forms of bias(e.g. PS matching)

* QBA-corrected log effect estimate (numeric)
* QBA-corrected log standard error of effect estimate (numeric)
* QBA-uncorrected log effect estimate (numeric)
* QBA-uncorrected log standard error of effect estimate (numeric)
* QBA-corrected exp effect estimate (numeric)
* QBA-corrected exp standard error of effect estimate (numeric)
* QBA-uncorrected log effect estimate (numeric)
* QBA-uncorrected log standard error of effect estimate (numeric)

### 7.10.3 Study characteristics

The following data will be extracted if applicable and available. Target, comparator, and outcome counts will be those reported after all study design restrictions implemented.

* Study type (categorical):
  + Clinical
  + Methods
* Design (categorical), if Study type = Clinical:
  + Cohort
  + Case-control
  + Self-controlled
* Design (categorical), if Study type = Methods:
  + Empirical
  + Simulated
  + Quasi-simulated
* Target definition (text)
* Target N (numeric)
* Comparator definition (text)
* Comparator N (numeric)
* Time-at-risk definition (text)
* Target time-at-risk days (numeric)
* Comparator time-at-risk days (numeric)
* Additional bias correction approaches (multiple choices categorical)
  + QBA for exposure misclassification
  + QBA for covariate misclassification
  + QBA for unmeasured confounding
  + QBA for selection bias correction
  + PS matching
  + PS stratification
  + Multiple regression outcome modeling
  + Negative control assessment
  + Other
* Outcome model and estimator (categorical)
  + Logistic regression/OR
  + Poisson regression/RR
  + Cox PH regression/HR
  + Other
* Distributed database network study (categorical)
  + Yes
  + No
* Data source (multiple choices categorical)
  + Administrative claims
  + EHR
  + Registry

### 7.10.4 QBA characteristics

* QBA type (categorical)
  + Simple
  + Multidimensional
  + Probabilistic
  + Multiple bias modeling
* Validation study (categorical)
  + Internal
  + External
  + Other
* Bias parameters (categorical)
  + Sensitivity/specificity
  + PPV/NPV
* Validation study population (categorical)
  + Same
  + Similar
  + Different
  + Unknowns

### 7.10.5 Quality assessment

The performance QBA for outcome misclassification correction will be quantified and evaluated using the main and additional outcomes. QBA reporting quality will be assessed by reporting the proportions of missing data from study and QBA characteristics.

### 7.10.6 Data synthesis

The bias difference and standard error of the correlated bias difference will be computed for each QBA-corrected and uncorrected estimate. Heterogeneity across studies will be assessed with the I2 statistic. A pooled estimate of bias difference will be calculated if I2<50%.

### 7.10.7 Subgroup analyses

No specific subgroups will be explicitly assessed. Rather, outcomes will be tested for association with study characteristics and QBA characteristics (Re: objective “Explore study characteristics and QBA methodological variables that may explain impact”). Main and additional outcomes will be characterized with mean and standard deviation by study characteristics and QBA characteristics. For each main and additional outcome, we maywill fit a multivariable predictive model assessing the predictive impact study characteristics and QBA characteristics on the outcomes. Not that the resulting model parameters reflect associations of predictive importance and are not intended to be interpreted as causal.

# 8 Existing reviews on similar topics

A review in 2016[[8](#ref-Hunnicutt2016-gn)] assessed the use of probabilistic bias analysis for correcting selection, confounding, and misclassification biases in pharmacoepidemiology and covered literature from 2010-01-01 to 2015-10-31. Another review in 2021[[9](#ref-Petersen2021-yk)] assessed the use of 4 methods of quantitative bias analysis for correcting selection, confounding, and misclassification biases in epidemiology and covered literature from 2006-01-01 to 2019-12-31. The current systematic review will assess the use of 4 types of quantitative bias analysis for outcome misclassification correction in pharmacoepidemiology and will cover literature from 2006-01-01 to YYYY-MM-DD.

* Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, Lapane KL. Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: a systematic review. Pharmacoepidemiol Drug Saf. 2016 Dec;25(12):1343-1353. doi: 10.1002/pds.4076. Epub 2016 Sep 5. PMID: 27593968; PMCID: PMC5272921.[[8](#ref-Hunnicutt2016-gn)]
  + Assessed probabilistic QBA for selection, confounding, and misclassification biases in pharmacoepidemiology
* Petersen JM, Ranker LR, Barnard-Mayers R, MacLehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. Int J Epidemiol. 2021 Apr 20:dyab061. doi: 10.1093/ije/dyab061. Epub ahead of print. PMID: 33880532.[[9](#ref-Petersen2021-yk)]
  + Assessed simple, multidimensional, and probabilistic QBA for selection, confounding, and misclassification biases in epidemiology

# References

1 Sherman RE, Anderson SA, Dal Pan GJ, *et al.* Real-World evidence - what is it and what can it tell us? *N Engl J Med* 2016;**375**:2293–7.

2 Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prev Sci* 2015;**16**:475–85.

3 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;**70**:41–55.

4 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;**46**:399–424.

5 Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;**127**:757–63.

6 Franklin JM, Eddings W, Austin PC, *et al.* Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. *Stat Med* 2017;**36**:1946–63.

7 Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol* 2018;**47**:2005–14.

8 Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, *et al.* Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: A systematic review. *Pharmacoepidemiol Drug Saf* 2016;**25**:1343–53.

9 Petersen JM, Ranker LR, Barnard-Mayers R, *et al.* A systematic review of quantitative bias analysis applied to epidemiological research. *Int J Epidemiol* 2021.

10 Lash TL, Fox MP, Fink AK. *Applying quantitative bias analysis to epidemiologic data*. Springer New York 2009.

11 Chu H, Wang Z, Cole SR, *et al.* Sensitivity analysis of misclassification: A graphical and a bayesian approach. *Ann Epidemiol* 2006;**16**:834–41.

12 Funk MJ, Landi SN. Misclassification in administrative claims data: Quantifying the impact on treatment effect estimates. *Curr Epidemiol Rep* 2014;**1**:175–85.

13 Lash TL, Fox MP, MacLehose RF, *et al.* Good practices for quantitative bias analysis. *Int J Epidemiol* 2014;**43**:1969–85.

14 VanderWeele TJ, Hernán MA. Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. *Am J Epidemiol* 2012;**175**:1303–10.

1. In most uses related to research, the term real-world is interchangeable with the term observational. For example, the terms real-world data and real-world research are equivalent to observational data and observational research, respectively. The term observational will be used for the remainder of this document. [↑](#footnote-ref-34)