

RESEARCH

Open Access



Quantitative bias analysis for mismeasured variables in health research: a review of software tools

Codie J. C. Wood^{1,2,3*}, Kate Tilling^{2,3}, Jonathan W. Bartlett⁴ and Rachael A. Hughes^{2,3}

Abstract

Background Measurement error and misclassification can cause bias or loss of power in epidemiological studies. Software performing quantitative bias analysis (QBA) to assess the sensitivity of results to mismeasurement are available. However, QBA is still not commonly used in practice, partly due to a lack of knowledge of these software implementations. The features and particular use cases of these tools have not been systematically evaluated.

Methods We reviewed and summarised the latest available software tools for QBA in relation to mismeasured variables in health research. We searched the electronic database Web of Science for studies published between 1st January 2014 and 1st May 2024 (inclusive). We included epidemiological studies that described the use of software tools for QBA in relation to mismeasurement. We also searched for tools catalogued on the CRAN archive, in Stata manuals, and via Stata's *net* command, available from within Stata or from the IDEAS/RePEc database. Tools were included if they were purpose-built, had documentation, and were applicable to epidemiological research. Data on the tools' features and use cases were then extracted from the full article texts and software documentation.

Results 17 publicly available software tools for QBA were identified, accessible via R, Stata, and online web tools. The tools cover various types of analysis, including regression, contingency tables, mediation analysis, longitudinal analysis, survival analysis and instrumental variable analysis. However, there is a lack of software tools performing QBA for misclassification of categorical variables and measurement error outside of the classical model. Additionally, the existing tools often require specialist knowledge.

Conclusions Despite the availability of several software tools, there are still gaps in the existing collection of tools that need to be addressed to enable wider usage of QBA in epidemiological studies. Efforts should be made to create new tools to assess multiple mismeasurement scenarios simultaneously, and also to increase the clarity of documentation for existing tools, and provide tutorials and examples for their usage. By doing so, the uptake of QBA techniques in epidemiology can be improved, leading to more accurate and reliable research findings.

Keywords Quantitative bias analysis, Measurement error, Misclassification, Software tools, Scoping review

*Correspondence:

Codie J. C. Wood
codie.wood@bristol.ac.uk

¹Institute of Statistical Sciences, School of Mathematics, University of
Bristol, Woodland Road, Bristol BS8 1UG, UK

²MRC Integrative Epidemiology Unit, University of Bristol, Oakfield Road,
Bristol BS8 2BN, UK

³Population Health Sciences, Bristol Medical School, University of Bristol,
Whately Road, Bristol BS8 2PS, UK

⁴Department of Medical Statistics, London School of Hygiene & Tropical
Medicine, Keppel Street, WC1E 7HT London, UK



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

In epidemiological and population health studies, we often aim to estimate the causal effect of an exposure or treatment on an outcome (referred to as the exposure effect) while adjusting for confounders or other variables [1]. Most methods of estimating an exposure effect rely on the assumption that sufficient confounders are known and have been included in the model, and that the included variables have been measured without error. When data are obtained for epidemiological studies, there is potential for some of the variables to be measured with error and so this assumption may not be plausible [2]. Where we have categorical or binary variables measured with error (as opposed to continuous variables), we refer to measurement error as misclassification. Throughout, the umbrella term “mismeasurement” is used to capture both scenarios.

It is a common misconception that non-differentially mismeasured variables will always bias the effect estimate towards the null [3, 4]. In fact, the impact of mismeasurement on an effect estimate depends on a number of factors, including the role of the variable(s) in which the mismeasurement occurs (i.e., whether it is the outcome, exposure, or other covariate), the type of the variable (i.e., whether it is binary, continuous, or categorical) [5, 6], whether errors in multiple variables are dependent on each other [7], the type of analysis being conducted, and whether the mismeasurement is differential (i.e., some aspect of the error distribution depends on another variable) [8].

Failing to account for mismeasurement can result in problems such as decreased statistical power, biased effect estimates (either towards or away from the null), and inaccurate representations of estimate uncertainty [9]. Any of these issues could result in the reporting of erroneous study conclusions. Inaccurate findings may not only influence government policies and the development of large-scale health interventions, but could also shape the direction of subsequent studies, impact the scientific evidence base, and introduce bias into meta-analyses. Therefore, it is important to account for and quantify the potential effects of mismeasurement. Although potential mismeasurement is sometimes mentioned as a study limitation, it is rarely investigated or adjusted for in practice [4, 10]. A recent review of measurement error in medical literature found that of 565 studies reviewed, only 44% mentioned measurement error at all, with 70% of those doing so only in the discussion section. Of the studies that mentioned mismeasurement, just 7% undertook any investigation or correction [10].

There exist many methods to adjust for mismeasurement, which have been described extensively in the literature [5, 6, 8, 11, 12]. These methods typically require some form of ancillary data, such as validation data

(either internal or external), or replication data [9]. However, ancillary data are often not readily available. In these cases, sensitivity analyses such as a quantitative bias analysis (QBA) can be used to evaluate the potential impact of mismeasurement on a study's conclusions.

QBA consists of a group of statistical methods for assessing uncertainty arising due to biases in a study [13]. It can be applied to various biases, including, but not limited to, unmeasured confounding [14] and selection bias [15]. Here, we focus on QBA for mismeasurement, which is used to quantify the potential impact of mismeasurement or to assess how severe it would need to be to change a study's conclusions. This allows researchers to assess the robustness of the study's conclusions to the assumption of no mismeasurement. See [Background on quantitative bias analysis](#) section for further information on QBA.

Currently, QBA methods are not employed as standard practice. A recent review found that QBA usage in epidemiology increased between 2006 and 2019 [15], but it was still relatively rare. Possible contributors to the limited use of QBA include the historical lack of available software, limited awareness of existing tools, and the relatively low profile of QBA methods in epidemiological training. Our review seeks to address these challenges by collating information on current software options, highlighting existing gaps, and increasing awareness of the tools available.

There have been several reviews of implementations of QBA methods for unmeasured confounding, misclassification and selection bias in epidemiology and health-related fields [15–20]. These reviews primarily focused on the methodological aspects of QBA—for example, describing available methods, summarising their application in practice, and evaluating their use in empirical studies. They did not review the availability or functionality of software tools used to implement these methods. One review did examine software implementations of QBA methods, but it was specific to tools addressing bias due to unmeasured confounding [14].

In this scoping review, we aim to identify the latest available software tools that implement a QBA for mismeasurement within epidemiological studies quantifying an exposure effect estimate, and provide details on their features and use cases. This will increase awareness of the software and, alongside developments in guidance for researchers on appropriate QBA implementations [19–21], promote its usage as standard practice for health research. We also aim to highlight potential future areas for software development.

Background on quantitative bias analysis

A QBA for mismeasurement quantifies the likely magnitude and direction of the bias under different plausible

assumptions about the mismeasurement process (assuming no other sources of bias). Generally, a QBA requires a model (known as a bias model) for the observed data and the measurement errors [13]. The bias model includes one or more parameters (known as bias or sensitivity parameters), which cannot be estimated from the observed data. These bias parameters encode the researcher's assumptions about the mismeasurement process, determining the magnitude and direction of the bias-adjustment. For example, in the case of misclassification, the bias parameters may include some combination of the sensitivity, specificity, positive predictive value and negative predictive values¹. For a continuous variable measured with error, bias parameters may include reliability metrics (such as the reliability ratio or intraclass correlation coefficient), or error quantities (such as error variance or mean squared error).

The observed data alone cannot be used to inform the values of the bias parameters. Researchers must pre-specify values or distributions of values for the bias parameters to enable estimation of the remaining parameters of the bias model and thus obtain a bias-adjusted estimate of the parameter of interest (e.g., the exposure effect). This information is usually obtained from external sources such as validation studies, prior research, expert elicitation, or theoretical constraints [21]. Although the observed data cannot by themselves determine bias parameters, in some circumstances they may provide information to rule out specific combinations of bias parameters as incompatible with the empirical data.

QBA methods can broadly be classified into two categories: deterministic and probabilistic [13]. A deterministic QBA specifies one or more values for each bias parameter. A "simple bias analysis" fixes each bias parameter to a single value (i.e., treating the bias parameter values as known) and outputs a single bias-adjusted estimate of the exposure effect [13].

Typically, the bias parameters are unknown and so the researcher will need to perform a "multidimensional bias analysis" where multiple values are specified for each bias parameter, and the bias model is repeatedly fitted for each combination of bias parameter values. For example, in the case of multidimensional bias analysis for non-differential misclassification of a binary variable, we could consider different pairs of sensitivity and specificity values². A multidimensional bias analysis then outputs multiple bias-adjusted estimates. When there is

limited information about plausible values for the bias parameters, a tipping point analysis can be conducted to explore which combinations of values of the bias parameters would overturn study conclusions. This frames QBA not only as a method for estimating the potential impact of bias, but also as a tool for assessing the robustness of study conclusions by identifying how extreme the bias would need to be to meaningfully alter inferences.

In a probabilistic QBA, the researcher specifies a prior probability distribution for each bias parameter. Using this prior distribution, the researcher can specify information about the range of plausible values of the bias parameters, the value combinations that are most likely to occur, and the researcher's uncertainty about this information. For example, in the case of misclassification, rather than assigning fixed values for sensitivity and specificity as is done in deterministic QBA, the researcher can specify probability distributions (such as Beta distributions) that reflect both plausible ranges and uncertainty. These distributions can differ between cases and non-cases to allow for differential misclassification. Additionally, correlations among parameters (e.g., between sensitivity and specificity for cases) can be incorporated to reflect dependencies. By simulating draws from these joint distributions, the uncertainty in the bias parameters can be propagated through the analysis, yielding adjusted effect estimates with associated uncertainty intervals.

Two main approaches to probabilistic QBA are Bayesian bias analysis (where the prior distribution of the bias parameters is combined with the likelihood function for the data) and Monte Carlo bias analysis (where values of the bias parameters are directly sampled from their distribution and then used to fix the bias parameters to enable estimation of the bias-adjusted exposure effect) [23].

Probabilistic QBA generates an empirical distribution of bias-adjusted effect estimates, which can be summarised using point and interval estimates. The point estimate typically reflects the central tendency (e.g., mean or median) of the distribution under the QBA's assumptions. The interpretation of the interval estimate depends on the QBA approach: in a Bayesian approach, the interval (a credible interval) can be interpreted as having a specified probability of containing the true exposure effect, conditional on the model and prior. In contrast, in a Monte Carlo approach, the interval (described as a simulation interval [13]) reflects the variability induced by the simulation process and does not necessarily have a direct probabilistic interpretation about the true effect.

A multiple QBA (also known as a multiple-bias analysis) assesses the sensitivity of study results to multiple sources of bias such as mismeasurement, unmeasured confounding, and selection bias. A sequential multiple QBA adjusts for one bias at a time, where the order of

¹Although agreement-based metrics such as Cohen's κ [22] are sometimes reported for categorical variables, they are not suitable for use as bias parameters in QBA, since they measure agreement rather than the relationship between observed and true values.

²Note that differential misclassification would require separate parameter values for different subgroups, such as cases and controls within a case-control study.

adjustment should be based on the reverse order in which the biases occurred during the data generation process [24]. Note that the order of bias adjustments can affect the results of the multiple QBA. When the order of adjustment is in doubt, the researcher should assess sensitivity of the conclusions of the multiple QBA to different ordering of the bias adjustments [13]. A simultaneous multiple QBA avoids this issue because it adjusts for the multiple sources of bias simultaneously [25].

Methods

We searched for QBA software described in health research articles, as well as those available in software databases, published between 1st January 2014 and 1st May 2024 (inclusive). We selected this 10-year time frame to focus on recent tools that are methodologically current and more likely to be actively maintained.

We define a QBA as a method that adjusts for mismeasurement using a model that includes one or more bias parameters. Also, “software” is defined as a web tool, package, or code that is publicly available to use, is not specific to a particular data example, and is accompanied by documentation. To be classified as having “documentation,” a tool must provide enough information for users to understand its function and implementation without reliance on an external publication. This includes a user guide or in-code comments that explain the syntax, input requirements, and expected outputs. Examples of tools not meeting our software definition would be tools that were not publicly available, code files for specific examples that the researcher had to manually edit to apply to their study, and software that lacked documentation describing how to use the code, such as raw code files without explanatory comments.

We searched published literature from health research and software databases, as this is how most health researchers would identify methods and tools they can use. We did not look for methods within textbooks as these are not publicly available and often cannot be easily searched by applied researchers.

This review was written following the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines [26], and the PRISMA-ScR checklist can be found in Additional File 1. Our search was conducted in three steps: search implementation, eligibility screening and data extraction.

Publication search

In our first step, we used Web of Science to identify papers that mentioned all of the terms “measurement error”, “bias analysis” and “software” (or some other synonym of these terms) in either the title, abstract or as keywords. The specific search terms used can be seen

in Fig. 1 [27]. The search was applied to databases “Web of Science Core Collection”, “BIOSIS Citation Index”, “KCI-Korean Journal Database”, “MEDLINE” and “Scielo Citation Index”. We excluded from our search any meeting abstracts, clinical trials or patents. In addition, we excluded any articles published in journals outside of the fields of statistics, medicine, population health, and epidemiology and so deemed out of the scope of health research. A list of the excluded journals is given in Fig. 2.

Software repository search

In order to capture software implementations not mentioned in published literature, we completed several additional searches outside of Web of Science. We searched The Comprehensive R Archive Network (CRAN) [28], R’s central software repository containing a large collection of quality-assured contributed packages. We also searched IDEAS/RePEc [29], an online database indexing items of economics research including articles as well as Boston College’s Statistical Software Components (SSC) archive, which contains user-written Stata commands and other code. In addition, using Stata’s *search* command, we conducted a search of the Stata manuals, the Stata Journal, and all Stata-related user-written commands that are available via Stata’s *net* command.

We limited our results to tools which were first made publicly available (or had updated versions with new features implementing a QBA to mismeasurement) between 1st January 2014 and 1st May 2024 (inclusive). To implement this criterion, we manually checked the publication or update date of each tool and excluded any that fell outside the specified period.

CRAN

For our search of CRAN, we identified packages that mentioned both “measurement error” and “bias analysis” (or synonyms of these terms) in either their title or description. The R code used to implement this search is included in Additional File 2.

We first used R’s built-in CRAN package repository tools to extract the names, titles and descriptions of all of the packages maintained on CRAN on the search date, 8th May 2024. After cleaning the extracted text, removing new line breaks and any multiple spaces, we then used the R *grep* function, which searches for pattern matches to its argument, to search for those packages which mentioned both “measurement error” and “bias analysis” in their title or description. Search terms and synonyms used were equivalent to those in Fig. 1, in order to maintain consistency between our publication search and our repository search.

(TS=(measure* NEAR/3 error*) OR TS=(measure*-error) OR TS=(mismeasure*)
OR TS=(insensitive NEAR/0 measure*) OR TS=(unspecific NEAR/0 measure*) OR
TS=(information NEAR/0 bias) OR TS=(misclassif*) OR TS=(classif* NEAR/3 error)
OR TS=(miscode*) OR TS=(coding NEAR/0 mistake) OR TS=(coding NEAR/0
error) OR TS=(error NEAR/3 assessment) OR TS=(bias NEAR/0 measur*) OR
TS=(errant* NEAR/3 measur*) OR TS=(Berkson NEAR/3 error) OR TS=(Berkson-
error) OR TS=(miscategor*) OR TS=(categor* NEAR/3 error))
AND
(TS=(bias NEAR/0 analy*) OR TS=(sensitivity NEAR/0 analy*) OR TS=(bias NEAR/0
model*) OR TS=(uncertainty NEAR/0 analy*) OR TS=(bias NEAR/3 correct*) OR
TS=(bias NEAR/3 adjust*))
AND
(TS=(software) OR TS=(freeware) OR TS=(shareware) OR TS=(groupware) OR
TS=(Stata) OR TS=(Excel) OR TS=(Shiny) OR TS=(S-Plus) OR TS=(SPSS)
OR TS=(Stat-JR) OR TS=(MLwiN) OR TS=(command*) OR TS=(package*) OR
TS=(spreadsheet*) OR TS=(calculator*) OR TS=(desktop) OR TS=(mathematica)
OR TS=(fortran) OR TS=(R NEAR/3 program*) OR TS=(SAS NEAR/3 pro-
gram*) OR TS=(web NEAR/3 program*) OR TS=(online NEAR/3 program*) OR
TS=(computer NEAR/3 program*) OR TS=(virtual NEAR/3 program*) OR TS=(R
NEAR/0 function*) OR TS=(SAS NEAR/3 function*) OR TS=(web NEAR/3 func-
tion*) OR TS=(online NEAR/3 function*) OR TS=(computer NEAR/3 function*)
OR TS=(R NEAR/3 procedure*) OR TS=(SAS NEAR/3 procedure*) OR TS=(web
NEAR/3 procedure*) OR TS=(online NEAR/3 procedure*) OR TS=(computer NEAR/3
procedure*) OR TS=(virtual NEAR/3 procedure*) OR TS=(R NEAR/3 tool*) OR
TS=(SAS NEAR/3 tool*) OR TS=(web NEAR/3 tool*) OR TS=(online NEAR/3
tool*) OR TS=(computer NEAR/3 tool*) OR TS=(virtual NEAR/3 tool*) OR TS=(R
NEAR/3 macro*) OR TS=(SAS NEAR/3 macro*) OR TS=(web NEAR/3 macro*) OR
TS=(online NEAR/3 macro*) OR TS=(computer NEAR/3 macro*) OR TS=(virtual
NEAR/3 macro*) OR TS=(R NEAR/3 module*) OR TS=(SAS NEAR/3 module*) OR
TS=(web NEAR/3 module*) OR TS=(online NEAR/3 module*) OR TS=(computer
NEAR/3 module*) OR TS=(virtual NEAR/3 module*) OR TS=(R NEAR/3 code*)
OR TS=(SAS NEAR/3 code*) OR TS=(web NEAR/3 code*) OR TS=(online NEAR/3
code*) OR TS=(computer NEAR/3 code*) OR TS=(virtual NEAR/3 code*) OR TS=(R
NEAR/3 script*) OR TS=(SAS NEAR/3 script*) OR TS=(web NEAR/3 script*) OR
TS=(online NEAR/3 script*) OR TS=(computer NEAR/3 script*) OR TS=(virtual
NEAR/3 script*))

Fig. 1 Clarivate Web of Science search terms for this software review [27]

IDEAS/RePEc and Stata

The “advanced search” tool of IDEAS was less flexible than the R functions used to search CRAN, and so for this database we simplified our search strategy. We searched IDEAS for software components that had both

“measurement error” and “bias analysis” or their synonyms in any of their title, abstract or key words using the search string given in Fig. 3. We also used this same set of terms to search the Stata manuals, the implementation of which can be found in Additional File 3. We

- 2014 11th International Multi-Conference on Systems, Signals & Devices (SSD)
- 2014 IEEE Conference on Control Applications (CCA)
- 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)
- 2018 Conference on Precision Electromagnetic Measurements (CPEM 2018)
- 2018 IEEE International Symposium on Circuits and Systems (ISCAS)
- 2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE 2021)
- 27th International Conference on Metallurgy and Materials (METAL 2018)
- ACM Transactions on Programming Languages and Systems
- Aerosol Science and Technology
- Applied Engineering in Agriculture
- Applied Sciences-Basel
- Chemie Ingenieur Technik
- Climate of the Past
- Communication Methods and Measures
- Comptes Rendus Geoscience
- Computers & Fluids
- Earth Observing Systems XXII
- Ecography
- Energy and Buildings
- Engineering Structures
- Geochemistry Geophysics Geosystems
- Groundwater for Sustainable Development
- IEEE Access
- IEEE Transactions on Instrumentation and Measurement
- IEEE Transactions on Power Systems
- IEEE Transactions on Antennas and Propagation
- Informatica
- International Journal of Advanced Manufacturing Technology
- International Journal of Automation Technology
- International Journal of Hydrogen Energy
- International Journal of Extreme Manufacturing
- Journal of Applied Crystallography
- Journal of Hydrology
- Journal of Infrared Millimeter and Terahertz Waves
- Journal of Intelligent Transportation Systems
- Journal of Measurements in Engineering
- Journal of Mechanical Science and Technology
- Journal of Vacuum Science & Technology B
- Journal of Verification, Validation and
- Uncertainty Quantification
- Journal of Aerospace Engineering
- Management Science
- Mathematics
- Measurement Science and Technology
- Metrologia
- Political Analysis
- Proceedings of 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC 2019)
- Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference (2017 ELCON-RUS)
- Product-Focused Software Process Improvement, PROFES 2015
- Review of Scientific Instruments
- Sensors
- Spectrochimica Acta Part B-Atomic Spectroscopy
- Structural Design of Tall and Special Buildings
- Theoretical Foundations of Chemical Engineering
- Transportation Research Part B-Methodological
- VII Brazilian Congress on Metrology (Metrologia 2013)
- Water

Fig. 2 Journals outside of the scope of health research that were removed from our search results

considered all results which referenced either “measurement error” or “bias analysis” or their synonyms.

Eligibility criteria

In our second step, the eligibility of identified abstracts and tools was assessed independently by two reviewers (CW and RH), with any disagreements resolved by consensus. Abstracts and tools were eligible for data extraction if they satisfied all of the following criteria:

1. the abstract mentioned purpose-built software,
2. the abstract discussed bias due to mismeasurement,
3. the software implemented a QBA for mismeasurement.

Examples of abstracts that would be excluded were those that only mentioned programming languages or code for examples rather than providing a purpose-built tool, abstracts where a QBA was not conducted but mentioned as further work, and abstracts which had software for purposes other than a QBA for mismeasurement.

Data extraction

In our third step, we examined the full texts of the included published papers and the documentation of the packages found via our CRAN, Stata manual, Stata *net* command and IDEAS/RePEc database searches to extract information about any software presented and its features. We excluded any R packages that had been removed from CRAN, software that could not be loaded, and software with example code that failed to run due to unhandled errors. We also excluded any sensitivity analysis implementations that did not meet our definition for software or could not be considered a QBA due to not including at least one bias parameter.

Information was extracted on several data domains, reflecting tool characteristics and capabilities. An overview and description of the data collected is given in Fig. 4.

When assessing the documentation and usability of the tools, two reviewers (CW and RH) independently evaluated the level of detail in the documentation and the level

(“sensitivity analysis” | “bias analysis” | “sensitivity analyses” | “bias analyses” |
 “uncertainty analysis” | “uncertainty analyses” | “tipping”)
 +
 (“measurement error” | “measurement bias” | misclassification | misclassified |
 “measured with error” | “information bias” | mismeasure | mismeasured | errors)’

Fig. 3 IDEAS/RePEc and Stata manual search terms for this software review

<p>General information - the name of the tool, the year of development (or online publication), and the year of the most recent recorded update (where relevant).</p> <p>Environments and outputs - the implementation environment (e.g., R, Stata, SAS, Excel, or web-based), and whether outputs were tabular or graphical.</p> <p>Mismeasurement - the mismeasurement mechanisms addressed (differential or non-differential), whether multiple variables could be mismeasured simultaneously, and which variables were affected by mismeasurement (e.g., exposure, outcome, or covariates).</p> <p>Bias analysis - whether the tool implemented deterministic or probabilistic QBA, and (for deterministic QBAs) whether it supported multidimensional bias analysis.</p> <p>Multiple bias analysis - whether the tool could account for multiple sources of bias (e.g., misclassification, unmeasured confounding, and selection bias), and whether these were addressed simultaneously or sequentially.</p> <p>Data requirements and analysis of interest - the type of data required (individual-level, aggregate counts, or summary statistics such as estimated regression coefficients), the type of analysis the tool supported, and the nature of the observed variables (binary, categorical, or continuous).</p> <p>Documentation and usability - whether usage examples were provided, the level of documentation (categorised as minimal, moderate, or extensive), and the level of QBA knowledge required to use the tool effectively (essential or specialist).</p>
--

Fig. 4 Domains of data extraction for each software tool

of QBA knowledge required to perform a multidimensional or probabilistic QBA using the tool. Any disagreements were resolved by consensus.

Documentation was categorised into three levels; minimal, moderate and extensive. Tools classified as having minimal documentation provided only a brief description of the tool's purpose, required inputs, and syntax (where applicable). Moderate documentation included a full description of each function, at least one usage example, a written explanation of the output, and a detailed description of the method implemented. Tools at this level also provided practice datasets where applicable. Tools classified as having extensive documentation offered additional tutorial materials such as vignettes, video tutorials, or an accompanying software journal article.

The level of QBA knowledge required was classified as either essential or specialist. Requiring essential knowledge indicated that the tool fully implemented a multidimensional or probabilistic bias analysis and displayed the results without researchers having to manually code these steps. Specialist knowledge was deemed to be required when researchers had to manually implement or visualize a multidimensional or probabilistic

bias analysis. Alternatively, the tool may have required expertise in Bayesian methods, such as defining priors or assessing the convergence of MCMC samplers.

Results

Publication search

After removal of duplicates, our initial Web of Science search returned 254 results. We then excluded 110 papers when restricting to publications made between 1st January 2014 and 1st May 2024. A further 63 papers were manually excluded that were published in journals outside of the scope of health research (as listed in Fig. 2). We were left with a total of 81 abstracts. This initial search step is illustrated in Fig. 5.

We excluded 37 abstracts which did not provide a purpose-built statistical software implementation, 10 abstracts that did not focus on bias due to mismeasurement, and nine abstracts where the software provided was not conducting a QBA for mismeasurement (e.g., the QBA was instead for an alternative form of bias). The abstract screening process is illustrated in Fig. 6. When reviewing the full text of the remaining 25 articles, we found references to 24 unique software tools.

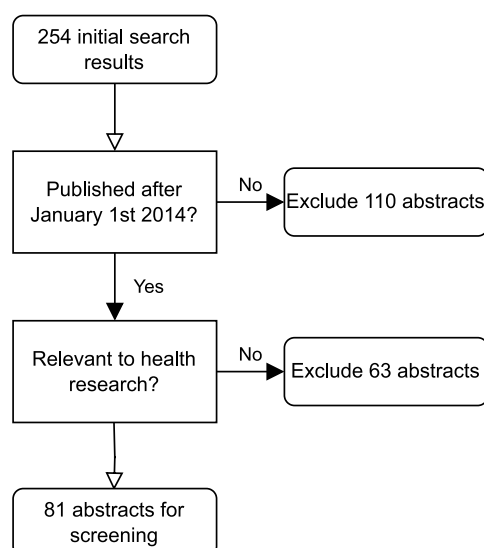


Fig. 5 Flowchart of the publication search step of the review

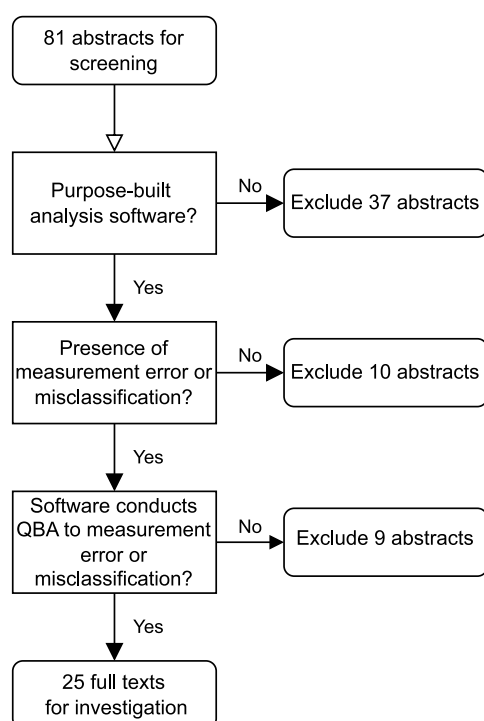


Fig. 6 Flowchart of the abstract screening step of the review

Software repository search

Our IDEAS search identified a single Stata command, *episens* [30, 31], however, the mismeasurement-related functions of this tool had not been updated since 2008, and so the tool fell outside of our date range for eligibility. We discuss this tool further in [Grey literature](#) section, and in Table 2.

Our CRAN search returned ten software tools, four of which had also been identified by our Web of Science

search. The tool *multibias* [32] also referenced an additional web tool implementation in its documentation, *multibias web tool* [33]. Thus, in total, our CRAN search provided an additional seven tools.

Our search using the Stata *search* command returned an initial 205 results, 136 of which were either duplicates or were outside of our date range and so were excluded. From the 69 remaining search results, four eligible tools were found.

Eligibility criteria

In total, 35 unique software tools were identified across all of our searches, the process of which is summarised in Fig. 7. Among these 35 tools, nine were excluded because they did not implement a QBA to mismeasurement, four were excluded because they did not provide sufficient documentation or were code for a specific example requiring user adjustment³, one tool was excluded because it failed to run example code due to an unresolved runtime error at the time of review, and one tool was excluded because it had been removed from CRAN. The remaining 20 met our inclusion criteria.

Of these 20 tools, web tool *APScalculator* [36], Stata command *bamm* [37], and R package *SensiPhy* [38] do not implement a QBA for an exposure effect estimate of an epidemiological study and so are excluded from the results presented in Table 1. The web tool *APScalculator* evaluates the impact of classical measurement error ([6], Chapter 1) on the categorization of a continuous variable into user-specified groups, rather than directly assessing bias in effect estimates. The R package *SensiPhy* estimates the impact of various sources of uncertainty in phylogenetic comparative methods used within ecology, which differs substantially from applications in health research. The Stata command *bamm* conducts a Bayesian bias analysis to investigate the distribution of a single misclassified categorical variable, which could be either nominal or ordinal.

Overview of included tools

Table 1 summarises the key features of the 17 total software programs we found that are applicable to health studies aiming to quantify bias in an effect estimate, in order of most recent update.

Environments and outputs

Of the tools reviewed, eight (47%) are implemented as R packages, six (35%) are web-based applications, and three (18%) are Stata commands.⁴

³One of these tools, *Short code* [34, 35], is discussed further in [Grey literature](#) section and included in Table 2, due to its prominence in core QBA literature [13].

⁴We note here that the R package *EValue* has a corresponding Stata implementation *evalue* [60] and web tool [61, 62]. At the time of writing, neither

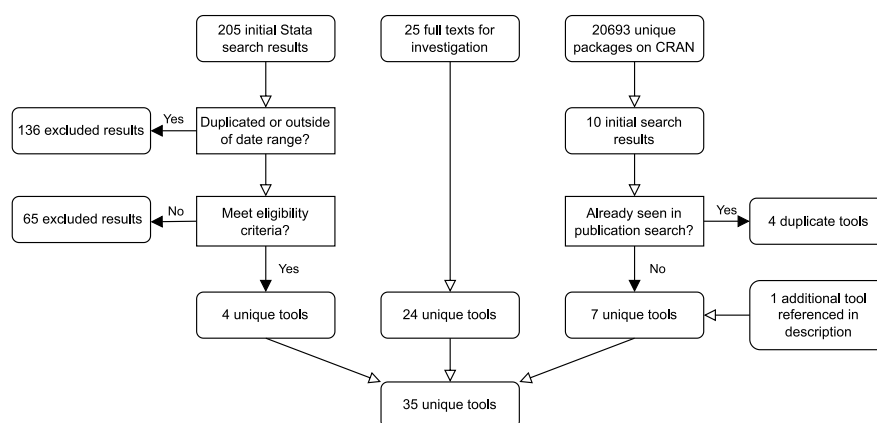


Fig. 7 Flowchart of the software tool search process

Approximately half of the tools (eight) provide both graphical and tabular outputs to aid the researcher in interpreting the results. Only three tools do not produce tables: web tool *SAMBA-EHR*, web tool *SensitivityAnalysis* and R package *ConMed*. Three R packages and three Stata commands do not provide graphical plots of their results.

Mismeasurement

Of the 17 tools, four implement a QBA in the case of measurement error in a continuous variable (R packages *miCoPTCM*, *ConMed* and *rcme*, and web tool *MediationSensitivityAnalysis*). Only R package *rcme* allows for multiplicative measurement error (i.e., error that scales with the true value of the variable), whilst the rest of the tools employ a classical additive measurement error model ([6], Chapter 1). A total of 12 tools apply a QBA for misclassification of a binary variable. Only one tool, R package *mgee2*, implements a QBA when the misclassified variable has more than two categories.

In total, 11 software tools handle cases of outcome mismeasurement, 12 handle exposure mismeasurement, and seven handle mismeasurement in other covariates (such as effect modifiers, mediators, or potential confounders). Among these, two tools (web tools *SAMBA-EHR* and *Outcome Misclassification*) focus solely on outcome mismeasurement, four (Stata commands *pvw* and *ivbounds*, R package *BayesSenMC*, and *multibias web tool*) exclusively handle exposure mismeasurement, and only the web tool *SensitivityAnalysis* is specific to misclassification of a confounder.

Among the tools, nine are applicable for both differential and non-differential mismeasurement, while seven are for non-differential mismeasurement only. The R package *EValue* is specific to differential misclassification. Multiple variables can be mismeasured simultaneously in seven (41%) of the tools.

Bias analysis

A deterministic QBA is implemented exclusively in 11 tools, two (web tool *SensitivityAnalysis* and R package *BayesSenMC*) support only a probabilistic QBA, and three include options to implement both a deterministic and probabilistic QBA (R package *episensr*, web tool *apisensr*, and *multibias web tool*⁵). Among the tools that implement a probabilistic QBA, only R package *BayesSenMC* performs a Bayesian bias analysis, and the remaining tools perform a Monte Carlo bias analysis. Among the tools that implement a deterministic QBA, only six (40%) perform a multidimensional analysis.

Multiple bias analysis

The Stata command *biasepi*, web tools *MediationSensitivityAnalysis* and *multibias web tool*, and the R packages *EValue*, *ConMed*, *episensr*, and *multibias* can all perform a multiple QBA for mismeasurement and unmeasured confounding. Additionally, Stata command *biasepi*, R package *EValue*, R package *multibias*, and *multibias web tool* can also adjust for selection bias. Note that all tools except R package *EValue*, R package *episensr*, and Stata command *biasepi* adjust for multiple sources of bias simultaneously; these tools instead apply a sequential approach.

of these tools implement a QBA to mismeasured variables; instead, they focus on bias due to unmeasured confounding, so we have not included them here. See [14] for further details on these tools for QBA to unmeasured confounding. The web tool *SAMBA-EHR* also exists alongside R package *SAMBA* [63]. This R package does not implement a QBA to mismeasured variables and so has not been reviewed here.

⁵The R package *multibias* does not perform a probabilistic QBA, but does provide guidance and example code demonstrating how a researcher can manually implement a probabilistic QBA using the tool. However, *multibias web tool* provides a function which performs a probabilistic QBA.

Table 1 Software programs implementing a quantitative bias analysis for mismeasurement published between 2014 and 2024

Software name (Year)	Environment	Output		Mismeasurement			Bias analysis	
		Table	Plot	Type	Differential	Multiple mismeasured variables	Type	Multi dimensional
<i>biasepi</i> (2019) [39, 40]	Stata command	✓	-	MC	D, ND	-	Det	-
<i>pvw</i> (2019) [41]	Stata command	✓	-	MC	D, ND	-	Det	-
<i>SAMBA-EHR</i> (2020) [42, 43]	Web tool	-	✓	MC	ND	-	Det	✓
<i>Outcome Misclassification</i> (2020) [44]	Web tool	✓	✓	MC	D, ND	-	Det	✓
<i>SensitivityAnalysis</i> (2020) [45]	Web tool	-	✓	MC	ND	-	Prob	-
<i>miCoPTCM</i> (2016, upd. 2020) [46]	R package	✓	-	ME	ND	✓	Det	-
<i>ivbounds</i> (2021) [47, 48]	Stata command	✓	-	MC	ND	-	Det	✓
<i>MediationSensitivityAnalysis</i> (2021) [49]	Web tool	✓	✓	ME	ND	✓	Det	✓
<i>BayesSenMC</i> (2021) [50]	R package	✓	✓	MC	D, ND	-	Prob	-
<i>EValue</i> (2017, upd. 2021) [51, 52]	R package	✓	-	MC	D	✓	Det	-
<i>ConMed</i> (2023) [53]	R package: non-CRAN	-	✓	ME	ND	✓	Det	-
<i>episensr</i> (2015, upd. 2023) [54]	R package	✓	✓	MC	D, ND	✓	Det, Prob	✓
<i>apisensr</i> (2021, upd. 2023) [55, 56]	Web tool	✓	✓	MC	D, ND	-	Det, Prob	✓
<i>mgee2</i> (2020, upd. 2023) [57, 58]	R package	✓	✓	MC	ND	✓	Det	-
<i>multibias web tool</i> (2023) [25, 33]	Web tool	✓	✓	MC	D, ND	-	Det, Prob	-
<i>multibias</i> (2023, upd. 2024) [25, 32]	R package	✓	-	MC	D, ND	✓	Det	-
<i>rcme</i> (2023, upd. 2024) [59]	R package: non-CRAN	✓	✓	ME	D, ND	-	Det	-

Software name (Year)	Analysis				
	Data type	Analysis of interest	Outcome	Exposure	Other covariates
<i>biasepi</i> (2019) [39, 40]	Individual, Aggregate	Contingency table	bin	bin	-
<i>pvw</i> (2019) [41]	Individual	Regression	bin	bin	bin, cat, cts
<i>SAMBA-EHR</i> (2020) [42, 43]	Summary	Regression	bin	cts, cat	-
<i>Outcome Misclassification</i> (2020) [44]	Summary	Contingency table	bin	bin	-
<i>SensitivityAnalysis</i> (2020) [45]	Summary	Regression	cts	bin	bin
<i>miCoPTCM</i> (2016, upd. 2020) [46]	Individual	Survival analysis	TTE	cts	cts
<i>ivbounds</i> (2021) [47, 48]	Individual	Instrumental variable analysis	bin, cat, cts	bin	bin, cat, cts
<i>MediationSensitivityAnalysis</i> (2021) [49]	Summary	Mediation analysis	cts	bin	Mediator (cts), other (cts)
<i>BayesSenMC</i> (2021) [50]	Aggregate	Contingency table	bin	bin	-
<i>EValue</i> (2017, upd. 2021) [51, 52]	Summary	Regression	bin	bin	-
<i>ConMed</i> (2023) [53]	Summary	Mediation analysis	cts	cts	Mediator (cts), other (bin, cat, cts)
<i>episensr</i> (2015, upd. 2023) [54]	Aggregate	Contingency table	bin, TTE	bin	bin
<i>apisensr</i> (2021, upd. 2023) [55, 56]	Aggregate	Contingency table	bin	bin	bin
<i>mgee2</i> (2020, upd. 2023) [57, 58]	Individual	Longitudinal analysis	cat (ordinal)	cat (ordinal), bin, cts	cat (ordinal), bin, cat, cts
<i>multibias web tool</i> (2023) [25, 33]	Individual	Regression	bin	bin	bin, cat, cts
<i>multibias</i> (2023, upd. 2024) [25, 32]	Individual	Regression	bin	bin	bin, cat, cts
<i>rcme</i> (2023, upd. 2024) [59]	Individual	Regression	cts	cts	cts

Software name (Year)	Usage		
	Level of documentation	Examples	QBA knowledge required
<i>biasepi</i> (2019) [39, 40]	Moderate	✓	Specialist
<i>pvw</i> (2019) [41]	Moderate	✓	Specialist
<i>SAMBA-EHR</i> (2020) [42, 43]	Extensive	✓	Essential
<i>Outcome Misclassification</i> (2020) [44]	Extensive	✓	Essential
<i>SensitivityAnalysis</i> (2020) [45]	Moderate	✓	Essential
<i>miCoPTCM</i> (2016, upd. 2020) [46]	Moderate	✓	Specialist

Table 1 (continued)

Software name (Year)	Usage		
	Level of documentation	Examples	QBA knowledge required
<i>ivbounds</i> (2021) [47, 48]	Extensive	✓	Essential
<i>MediationSensitivityAnalysis</i> (2021) [49]	Minimal	-	Essential
<i>BayesSenMC</i> (2021) [50]	Extensive	✓	Specialist
<i>EValue</i> (2017, upd. 2021) [51, 52]	Extensive	✓	Essential
<i>ConMed</i> (2023) [53]	Moderate	✓	Essential
<i>episensr</i> (2015, upd. 2023) [54]	Extensive	✓	Essential
<i>apisensr</i> (2021, upd. 2023) [55, 56]	Extensive	✓	Essential
<i>mgee2</i> (2020, upd. 2023) [57, 58]	Moderate	✓	Specialist
<i>multibias web tool</i> (2023) [25, 33]	Minimal	-	Essential
<i>multibias</i> (2023, upd. 2024) [25, 32]	Extensive	✓	Specialist
<i>rcme</i> (2023, upd. 2024) [59]	Minimal	✓	Specialist

Bold text for the “Outcome”, “Exposure” and “Other covariates” columns indicates the variable can be considered mismeasured by the tool. “Aggregate” data here means any non-individual level data e.g., count data. “Summary” data here means statistics calculated from data, e.g., regression coefficients or standard deviations

Abbreviations used (in alphabetical order): *D* Differential, *Det* Deterministic, *MC* Misclassification, *ME* Measurement error, *ND* Non-differential, *Prob* Probabilistic, *bin* Binary, *cat* Categorical, *cts* Continuous, *TTE* Time-to-event, *upd* Updated

Data requirements and analysis of interest

All tools except Stata command *biasepi*, which can take either individual or aggregate data as inputs, are specific in the data type required. Individual-level data is required by seven of the tools, aggregated count data by four tools, and summary statistics (such as regression coefficients or other statistics derived from the data) are required by six of the tools.

When the analysis of interest is a mediation analysis, two tools are applicable: web tool *MediationSensitivityAnalysis* and R package *ConMed*. The web tool *MediationSensitivityAnalysis* performs bias analysis for measurement error in the outcome, mediator, or other observed covariates (effect modifiers or potential confounders), with the assumption that the binary exposure is measured without error. The R package *ConMed* adjusts for measurement error in the mediator or outcome, and is applicable specifically when there is unmeasured confounding as well as measurement error.

For other types of analysis of interest, R package *miCoPTCM* accounts for measurement error of a continuous covariate or exposure in survival analysis (specifically a promotion time cure model), where the outcome is a time-to-event variable. Stata command *ivbounds* applies QBA for an instrumental variable analysis, allowing for a binary or categorical instrumental variable. R package *mgee2* conducts QBA for a longitudinal analysis, where there are changes within the same individuals or groups over time. Of the remaining tools, five are applicable for the analysis of contingency tables and seven for logistic or linear regression.

Most tools require the outcome variable of the analysis of interest to be either binary (nine programs) or continuous (five programs). Only R package *mgee2* and Stata command *ivbounds* allow for a discrete outcome variable with more than two categories. The exposure variable of

the analysis of interest is typically required to be exclusively either binary (nine programs) or continuous (four programs), but two programs (web tool *SAMBA-EHR* and R package *mgee2*) allow exposure variables to be of multiple types including both discrete or continuous options. Of all of the tools, 12 (71%) allow for the inclusion of other covariates than just the exposure and outcome in the analysis.

Documentation and usability

Of the 17 tools reviewed, only two do not include usage examples, both of which are web tools. Documentation quality varies: eight tools have extensive documentation, six have moderate documentation, and three have minimal documentation. Ten tools fully implement a multidimensional or probabilistic QBA (i.e., user only requires essential QBA knowledge as the software implements all steps of the QBA including summaries of the results). However, seven tools require users to have specialist knowledge (e.g., a tool only performs a simple bias analysis and so a user must write their own code to conduct a probabilistic QBA using this tool).

Grey literature

In our formal search we focused on software described in the published literature between January 1st 2014 and May 1st 2024, or software made available during this period via CRAN, the Stata manuals, the IDEAS/RePEc database, or other Stata user-written commands available using Stata's *net* command. This approach ensured we captured the latest tools which applied researchers could readily locate. However, we recognise that additional software exists that was not identified through this search. For example, some tools have not been mentioned in journal articles and are hosted in alternative environments such as web-based platforms. Others

Table 2 Software tools for mismeasurement correction and QBA not captured by our formal search strategy

Tool	Environment	Brief description
<i>Quantitative Bias Analysis</i> [64]	Web tool	Conducts simple and multidimensional bias analyses for misclassification and unmeasured confounding.
<i>CMAverse</i> [65, 66]	R package (non-CRAN)	Conducts sensitivity analyses for unmeasured confounding, measurement error, and selection bias in causal mediation analysis.
<i>episens</i> [30, 31]	Stata	Provides basic sensitivity analysis of the observed relative risks, adjusting for unmeasured confounding and misclassification of the exposure.
<i>eivtools</i> [67]	R package	Functions for analysis with error-prone covariates.
<i>Prob Bias Analysis for Information Bias</i> [68]	Web tool	Probabilistic bias analysis for misclassification; an R based web interface for <i>episensr</i> [54].
<i>simplemba</i> [69]	Web tool	Odds ratio calculator with misclassification, selection bias, and confounding adjustment.
<i>sensmac</i> [70, 71]	SAS	Implements probabilistic sensitivity analysis to misclassification of a binary variable.
<i>Multiple bias model</i> [72, 73]	SAS	Code and dataset for conducting multiple bias modelling.
<i>Mecor</i> [74]	R package	Functions to perform covariate measurement error correction.
<i>Short code</i> [34, 35]	R and SAS code	Code to perform record and summary level QBA, for misclassification and confounding.
<i>Misclassification spreadsheet</i> [13, 75]	Excel spreadsheet	Spreadsheet to perform simple bias analysis for misclassification in contingency tables.

were developed before 2014 and have not been significantly updated since but remain in use. Table 2 gives a non-exhaustive brief overview of some tools known to the authors that were not captured by our formal search strategy, but that may be of interest to researchers.

Short code [34, 35], which consists of SAS and R code, is included in Table 2 despite having been initially captured in our formal search and later excluded based on our inclusion criteria; specifically, due to a lack of adequate documentation (see [Eligibility criteria](#) section). It is the only tool in Table 2 that passed initial screening stages but was actively excluded from the final review; the rest were not captured by our formal search. We include it here due to its prominence in the QBA literature and its potential value for learners, particularly when used alongside [13].

Another notable inclusion in Table 2 is the *Misclassification spreadsheet* Excel tool [75], which accompanies [13]. Although the spreadsheet was not captured by our

formal search due to its distribution via a textbook website rather than publication or software repositories, we note that it has been widely disseminated (including, since this review was conducted, being directly linked in methodological guidance [76]). This spreadsheet remains a valuable and accessible tool for introducing researchers to the mechanics of QBA. Similarly, the *episens* command for Stata, originally released in 2008 [30], remains relatively widely used in the Stata community for QBA to exposure misclassification.

Discussion

We have conducted an up-to-date review of software implementations of QBA to mismeasurement described in the published literature, R packages available on CRAN, and Stata commands, including user-written commands available from the SSC archive or via Stata's *net* command. All software tools were developed or significantly updated post-2019, with most (65%) having been developed or updated since 2021. The software tools were either R packages, Stata commands or online web tools and were available for routine analyses of interest such as linear regression and contingency tables, and for more specialized analyses such as mediation analysis, instrumental variable analysis and survival analysis. All but one software tool implemented a QBA to non-differential mismeasurement with just over half applicable for differential and non-differential mismeasurement. Also, more than half of the software tools implement a QBA for misclassification of a binary variable. Although most software tools implemented a deterministic bias analysis, only six (40%) of these tools included features to allow the user to perform a multidimensional bias analysis. Most tools provided usage examples, but documentation quality varied from minimal to extensive. While several tools offered comprehensive guidance, including tutorials and vignettes, others provided only brief descriptions of inputs and outputs. Just over half of the tools implemented all the steps of a multidimensional or probabilistic QBA for the user, but a subset required specialist QBA knowledge, such as understanding Bayesian priors or manually implementing a multidimensional bias analysis.

Remarks on review limitations and future work

Our review did not assess the extent to which each tool is used in practice. While citation counts or software download statistics (e.g., using the R package *cranlogs*) could offer useful proxies for uptake, such metrics are not consistently available across different environments. Future work could explore quantifying the uptake of QBA tools in applied research and identifying factors associated with broader usage.

This review did not set out to evaluate the performance of the software tools or verify their outputs and

so we cannot comment on their runtime behaviour. Tools which failed to produce outputs due to unhandled errors were excluded. However, we note that for some probabilistic QBAs, particularly those addressing misclassification, admissible bias-adjusted results may not always be obtainable depending on the combination of input assumptions. In such cases, some tools (e.g., R package *episensr*) are designed to filter out inadmissible results, providing informative error messages. Other tools may not include such error handling or may return non-specific error messages. This highlights an important practical consideration for users: the interpretability and robustness of software feedback can vary and may impact the usability of tools in real-world settings.

Although the focus of this review is on software implementation, we also acknowledge that another potential barrier to wider uptake of QBA is the challenge of specifying plausible bias parameter values and priors when they are not clearly identifiable from existing literature or validation data [15].

While these issues lie outside the scope of our review, applied examples that demonstrate not only the use of the software but also how plausible parameter values and priors might be selected could help reduce both barriers. We suggest that future works could build on this review by including such examples, clarifying in which cases each tool is most appropriate and highlighting differences in error handling and user support features.

Another area for expansion of this work would be broadening the scope of our searches. A limitation of this review is that we restricted our software repository searches to R packages available on CRAN and Stata commands available via the SSC archive or the *net* command. As such, we may have missed some tools available only on platforms such as GitHub or implemented in other programming environments such as Python, and without accompanying publications. These tools could be of interest to computationally oriented researchers, but may be unlikely to be widely adopted in applied epidemiological settings without mention in a formal publication or inclusion in repositories.

We also restricted our publication search to epidemiology, statistics, and health journals. This may have excluded software from other disciplines that could be applicable to health research. The numbers of these tools would likely be small, however, as different fields face very different complications and considerations to health research. Further, it is unlikely that health researchers would look outside of their field in order to find tools for use. Conducting similar reviews in domains such as psychology, engineering, or computational biology could increase awareness of potentially useful tools across disciplines.

Additionally, as discussed in [Grey literature](#) section, our review focused on tools developed or updated between 2014 and 2024, which may have led to the exclusion of older software still in use. This time period was chosen to prioritise tools likely to remain supported and aligned with current statistical practice. However, useful legacy tools may have been missed. While Table 2 provides a brief overview of additional tools not captured in our formal search, including some which fell outside of this time period (e.g., popular Stata command *episens*), we did not conduct a full evaluation of these tools. This reflects a decision to focus on tools that met our inclusion criteria; however, we acknowledge that some of these additional tools may still be useful in practice.

Future work could build on our review by systematically identifying and reviewing in detail tools that fall outside standard dissemination channels or have not been recently updated.

Remarks on implications for QBA uptake

Although previous work has suggested that implementation challenges in QBA have largely been addressed [21], our findings indicate significant gaps remain. Existing tools rarely support QBA for categorical variables with more than two levels, and few address measurement error in continuous variables beyond classical error models. Further, many of the tools reviewed were designed for specific use cases, often allowing only a single data type or a single type of bias analysis. A notable proportion of the tools only handle non-differential mismeasurement, despite growing emphasis in the literature that differential mismeasurement is common and may introduce complex bias in real-world studies [3, 4, 8, 77]. Tools capable of addressing differential mismeasurement may therefore be of greater practical value, particularly in applied epidemiologic research. Developing or expanding software to cover these scenarios, and to handle multiple potential mismeasurement types, would improve accessibility and could lead to greater uptake by applied researchers.

While our formal search identified tools implemented in R, Stata and web environments, none were implemented in SAS (see Table 2 for details of some SAS macros not captured by our formal search strategy). Given the use of SAS in many organisations (such as government agencies, healthcare institutions, and other applied research settings), this gap in platform availability may limit the accessibility of QBA tools to some users. Increasing software availability across platforms could support broader adoption in applied research contexts.

In addition to gaps in available methods and environments, level of documentation may pose a barrier to effective tool use. While we assessed documentation quality and the level of QBA knowledge required

separately, these aspects of tools are likely to be related. More extensive, good quality documentation may help to lessen the need for specialist expertise by supporting learning and correct use, while minimal documentation can make a tool inaccessible even if specialist knowledge is not required. This suggests that improving documentation could make existing tools more accessible to a broader range of researchers, especially those less familiar with QBA.

Further, although all tools met our definition of “software” by including at least minimal documentation, we found that documentation often did not explicitly state key assumptions about the data. For example, despite eight tools allowing multiple variables to be simultaneously measured with error, only R package *episensr* explicitly stated that errors were required to be independent. The lack of clear statements about underlying assumptions forces researchers to rely on prior methodological knowledge or manual code inspection, increasing the potential risk of misuse or misinterpretation. This not only exacerbates the broader issue of unacknowledged dependent error in epidemiological studies [78] but may also create a barrier to the adoption of tools by applied researchers. Addressing these gaps by explicitly stating assumptions and tool limitations in documentation could facilitate wider adoption and correct application of QBA methods.

The substantial number of tools in Table 2 that were not identified through our formal search highlights the challenge for applied researchers in discovering relevant QBA software. Many of these tools would be difficult for researchers unfamiliar with mismeasurement and quantitative bias analysis to discover, reinforcing the need for greater visibility of tools. Publishing software tools in widely recognized repositories, maintaining clear documentation of updates and expansions, and encouraging researchers to cite software in their outputs would help bridge this gap.

Conclusions

Our review highlights an increase in the number of software tools for QBA to mismeasurement but also reveals important gaps that may limit their accessibility and applicability. While many tools support common analyses and provide extensive documentation, others lack clarity on key assumptions, require specialist knowledge, or are restricted to specific use cases. There is a lack of tools for handling misclassification of categorical variables and for addressing non-classical measurement error in continuous variables. Improved documentation, broader methodological coverage, and increased visibility through publication in software journals and good citation practices could enhance the usability and adoption of these tools. Future efforts should focus on developing

more comprehensive software tools and ensuring that researchers can easily identify and apply appropriate programs for addressing mismeasurement in their studies.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-025-02635-w>.

Additional file 1: PRISMA-ScR Checklist. Filled PRISMA-ScR checklist for our scoping review methodology.

Additional file 2: CRAN Search Code. R code used to conduct the CRAN search stage.

Additional file 3: Stata Search Code. Stata code used to conduct the search of the Stata manual, Stata journal and user-written code.

Authors' contributions

RAH proposed and designed the study. CJCW carried out the statistical analyses with participation from RAH. CJCW, RAH, KT, and JWB drafted the manuscript. All authors read and approved the final version of the manuscript.

Funding

CJCW is supported by the Engineering and Physical Sciences Research Council (EPSRC) (grant EP/S023569/1). RAH is supported by a Sir Henry Dale Fellowship that is jointly funded by the Wellcome Trust and the Royal Society (grant 215408/Z/19/Z). KT works in the MRC Integrative Epidemiology Unit, which is supported by the University of Bristol and the UK Medical Research Council (grant MC_UU_00032/2). JWB is supported by the UK Medical Research Council (grant MR/T023953/1).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 11 April 2025 / Accepted: 8 July 2025

Published online: 01 August 2025

References

1. Shaw PA, Deffner V, Keogh RH, Tooze JA, Dodd KW, Küchenhoff H, et al. Epidemiologic Analyses with Error-Prone Exposures: Review of Current Practice and Recommendations. *Ann Epidemiol*. 2018;28(11):821–8. <https://doi.org/10.1016/j.annepidem.2018.09.001>.
2. Lash TL, VanderWeele TJ, Haneuse S, Rothman K. *Modern Epidemiology*. 4th ed. Philadelphia: Wolters Kluwer Health; 2021.
3. Jurek AM, Greenland S, Maldonado G, Church TR. Proper Interpretation of Non-Differential Misclassification Effects: Expectations vs Observations. *Int J Epidemiol*. 2005;34(3):680–7. <https://doi.org/10.1093/ije/dyi060>.
4. Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-Measurement Error Is Frequently Ignored When Interpreting Epidemiologic Study Results. *Eur J Epidemiol*. 2006;21(12):871–6. <https://doi.org/10.1007/s10654-006-9083-0>.
5. Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Interdisciplinary Statistics Series. Boca Raton: Chapman & Hall/CRC; 2004.

6. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement Error in Nonlinear Models: A Modern Perspective. 2nd ed. New York: Chapman and Hall/CRC; 2006. <https://doi.org/10.1201/9781420010138>.
7. Kristensen P. Bias from Nondifferential but Dependent Misclassification of Exposure and Outcome. *Epidemiology* (Cambridge, Mass). 1992;3(3):210–5. <https://doi.org/10.1097/00001648-199205000-00005>.
8. Keogh RH, Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, et al. STRATOS Guidance Document on Measurement Error and Misclassification of Variables in Observational Epidemiology: Part 1—Basic Theory and Simple Methods of Adjustment. *Stat Med*. 2020;39(16):2197–231. <https://doi.org/10.1002/sim.8532>.
9. Innes GK, Bhondokhan F, Lau B, Gross AL, Ng DK, Abraham AG. The Measurement Error Elephant in the Room: Challenges and Solutions to Measurement Error in Epidemiology. *Epidemiol Rev*. 2021;43(1):94–105. <https://doi.org/10.1093/epirev/mxab011>.
10. Brakenhoff TB, Mitroiu M, Keogh RH, Moons KGM, Groenwold RHH, Van Smeden M. Measurement Error Is Often Neglected in Medical Literature: A Systematic Review. *J Clin Epidemiol*. 2018;98:89–97. <https://doi.org/10.1016/j.jclinepi.2018.02.023>.
11. Buonaccorsi JP. Measurement Error: Models, Methods, and Applications. New York: Chapman and Hall/CRC; 2010. <https://doi.org/10.1201/9781420066586>.
12. Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, Keogh RH, et al. STRATOS Guidance Document on Measurement Error and Misclassification of Variables in Observational Epidemiology: Part 2—More Complex Methods of Adjustment and Advanced Topics. *Stat Med*. 2020;39(16):2232–63. <https://doi.org/10.1002/sim.8531>.
13. Fox MP, MacLehose RF, Lash TL. Applying Quantitative Bias Analysis to Epidemiologic Data. 2nd ed. Statistics for Biology and Health. Cham: Springer International Publishing; 2021. <https://doi.org/10.1007/978-3-030-82673-4>.
14. Kawabata E, Tilling K, Groenwold RHH, Hughes RA. Quantitative Bias Analysis in Practice: Review of Software for Regression with Unmeasured Confounding. *BMC Med Res Methodol*. 2023;23(1):111. <https://doi.org/10.1186/s12874-023-01906-8>.
15. 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;50(5):1708–30. <https://doi.org/10.1093/ije/dyab061>.
16. Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, Lapane KL. Probabilistic Bias Analysis in Pharmacoepidemiology and Comparative Effectiveness Research: A Systematic Review. *Pharmacoepidemiol Drug Saf*. 2016;25(12):1343–53. <https://doi.org/10.1002/pds.4076>.
17. Leahy TP, Kent S, Sammon C, Groenwold RH, Grieve R, Ramagopal S, et al. Unmeasured Confounding in Nonrandomized Studies: Quantitative Bias Analysis in Health Technology Assessment. *J Comp Eff Res*. 2022;11(12):851–9. <https://doi.org/10.2217/ce-2022-0029>.
18. Leahy TP, Durand-Zaleski I, Sampietro-Colom L, Kent S, Zöllner Y, Coyle D, et al. The Role of Quantitative Bias Analysis for Nonrandomized Comparisons in Health Technology Assessment: Recommendations from an Expert Workshop. *Int J Technol Assess Health Care*. 2023;39(1):e68. <https://doi.org/10.1017/S0266462323002702>.
19. Bond JC, Fox MP, Wise LA, Heaton B. Quantitative Assessment of Systematic Bias: A Guide for Researchers. *J Dent Res*. 2023;102(12):1288–92. <https://doi.org/10.1177/00220345231193314>.
20. Brown JP, Hunnicutt JN, Ali MS, Bhaskaran K, Cole A, Langan SM, et al. Quantifying possible bias in clinical and epidemiological studies with quantitative bias analysis: common approaches and limitations. *BMJ*. 2024;385. <https://doi.org/10.1136/bmj-2023-076365>.
21. 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. <https://doi.org/10.1093/ije/dyu149>.
22. McHugh ML. Interrater Reliability: The Kappa Statistic. *Biochemia Med*. 2012;22(3):276–82. <https://doi.org/10.11613/BM.2012.031>.
23. McCandless LC, Gustafson P. A Comparison of Bayesian and Monte Carlo Sensitivity Analysis for Unmeasured Confounding. *Stat Med*. 2017;36(18):2887–901. <https://doi.org/10.1002/sim.7298>.
24. Phillips CV. Quantifying and Reporting Uncertainty from Systematic Errors. *Epidemiology*. 2003;14(4):459. <https://doi.org/10.1097/01.eid.0000072106.65262.ae>.
25. Brendel P, Torres A, Arah OA. Simultaneous Adjustment of Uncontrolled Confounding, Selection Bias and Misclassification in Multiple-Bias Modelling. *Int J Epidemiol*. 2023;52(4):1220–30. <https://doi.org/10.1093/ije/dyad001>.
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467–73. <https://doi.org/10.7326/M18-0850>.
27. Wood C, Hughes R, Tilling K, Bartlett J. Software Tools for Quantitative Bias Analysis on Measurement Error or Misclassification (Web of Science Core Collection). searchRxiv. 2023. <https://doi.org/10.1079/searchRxiv.2023.00202>.
28. R Core Team. The Comprehensive R Archive Network. <https://cran.r-project.org/>. Accessed 08 Aug 2024.
29. Zimmermann C. Economics and Finance Research| IDEAS/RePEc. <https://ideas.repec.org/>. Accessed 07 Apr 2025.
30. Orsini N, Bellocco R, Bottai M, Wolk A, Greenland S. A Tool for Deterministic and Probabilistic Sensitivity Analysis of Epidemiologic Studies. *Stata J*. 2008;8(1):29–48. <https://doi.org/10.1177/1536867X0800800103>.
31. Orsini N, Bellocco R, Greenland S. EPISENS: Stata Module for Basic Sensitivity Analysis of Epidemiological Results. Statistical Software Components, Boston College Department of Economics. 2024. <https://ideas.repec.org/c/boc/bocode/s456792.html>. Accessed 10 Mar 2025.
32. Brendel P. Multibias: Simultaneous Multi-Bias Adjustment. 2024. R package version: 1.5.0. <https://doi.org/10.32614/CRAN.package.multibias>. Accessed 22 May 2024.
33. Brendel P. Simultaneous Multi-bias Analysis. <https://pcbrendel.shinyapps.io/multibias/>. Accessed 10 Mar 2025.
34. Fox MP, MacLehose RF, Lash TL. SAS and R code for probabilistic quantitative bias analysis for misclassified binary variables and binary unmeasured confounders. *Int J Epidemiol*. 2023;52(5):1624–33. <https://doi.org/10.1093/ije/dyad053>.
35. Fox MP, MacLehose RF, Lash TL. Short code. <https://sites.google.com/site/biasanalysis/short-code>. Accessed 10 Mar 2025.
36. Çubukçu HC, Vanstapel F, Thelen M, Lantman MVS, Bernabeu-Andreu FA, Brguljan PM, et al. APS Calculator: A Data-Driven Tool for Setting Outcome-Based Analytical Performance Specifications for Measurement Uncertainty Using Specific Clinical Requirements and Population Data. *Clin Chem Lab Med* (CCLM). 2024;62(4):597–607. <https://doi.org/10.1515/cclm-2023-0740>.
37. Parish WJ, Aldridge A, van Hasselt M. A Bayesian Method for Addressing Multinomial Misclassification with Applications for Alcohol Epidemiological Modeling. *Stata J*. 2024;24(1):113–37. <https://doi.org/10.1177/1536867X241233671>.
38. Paterno GB, Penone C, Werner GDA. sensiPhy: An R-Package for Sensitivity Analysis in Phylogenetic Comparative Methods. *Methods Ecol Evol*. 2018;9(6):1461–7. <https://doi.org/10.1111/2041-210X.12990>.
39. Wu C. Quantitative Bias Analysis in Epidemiology. *Med Stat Data Sci*. <https://medical-statistics.dk/MSDS/epi/bias/bias.html>. Accessed 07 Apr 2025.
40. Wu C. BIASEPI: Stata module to perform simple bias analysis, multidimensional bias analysis, and multiple bias modeling. 2019. Statistical Software Components, Boston College Department of Economics. <https://ideas.repec.org/c/boc/bocode/s458617.html>. Accessed 10 Mar 2025.
41. Bartlett J. PVW: Stata module to perform predictive value weighting for covariate misclassification in logistic regression. 2014. Statistical Software Components, Boston College Department of Economics. <https://ideas.repec.org/c/boc/bocode/s457825.html>. Accessed 10 Mar 2025.
42. Beesley LJ, Fritsche LG, Mukherjee B. An Analytic Framework for Exploring Sampling and Observation Process Biases in Genome and Phenome-Wide Association Studies Using Electronic Health Records. *Stat Med*. 2020;39(14):1965–79. <https://doi.org/10.1002/sim.8524>.
43. Beesley LJ. SAMBA-EHR: Sampling And Misclassification Bias Analysis in Genome and Phenome-wide Association Studies Using Electronic Health Records. <https://dsde.shinyapps.io/samba-ehr/>. Accessed 10 Mar 2025.
44. Hall GC, Lanes S, Bollaerts K, Zhou X, Ferreira G, Gini R. Outcome Misclassification: Impact, Usual Practice in Pharmacoepidemiology Database Studies and an Online Aid to Correct Biased Estimates of Risk Ratio or Cumulative Incidence. *Pharmacoepidemiol Drug Saf*. 2020;29(11):1450–5. <https://doi.org/10.1002/pds.5109>.
45. Nab L, Groenwold RHH, van Smeden M, Keogh RH. Quantitative Bias Analysis for a Misclassified Confounder: A Comparison Between Marginal Structural Models and Conditional Models for Point Treatments. *Epidemiology*. 2020;31(6):796. <https://doi.org/10.1097/EDE.0000000000001239>.
46. Bertrand A, Legrand C, Léonard D, Van Keilegom I. Robustness of Estimation Methods in a Survival Cure Model with Mismeasured Covariates. *Comput Stat Data Anal*. 2017;113:3–18. <https://doi.org/10.1016/j.csda.2016.11.013>.
47. Lin A, Tommasi D, Zhang L. IVBOUNDS: Stata module providing instrumental variable method to bound treatment-effects estimates with potentially misreported and endogenous program participation. 2021. Statistical Software

- Components, Boston College Department of Economics. <https://ideas.repec.org/c/boc/bocode/s458967.html>. Accessed 22 Mar 2025.
48. Lin A, Tommasi D, Zhang L. Bounding Program Benefits When Participation Is Misreported: Estimation and Inference with Stata. *Stata J.* 2024;24(2):185–212. <https://doi.org/10.1177/1536867X241257347>.
49. Liu X, Wang L. The Impact of Measurement Error and Omitting Confounders on Statistical Inference of Mediation Effects and Tools for Sensitivity Analysis. *Psychol Methods.* 2021;26(3):327–42. <https://doi.org/10.1037/met0000345>.
50. Yang J, Lin L, Chu H. BayesSenMC: An R Package for Bayesian Sensitivity Analysis of Misclassification. *R Journal.* 2021;13(2):228–38. <https://doi.org/10.32614/RJ-2021-097>.
51. Smith LH, Mathur MB, VanderWeele TJ. Multiple-Bias Sensitivity Analysis Using Bounds. *Epidemiology.* 2021;32(5):625. <https://doi.org/10.1097/EDE.0000000000001380>.
52. Mathur MB, Smith LH, Ding P, VanderWeele TJ. EValue: Sensitivity Analyses for Unmeasured Confounding and Other Biases in Observational Studies and Meta-Analyses. 2021. R package version 4.1.3. <https://doi.org/10.32614/CRAN.package.EValue>. Accessed 25 Apr 2024.
53. Lin Q, Nuttall AK, Zhang Q, Frank KA. How Do Unobserved Confounding Mediators and Measurement Error Impact Estimated Mediation Effects and Corresponding Statistical Inferences? Introducing the R Package ConMed for Sensitivity Analysis. *Psychol Methods.* 2023;28(2):339–58. <https://doi.org/10.1037/met0000567>.
54. Haine D. The episensr package: basicsensitivity analysis of epidemiological results. 2023. R package version 1.3.0. <https://dhaine.github.io/episensr/>. <https://doi.org/10.5281/zenodo.8299430>. Accessed 07 Sept 2023.
55. Denis Haine. The apisensr Shiny app package: interface to episensr for sensitivity analysis of epidemiological results. 2021. <https://doi.org/10.5281/zenodo.8322272>. R package version 1.0.0. <https://github.com/dhaine/apisensr/>. Accessed 17 Jan 2025.
56. Banack HR, Smith SN, Bodnar LM. Application of a Web-based Tool for Quantitative Bias Analysis: The Example of Misclassification Due to Self-reported Body Mass Index. *Epidemiology (Cambridge, Mass).* 2024;35(3):359–67. <https://doi.org/10.1097/ede.0000000000001726>.
57. Xu Y, Liu S, Yi GY. Mgee2: An R Package for Marginal Analysis of Longitudinal Ordinal Data with Misclassified Responses and Covariates. *R Journal.* 2021;13(2):471–84. <https://doi.org/10.32614/rj-2021-093>.
58. Xu Y, Chen Z, Liu SS, Yi G. Mgee2: Marginal Analysis of Misclassified Longitudinal Ordinal Data. 2023. R package version 0.6. <https://doi.org/10.32614/CRAN.package.mgee2>. Accessed 7 Apr 2025.
59. Pina-Sánchez J, Brunton-Smith I, Buil-Gil D, Cernat A. Exploring the Impact of Measurement Error in Police Recorded Crime Rates through Sensitivity Analysis. *Crime Sci.* 2023;12(1):14. <https://doi.org/10.1186/s40163-023-00192-5>.
60. Linden A, Mathur MB, VanderWeele TJ. Conducting Sensitivity Analysis for Unmeasured Confounding in Observational Studies Using E-values: The EValue Package. *Stata J.* 2020;20(1):162–75. <https://doi.org/10.1177/1536867X20909696>.
61. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med.* 2017;167(4):268–74. <https://doi.org/10.7326/M16-2607>.
62. Mathur MB, Ding P, Riddell CA, VanderWeele TJ. Web Site and R Package for Computing E-values. *Epidemiology.* 2018;29(5):e45. <https://doi.org/10.1097/DE.0000000000000864>.
63. Beesley LJ, Mukherjee B. Statistical Inference for Association Studies Using Electronic Health Records: Handling Both Selection Bias and Outcome Misclassification. *Biometrics.* 2020;78(1):214–26. <https://doi.org/10.1111/biom.13400>.
64. Jiang T, Fox MP. Quantitative Bias Analysis. 2021. https://jiangtammy.shinyapp.io/quantitative_bias_analysis/. Accessed 17 Jan 2025.
65. Shi B, Choirat C, Coull BA, VanderWeele TJ, Valeri L. CMAverse: A Suite of Functions for Reproducible Causal Mediation Analyses. *Epidemiology.* 2021;32(5):e20. <https://doi.org/10.1097/EDE.0000000000001378>.
66. Shi B, Wang Z, Choirat C, Valeri L. CMAverse: a suite of functions for causal mediation analysis. <https://bs1125.github.io/CMAverse/>. Accessed 07 Apr 2025.
67. Lockwood JR. Eivtools: Measurement Error Modeling Tools. Comprehensive R Archive Network. 2018. R package version 0.1-8. <https://CRAN.R-project.org/package=eivtools>. <https://doi.org/10.32614/CRAN.package.eivtools>. Accessed 17 Feb 2025.
68. Prentice-Mott G. Probabilistic Bias Analysis - Information Bias. https://graemepmott.shinyapps.io/prob_bias_analysis/. Accessed 07 Apr 2025.
69. MacLehose RF. Odds Ratio Calculator with Misclassification, Selection Bias, and Confounding Adjustment. <https://maclehose.shinyapps.io/simplemba/>. Accessed 07 Apr 2025.
70. Fox MP, Lash TL, Greenland S. A Method to Automate Probabilistic Sensitivity Analyses of Misclassified Binary Variables. *Int J Epidemiol.* 2005;34(6):1370–6. <https://doi.org/10.1093/ije/dyi184>.
71. Fox MP, Lash TL, Greenland S, Cramer M. sensmac SAS Macro. <https://sites.google.com/site/biasanalysis/sensmac>. Accessed 07 Apr 2025.
72. Lash TL, Fink AK. Semi-Automated Sensitivity Analysis to Assess Systematic Errors in Observational Data. *Epidemiology.* 2003;14(4):451. <https://doi.org/10.1097/01.EDE.0000071419.41011.cf>.
73. Lash TL, Fink AK. Multiple bias model. <https://sites.google.com/site/biasanalysis/multiple-bias-model-lash-tl-fink-a>. Accessed 10 Mar 2025.
74. Nab L, van Smeden M, Keogh RH, Groenwold RHH. Mecor: An R Package for Measurement Error Correction in Linear Regression Models with a Continuous Outcome. *Comput Methods Prog Biomed.* 2021;208:106238. <https://doi.org/10.1016/j.cmpb.2021.106238>.
75. Fox MP, MacLehose RF, Lash TL. Misclassification Spreadsheet. <https://sites.google.com/site/biasanalysis/>. Accessed 30 May 2025.
76. A BdG, DB R, MK SB, editors. Bias Assessment in Case–Control and Cohort Studies for Hazard Identification. vol. V of Statistical Methods in Cancer Research. Lyon: IARC Scientific Publication; 2024.
77. Jurek AM, Greenland S, Maldonado G. How Far from Non-Differential Does Exposure or Disease Misclassification Have to Be to Bias Measures of Association Away from the Null? *Int J Epidemiol.* 2008;37(2):382–5. <https://doi.org/10.1093/ije/dym291>.
78. Ranker LR, Petersen JM, Fox MP. Awareness of and Potential for Dependent Error in the Observational Epidemiologic Literature: A Review. *Ann Epidemiol.* 2019;36:15–19.e2. <https://doi.org/10.1016/j.annepidem.2019.06.001>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.