Unmeasured confounding in nonrandomized studies: quantitative bias analysis in health technology assessment

Thomas P Leahy[1], Seamus Kent[2], Cormac Sammon[1], Rolf HH Groenwold[3], Richard Grieve[4], Sreeram Ramagopalan*[,5] & Manuel Gomes[6]

[1]PHMR Ltd., Westport, F28 ET85, Ireland

[2]National Institute for Health & Care Excellence, Manchester, M1 4BT, UK

[3]Department of Clinical Epidemiology & Department of Biomedical Data Sciences, Leiden University Medical Centre, Einthovenweg 20, Leiden, 2333, The Netherlands

[4]Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK

[5]Global Access, F. Hoffmann-La Roche, Grenzacherstrasse 124 CH-4070, Basel, Switzerland

[6]Department of Applied Health Research, University College London, London, WC1E 6BT, UK

*Author for correspondence: sreeram.ramagopalan@roche.com

Evidence generated from nonrandomized studies (NRS) is increasingly submitted to health technology assessment (HTA) agencies. Unmeasured confounding is a primary concern with this type of evidence, as it may result in biased treatment effect estimates, which has led to much criticism of NRS by HTA agencies. Quantitative bias analyses are a group of methods that have been developed in the epidemiological literature to quantify the impact of unmeasured confounding and adjust effect estimates from NRS. Key considerations for application in HTA proposed in this article reflect the need to balance methodological complexity with ease of application and interpretation, and the need to ensure the methods fit within the existing frameworks used to assess nonrandomized evidence by HTA bodies.

Guidance from regulatory and health technology assessment (HTA) agencies recommend that in many settings, evidence on relative treatment effectiveness is generated from well-conducted randomised controlled trials (RCTs). However, the conduct of well-powered, unbiased RCTs is not always feasible, as can sometimes be the case with very rare diseases. In the HTA setting, the use of RCTs is further complicated by the fact they often provide comparison against technologies that do not reflect the standard of care relevant to the decision problem of interest [1]. Notably, the growth of personalised medicine and diversification of treatment options appear to be increasing the prominence of these issues. These initiatives have been accompanied by an increasing number of single-arm trials and a resultant increase in the use of estimates of treatment effectiveness from nonrandomized studies (NRS) to inform submissions to regulatory bodies and HTA agencies [2–4]. One recent review reported the number of HTA submissions including single-arm trials to have increased from 8 in 2011 to 102 in 2019, across several HTA agencies including the Canadian Agency for Drugs and Technologies in Health, the French National Authority for Health, the National Institute for Health and Care Excellence in the United Kingdom, the German Federal Joint Committee and Australia's Pharmaceutical Benefits Advisory Committee [4].

For medicines, given the initiatives towards faster access, particularly for obtaining marketing authorisation for oncology products, NRS submitted to regulators and payers are typically formed of a single-armed trial of the experimental treatment compared with another source of data reflecting the comparator treatment or control. In these cases, the control data can be obtained from a historic trial, from an observational cohort identified in real-world data sources such as electronic health record databases or patient registries, or a prospective observational study among other sources. However, NRS based solely on real-world data sources may also be submitted, particularly in HTA reassessments.

Future Medicine

A key concern with all NRS is confounding [5,6]. Rigorous design of NRS [7], in combination with the appropriate analytical methods, should be the primary approach to address confounding, particularly when complete, high-quality data on confounders are available [8,9]. Even in well-designed NRS, it is impossible to exclude the possibility of unmeasured confounding, and most NRS typically assume 'no unmeasured confounding' [10]. If the data on confounding variables are not accurate or are not collected in the data source, residual confounding may result in biased treatment effect estimates, which may lead to incorrect conclusions about the effectiveness and cost–effectiveness of the health technology. As one can never be certain that all confounders have been captured in an analysis, unmeasured confounding is a concern in all NRS.

Given the potential for unmeasured confounding in NRS, an important step in the design and analysis of such studies is to quantitatively explore the potential impact of unmeasured confounding on results [6,11]. There are a variety of so-called quantitative bias analysis (QBA) methods that have been developed to assess the impact of confounding and other sources of systematic error [11–13]. QBA offers a potentially powerful tool to support decision makers in utilising NRS and has been discussed recently in the context of regulatory decision making [13]. However, its use in HTA has received little attention [6,14]. This may be due to a lack of knowledge of the methods and/or uncertainties in how best to utilise them within existing HTA frameworks. This article introduces QBA methodologies that can be used to assess the potential impact of unmeasured confounding on the treatment effect estimates and the resultant impact on cost–effectiveness analyses, and discusses important considerations related to their use in HTA settings.

This article focuses on QBA methods that rely on external data and expert or analyst opinion to estimate the treatment effect typically after measured confounders have been adjusted for within a prespecified primary analysis. QBA methods that aim to account for unmeasured confounding using internal data (i.e., data on a subgroup of the population) and methods that are applied during the design or primary analysis of the study such as instrumental variable estimation and propensity score calibration are not considered within this article [15,16].

Overview of QBA methods for confounding

A large number of QBA methods have been developed over the past number of decades [17] and can be linked back to work by Cornfield *et al.* [12]. The methods are all based on a similar premise, that additional information describing the nature of unmeasured confounding can be utilised to adjust the estimates of interest to better reflect the 'true' estimates that would have been observed had the confounder(s) in question been accounted for.

Table 1 summarises several of these methods, dividing them into groups based on the analytical methods used. These include but are not limited to Bayesian twin-regression/hierarchical modelling, a simulation-based approach, Rosenbaum's approach, Rosenbaum-Rubin sensitivity analysis, and directly derived formulae approaches such as the calculation of bias factors, E-value and the use of contingency tables. Further commentary on each group of methods is provided in a supplementary file.

A common assumption of the developed methods is independence of unmeasured confounders from measured confounder(s). Under this assumption, there is a large number of methods that could be applied [31,34,45]. Typically, this assumption cannot be validated; in addition, there are potential unknown unmeasured confounders, in which case this assumption is even harder to justify. However, there are several approaches that do not make this assumption that may better reflect realistic scenarios [25,35–37].

In general, most QBA approaches assume a single hypothetical unmeasured confounder, which is a strong assumption due to the possibility that multiple unmeasured confounders could have a large joint effect. However, some methods, such as the simulation-based approach introduced in the work of Groenwold *et al.* [32] is one such approach that can adjust for multiple unmeasured confounders. Other methods assume particular forms of the unmeasured confounder. For example, several approaches have assumed that the single unmeasured confounder is binary in nature [18,26,38,45]. These approaches can still be applicable to settings with continuous unmeasured confounders by dichotomising the continuous variable, although this often leads to significant loss of information.

The value of these QBA methods is highly dependent on the quality of the external information available to inform them. External data is ideally obtained from a data source in which the relationship between the unmeasured confounder and other key variables has been (or can be) measured in a similar study population. The choice of QBA method may differ according to whether patient level or aggregate data are available. The methods that specifically use external data implicitly assume transportability between the external data and the original population – that is, that the external data are representative of the study sample regarding the joint distribution of the exposure, outcome and confounding variables. This assumption can be difficult to validate, typically due to the lack of

Method type	Description	Effect measure	Limitations	Ref.
Rosenbaum's approach	Finding the thresholds of association between the unmeasured confounder and the exposure and/or outcome variable that would result in the test statistic becoming insignificant	OR, HR	Lack of treatment effect estimate, assumes single unmeasured confounder, assumes ignorability	[18–21]
Rosenbaum-Rubin	Estimating the average effect of a treatment on a binary outcome after adjusting for covariates and an unobserved binary covariate	OR, RR, HR	Setting simulation parameters, assumes only rare outcomes for survival data, assumes ignorability, potential for investigator bias	[22–24]
Bayesian hierarchical/twin-regression modelling	A model is defined in hierarchies and the parameters of the respective posterior distributions are estimated through the Bayesian framework, and the unmeasured confounder is modelled as missing data; these methods avoid presenting many equally likely scenarios ('multiverse approach')	OR, HR, difference of continuous measures	Assumes a single unmeasured confounder, setting prior distribution, prior distribution dependency, potential for investigator bias	[23,25–30]
Simulation based	Hypothetical unmeasured confounders are simulated through assumptions and correlation with measured confounders and the outcome, and the treatment effect is estimated under different scenarios	OR, RR, RD, HR, difference of continuous measures	Assumes ignorability, simulation parameter settings, potential for investigator bias	[31–33]
Derived bias formulas (e.g., E-value, bias/bounding factor, contingency table, array/rule-out approach)	Use derived formulae from a statistical model typically with some assumptions to adjust the observed point estimate and confidence interval for a set of sensitivity parameters	OR, RR, RD, HR, difference of continuous measures	Defining sensitivity parameters, assumes rare outcomes for survival data, confounders independent to unmeasured confounder, potential for investigator bias	[34–44]

This table is not exhaustive; rather, it provides an indication of the types of QBA methods available, the settings in which they can be used and their limitations. There are other QBA methods available that have not been accounted for in this table.
HR: Hazard ratio; OR: Odds ratio; QBA: Quantitative bias analysis; RD: Risk difference; RR: Relative risk.

information on the correlation between the measured and unmeasured confounders. In the absence of data from an external source, one can use subjective evidence to assign values to the parameters of these methods. In such cases, evidence should ideally be provided by external experts with no vested interest in the outcome of the analysis and prespecified in study protocols to mitigate investigator bias.

In many settings, such high-quality external data is not available, and approaches that focus on threshold analysis, such as the E-value, and Rosenbaum's approaches are often preferred. Essentially, these methods assess how strong (association with treatment assignment and outcome) the unmeasured confounder would need to be, to change the conclusions of the study. This type of QBA method is relatively straightforward to implement and therefore tends to be more prevalent in practice [46–48] and meaningful interpretation of the results typically requires information of the nature required to parameterise other methods. As a result, it places the onus on the investigator and decision maker to interpret the plausibility of the relationships between the hypothetical unmeasured confounder, treatment and outcomes [49].

In studies with time-to-event outcomes, such as– survival or competing risk outcomes, the proportional hazards assumption is often required for QBA methods. The proportional hazards assumption assumes that each covariate has a multiplicative effect in the hazard function and is constant over time – that is, the ratio of any two hazards is constant over time. When this assumption is not met, QBA methods that assume proportional hazards [19,22,27,33,39,50] have limited applicability.

Key considerations for application in HTA
General considerations for good practice in the application of QBA methods have been proposed by Lash and colleagues [11]. Their recommendations broadly focus on the thorough and transparent selection and reporting of QBA methods and highlight the importance of pre-specification of any QBA in a study protocol. Specific topics discussed by Lash *et al.* [11] include how to select the sources of biases that ought to be addressed, how to select a method to model biases, how to assign bias parameter models, the use of transparent and credible methods to test the sensitivity of results to the choice of input parameters, and guidance on when QBA is advisable versus when it is essential. These recommendations were developed from an epidemiological perspective, and some important aspects relevant to the HTA setting were not covered.

Consideration	Summary
Table 2. Summary of the key considerations for the application of quantitative bias analysis methods in health technology assessment.	
Responsibility for conducting QBA	• The primary responsibility for carrying out QBA for HTA should lie with the manufacturer • However, there is a need for HTA agencies to provide clear guidelines for such analyses • In the absence of a manufacturer-submitted QBA, HTA bodies should carry out their own simple, conservative QBA
Identification of confounders	• Systematic and transparent approach to confounder identification needed • Based on search of the published literature, *de novo* studies and/or elicitation of expert opinion • Consider use of directed acyclic graphs in conceptualizing confounders and other sources of bias
Choice of QBA method	• Simple, threshold-based methods represent an accessible starting point for the use of QBA in HTA and possibly the minimum required in submissions that derive treatment effects from NRS • A requirement for more complex approaches will require more advanced knowledge and experience of the methods • The use of Bayesian approaches should be considered particularly when relevant prior evidence is available • In cost–effectiveness settings, the compatibility of the QBA method with cost–effectiveness decision models should be considered • The choice of method may depend on the type of outcome of interest
Valuation of QBA parameters	• Critically review the literature to inform the values of the sensitivity parameters of a QBA • In the absence of published data, the use of formal expert elicitation to value parameters may be acceptable to HTA bodies if best practices for elicitation are followed • Thorough valuation of QBA parameters should be carried out irrespective of the approach considered
Incorporation into evidence synthesis	• QBA can be applied to individual study results prior to evidence syntheses • QBA could potentially be used alongside methods such as design-adjusted methods to further mitigate the bias in NRS
HTA: Health technology assessment; NRS: Nonrandomized studies; QBA: Quantitative bias analysis.	

In the following, we discuss several additional considerations required for the application of QBA for confounding in the HTA context (Table 2).

Responsibility for conducting QBA

As with other evidence generation activities for HTA, we believe the primary responsibility for carrying out QBA should lie with the manufacturer. However, there is a need for HTA agencies to support manufacturers in conducting QBA by providing clear guidelines on their preferred QBA methods. Additionally, where a manufacturer does not provide sufficient consideration of confounding in a submission containing a NRS or claims there to be little or no evidence suggestive of unmeasured confounding, it may be prudent for HTA bodies to utilise QBA to ensure the results of a NRS demonstrates robustness to some default hypothetical level(s) of confounding. Notably, the knowledge that a default, relatively conservative approach to QBA would be applied in the absence of more considered approaches may better incentivise manufacturers to carry out thorough work to identify and address confounding, both in their main analysis and in any QBA. The subsequent considerations are provided in this context.

Identification of confounders

As with general guidance on QBA practices, and NRS more generally, we recommend a systematic and transparent approach to the identification of relevant confounders during study design [7,11]. Manufacturers should include a systematic search of the published literature and elicitation of expert opinion. The use of directed acyclic graphs is becoming increasingly common in the design of NRS, and they are a simple and transparent approach to identify potential causal relationships between variables of interest [11,51]. We would strongly advocate for their use in conceptualising and communicating potential sources of bias in effect estimates from NRS for submission to HTA bodies.

Where it is anticipated that NRS will form a key part of a HTA submission, discussion and alignment on potential sources of bias and the need for QBA, should be considered during early engagement with HTA bodies and other key stakeholders. All details regarding the potential sources of bias and planned QBA analyses should be prespecified in the study protocol and statistical analysis plan.

Choice of QBA method

Determining the most appropriate QBA method in practice is not trivial. Here, we provide some important considerations to the use of QBA methods in a HTA setting.

The relative simplicity of threshold-based methods such as the E-value means that they have the potential to be widely used in HTA as has been the case in epidemiological settings [46–48,52]. These approaches are akin to the

'dramatic effects' approach currently taken by the independent Institute for Quality and Efficiency in Health Care (IQWiG) in Germany, under which nonrandomized treatment effects may be considered acceptable if they are of a sufficient magnitude and meet a number of other key criteria [53]. Like the IQWiG approach, these approaches could utilise a relatively conservative set of assumptions/thresholds to ensure that confounding is not inappropriately ruled out as a potential explanation. In cost–effectiveness settings, threshold-based approaches could also be anchored against a cost–effectiveness threshold, rather than a null clinical effect, to better support decision making.

Although their simplicity and interpretation are appealing, there are limitations to the threshold-based methods. Notably, some of their simplicity comes at the cost of placing a greater responsibility on the end user (in this case the HTA body) in interpreting the plausibility of the results. In some cases, the use of the nonthreshold-based methods may allow QBA to be formulated in a way that is more readily understood by HTA agencies. Further, HTA decision making is often based on a consideration of the magnitude of relative treatment effect(s), in contrast to regulators who are more interested in establishing the simple presence or absence of the effect(s). As a result, QBA approaches that focus on prespecification of scenarios of unmeasured confounding and assessment of their impact on the observed treatment effect are likely to be better aligned with decision problems of this nature than threshold-based methods.

The use of threshold-based approaches perhaps represents an accessible starting point for the uptake of QBA methods in the HTA setting and a minimum level of QBA that should be required in HTAs involving NRS. In contrast, the approaches in which the impact of prespecified scenarios is modelled perhaps represents an ideal practice which manufacturers could be increasingly encouraged to pursue as familiarity with the methods within the HTA community grows. Irrespective of the approach used, the choice of QBA should be justified and informed by subject matter expertise in the relevant clinical context.

Another key question in selecting a QBA method relates to the choice of a frequentist or Bayesian approach. QBA typically involves exploring a range of plausible sensitivity parameter values, and frequentist-type QBA methods often assume that all different sensitivity parameter values, and the resultant set of treatment effect estimates, are equally likely. This typically leads to a multiverse approach, where all results of many different possible analyses are presented to a reader. Although broadly informative, particularly when coupled with transparent reporting of assumptions, multiverse approaches may leave HTA decision makers with challenges in leveraging the results to support their decision making. In contrast, Bayesian approaches typically evaluate the entire posterior distribution and as such provide the reader with a more succinct synthesis to support their decision problem. The advantages of Bayesian methods would only be fully realised if sufficient historical data are available to inform the specification of the prior distribution and other model parameters. The Bayesian approach also requires that HTA decision makers are familiar with the review and interpretation of Bayesian inference, which may not be the case in all jurisdictions.

Another important consideration to the choice of QBA method is the extent to which different approaches can facilitate the incorporation of the uncertainty directly into cost–effectiveness decision models. Methods which can be integrated into a cost–effectiveness model would be likely to allow for the better integration of the QBA results into many of the key outputs used to support decision making by these HTA bodies. For example, some approaches could be incorporated into the existing deterministic and probabilistic sensitivity analysis approaches applied in cost–effectiveness models thereby allowing for the consideration of the structural uncertainty introduced by these biases alongside other sources of uncertainty. Although QBA methods that make use of individual patient-level data have some methodological advantages over those that are applied to aggregate data, such as allowing for the alignment of outcomes and greater scope to assess the comparability of populations, they may be more challenging to apply directly in a decision model and would have the added complexity of requiring review of a separate patient-level analysis. As a result, some HTA bodies may have a preference for those that can be applied to aggregate data.

The choice of method will also be determined by the outcomes under study and the extent to which the available methods have been developed to handle them. Although most QBA methods can be adapted for different outcome types and effect estimates, their application to time-to-event end points is more challenging and the options available are more limited.

Additionally, other than a bias induced due to unmeasured confounding, there are other potential biases that may impact the results of NRS, such as information bias (i.e., measurement error, misclassification) and selection bias. Although we have not reviewed QBA methods related to other biases within the scope of this article, guidelines for QBA suggest that all potential sources of bias should be discussed and, if possible, assessed using QBA. As such,

in choosing or recommending a QBA method for use in HTA settings, it may be prudent to consider the extent to which a method allows for multiple types of biases to be addressed within a single analysis [53].

Whichever QBA method is chosen, it is crucial that its choice is clearly justified and transparently reported, and that the interpretation of the results are consistent with its underlying assumptions. Any corresponding software code used in the QBA should be made available to aid in replicability.

Valuation of QBA parameters

A literature review used to identify potential confounders can also typically serve the purpose of providing some data to inform the values to assign the sensitivity parameters of a QBA. Where sufficient data is not identified in the literature, the conduct of *de novo* epidemiological studies should be considered [54]. When external data are available, assumptions regarding transportability of relationships across populations should be carefully considered and critically examined as highlighted previously.

In practice, the data required to support data-driven approaches to QBA are often unavailable, and more subjective approaches to assigning parameter values are commonly necessary. For example, formal expert elicitation methods are already commonly used to support HTA, and it is the view of many HTA agencies that when no other empirical evidence is available, expert elicitation can be considered. However, given the centrality of relative efficacy/effectiveness estimates to the HTA process, the potential for inaccuracy in the elicited information is likely to be a major concern. The potential for conflicts of interest to influence expert inputs is likely to be a further concern. As a result, it is vital that any QBA based on expert elicitation follow formal, protocol-driven best practices.

Notably, although the threshold-based approaches to QBA do not require evidence of this nature to inform their application, such information is needed to inform the plausibility of the sensitivity analysis scenarios. Therefore, we would advocate for collection of such data to inform QBA regardless of the approach considered.

Incorporation into evidence synthesis

In the HTA setting, it may be the case that there are multiple studies informing an estimate of treatment effect, and there is a requirement to synthesise these in an indirect treatment comparison before incorporating them into a cost–effectiveness model. Recent guidance from NICE on evidence synthesis appears to group QBA methods under 'design-adjusted' approaches – one of several possible options they suggest to consider when incorporating NRS into indirect treatment comparisons [55]. An advantage of QBA or design-adjusted approaches is that they require explicit consideration of the sources and nature of biases. This is in contrast to other available methods that synthesise the results of NRS focusing on broadly decreasing the extent to which the NRS influence results relative to RCTs. An approach which both applies QBA/design-adjustment methods and utilises methods which limit the influence of NRS in the synthesis may allow one to harness the advantages of both groups of approaches.

It is expected that in the short term, the majority of situations in which nonrandomized data are submitted to HTA bodies will be in settings where there is quite limited evidence available, and therefore evidence synthesis may not be relevant. However, it will be important to consider the role of QBA as approaches to synthesis of NRS for HTA are further developed.

Conclusion

A variety of QBA methods have been proposed to adjust for unmeasured confounding in the estimation of treatment effects in NRS. These methods can be applied to a range of different types of end points, using independent patient-level or aggregate data, adopting either a Bayesian or frequentist perspective. Although QBA should not be viewed as the only approach to address biases in NRS, when conceived and implemented as part of a well-designed and analysed NRS, these methods can play an important role in ensuring that estimates of treatment effects from NRS are appropriately considered in HTA.

This article highlights several methodological and practical considerations which must be examined if the methods are to be successfully used in the HTA setting. These points reflect the need to balance methodological complexity with ease of application and interpretation, the need to ensure objectivity and transparency in their application and the need to determine suitability of the methods to fit within the existing frameworks used to assess evidence by HTA bodies. Clear methodological guidelines on the application of QBA in the HTA setting will aid the uptake of these methods. The development of any such guidelines should include the input of various stakeholders including HTA bodies, manufacturers, and clinical and methodological experts.

Future perspective

Quantitative bias analysis (QBA) for unmeasured confounding represents a potentially useful tool for decision makers, particularly when assessing outcomes from non-randomised studies. Their application in the health technology assessment (HTA) setting raises a number of unique challenges. Collaboration between academia, industry and HTA agencies to further increase the awareness of the potential of QBA in HTA and to develop best practices for their use, will be key to increasing their implementation in this setting.

Executive summary

- Quantitative bias analysis for unmeasured confounding is well placed to support the use of nonrandomized studies in health technology assessment (HTA).
- Key considerations for application in HTA reflect the need to balance methodological complexity with ease of application and interpretation.
- Ensuring quantitative bias analysis, the methods fit within the existing frameworks used to assess evidence by HTA bodies will increase transparency and application of such methods.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection, analysis and writing of the first draft of the manuscript were performed by TP Leahy and C Sammon. All authors commented on previous versions of the manuscript and read and approved the final manuscript.

References

1. Ayling K, Brierley S, Johnson B, Heller S, Eiser C. How standard is standard care? Exploring control group outcomes in behaviour change interventions for young people with type 1 diabetes. *Psychol. Health* 30(1), 85–103 (2015).

2. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. *BMJ Open* 6(6), e011666 (2016).

3. Griffiths EA, Macaulay R, Vadlamudi NK, Uddin J, Samuels ER. The role of noncomparative evidence in health technology assessment decisions. *Value Health* 20(10), 1245–1251 (2017).

4. Patel D, Grimson F, Mihaylova E *et al.* Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health* 24(8), 1118–1125 (2021).

5. Vanderweele TJ, Arah OA. Unmeasured confounding for general outcomes, treatments, and confounders: bias formulas for sensitivity analysis. *Epidemiology (Cambridge, Mass.)* 22(1), 42 (2011).

6. Sammon CJ, Leahy TP, Gsteiger S, Ramagopalan S. Real-world evidence and nonrandomized data in health technology assessment: using existing methods to address unmeasured confounding? *J. Comp. Eff. Res.* 9(14), 969–972 (2020).

7. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* 183(8), 758–764 (2016).

8. Faria R, Alava MH, Manca A, Wailoo AJ. *NICE DSU Technical Support Document 17: The Use of Observational Data to Inform Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data.* NICE Decision Support Unit, Sheffield, UK (2015).

9. Phillippo DM, Ades AE, Dias S. *NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE.* NICE Decision Support Unit, Sheffield, UK (2016).

10. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health Econ.* 22(4), 486–500 (2013).

11. Lash TL, Fox MP, Maclehose RF, Maldonado G, Mccandless LC, Greenland S. Good practices for quantitative bias analysis. *Int. J. Epidemiol.* 43(6), 1969–1985 (2014).

12. 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. Nat. Cancer Inst.* 22(1), 173–203 (1959).

13. Lash TL, Fox MP, Cooney D, Lu Y, Forshee RA. Quantitative bias analysis in regulatory settings. *Am. J. Public Health* 106(7), 1227–1230 (2016).

14. National Institute for Health Care and Excellence. *Appendix I: Real World Evidence Framework*. NICE. Sheffield, UK (2021).

15. Brumback BA, Hernán MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat. Med.* 23(5), 749–767 (2004).

16. Ertefaie A, Small DS, Flory JH, Hennessy S. A tutorial on the use of instrumental variables in pharmacoepidemiology. *Pharmacoepidemiol. Drug Saf.* 26(4), 357–367 (2017).

17. Rosenbaum PR. Sensitivity to hidden bias. In: *Observational Studies*. Springer, NY, USA, 105–170 (2002).

18. Nattino G, Lu B. Model assisted sensitivity analyses for hidden bias with binary outcomes. *Biometrics* 74(4), 1141–1149 (2018).

19. Lu B, Cai D, Tong X. Testing causal effects in observational survival data using propensity score matching design. *Stat. Med.* 37(11), 1846–1858 (2018).

20. Hasegawa R, Small D. Sensitivity analysis for matched pair analysis of binary data: from worst case to average case analysis. *Biometrics* 73(4), 1424–1432 (2017).

21. Rosenbaum PR. Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika* 74(1), 13–26 (1987).

22. Lin NX, Logan S, Henley WE. Bias and sensitivity analysis when estimating treatment effects from the Cox model with omitted covariates. *Biometrics* 69(4), 850–860 (2013).

23. McCandless LC, Gustafson P, Levy AR, Richardson S. Hierarchical priors for bias parameters in Bayesian sensitivity analysis for unmeasured confounding. *Stat Med* 31(4), 383–396 (2012).

24. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J R Stat Soc Series B Stat Methodol* 45(2), 212–218 (1983).

25. Gustafson P, McCandless LC, Levy AR, Richardson S. Simplified Bayesian sensitivity analysis for mismeasured and unobserved confounders. *Biometrics* 66(4), 1129–1137 (2010).

26. McCandless LC, Gustafson P. A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding. *Stat. Med.* 36(18), 2887–2901 (2017).

27. Huang R, Xu R, Dulai PS. Sensitivity analysis of treatment effect to unmeasured confounding in observational studies with survival and competing risks outcomes. *Stat. Med.* 39(24), 3397–3411 (2020).

28. McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med* 26(11), 2331–2347 (2007).

29. McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding. *J Clin Epidemiol* 61(3), 247–255 (2008).

30. Zhang X, Faries DE, Boytsov N, Stamey JD, Seaman JW. A Bayesian sensitivity analysis to evaluate the impact of unmeasured confounding with external data: a real world comparative effectiveness study in osteoporosis. *Pharmacoepidemiol Drug Saf* 25(9), 982–992 (2016).

31. Dorie V, Harada M, Carnegie NB, Hill J. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Stat. Med.* 35(20), 3453–3470 (2016).

32. Groenwold RHH, Sterne JAC, Lawlor DA, Moons KGM, Hoes AW, Tilling K. Sensitivity analysis for the effects of multiple unmeasured confounders. *Ann. Epidemiol.* 26(9), 605–611 (2016).

33. Barrowman MA, Peek N, Lambie M, Martin GP, Sperrin M. How unmeasured confounding in a competing risks setting can affect treatment effect estimates in observational studies. *BMC Med. Res. Methodol.* 19(1), 166 (2019).

34. Corrao G, Nicotra F, Parodi A *et al.* External adjustment for unmeasured confounders improved drug-outcome association estimates based on health care utilization data. *J. Clin. Epidemiol.* 65(11), 1190–1199 (2012).

35. Ding P, Vanderweele TJ. Sensitivity analysis without assumptions. *Epidemiology* 27(3), 368–377 (2016).

36. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22(1), 42–52 (2011).

37. Vanderweele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* 167(4), 268–274 (2017).

38. Greenland S. Basic methods for sensitivity analysis of biases. *Int. J. Epidemiol.* 25(6), 1107–1116 (1996).

39. Vanderweele TJ. Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects. *Eur. J. Epidemiol.* 28(2), 113–117 (2013).

40. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol* 18(8), 637–646 (2008).

41. Cusson A, Infante-Rivard C. Bias factor, maximum bias and the E-value: insight and extended applications. *Int J Epidemiol* 49(5), 1509–1516 (2020).

42. Mathur MB, Vanderweele TJ. Robust metrics and sensitivity analyses for meta-analyses of heterogeneous effects. *Epidemiology* 31(3), 356–358 (2020).

43. Mittinty MN. Estimating bias due to unmeasured confounding in oral health epidemiology. *Community Dent Health* 37(1), 84–89 (2020).

44. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 15(5), 291–303 (2006).

45. Groenwold RHH, Nelson David BDB, Nichol Kristin LKL, Hoes AW, Hak E. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *Int. J. Epidemiol.* 39(1), 107–117 (2010).

46. Gong CL, Song AY, Horak R *et al.* Impact of confounding on cost, survival, and length-of-stay outcomes for neonates with hypoplastic left heart syndrome undergoing stage 1 palliation surgery. *Pediatr. Cardiol.* 41(5), 996–1011 (2020).

47. Hay JW, Gong CL, Jiao X *et al.* A US population health survey on the impact of COVID-19 using the EQ-5D-5L. *J. Gen. Intern. Med.* 36(5), 1292–1301 (2021).

48. Schultze A, Walker AJ, Mackenna B *et al.* Risk of COVID-19-related death among patients with chronic obstructive pulmonary disease or asthma prescribed inhaled corticosteroids: an observational cohort study using the OpenSAFELY platform. *Lancet Respir. Med.* 8(11), 1106–1120 (2020).

49. Barberio J, Ahern TP, MacLehose RF *et al.* Assessing techniques for quantifying the impact of bias due to an unmeasured confounder: an applied example. *Clin. Epidemiol.* 13, 627 (2021).

50. Klungsoyr O, Sexton J, Sandanger I, Nygard JF. Sensitivity analysis for unmeasured confounding in a marginal structural Cox proportional hazards model. *Lifetime Data Anal.* 15(2), 278–294 (2009).

51. Tennant PW, Murray EJ, Arnold KF *et al.* Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int. J. Epidemiol.* 50(2), 620–632 (2021).

52. Blum MR, Tan YJ, Ioannidis JP. Use of E-values for addressing confounding in observational studies – an empirical assessment of the literature. *Int. J. Epidemiol.* 49(5), 1482–1494 (2020).

53. Institute for Quality and Efficiency in Health Care (IQWiG). *IQWiG General Methods: Version 6.0* IQWiG, Nordrhein-Westfalen, Germany (2020).

54. Leahy TP, Ramagopalan S, Sammon C. The use of UK primary care databases in health technology assessments carried out by the National Institute for Health and Care Excellence (NICE). *BMC Health Serv. Res.* 20(1), 1–9 (2020).

55. Abrams K. *CHTE2020 Sources and Synthesis of Evidence; Update to Evidence Synthesis Methods.* NICE Decision Support Unit, Sheffield, UK (2020).