



Education Corner

Good practices for quantitative bias analysis

**Timothy L Lash,^{1*} Matthew P Fox,² Richard F MacLehose,³
George Maldonado,⁴ Lawrence C McCandless⁵ and Sander Greenland⁶**

¹Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA,

²Department of Epidemiology and Center for Global Health & Development, Boston University School of Public Health, Boston, MA, USA, ³Division of Epidemiology and Community Health and ⁴Division of Environmental Health Sciences, University of Minnesota School of Public Health, Minneapolis, MN, USA, ⁵Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada and ⁶Department of Epidemiology and Department of Statistics, University of California Los Angeles, Los Angeles, CA, USA

*Corresponding author. Department of Epidemiology, Rollins School of Public Health, Emory University, 1518-002-3BB, 1518 Clifton Rd NE, Atlanta, GA 30322, USA. E-mail: tlash@emory.edu

Accepted 7 July 2014

Abstract

Quantitative bias analysis serves several objectives in epidemiological research. First, it provides a quantitative estimate of the direction, magnitude and uncertainty arising from systematic errors. Second, the acts of identifying sources of systematic error, writing down models to quantify them, assigning values to the bias parameters and interpreting the results combat the human tendency towards overconfidence in research results, syntheses and critiques and the inferences that rest upon them. Finally, by suggesting aspects that dominate uncertainty in a particular research result or topic area, bias analysis can guide efficient allocation of sparse research resources.

The fundamental methods of bias analyses have been known for decades, and there have been calls for more widespread use for nearly as long. There was a time when some believed that bias analyses were rarely undertaken because the methods were not widely known and because automated computing tools were not readily available to implement the methods. These shortcomings have been largely resolved. We must, therefore, contemplate other barriers to implementation. One possibility is that practitioners avoid the analyses because they lack confidence in the practice of bias analysis.

The purpose of this paper is therefore to describe what we view as good practices for applying quantitative bias analysis to epidemiological data, directed towards those familiar with the methods. We focus on answering questions often posed to those of us who advocate incorporation of bias analysis methods into teaching and research. These include the following. When is bias analysis practical and productive? How does one select the biases that ought to be addressed? How does one select a method to model biases? How does one assign values to the parameters of a bias model? How does one present and interpret a bias analysis?

We hope that our guide to good practices for conducting and presenting bias analyses will encourage more widespread use of bias analysis to estimate the potential magnitude and direction of biases, as well as the uncertainty in estimates potentially influenced by the biases.

Key words: Epidemiological biases, analysis, best practice

Key Messages

- Quantitative bias analysis provides an estimate of uncertainty arising from systematic errors, combats overconfidence in research results and guides future research.
- Methods of bias analysis have been well known for decades and endorsed for widespread use, yet bias analysis is rarely implemented.
- One possible barrier to implementation is lack of guidance focused more on practice and less on bias models or methods. The purpose of this paper is to provide this missing guidance, and thereby to encourage more widespread use of bias analysis.

Introduction

Quantitative bias analysis models nonrandom errors that may distort results of epidemiological research. The primary objective of bias analysis is to estimate the potential magnitude and direction of biases, and to quantify the uncertainty about these biases. Models to quantify the direction and magnitude of biases have been known for decades.^{1–10} There have been hundreds of articles on adjustment methods for measured bias sources such as confounders, measurement error (including misclassification) and missing data, resulting in several textbooks dealing with these topics.^{11–15} Most textbook methods assume that data are available to allow an analyst to estimate parameters used in an adjustment method, for example by imputation of the missing correct values.^{11,13,16}

Only a small proportion of the literature deals with cases in which available data are inadequate to support these methods, although these cases are probably more often encountered in practice. This problem has led to development of methods for sensitivity analysis^{17–23} and extensions for simulation of bias effects under scenarios deemed plausible based on background information.^{15,23–32} There has, however, been only limited guidance on when particular bias analysis methods are helpful and on what constitutes good practices in conducting such analyses.^{15,23,28,31} This lack of guidance may partly explain the relative dearth of applications in published research.

There are many parallels between good practices for epidemiological research and good practices for applying bias analysis to epidemiological data.¹⁵ For example, good

research practices and good bias analysis practices both include: (i) development of a protocol to guide the work; (ii) documentation of revisions to the protocol that are made once the work is under way, along with reasons for (Color online) these revisions; (iii) detailed description of the data used; (iv) a complete description of all analytical methods used and their results, along with reasons for emphasizing particular results for presentation; and (v) discussion of underlying assumptions and limitations of the methods used. Good practices in presentation provide (i)–(v) along with (vi), description of possible explanations for the results. If inferences beyond the study are attempted, they should be prudent, circumspect and integrated with prior knowledge on the topic at hand; inferences based on single studies can be especially misleading given that most inferences require careful synthesis of diverse and extensive literature.^{33–35}

Even if everyone agreed on certain principles, however, both good research practices and good bias analysis practices would require a presumption that researchers, analysts, authors and reviewers have made in good faith an effort to follow these principles. This presumption can never be guaranteed, but can be bolstered by transparent declaration of competing interests, by symmetrical consideration of bias sources and by other evidence of attempts at neutrality.³⁶

The purpose of this paper is not to review the methods of bias analysis or ethical research practices, however, but rather to describe what we view as good practices for applying quantitative bias analysis to epidemiological data. Thus we will presume that the data to which these

methods will be applied have been gathered and analysed according to good research practices and ethical research conduct. Our focus will instead be on answering questions often posed to those of us who advocate incorporation of bias analysis methods into teaching and research. These questions include the following.

- When is bias analysis practical and productive?
- How does one select the biases that ought to be addressed?
- How does one select a method to model biases?
- How does one assign values to the parameters of a bias model?
- How does one present and interpret a bias analysis?

Box 1 summarizes our recommendations in reply to these questions. We do not intend to provide absolute or complete rules of conduct or a definitive checklist to evaluate the quality of a bias analysis. Instead, we provide some initial guidelines for answering the above questions, with the goals of easing the task for those interested in applying bias analysis and encouraging others to view bias analysis as a viable and desirable tool in their own work. Another benefit would be for these guidelines to improve the quality of bias analyses. In turn, we hope that these guidelines will themselves be improved by feedback from readers and users. Eventually, such feedback along with wider experience may lead to more detailed and extensive collaborative guidelines, perhaps along the lines of CONSORT, STROBE and other community efforts to improve research conduct and reporting.

When is bias analysis practical and productive?

Bias analysis covers a broad range of methods, from simple sensitivity analyses to in-depth probabilistic analyses requiring considerable labour. Choosing a method depends on judging the point at which the likely benefit of doing further analyses no longer justifies the labour. This question is complicated by the fact that we rarely have more than vague ideas of the cost or benefit of further analyses. The benefit in particular may be largely unknown until we do the analysis. Fortunately, our analysis decisions are subject to challenge and revision as long as the data remain available.

Later sections will outline what we think is needed for a 'good' bias analysis, which will provide some sense of cost. As for benefits, a good bias analysis provides an effect estimate that accounts for plausible sources of bias, aiding in scientific inference. Further, it can provide a sense of the uncertainty warranted given the assumptions incorporated into the bias analysis. As with any methodology, however,

bias analysis is not foolproof: poor choice of models or parameter values could harm inferences. To aid decisions on whether and how much bias analysis is needed, we have created a rough classification scheme of situations, ranging from those in which bias analysis seems unnecessary to those in which it appears essential.

Cases in which bias analysis is not essential

Bias analysis is not essential when a research report strictly limits itself to description of its motivation, conduct and data, and stops short of discussing causality or other inferences beyond the observations. Although such purely descriptive reports are unusual and are even discouraged by many, they have been recommended as preferable to opposite extremes in which single studies attempt to argue for or against causality without regard to studies of the same topic or other relevant research.³⁵

Bias analysis may be helpful, but not necessary, when a report stops short of drawing inferences about causality or other targets beyond the observations, and instead offers alternative explanations for observations. This sort of report is among the most cautious seen in the literature, focusing on data limitations and needs for further research but refraining from substantive conclusions.

Bias analysis may be unnecessary when ordinary statistical analyses, encompassing only random error, show the study is incapable of discriminating among the alternative hypotheses under consideration within the broader topic community. Common examples include studies where the precision of the effect estimate is so poor that the confidence interval includes all associations that are taken seriously by the topic community as possible effect sizes. This situation commonly arises when modest point estimates (e.g. relative risks around 2) are reported, the null value is included in the conventional frequentist confidence interval and no one seriously argues that the effect, if any, could be large (e.g. relative risks above 5). Attempts to argue for or against causality in such cases would be ill-advised even if bias were absent, and discussion may be adequately restrained by considering both limits of the interval estimates with equal weight.³⁷ In the above situations, however, bias analysis becomes necessary if a reader attempts to draw substantive conclusions beyond those of the original study report, such as in public health policy, legal and regulatory settings.

Bias analysis may also be unnecessary when the observed associations are dramatic, consistent across studies and coherent to the point that bias claims appear unreasonable or motivated by obfuscation goals. Classic examples include associations between smoking and lung cancer, occupational exposure to vinyl chloride and

Box 1. Summary description of good practices for quantitative bias analysis**When is bias analysis practical and productive?**

- Advisable when a report of an association that is not dramatic goes beyond description and alternative explanations for results, and attempts to draw inferences about causality.
- Essential when a report makes action or policy recommendations, or has been developed specifically as a synthesis for decision making, and the decisions are sensitive to biases.

How does one select the biases that ought to be addressed?

- Begin with a review of selection and retention of study subjects, data collection methods and opportunities for confounding, selection bias and measurement error.
- Create a graphical display of presumed causal relations among variables and their measurements, and present these DAGs to display the underlying assumptions.
- Complete simplified bias calculations to prioritize biases likely to have the greatest influence.
- Biases that could credibly explain a finding merit more attention than biases that could not.

How does one select a method to model biases?

- A realistic model of bias sources is likely to be complex. Balance realistic modelling and practicality, just as with conventional epidemiological data analysis.
- Transparency and credibility are essential. Increasing complexity can reduce transparency and hence the credibility of an analysis. Publish code to improve transparency, aid future implementations and identify algorithmic errors.

How does one assign values to the parameters of a bias model?

- Assign credible values and distributions to bias parameters that reflect available data, including internal sub-studies, external validation data and expert judgment.
- Choose a range of plausible values to yield a grid of adjustments that can be examined for consistency and to understand the dependence of results on values or their combinations.
- Assign distributions to bias parameters, rather than sets of values, for probabilistic bias analysis and Bayesian analysis to describe the central tendency and spread of adjustments.
- Include a sensitivity analysis of the bias analysis to evaluate the dependence of the results on underlying assumptions.

How does one present and interpret a bias analysis?

- Begin with a clear statement of objectives, which should relate directly to some aspect of the conventional methods description.
- Link objectives to a bias model, which relates measured variables to the bias analysis result through the bias parameters. Give values or distributions assigned to these bias parameters, the basis for the assignments and reasons for rejecting alternatives.
- Provide an example of the calculations completed using the bias model.
- Use diagnostics to assure bias modelling results correspond to the intended model.
- Use tables and figures to depict the complete set of bias analysis results.
- Interpret bias analysis results beginning with a restatement of the underlying assumptions, and concluding with a description of any resulting change in the study inferences.

angiosarcoma, estrogen replacement therapy and endometrial cancer, and outbreaks from infectious or toxic sources. In these situations, bias analysis may still be helpful to improve accuracy of uncertainty assessment. It may also be helpful for policy makers seeking to incorporate the size of an effect estimate and its total uncertainty into hazard

prioritization and regulation. Finally, bias analysis may be useful in this setting to demonstrate the unreasonableness of denialist claims, as did Cornfield *et al.*³ in response to claims that the smoking-lung cancer association could be attributed to a genetic factor affecting both tendency to smoke and cancer risk. As a historical note, this paper

is often cited as the first sensitivity analysis, although Berkson¹ is an earlier example of quantitative bias analysis.

Cases in which bias analysis is advisable

Bias analysis is advisable when a report of an association that is not dramatic goes beyond description and possible alternative explanations for results, and attempts to draw inferences about causality or other targets beyond the immediate observations. In these cases, the inferences drawn from conventional statistics may not hold up under the scrutiny afforded by bias analysis, especially when conventional statistical analyses make it appear that the study is capable of discriminating among importantly different alternatives or there is any attempt to interpret the study as if it does so. In public health policy, legal and regulatory settings involving hazards, this situation frequently arises when the lower relative-risk confidence limit is above 1 or the upper limit is below 2.

When conventional statistics appear decisive to some in the topic community, discussion needs to be adequately restrained by considering the potential impact of bias. Simple bias-sensitivity analyses will often suffice to adequately demonstrate robustness or sensitivity of inferences to specific biases. The aforementioned Cornfield *et al.*³ paper is an example that addressed an extreme and unreasonable bias explanation (complete genetic confounding) for an extreme and consistent association (which was being promoted as calling for policy change). The analysis by Cornfield *et al.* demonstrated the extremity of the bias explanation relative to what was known at the time (and has since been borne out by genetic and twin studies). It is uncertain whether they would have gone through this exercise had not a highly influential scientist raised this challenge, but the paper established the notion that one could not explain away the association between smoking and lung cancer as confounding alone without invoking associations and effects at least as large as the one in question.

Cases in which bias analysis is arguably essential

Bias analysis becomes essential when a report makes action or policy recommendations, or has been developed specifically as a research synthesis for decision making, and the decisions (as opposed to the statistical estimates) are sensitive to biases considered reasonable by the topic community. As with Cornfield *et al.*,³ simple bias-sensitivity analyses might suffice to demonstrate robustness or sensitivity of inferences. Nonetheless, multiple-bias analysis might be necessary in direct policy or decision settings, and that in turn usually requires probabilistic inputs to deal with the large number of bias parameters.

As an example, by the early 2000s, over a dozen studies exhibited relatively consistent but weak (relative-risk estimates dispersed around 1.7) associations of elevated residential electromagnetic fields (EMFs) and childhood leukaemia. Conventional meta-analyses gave relative-risk interval estimates in the range of 1.3 to 2.3 ($P=0.0001$).^{38,39} Consequently, there were calls by some groups for costly remediation (e.g. relocation of power lines). Probabilistic bias analysis found that more credible interval estimates could easily include the null (no effect),²⁸ as well as very large effects that were inconsistent with surveillance data.⁴⁰ Thus, bias analysis showed that the evidence provided by the conventional meta-analysis should be downweighted when considering remediation. In settings where immediate policy action is not needed, bias analysis results can provide a rationale for continued collection of better evidence and can even provide a guide for further research.⁴¹

In summary, simple bias analyses seldom strain resources and so are often worthwhile. They are, however, not necessary until research reports contemplate alternative hypotheses and draw inferences. At this point, and certainly once policy decisions are contemplated, bias quantification by simple bias modelling becomes essential and more complex modelling may also be needed.

How does one select the biases that ought to be addressed?

When a bias analysis is advisable, the next order of business is to decide which sources of bias to examine. Most bias analyses will have to consider the possibility that results are affected by uncontrolled confounding, selection bias and measurement error (including misclassification) because most epidemiological studies are susceptible to these biases. Which biases to account for with quantitative analysis will depend on the goals of the analysis (e.g. full quantification of study error vs bounding the impact of a single source of bias) and which biases, if any, were ruled out by study features (e.g. a study with mortality as outcome may have no concern about outcome misclassification).

After defining a clear causal question, the analyst should describe the bias sources. This description begins with a detailed review of selection and retention of study subjects in comparison with the source population they are meant to represent, data collection methods, and opportunities for confounding, selection bias and measurement error. Although these descriptions provide a sound foundation, they may miss certain types of biases such as bias from conditioning on colliders.⁴² Directed acyclic graphs (DAGs)^{42–46} can be useful for identifying potential bias

sources, hence we recommend, as a preliminary step to guide the analysis, creating a graphical display of presumed causal relations among analysis variables and their measurements. Further, we recommend presenting these DAGs along with the analysis to help display the assumptions underlying the methods used.

In terms of effort, biases that could credibly explain a finding may merit more attention than biases that could not. For example, in a null study of vaccination and autism risk, an analysis that examined misclassification would be critical if the inference is one of no association. Nondifferential misclassification is typically expected to produce bias toward the null, but small departures from nondifferentiality may lead to bias away from the null,⁴⁷ and some forms of differential misclassification may lead to bias toward the null.⁴⁸ In contrast, in a non-null study of the association between neighbourhood quality and physical function, correlation of errors between measures of neighbourhood quality and measures of physical function may be most important to evaluate before inferring that poor neighbourhood quality causes poor physical function.⁴⁹

Finally, there will often be restrictions on what can be done given available software. Missing-data and Bayesian software can sometimes effectively be used^{30,50} and procedures for Excel, SAS and Stata have been published.^{15,27,29}

Potential sources of bias to be considered

Uncontrolled confounding arises from failure to adjust for important confounders that account, in part, for lack of exchangeability between groups. Failure to adjust properly is due to either failure to measure these confounders, or inappropriate use of statistical adjustment, or variable-selection procedures such as stepwise regression. Incomplete adjustment may also arise from use of inaccurately measured or categorized confounders or from misspecification of the functional form of the relationship between the confounder and the outcome variable (e.g. smoker/non-smoker vs full smoking history). Many bias analysis methods assume no effect-measure modification by the unmeasured confounder, although methods to account for effect-measure modification are available.^{15,28}

A mirror problem to uncontrolled confounding is over-adjustment bias, which arises from adjustment for inappropriate variables (such as intermediates and other variables affected by exposure). Failure to adjust for well-measured confounders and overadjustment bias can be remedied given the original data by adding and deleting adjustment variables as appropriate, but these data are usually unavailable for subsequent meta-analyses or risk analyses.

Selection bias arises from biased subject sampling, losses to follow-up, subject nonresponse, subject selection after susceptibles have left the pool of subjects and other mechanisms. Selection bias is often a major concern in selection of controls in case-control studies, but can also arise in case-control and cohort studies when loss to follow-up is related to both the exposure and the outcome, when follow-up begins after the onset of exposure^{51,52} or when there is matching on inappropriate variables (over-matching). Information on the relation of both exposure and outcome to selection is rarely available. Validation studies of selection proportions are difficult to conduct, because the subjects under consideration are not, and may never have been, in the study. Even when such validation studies are done, as when data from a study can be compared with population registries, the results may not easily translate to the source population for the study subjects. Nonetheless, available information can be used to bound the magnitude of bias due to nonrandom subject selection.

Mismeasurement of variables can be expected in almost all studies. Exposure mismeasurement is common in most nonexperimental designs because of the nature of data collection. Self-reports, medical records, laboratory tests etc. can all result in measurement errors. Approximately nondifferential mismeasurement of exposures and covariates with respect to the study outcome may be plausible when measurements are collected before the outcome occurs. Differential mismeasurement may arise, however, when exposure and covariate measurements are influenced by or share influences with the outcome. Classic examples arise in studies that interview subjects about exposure history after the study outcome, as knowledge of the outcome can influence recall of exposures (recall bias). Contrary to common lore, the net bias that results need not be away from the null.⁴⁸ When independent of other errors, nondifferential confounder mismeasurement usually leads to bias in the direction of the original confounding.^{7,53} Regardless of nondifferentiality, mismeasurement of a covariate that is a strong confounder can lead to substantial bias.

Measurement errors in one variable may also be correlated with measurement errors in other variables.^{54,55} Such correlated or dependent errors should be expected whenever measurements are obtained or constructed using the same instrument or data. For example, a survey of self-perceived physical function and neighbourhood quality may yield an association between them, even if none exists, because respondents who overstate or understate the true quality of their neighbourhood may do the same with regard to the quality of their own physical function.⁴⁹ Errors in occupational exposure histories constructed from the same job-exposure matrix will be correlated since they will incorporate the same errors in the job histories. Similarly,

errors in nutrient intakes calculated from the same food-nutrient table will be dependent since they will incorporate the same errors in the diet histories. Less extreme but nonetheless important error dependence can arise among questionnaire responses, especially within related items (e.g. long-term recall of life events or habits). Even when errors are nondifferential, the presence of dependent error between the exposure and the outcome variable can create bias away from the null, and may even create the appearance of a strong association when there is no association at all.⁵⁴

Which sources of bias to model

Once the sources of bias have been identified, one must prioritize which biases to include in the analysis. We recommend prioritizing biases likely to have the greatest influence on study results. Judging this often requires relatively quick, simplified bias calculations (described in the next section) based on review of the subject literature and expert subject knowledge. Each of the sources of bias described above may be evaluated tentatively using simple bias analyses. Such an approach will often require a fair amount of labour, but is essential to informing the main part of the bias analysis and any conclusions that follow from it.

As an example, if little or no association has been observed, priority might be given to analysing single biases or combinations of biases that are likely to be toward the null (e.g. independent nondifferential misclassification) and thus might explain the observation. In this regard, signed DAGs^{56,57} can sometimes indicate the direction of bias and thus help to identify explanatory biases. A danger, however, is that by selecting biases to analyse based on expected direction, one will analyse a biased set of biases and thus reach biased conclusions. We thus advise that any bias that may be of substantively important magnitude be included in the final analyses, without regard to its likely direction.

Investigators may think that a source of bias is present, but that the magnitude of the bias is unimportant relative to the other errors present. For example, if the literature indicates that the association between an uncontrolled confounder and the exposure or outcome is small (e.g. as with socioeconomic status and childhood leukaemia), then the amount of uncontrolled bias from this confounder is also likely to be small.^{3,58} A number of authors give bounds on the magnitude of bias due to uncontrolled confounding based on bounds for the component associations,^{17,59–61} which allow the analyst to judge whether that bias is important in their application.

Soliciting expert opinion about possible bias sources can be a useful complement to, but no substitute for, the process described above in conjunction with a full

literature review. Experts in the field may be aware of sources of bias that are not commonly mentioned in the literature. It is unlikely, however, that one will be able to obtain a random sample of expert opinions, a concern of special importance in controversial topic areas where experts may disagree vehemently.

How does one select a method to model biases?

Balancing computational intensity and sophistication

Quantitative bias analysis encompasses an array of methods ranging from the relatively simple to the very complex (Table 1). Bias analysts consider such factors as computational intensity and the sophistication needed to implement the method when selecting from among the options. All methods require specifying a bias model and its parameters, but the method's computational intensity is dictated, in part, by how the bias parameters are specified.

In simple bias-sensitivity analysis, the user treats the bias parameters as fixed quantities which are then sometimes varied systematically together (multidimensional bias analysis^{15,23}). For example, to study bias due to confounding by an unmeasured covariate, the analyst may examine many combinations of the confounder distribution and its relations to exposure and to the outcome. Similarly, to study bias from exposure misclassification, the analyst might explore different pairs of sensitivity and specificity.^{15,23} These analyses can be computationally straightforward and require no detailed specification of a distribution for the bias parameters. Once the bias model and its initial values have been coded in a spreadsheet, for example, it is usually a small matter to change the values assigned to the bias parameters to generate a multidimensional analysis. However, such analyses do not explicitly incorporate uncertainty about the bias parameters in interval estimates or tests of the target parameter. Whereas an analyst may wish to begin with simple and multidimensional methods, we recommend formal sensitivity analysis in cases where plausible changes in values of bias parameters result in drastic changes in the bias-adjusted estimate, as often occurs in exposure-misclassification problems^{6,62} or when more complete depictions of uncertainty are indicated.²⁷

One way to incorporate this uncertainty into statistical results is to use probabilistic bias analysis (PBA). PBA is a generalization of simple bias analysis in which the bias parameters are assigned a joint probability distribution. This distribution is known as a bias-parameter distribution or, in Bayesian terms, a joint prior distribution, and is

Table 1. Summary of quantitative bias analysis techniques

Analytical technique	Treatment of bias parameters	Number of biases analysed	Output	Combines random error?	Computationally intensive?
Simple sensitivity analysis	One fixed value assigned to each bias parameter	One at a time	Single revised estimate of association	Usually no	No
Multidimensional analysis	More than one value assigned to each bias parameter	One at a time	Range of revised estimates of association	No	No
Probabilistic analysis	Probability distributions assigned to each bias parameter	One at a time	Frequency distribution of revised estimates of association	Yes	Yes
Multiple bias modelling	Probability distributions assigned to bias parameters	Multiple biases at once	Frequency distribution of revised estimates of association	Yes	Yes

Reprinted from Lash, Fox and Fink (2009).¹⁵

supposed to represent the analyst's uncertainty regarding the true value of a bias parameter.

PBA can be implemented in various ways.^{15,23,24} The simplest approach (sometimes called Monte-Carlo sensitivity analysis, or MCSA) is to repeatedly sample bias parameters from their joint distribution and to use the sampled values in the same basic sensitivity formulas as used in simple fixed-value analysis. Unlike simple bias analysis, however, summaries of the adjusted estimates (e.g. histograms) from PBA reflect the uncertainty about the target parameter due to uncertainty about the bias parameters, provided the latter uncertainty is properly captured by the distribution used.

When multiple sources of bias are of concern, effect estimates can be adjusted for each source simultaneously using multiple bias modelling (MBM). In these situations there are usually far too many bias parameters to carry out simple fixed-value bias analysis, and PBA becomes essential.²⁸ This type of analysis is more realistic since it can incorporate all biases that are of serious concern, but there is little distributed software to do it. It is possible to combine single-bias algorithms to create a multiple bias adjustment, but care is needed in doing so. In particular, the order of adjustment is important. Adjustments need to be modelled in the reverse order from that in which the biases actually occur,^{15,23} which depends on the study design. Often confounding occurs first in the source population, selection bias second as the researcher selects subjects, and measurement error last as exposure, covariates and outcomes are measured. These biases should be analysed in the reverse order. Exceptions to this order are also common. For example, when a study sample is drawn from a database, and the inclusion criteria are based on measurements in the database, then selection-bias adjustment should precede adjustment for measurement error. If

subsequent measurements are made on patients (including interviews), then adjustment for errors in those measurements should precede selection-bias adjustment. It is essential that analysts report and explain the order in which the adjustments were made to allow evaluation by interested parties.

In typical probabilistic multiple-bias analyses, each bias source receives its own distribution. This modelling implicitly assumes that the distributions, and hence the biases, are independent (e.g. that selection probabilities tell us nothing about misclassification probabilities, and vice versa), which may not always be accurate. Dependencies can, however, be introduced directly as prior correlations⁶³ or indirectly by using hierarchical models.^{28,32} A more statistically refined approach is to use the bias-parameter distributions as prior distributions in Bayesian posterior computations.^{30,50,64–67} Fully Bayesian bias analysis can be difficult to implement, however, requiring special software packages and special checks for convergence of the fitting algorithm, which may fail more easily than in conventional analyses. Fortunately, MCSA appears to provide a good approximation to a partial-Bayesian analysis in which only the bias parameters are given prior distributions, provided that these distributions do not include values that are in outright conflict with the data being analysed.^{23,28,68,69} In particular, if the bias parameters are completely unidentified from the data (there is no data information about them) and the priors used apply only to these parameters, the resulting MCSA procedure can be viewed as a method of generating samples from the posterior distribution.^{30,70}

Although needless complexity should be avoided, there are areas in which too much simplification should also be avoided. When examining misclassification parameters, it is unreasonable to assume that the parameters are independent of one another. For example, when examining exposure

misclassification in a case-control study, we should ordinarily expect the sensitivity among the cases and the sensitivity among the controls to be similar, at least if the same instrument is used to collect exposure information in each group. That is, a higher sensitivity in one group will usually imply a higher sensitivity in the other group and this is modelled by specifying a high positive correlation in the joint prior distribution. The same is true for specificity. In fact, under nondifferentiality these correlations will be 1 (although perfect correlation does not by itself imply nondifferentiality). Failure to include correlations among related bias parameters can result in the sampling of unlikely parameter combinations, which in turn could result in adjusted-estimate distributions that are more misleading than the original unadjusted confidence interval for the target parameter.

Balancing realistic modelling against simplicity and ease of presentation

Any realistic model of bias sources is likely to be complex. As with conventional epidemiological data analysis, trade-offs must be made between realistic modelling and practicality. Simplifying assumptions are always required, and it is important that these assumptions are made explicit. For example, if the decision is made to omit some biases (perhaps because they are not viewed as being overly influential), the omissions and their rationales should be reported.

We encourage researchers using complex models to also examine simpler approximations to these models as a way to both check coding and gain intuition about the more complex model. For instance, multiple bias models can provide realistic estimates of total study error but may obscure the impacts of distinct bias sources. We thus advise researchers implementing a multiple-bias model to examine each source of bias individually, which helps identify adjustments with the greatest impact on results. One can also compare estimates obtained from probabilistic analysis with the estimate obtained when the bias parameters are fixed at the modes, medians or means of their prior distributions. In the event that the results of the simpler and more complex analyses do not align, the author should provide an explanation as to why.

Implications regarding transparency and credibility

Transparency and credibility are integral to any quantitative bias analysis. Unfortunately, increasing model complexity can lead to less transparency and hence, reduce the credibility of an analysis. Researchers should take several steps to increase the transparency of the methods they use. As with all analyses, researchers should avoid using models that they

do not fully understand. Giving a full explanation of why the model specification produced the given results can increase transparency. We also encourage authors to make the data and code from their bias analyses publicly available. With the advent of electronic appendices in most major journals, providing bias analysis code as web appendices poses little problem. Published code will aid future researchers who need to implement bias analyses. Further, quantitative bias modelling is complex and public dissemination of code can help to identify and correct algorithmic errors.

Using available resources versus writing a new model

Numerous resources are available to help researchers implement quantitative bias analysis. Many sources we cite contain detailed examples that illustrate the analyses. Several have provided code so that future researchers could implement their analyses as well.^{15,27,68,71} When possible, we encourage authors to adopt code that has been previously developed, because it should help to identify and reduce coding errors. Existing resources may be difficult to adapt to new situations, however, particularly for multiple bias models. In that case, researchers have to write their own programs.

How does one assign values to the parameters of a bias model?

After choosing a bias model that is specified by a collection of bias parameters, the next step is to assign values or distributions to the bias parameters. Here one must wrestle with which value assignments are reasonable, based on the subject matter literature and on experience, and what other considerations should be made when assigning values to the bias parameters. Sometimes only summary data from a publication are available, whereas the original authors would have access to record level data.

Sources of information about bias parameters

Internal validation data

Credible values and distributions assigned to bias parameters should reflect relevant available data. Some studies obtain bias parameter information from an internal second-phase or validation sub-study, in which additional data are collected that allow adjustment for bias or estimation of values or distributions to assign to bias parameters (e.g. measurements of confounders that are not recorded in the full data set, such as full smoking histories, or laboratory measurements that are collected from a subsample to

validate self-reported exposure status).^{14,15} Internal validation may be the best source of data on the bias parameters in the study in which it was conducted, which implies that a substantial proportion of study resources should be expended on validation sub-studies, even if it requires a reduction in total sample size. The results of such studies often do more to improve the yield from the research than expending these resources on a larger sample size or longer follow-up.

Many statistical methods are available for joint analysis of primary data with internal validation data, including missing-data and measurement-error correction methods.^{13,16,72} Nonetheless, these methods assume that the internal validation data are themselves free of bias. This assumption is often unreasonable, and if violated will result in bias in the bias-adjusted estimate. For example, to adjust for bias due to non-response, after initial requests we could ask all original invitees (including non-responders) to answer a brief supplementary questionnaire. Data provided by those initial non-responders who responded to this call-back survey might provide individual-level information about basic confounders like age and sex, and perhaps exposure and disease status, to identify the determinants of non-response. We should expect however that many initial non-responders will also not respond to this survey, and those that do are unlikely to be a random sample of all initial non-responders. Similarly, internal measurement-validation studies are themselves prone to selection bias when they place an additional burden on study participants, such as filling out further questionnaires, keeping diaries or supplying biological specimens. Those who agree to this additional burden are likely to differ from those who refuse, and these differences may relate to the size of the measurement errors characterized by the validation sub-study. The validation data they supply and adjustments based on them may therefore also be subject to unknown degrees of bias. Consequently, although a validation sub-study can supply valuable information, that information may have to be analysed with allowance for sources of bias in the sub-study.

External validation data

External validation data and external adjustment describe the scenario where we obtain bias parameter information from data outside of the study.^{15,23} Data from external validation studies can supplement internal validation data (which are often sparse) and are often the only direct source of information about bias parameters. Examples include individual-level data from a second study population, or parameter estimates obtained from a systematic review or meta-analysis. For example, to adjust for bias from an unmeasured confounder, we could conduct a

review of the literature to identify published estimates of the distribution of the confounder in the population and the associations between the confounder and the exposure and outcome variables.

As described above, internal and external validation data are themselves subject to systematic as well as random errors, and thus provide imperfect estimates of bias parameters. Nonetheless, such data can help set the range and distribution of values to assign those parameters. Uncertainty about the resulting bias parameter estimates can be incorporated into bias adjustments via sensitivity analyses, as described below.

Input from experts

Validation data are often unavailable, forcing reliance on expert opinion and educated guesses to specify the bias parameters and their distributions. Formulating knowledge or beliefs about unknown parameters into a joint probability distribution is called elicitation of the prior distribution.⁷³ One formal approach is to ask each expert for an interval within which they think the parameter falls and the odds or percentage they would bet on the parameter falling in this interval. From this interval one may specify a member of a convenient parametric family, such as a log-normal or normal distribution. For example, suppose an expert would give a certain odds or probability that a false-positive probability p ($p = 1 - \text{specificity}$) falls between 0.05 and 0.20. If we modelled this expert's bet as arising from a distribution for $\text{logit}(p)$ that was symmetrical (thus having a mean equal to its median), the expert's implied prior median for p would be $\text{expit}[(\text{logit}(0.20) + \text{logit}(0.05))/2] = 0.10$. Further modelling the expert's uncertainty as roughly normal on the logit scale, we would deduce that the standard deviation of this normal distribution is $(\text{logit}(0.20) - \text{logit}(0.05))/(2 \times 1.96) = 0.40$.

There is little evidence about which methods of constructing priors are more accurate; research on the quality of reasoning under uncertainty in general suggests that direct expert elicitations are unlikely to provide reliably accurate estimates of values or distributions for assignment to bias parameters.^{74,75} Of great concern is that expert opinions are highly susceptible to bias. Experts are often influenced by their selective knowledge, reading and interpretation of the literature, as well as personal preferences ('wish bias'). They can also be overconfident and understate the uncertainty about bias that would be warranted by available evidence,⁷⁶ which in turn results in overconfidence about the size of effect under study.^{15,23} Furthermore, experts may seriously misjudge the quality of the literature and the extent to which bias accounts for previous findings. Such misjudgments may be aggravated by expert overconfidence or poor judgment about the

reliability or quality of articles (e.g. over-rating their own studies or those that agree with their views, and under-rating those that conflict with what they expect). As a result, we recommend that analysts inspect the literature directly rather than rely solely on expert opinions. In doing so the analyst should bear in mind that, like reviews, judgment may also be distorted by publication bias and by lack of information on study problems in published reports.

Assigning values and distributions to bias parameters

A parsimonious strategy that does not require specifying the bias parameter values or distributions is to use target-adjustment sensitivity analysis.²⁶ In this approach, one back-calculates from conventional results to find combinations of bias-parameter values that would qualitatively change or explain the conventional statistics (e.g. that would shift an estimated effect measure to the null value or to a doubling of risk). Target-adjustment sensitivity analysis can be easier to implement and understand than bias modelling with best estimates assigned as values for the bias parameters, for it demands only qualitative assumptions about the bias parameters.

Nonetheless, there are several objections to target adjustment. Most obviously, it only examines how the difference between the conventional estimate and the targeted value might be entirely an artefact of bias²⁶ and thus is of little use if the goal is to estimate plausible ranges for the effect measure. Target-adjustment sensitivity analysis is also difficult to evaluate when there are multiple biases, for then many plausible as well as implausible patterns of bias could explain the difference between estimate and target. Finally, target adjustment incurs a risk of contaminating subsequent analyses, since once one knows what values would change a conventional estimate to a targeted value, that knowledge can bias one's view of the plausibility (and hence probability) of such parameter combinations. Thus, target adjustment may be useful only when one bias source is to be evaluated and the only question is whether plausible values for the bias parameters might explain the difference between the study's result and a targeted effect size of particular interest.

Instead of focusing on a value of the target parameter, one may assign one or more values to the bias parameters based on estimates drawn from external validation studies, internal validation studies or the investigator's experience working in the topic area. This process may be called fixed bias-parameter analysis (FBA). It is crucial to explain the basis for the selected values. Investigators often choose a range of plausible values. The extreme limits of plausibility may also be selected to avoid understating the uncertainty.

When the bias model involves more than one bias parameter, this method ultimately yields a grid of adjustments corresponding to combinations of values assigned to the different parameters of the bias model. The resulting adjusted estimates can be examined for consistency and to understand the dependence of results on different values, or combinations of values, assigned to the bias parameters.

Instead of focusing on fixed sets of values, probabilistic bias analysis (PBA) assigns distributions to the bias parameters. The location and spread of the distributions may be determined by the same information used to assign sets of values for simple and multidimensional bias analysis. For example, suppose we wish to restrict the sensitivity of exposure classification to fall between a and c , and b is considered a most likely value (mode). Among other possibilities, one could then assign: (i) a uniform distribution ranging between a and c ; (ii) a triangular distribution with minimum a , maximum c and mode b ; (iii) a trapezoidal distribution with minimum a , maximum c and lower and upper modes equidistant from b ; (iv) a distribution that is normal on the logit scale, translated to fall between a and c , with mode b ; or (v) a beta distribution, again translated to fall between a and c , with mode b .

A simplicity advantage of the uniform and triangular distributions is that they are determined completely by the specified range a to c and most likely value b . The uniform distribution is exceptionally unrealistic, however, because it has a sudden drop in probability at its boundaries and makes no distinction within those boundaries; for example, if $a = 0.6$, $b = 0.9$, it states that 0.599 is impossible yet 0.601 is as probable as any other possibility including 0.7 and 0.8. Among more mild criticisms of the triangular and trapezoid distributions is that they are not smooth (although they entail no sudden change in probability), whereas logit-normal and beta-distributions may become bimodal with low-precision parameter settings. Thus, to help avoid implausible distributional features, we recommend that distributions be graphed before use. Nonetheless, form (shape) can be particularly difficult to judge visually and intuitively; for example, normal, logistic, and t -distributions are all unimodal symmetrical and are not strikingly different in appearance, yet switching from a normal to a logistic distribution triples the prior probability that the true parameter is over 3 standard deviations from the mean.

An objection to all range-restricted distributions is that we may have no basis for being completely certain a parameter is within the boundaries a , b unless those are purely logical limits (e.g. 1 is the upper limit for a sensitivity and specificity). This problem can be addressed by extending the range between a and b (e.g. to the logical limits of 0 and 1 for sensitivity and specificity). However, this extension can create another problem: when the data convey

some information about parameters in a bias model, some values for those parameters inside the chosen range may conflict with the original data, as manifested by impossible adjusted data such as negative adjusted cell counts. This can easily occur, for example, when adjusting for misclassification using sensitivity and specificity, and the minimum allowed value a is too low or the maximum value b is too high, creating an incompatibility between the observed data and the proposed values of sensitivities and specificities.^{15,23,27} It is important to describe these bias-parameter values and see why they produce impossible data. It is also important that the estimates from such values are not used in subsequent inferential statistics, especially when aggregating estimates into simulation summaries (as in MCSA). If only a small proportion of values result in impossible adjusted results, there may be little harm from simply discarding these values and using summaries based on the remaining bias-parameter values, a strategy that truncates the bias-parameter distribution away from values that produce impossible adjustments.^{23,27,28,69}

One may avoid impossible adjustments by using the priors in proper Bayesian procedures, or by using a bias model whose parameters are independent of the data.^{30,70} Nonetheless, encountering impossible adjusted data is often of substantive importance, as it may represent a fundamental disconnect between the priors and the data or data model, and may signal poor prior information, poor data modelling or unrecognized data problems.

Sensitivity analysis of the bias analysis

The values assigned to the location and spread of a given bias-parameter distribution can greatly influence the results of a bias analysis. We thus recommend that a sensitivity analysis of the bias analysis, at least to location and spread, should be included where space permits, for example as supplementary appendices. Increasing the spread of a prior distribution (e.g. the prior variance) will usually increase the spread of the bias-adjusted effect measures, and it can be crucial to assess this increase.

Other potentially important sources of sensitivity in prior distributions, and hence uncertainty about final results, include form (e.g. trapezoidal or beta), and dependencies (e.g. correlations) among parameters. Few attempts have been made to compare bias analysis results when different distribution types are assigned to the bias parameters of a bias model, holding the location and spread (e.g. mean and variance) constant. The one published example we are aware of found little difference from use of different distributions with the same location and spread,¹⁵ but more study is needed of sensitivity of bias analysis to distributional forms.

Prior dependencies among bias parameters can be of special concern because there is rarely any validation data to support choices, and yet typical statistical default values (such as zero correlation between case and control misclassification probabilities) may be contextually nonsensical, as discussed above.^{23,27} Nonetheless, it may be possible to reparameterize the bias model so that its parameters are approximately independent;^{30,77} comparisons between results from an independent-parameter model and the original model can reveal sensitivity of results to parameterization.

Diagnostics

An important element of bias analysis, and especially probabilistic bias analysis, is model diagnostics. If the analyst assigns distributions to the parameters of a bias model, then it is good practice to generate histograms of the values selected from the distributions and used in the analysis, and to plot these histograms against the probability densities of the distributions to assure that the sampling corresponds well enough to the expectation given the density. Among other problems, in some situations (e.g. when combinations of sensitivity and specificity lead to negative adjusted cell counts as described earlier) the histogram of values used in MCSA may not correspond to the assigned probability density. Graphical diagnostics are also essential in a full Bayesian bias analysis because of the risk of poor convergence of the fitting algorithm.

Presentation of probabilistic bias analysis results may focus on the median, 2.5th percentile and 97.5th percentile of the modelling results, but the analyst should examine the entire histogram of adjusted estimates for implausible results and unexpected shapes. If results from some modelling iterations were discarded, for example due to negative cell frequencies in contingency tables, then the frequency of discarded iterations should be presented. If discarded results influenced the selection of values or distributions assigned to the bias parameters, then this influence should be described. Complete diagnostic results may be too detailed for presentation in a publication, but the description of the methods should explain which diagnostics were undertaken and that the model and computing code were found to perform adequately.

How does one present and interpret a bias analysis?

Presenting bias analysis methods

Bias analysis methods are unfamiliar to many readers of epidemiological research, so presentations of these methods should be as complete and detailed as reasonably

possible. A good presentation of a bias analysis should begin with a clear statement of its objectives, which should relate directly to some aspect of the conventional methods description. That is, the conventional methods section should foreshadow the bias analysis methods. The stated objective should then link to a bias model, such as an equation that links measured variables to the bias analysis result through the non-identifiable bias parameters. The presentation should then give values or distributions assigned to these bias parameters, explain the basis for the assignments in terms of plausibility with respect to background literature and give reasons for rejecting other reasonable alternatives that were explicitly considered. A good methods presentation should also provide an example of the calculations completed using the bias model. For multiple bias analysis, this presentation should be repeated for each bias and the order of analysis should be described and explained.

To illustrate these recommendations, consider a bias analysis to address misclassification of a binary exposure by varying assumed sensitivity and specificity. It should state that the objective of the bias analysis is to evaluate the influence of exposure misclassification. The bias model equations link the measured cell frequencies to the adjusted cell frequencies as a function of the sensitivities and specificities. Values assigned to these parameters might come from internal or external validation data, or probability distributions may be assigned using the methods described above. A 2×2 table of the measured frequencies should be linked to a 2×2 table of the adjusted frequencies with the bias model equation, where the terms of the model are replaced by the measured frequencies, adjusted frequencies and single values drawn from the assigned distributions (e.g. their ranges and modes). This presentation of the methods allows the reader to trace from the objective, to the bias model, to the information and judgments used to assign values to the bias model and finally to the output that provides one possible answer to the objective, conditional on the bias model and assigned values. The example calculation, although perhaps extraneous, ties all of the elements together for the reader.

Presenting bias analysis results

Presentation of bias analysis results might be as simple as presenting the adjusted estimate when only a single simple bias analysis has been completed. It is only when multiple values or multiple combinations of values are assigned to bias parameters, or when distributions are assigned to the bias parameters, that presentation of the results of a bias analysis become more difficult. In general, one should

present results of all bias analyses that were completed, not just those with interesting results. The main problem then becomes presentation of complete results in a manner that respects the word and space limitations enforced by most journals. Online supplements provide one common alternative to assure completeness.

Using tables allows the reader to evaluate different bias scenarios created by different assignment values or different combinations of assigned values, which is especially important for presenting the results of multidimensional bias analyses. The disadvantage of using tables is that data reduction is often necessary to deal with complexity, and tables provide no summary of the final uncertainty that arises from uncertainty about the bias parameters. For example, in a multiple-bias analysis there may be several equally important sources of bias. If so, the results need to be presented using high-dimensional tables that are unwieldy, difficult to interpret and which may needlessly highlight implausible parameter combinations. Further, it is cumbersome to incorporate uncertainty from random error into such tables in enough detail so that someone can repeat the bias analysis under different assumptions.

When table complexity overwhelms comprehension, figures usually provide a workable alternative. Three-dimensional column charts with multiple bars along each axis can present the results of even complex multidimensional bias analyses (see Flanders and Khoury,¹⁷ for example). For PBA, one can use tornado diagrams to compare multiple 95% interval estimates that are computed by incorporating uncertainty from each different bias source individually or in subsets (see Stonebraker *et al.*,⁷⁸ for example). Histograms that depict the frequency of adjusted estimates from the iterations of a probabilistic bias analysis can be used to compare bias analysis results with the conventional results, the results of various bias models with one another, and the progression of results across the sequence of models applied in multiple bias analysis.^{15,23}

Space and word count limitations may preclude presentation of all important results in tables or figures. In this case, bias analysis results can be presented in text: as single values yielded by the model and a single set of values assigned to the bias parameters (simple bias analysis); as a range of values yielded by the model and multiple values or multiple combinations of values assigned to the bias parameters (multidimensional bias analysis); or as median and simulation intervals (2.5th and 97.5th percentiles of the adjusted estimates) and medians yielded by a probabilistic bias analysis. Good practice will usually require a more complete presentation of results online or, less preferably, as a posting on the author's own internet site. In no case should concerns about space limitations or word limits

deter the most suitable bias analysis from being undertaken.

Interpreting bias analysis results

One of the advantages of bias analysis is to counteract the human tendency towards overconfidence in research results and inferences based on them.⁷⁶ It would be counterproductive, therefore, if the interpretation of a bias analysis exaggerated that overconfidence rather than diminished it, or if it substituted overconfidence in the bias analysis for overconfidence in the conventional analysis. We encourage interpretations of the bias analysis results to begin with a clear restatement of the assumptions underlying it, including the choice of biases to examine, the choice of bias models used in the analysis and the choice of values or distributions assigned to the non-identifiable parameters of the bias model. This restatement should then be summarized in any further presentation or interpretation of the bias analysis results with a preface such as: 'given the methods used and the assumptions of the analysis'.

The focus of the bias analysis interpretation should then turn to a description of any change in the inferences that might result from the bias analysis. One might write, for example, that the bias analysis suggests that confounding by an unmeasured variable might, or might not, plausibly account for the association observed in the conventional result, conditional on the accuracy of the bias model. Similar statements could be made regarding selection bias, measurement errors or combinations of biases. Recommendations for interpreting simple, multidimensional or and probabilistic bias analyses have been made elsewhere.¹⁵ We recommend against interpreting bias analysis results as proving or otherwise definitively answering whether a bias might, or might not, account for a given conventional result, because of the dependence on the accuracy of the underlying and non-verifiable assumptions.

By identifying the largest sources of uncertainty, sensitivity analyses of the bias analysis results, or the bias analysis results themselves, offer an opportunity for discussion of productive avenues for research improvement, such as where more accurate measurements, validation studies or more confounder measurements are needed. We recognize that general calls for further research are of little utility, but these specific avenues for further research are a direct product of bias analysis, so somewhat different from general statements.

Conclusions

Quantitative bias analysis serves several important goals in epidemiological research. First, it provides a quantitative

estimate of the direction, magnitude and uncertainty arising from systematic errors.^{15,23,26,28,79} Second, the very acts of identifying sources of systematic error, writing down models to quantify them, assigning values to the bias parameters and interpreting the results, combat the human tendency towards overconfidence in research results and the inferences that rest upon them.^{76,80,81} Finally, in an era of diminishing research funds, efficient allocation of sparse resources is becoming even more important. By suggesting aspects that dominate uncertainty in a particular research result or topic area, quantitative bias analysis can provide a guide for productive avenues of additional research;⁴¹ and, as happened with smoking and lung cancer³ and exogenous estrogens and endometrial cancer,⁸² quantitative bias analysis may reinforce more basic results by showing that a particular bias is probably not as important as some claim.

We advocate transparency in description of the methods by which biases were identified for analysis, models were developed and values were assigned to the model parameters. We also encourage bias analysts to make their data and computer code available for use by others, so that the results can be challenged by modifications to the model or by different choices for the values assigned to the model parameters. When data cannot be made freely available, bias analysts at a minimum should offer to incorporate credible modelling modifications and changes to the values assigned to the model parameters, when these are suggested by other stakeholders, and to report completely the results of these revised analyses. Bias models cannot be verified as empirically correct, nor are values assigned to the model parameters identifiable. It is, therefore, crucial that credible alternatives be given thorough examination.

Bias analysis is not a panacea. It cannot resolve fundamental problems with poor epidemiological research design or reporting, although it can account for uncertainties arising from design limitations. If there is investigator bias that introduces fraud into the data collection or analysis,³⁶ or incompletely represents the data collection and analysis process,⁸³ then no analysis can be expected to correct the resulting bias. Because the bias analyses we have discussed are designed for unselected analyses of individual studies, they cannot resolve inferential errors arising from selective reporting of research results, whether this is due to selective reporting of 'significant' associations or suppression of undesired associations.^{84–86} Methods of publication-bias analysis^{84–86} and forensic statistics^{87–89} can help to investigate these problems.

We hope to increase the proportion of health research that includes a quantitative estimate of the influence of systematic errors on research results. This quantification has

been long advocated.^{18,90} There was a time when some of us believed that such quantification was rarely undertaken because the methods were not widely known and because automated computing tools were not readily available to implement the methods. These shortcomings have been largely resolved.^{15,25,27,29,71} We must, therefore, contemplate other barriers to implementation.

One sociological barrier is the lack of demand for quantitative bias analysis by reviewers and editors of peer-reviewed journals.¹⁵ So long as the authors confine themselves to description of their study and resulting data, along with thorough discussion of possible explanations for the results and contrasts to other studies, this may not be a problem. But many reports still attempt to extend their own results into general inferences about causes, effects and their policy implications, often overweighting their results relative to other relevant research. In such cases, reviewers and editors are often too willing to excuse study imperfections, if they are confessed in the discussion section,²⁶ providing little motivation for researchers to use quantitative bias analysis. With rare exceptions,⁹¹ such analyses will only expand the uncertainty interval and call into question the validity of the inferences; that is, after all, a major point of quantitative bias analysis. Researchers have little motivation, aside from scientific integrity, to call their own inferences into question in this way, so the demand must come from the gatekeepers to publication. We hope that our guide to good practices for conducting and presenting bias analyses will make it easier for editors and reviewers to request quantitative bias analysis in lieu of narrative description of study imperfections when investigators insist on drawing broad conclusions about general relations and policy implications.

Acknowledgments

The authors thank Charles Poole, Paul Gustafson and the reviewers for their valuable comments and suggestions about earlier drafts of the manuscript. Any errors that remain are the sole responsibility of the authors.

Conflict of interest: None declared.

References

- Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biomet Bull* 1946;2:47–53.
- Bross I. Misclassification in 2x2 tables. *Biometrics* 1954;10:478–89.
- Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 1959;22:173–203.
- Bross ID. Spurious effects from an extraneous variable. *J Chron Dis* 1966;19:637–47.
- Barron B. The effects of misclassification on the estimation of relative risks. *Biometrics* 1977;33:414–18.
- Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105:488–95.
- Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980;112:564–69.
- Kupper LL, Karon JM, Kleinbaum DG, Morgenstern H, Lewis DK. Matching in epidemiological studies: validity and efficiency considerations. *Biometrics* 1981;37:271–91.
- Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched-pair studies. *Int J Epidemiol* 1983;12:93–97.
- Gail MH, Wacholder S, Lubin JH. Indirect corrections for confounding under multiplicative and additive risk models. *Am J Ind Med* 1988;13:119–30.
- Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: Wiley, 2002.
- Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton, FL: Chapman and Hall/CRC, 2003.
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C. *Measurement Error in Nonlinear Models*. Boca Raton, FL: Chapman and Hall, 2006.
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins, 2008.
- Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiological Data*. New York: Springer, 2009.
- Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol* 2006;35:1074–81.
- Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology* 1990;1:239–46.
- Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996;25:1107–16.
- Scharfstein D, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc* 1999;94:1096–120.
- Robins JM, Rotnitzky A, Scharfstein D. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran E, Berry D (eds). *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. New York: Springer, 1999.
- Rosenbaum PR. *Observational Studies*. 2nd edn. New York: Springer, 2002.
- Checkoway H, Pearce N, Kriebel D. *Research Methods in Occupational Epidemiology*. 2nd edn. New York: Oxford University Press, 2004.
- Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins, 2008.
- Eddy DM, Hasselblad V, Schachter R. *Meta-analysis by the Confidence Profile Method*. New York: Academic Press, 1992.
- Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003;14:451–58.
- Phillips CV, LaPole LM. Quantifying errors without random sampling. *BMC Med Res Methodol* 2003;3:9.

27. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol* 2005;34:1370–76.
28. Greenland S. Multiple-bias modeling for analysis of observational data. *J R Stat Soc Ser A* 2005;168:267–308.
29. Orsini N, Bellocco R, Bottai M, Wolk A, Greenland S. A tool for deterministic and probabilistic sensitivity analysis of epidemiological studies. *Stata J* 2008;8:29–48.
30. Greenland S. Bayesian perspectives for epidemiological research: III. Bias analysis via missing-data methods. *Int J Epidemiol* 2009;38:1662–73.
31. Greenland S. Sensitivity analysis and bias analysis. In: Ahrens W, Pigeot I (eds). *Handbook of Epidemiology*. 2nd edn. New York: Springer, 2013.
32. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 2003;98:47–54.
33. Lanes SF, Poole C. 'Truth in packaging?' The unwrapping of epidemiological research. *J Occup Med* 1984;26:571–74.
34. Rothman KJ, Poole C. Science and policy making. *Am J Public Health* 1985;75:340–41.
35. Greenland S, Gago-Dominguez M, Castela JE. The value of risk-factor ('black-box') epidemiology. *Epidemiology* 2004;15:529–35.
36. Greenland S. Transparency and disclosure, neutrality and balance: shared values or just shared words? *J Epidemiol Community Health* 2012;66:967–70.
37. Poole C. Low P-values or narrow confidence intervals: which are more durable? *Epidemiology* 2001;12:291–94.
38. Ahlbom A, Day N, Feychting M *et al*. A pooled analysis of magnetic fields and childhood leukaemia. *Br J Cancer* 2000;83:692–98.
39. Greenland S, Sheppard AR, Kaune WT, Poole C, Kelsh MA. A pooled analysis of magnetic fields, wire codes, and childhood leukemia. Childhood Leukemia-EMF Study Group. *Epidemiology* 2000;11:624–34.
40. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol* 2006;35:765–75.
41. Lash TL, Ahern TP. Bias analysis to guide new data collection. *Int J Biostat* 2012;8:4679.
42. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiological research. *Epidemiology* 1999;10:37–48.
43. Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002;31:163–65.
44. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
45. Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins, 2008.
46. Hernan MA, Cole SR. Invited Commentary: Causal diagrams and measurement bias. *Am J Epidemiol* 2009;170:959–62; discussion 63–64.
47. 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;37:382–85.
48. Drews CD, Greenland S. The impact of differential recall on the results of case-control studies. *Int J Epidemiol* 1990;19:1107–12.
49. Lash TL, Fink AK. Re: 'Neighborhood environment and loss of physical function in older adults: evidence from the Alameda County Study'. *Am J Epidemiol* 2003;157:472–73.
50. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. *Ann Epidemiol* 2006;16:834–41.
51. Hernan MA. The hazards of hazard ratios. *Epidemiology* 2010;21:13–35.
52. Flanders WD, Klein M. Properties of 2 counterfactual effect definitions of a point exposure. *Epidemiology* 2007;18:453–60.
53. Ogburn EL, VanderWeele TJ. On the nondifferential misclassification of a binary confounder. *Epidemiology* 2012;23:433–39.
54. Kristensen P. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology* 1992;3:210–15.
55. Chavance M, Dellatolas G. Bias from dependent misclassification. *Epidemiology* 1993;4:180–82.
56. VanderWeele TJ, Robins JM. Signed directed acyclic graphs for causal inference. *J R Stat Soc Ser B* 2010;72:111–27.
57. VanderWeele TJ, Hernan 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.
58. Bross ID. Pertinency of an extraneous variable. *J Chron Dis* 1967;20:487–95.
59. Yanagawa T. Case-control studies: Assessing the effect of a confounding factor. *Biometrika* 1984;71:191–94.
60. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol* 2008;18:637–46.
61. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 2011;22:42–52.
62. Gustafson P, Le ND, Saskin R. Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics* 2001;57:598–609.
63. Greenland S. Generalized conjugate priors for Bayesian analysis of risk and survival regressions. *Biometrics* 2003;59:92–99.
64. Gustafson P. On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Stat Sci* 2005;20:111–29.
65. Gustafson P. On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables - Rejoinder. *Stat Sci* 2005;20:137–40.
66. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. *J R Stat Soc Ser A* 2009;172:21–47.
67. Gustafson P, McCandless LC. Probabilistic approaches to better quantifying the results of epidemiological studies. *Int J Environ Res Public Health* 2010;7:1520–39.
68. Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 2004;160:384–92.

69. MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian? *Epidemiology* 2012;**23**:151–58.
70. Greenland S. Relaxation penalties and priors for plausible modeling of nonidentified bias sources. *Stat Sci* 2009;**24**:195–210.
71. MacLehose RF, Olshan AF, Herring AH *et al.* Bayesian methods for correcting misclassification: an example from birth defects epidemiology. *Epidemiology* 2009;**20**:27–35.
72. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc* 1994;**89**:846–66.
73. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. *J Am Stat Assoc* 2005;**100**:680–700.
74. Kahneman D, Slovic P, Tversky A. *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press, 1982.
75. Gilovich T, Griffin D, Kahneman D. *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 2002.
76. Lash TL. Heuristic thinking and inference from observational epidemiology. *Epidemiology* 2007;**18**:67–72.
77. Greenland S. Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Analysis* 2001;**21**: 579–83.
78. Stonebraker JS, Farrugia A, Gathmann B, Party ERW, Orange JS. Modeling primary immunodeficiency disease epidemiology and its treatment to estimate latent therapeutic demand for immunoglobulin. *J Clin Immunol* 2014;**34**:233–44.
79. Maldonado G. Adjusting a relative-risk estimate for study imperfections. *J Epidemiol Community Health* 2008;**62**:655–63.
80. Maclure M, Schneeweiss S. Causation of bias: the episcopo. *Epidemiology* 2001;**12**:114–22.
81. Savitz DA. *Interpreting Epidemiological Evidence Strategies for Study Design and Analysis*. Oxford, UK: Oxford University Press, 2003.
82. Greenland S, Neutra R. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. *J Chron Dis* 1981;**34**:433–3-8.
83. Phillips CV. Publication bias in situ. *BMC Med Res Methodol* 2004;**4**:20.
84. Henmi M, Copas JB, Eguchi S. Confidence intervals and P-values for meta-analysis with publication bias. *Biometrics* 2007;**63**:475–82.
85. Copas J, Dwan K, Kirkham J, Williamson P. A model-based correction for outcome reporting bias in meta-analysis. *Biostatistics* 2014;**15**:370–83.
86. Greenland S, O'Rourke K. Meta-analysis. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins; 2008.
87. Al-Marzouki S, Evans S, Marshall T, Roberts I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ* 2005;**331**:267–70.
88. Baggerly KA, Coombes KR. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Ann Appl Stat* 2009;**3**:1309–34.
89. Pogue JM, Devereaux PJ, Thorlund K, Yusuf S. Central statistical monitoring: detecting fraud in clinical trials. *Clin Trials* 2013;**10**:225–35.
90. Blair A, Steenland K, Shy C, O'Berg M, Halperin W, Thomas T. Control of smoking in occupational epidemiological studies: methods and needs. *Am J Ind Med* 1988;**13**:3–4.
91. Gustafson P, Greenland S. Curious phenomena in Bayesian adjustment for exposure misclassification. *Stat Med* 2006;**25**:87–103.