



Application of quantitative bias analysis for unmeasured confounding in cost–effectiveness modelling

Journal of **Comparative Effectiveness Research**

Thomas P Leahy¹, Stephen Duffield², Seamus Kent², Cormac Sammon¹, Dimitris Tzelis³, Joshua Ray⁴, Rolf HH Groenwold⁵, Manuel Gomes⁶, Sreeram Ramagopalan^{*4} & Richard Grieve⁷

¹PHMR LTD, Westport, F28 ET85, Ireland

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

³PHMR LTD, London, NW1 8XY, England

⁴Global Access, F. Hoffmann-La Roche, Grenzacherstrasse 124, CH-4070, Basel, Switzerland

⁵Leiden University Medical Centre, Department of Clinical Epidemiology & Department of Biomedical Data Sciences, Einthovenweg 20, 2333, ZC Leiden, The Netherlands

⁶University College London, London, WC1E 7HB, UK

⁷London School of Hygiene & Tropical Medicine, London, WC1H 9SH, UK

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

Due to uncertainty regarding the potential impact of unmeasured confounding, health technology assessment (HTA) agencies often disregard evidence from nonrandomised studies when considering new technologies. Quantitative bias analysis (QBA) methods provide a means to quantify this uncertainty but have not been widely used in the HTA setting, particularly in the context of cost–effectiveness modelling (CEM). This study demonstrated the application of an aggregate and patient-level QBA approach to quantify and adjust for unmeasured confounding in a simulated nonrandomised comparison of survival outcomes. Application of the QBA output within a CEM through deterministic and probabilistic sensitivity analyses and under different scenarios of knowledge of an unmeasured confounder demonstrates the potential value of QBA in HTA.

First draft submitted: 14 February 2022; Accepted for publication: 24 May 2022; Published online: 9 June 2022

Keywords: cost–effectiveness • HTA • nonrandomised • quantitative bias analysis • unmeasured confounding

Health technology assessment (HTA) bodies systematically evaluate a wide range of evidence. The type of evidence considered depends on the reimbursement process of that country; for example, in England, the National Institute for Health and Care Excellence (NICE) considers both clinical and cost–effectiveness (CE), as well as budget impact [1]. A cost–effectiveness decision model (CEM) synthesises information about estimated costs and effects; hence, a critical component is information on the clinical effectiveness.

Randomised controlled trials (RCTs) are generally considered to be the gold standard of evidence when assessing effectiveness. However, in some cases, an RCT is not feasible, such as when there are ethical concerns or for rare outcomes (requiring a large sample size to achieve sufficient power). Therefore, HTA bodies are increasingly presented with nonrandomised evidence, including unanchored (indirect) treatment comparisons. Comparator data commonly includes previously conducted trials and real-world data (RWD).

One of the key issues with the use of results from nonrandomised studies (NRS) is confounding [2]. Although observed confounding variables can be controlled for in the analysis, such as through matching, unmeasured confounding variables cannot generally be controlled for. Unmeasured confounding may bias treatment effect estimates and consequently lead to misleading CE results. For these reasons, HTA agencies have remained sceptical about the value of real-world evidence for informing healthcare decision making [3].

Recent recommendations have suggested that it is essential to perform a broad range of sensitivity analyses, including quantitative bias analyses (QBAs), to assess the various types of biases that could occur by using nonrandomised data [3–5]. In a companion paper, we discuss which approaches may be the most appropriate in

Table 1. Nomenclature for the level of knowledge of the unmeasured confounder to guide the parameterisation of the quantitative bias analysis methods.

Term	Real-world scenario
Good knowledge	Where one obtains high-quality external data or elicits expert opinion that broadly reflects the true associations with the unmeasured confounder
Poor knowledge	Where one obtains low-quality external data or elicits poorly informed expert opinion that does not fully reflect the true associations with the unmeasured confounder
Incorrect knowledge	Where one elicits expert opinion that is incorrect or there was no information and incorrect assumptions were made that in no way reflected the true associations with the unmeasured confounder

different scenarios and the practicalities of implementing such approaches in a CEM [2]. However, these methods have not yet been demonstrated in practice to illustrate their impact on CE results.

The focus of this study is to demonstrate that the use of QBA methods in NRS which can support the assessments of uncertainty in CEM outputs and facilitate informed reimbursement decisions. The presented sensitivity analysis approaches adjust a simulated time-to-event effect estimate for unmeasured confounding after adjusting for measured confounders and incorporate the results into a CEM through both deterministic sensitivity analysis (DSA) and probabilistic sensitivity analysis (PSA). Simulated clinical data was used to investigate the impact of the methods under various real-world scenarios based on the quality of data available to inform the QBA. Specifically, scenarios in which the parameterisation of the QBA methods (based on the knowledge of the unmeasured confounder) is relatively ‘good’/relatively ‘poor’ compared with settings where the parameterisation is totally incorrect.

Methods

Simulating survival data

Artificial data were created to illustrate the QBA methods. A single dataset was generated to reflect a plausible scenario that may arise in an oncology setting, particularly for certain oncogene-positive populations where limited information may be available and single-arm trials were conducted, such as an ALK+ or EGFR+ metastatic non-small-cell lung cancer population. Specifically, individual patient-level overall survival (OS) data were simulated, representing information about $n = 300$ hypothetical study participants, using the approach described in [Supplementary Material Appendix 1](#). Two marginally independent measured baseline characteristics, a binary treatment variable and a binary unmeasured confounder were simulated. The simulated patient dataset was constructed such that it represented a hypothetical single-arm trial and RWD control arm with realistic representation of patient numbers and survival outcomes. The effect of the unmeasured confounder on the outcome was set at 3.43 (HR), a magnitude of association which appears plausible based on the strength of prognostic factors often unavailable in real-world oncology datasets such as the performance status (e.g., Eastern Cooperative Oncology Group) and tumour stage [6,7]. All patients were right censored after 5 years of follow-up, and an independent censoring mechanism was assumed over follow-up to achieve a proportion of patients censored between 15 and 25%. This single dataset of artificial patients was used to investigate the impact of the QBA methods with varying levels of knowledge of the unmeasured confounder to parameterise the methods (Table 1).

Additionally, the CEM required corresponding and clinically plausible progression-free survival (PFS) data. The algorithm to simulate the PFS data took a patient-level approach and generated progression events that either occurred before an OS event or at the same time as the OS event, or alternatively were censored.

Analysis

The Cox model was used to estimate the treatment effect and the parameters of the covariates [8]. The hazard ratio (HR) of the treatment effect was the outcome of interest and used as an input into the CEM. The measured confounders were adjusted for in the observed model, and both the measured and unmeasured confounders were adjusted for in the true model, which served as a reference.

Sensitivity analysis methods for unmeasured confounding

Several QBA methods have been described in the literature. In this study, two QBA approaches were used to adjust the treatment effect HR under varying assumptions of the unmeasured confounder. The first approach required patient-level data [9], whereas the second approach adjusted the HR on the aggregate level [10].

The first approach, described by Huang *et al.* [9] was work that was extended from Carnegie *et al.* [11] whereby an unmeasured confounder is simulated given an estimated marginal probability of the unmeasured confounder, and sensitivity parameters specified to quantify the assumed association between the unmeasured confounder and the treatment received, and the log hazard of the outcome. This approach assumes a single binary unmeasured confounder. The R package *survSens* was used for implementation with the expectation-maximisation (EM) algorithm method option [9].

The second approach used was introduced by Ding *et al.* [10] and adjusts the point estimate and corresponding CI on the aggregate level. The sensitivity assumptions required to adjust the observed treatment effect included an association between the unmeasured confounder and the outcome, and between the unmeasured confounder and the treatment. Under the assumption of a rare outcome and a binary unmeasured confounder, a bounding factor can be defined and used to adjust the estimated treatment effect

$$\text{Bounding factor} = \frac{RR_{EU} \times HR_{UT}}{RR_{EU} + HR_{UT} - 1}$$

where RR_{EU} is the estimated relative risk (RR) between the exposure (E) and the unmeasured confounder (U), and HR_{UT} is the estimated HR between the unmeasured confounder and the outcome (T). An outcome is considered relatively rare when it occurs at a rate <15% by the end of follow-up [12]. However, in this study, the assumption of a rare outcome will be violated given the nature of the simulated population reflecting a realistic oncology scenario. To overcome this limitation, the square root transformation was used to estimate an RR from a HR for a nonrare outcome, then the bounding factor formula defined for RR was used to adjust the estimated treatment effect [10,12,13].

Three hypothetical scenarios each representing different levels of knowledge of the unmeasured confounder to parameterise QBA methods were examined. These scenarios are defined in Table 1. The sensitivity parameters used in each scenario for each approach and resultant adjusted treatment effect estimates are provided in Supplementary Material Appendix 2 and reflect the assumptions one may make under uncertainty of the true association with an unmeasured confounder. For each scenario, we define a range of sensitivity parameters to reflect the likely uncertainty around parameter values one would encounter in real-life application of these methods. Notably, all parameters are selected to ensure that combinations are statistically feasible [14]. For the good scenario, the parameters were selected close to the true values with some uncertainty. For the poor scenario, the parameters were offset from the true parameters with some uncertainty, and the direction of association reflected the true direction of association. In the incorrect scenario, the parameters were offset to a greater extent from the true parameters compared with the poor scenario with some uncertainty and includes one parameter with a direction of association opposite to that of the true association.

Cost–effectiveness model

To illustrate how the application of QBA methods within the context of a CEM can be used to investigate uncertainty in the model outputs, a stylised CE three-state-partitioned survival analysis (PartSA) model was developed. The model was loosely based on a real case study of a novel therapy in lung cancer, for which the modelling has been substantially simplified for the purposes of illustration and therapies under consideration remain anonymised. Full details of the CEM can be found in Supplementary Material Appendix 3.

For the implementation into the CEM, all unmeasured confounding-adjusted HRs were generated and then input into the CEM for use in the evaluation of the CE outcomes. Firstly, the ‘true’ (albeit unknown) HR was estimated using the entirety of the available retrospective data. Then analyses were conducted using a comparison adjusted for measured confounders of survival from the simulated data, as well as an additional comparison using the two methods to adjust for assumed unmeasured confounding [9,10].

Both costs and utilities were assigned to three health states (preprogression, progressed disease and death) in the CEM, representing ongoing care and drug costs, as well as quality of life associated with each health state and treatment arm.

The outcomes produced by the CE analysis were the mean cost per patient and mean quality-adjusted life-years (QALYs) gained per patient. Standard composite outcomes such as the incremental cost–effectiveness ratio (ICER) between the intervention and the comparator, and the net monetary benefit (NMB), were also reported. The latter uses the costs (C), QALYs (Q) and an incremental CE threshold (λ) to express differences between the intervention

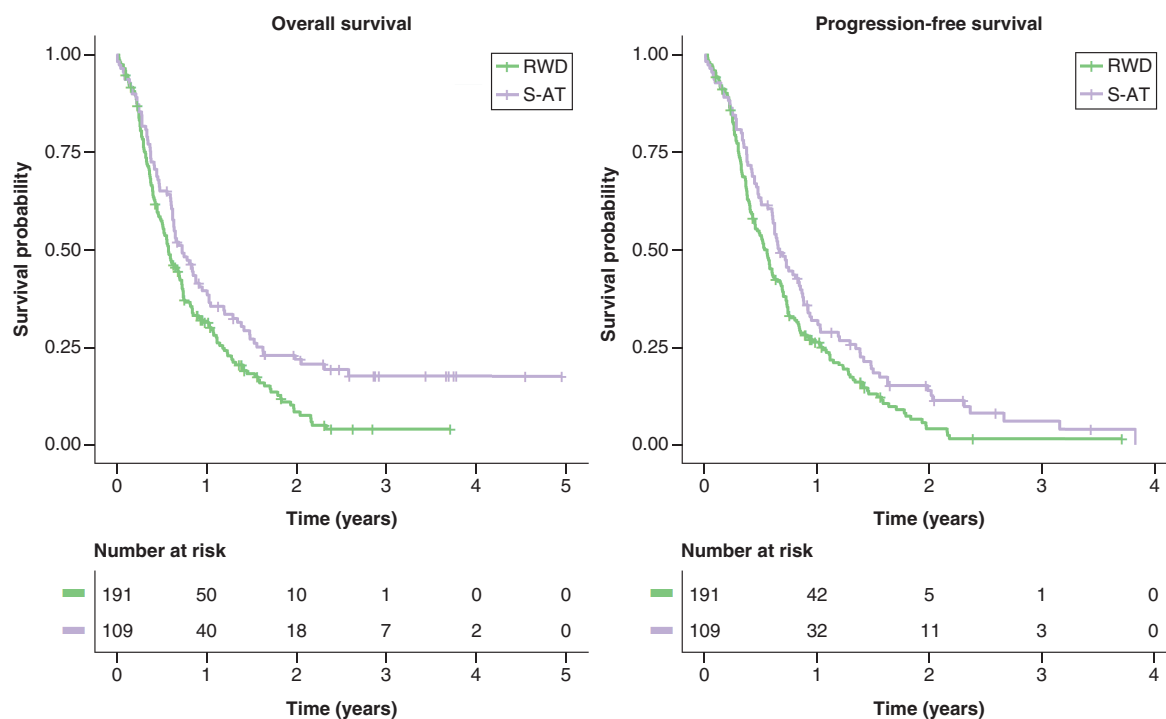


Figure 1. Kaplan-Meier curves of the simulated population for overall survival and progression-free survival. RWD: Real-world data; S-AT: Single-arm trial.

and the comparator in a single monetary metric:

$$NNM = Q \times \lambda - C$$

Point estimates and 95% credibility intervals were reported for all outcomes estimated by the CEM.

DSA and PSA were conducted to assess the sensitivity of the CEM outcomes to the treatment effect estimate under various assumptions of unmeasured confounding. For the PSA, a Monte Carlo simulation was conducted in which all parameter values were sampled from predefined sampling distributions over 30,000 iterations.

Results

The simulated patient-level dataset consisted of 300 patients. Of the 300 patients, there were 109 (36.3%) and 191 (63.7%) in the hypothetical trial and RWD arms, respectively. The simulated number of patients and respective proportions of treatment allocation represent a realistic scenario in a real-world setting for some oncogene-positive subgroups.

For OS and PFS, 17% and 10% of patients, respectively, were censored before an outcome was observed. The Kaplan-Meier curves for both simulated arms of OS and PFS outcome are shown in Figure 1.

For OS, the HR for the true treatment effect (i.e., accounting measured and unmeasured confounders) was 0.53 (0.40–0.70). The observed treatment effect (i.e., adjusted for the two measured confounders but not for the unmeasured confounder) was HR 0.69 (0.53–0.90). The proportional hazards assumption was tested for OS and PFS simulated datasets using a Chi-square test on the scaled Schoenfeld residuals. In both cases, the proportional hazards assumption held. Using the Akaike information criterion and Bayesian information criterion, the Weibull PH was found to be the best-fitting proportional hazards model to the OS and PFS outcomes when adjusting for measured confounders only.

The resulting adjusted HRs and their respective CIs after applying the Huang *et al.* [9] and Ding *et al.* [10] approaches are shown in Supplementary Material Appendix 2. In addition, the sensitivity parameters used for all scenarios (i.e., varying levels of knowledge of the unmeasured confounder) are given in Supplementary Material Appendix 2. Figure 2 displays boxplots of the median of the adjusted HRs after applying the QBA methods relative to the observed and true HR. Both methods perform similarly across all scenarios, and as the level of knowledge

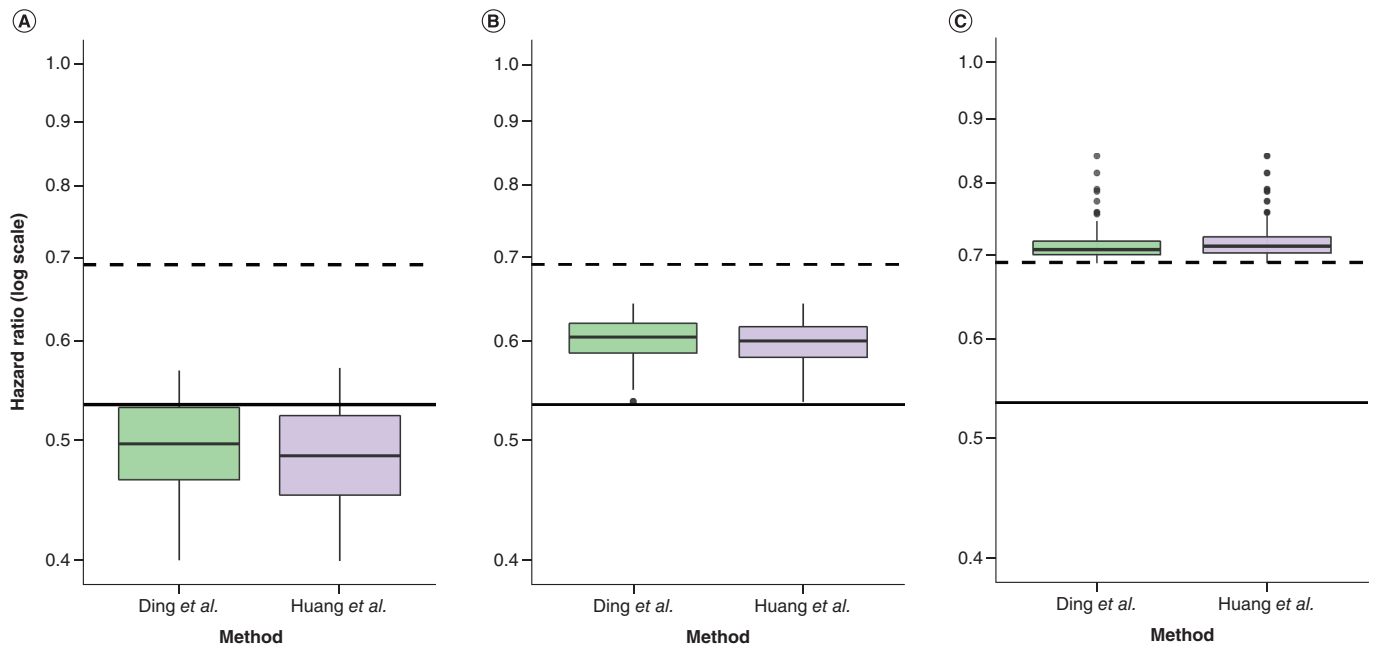


Figure 2. Adjusted hazard ratios. Boxplots of the adjusted HR using the Ding *et al.* and Huang *et al.* QBA methods under the varying levels of knowledge (A) good, (B) poor and (C) incorrect of the unmeasured confounder used to parameterise the methods; the dashed line is the observed HR, and the solid line is the true HR. HR: Hazard ratio.

of the unmeasured confounder becomes poorer and hence the parameterisation becomes poorer, the estimated HR moves further from the true treatment effect.

The results of the DSA assessing the treatment effect estimate sensitivity to unmeasured confounding are given in Figure 3. Under the scenario where assumptions were relatively good regarding the unmeasured confounder, the absolute percentage reduction in the ICER bias compared with the estimate when unmeasured confounding was not adjusted for was 73.6% for Huang *et al.* [9] and 81.0% for Ding *et al.* [10] using the median of the adjusted point estimates. Similarly, for the NMB outcome, the absolute percentage reduction in bias was 49.6% and 65.5% using the Huang *et al.* [9] and Ding *et al.* [10] methods, respectively. A further reduction in bias could be achieved if we had more precise knowledge of the relationships with the unmeasured confounder. When there was poor knowledge of the unmeasured confounder, the absolute percentage reduction in the ICER bias was 65.8% and 51.9%, and similarly a reduction in the NMB outcome of 54.5% and 40.2%, using the Huang *et al.* [9] and Ding *et al.* [10] methods, respectively. The parameters used in the incorrect knowledge scenario reflected very poor information on the unmeasured confounder, including opposite relationships to the true associations, and as a result increased the bias relative to the unadjusted estimate.

The results from the PSA for the change in costs and QALYs are shown in Figure 4 under all three scenarios of knowledge about the unmeasured confounder with the solid black line a nominal willingness-to-pay threshold of €30,000. The CE plane shows that when we have good knowledge of the unmeasured confounder, the samples generated from the two QBA methods [9,10] shift closer to the true samples, as shown by the marginal distributions of the change in QALYs in Figure 4A. Similarly, as also shown in the DSA, when incorrect knowledge is used for parameterisation of the QBA methods, the cloud of points generated in the CE plane by the QBA methods are further from the truth than the unadjusted observed results as indicated by the associated densities (Figure 4C). In the scenario where only poor knowledge of the unmeasured confounder was available, it resulted in the cloud of points generated by the two QBA approaches shifting towards the true samples but to a lesser extent than when good knowledge was available (Figure 4B). An additional representation of the results that estimates the proportion of iterations leading to potential misallocation of resources is given in Supplementary Material Appendix 4.

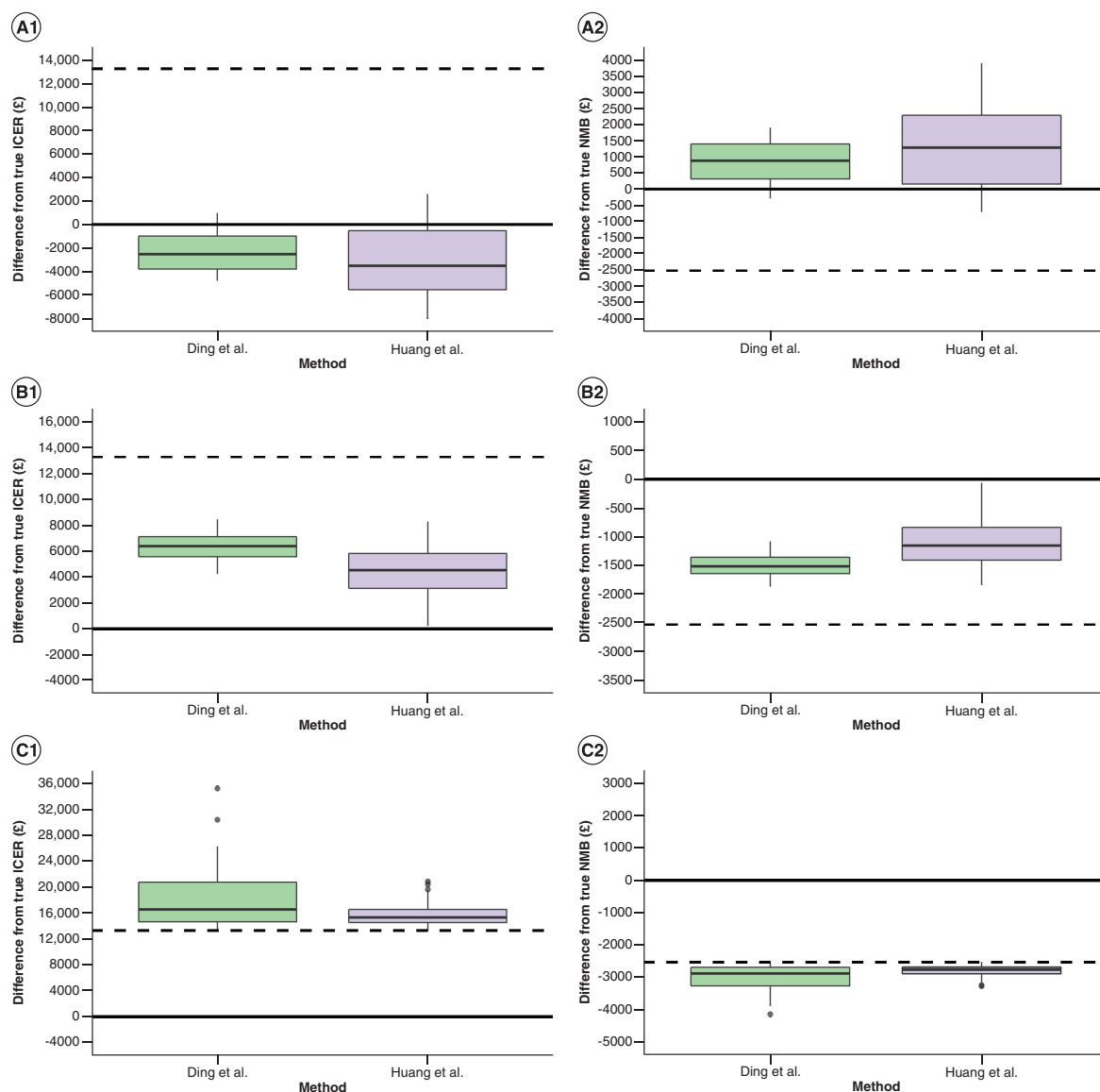


Figure 3. Results from the deterministic sensitivity analysis. (A1, B1 & C1) show the difference in GBP from the true ICER to the median of the adjusted point estimates as estimated by the Huang *et al.* and the Ding *et al.* methods with good, poor and incorrect knowledge of the unmeasured confounder, respectively, and the dashed line represents the observed or unadjusted estimate, **(A2, B2 & C2)** represent the difference in GBP from the true NMB estimated by the Huang *et al.* and the Ding *et al.* methods with good, poor and incorrect knowledge of the unmeasured confounder, respectively, and the dashed line represents the observed or unadjusted estimate. Huang *et al.* [9] requires individual patient-level data, whereas Ding *et al.* [10] requires aggregate data. GBP: British pound sterling; ICER: Incremental cost–effectiveness ratio; NMB: Net monetary benefit.

Discussion

Confounding is a concern in all NRS and continues to discourage HTA agencies from utilising nonrandomised evidence presented in submissions to inform treatment effectiveness. Although methodological and practical guidelines on QBA methods are widely available in causal inference literature [3,15], these have not yet permeated HTA practice. This study simulated survival data that is considered plausible in the context of previous studies [16,17] to illustrate how to incorporate the results of QBA methods into the DSA and/or PSA of a CEM to investigate the impact of biases due to unmeasured confounding in NRS. In addition, this article illustrates that the usefulness of these QBA methods may depend on the quality of the data available to parameterise them. Although the focus in

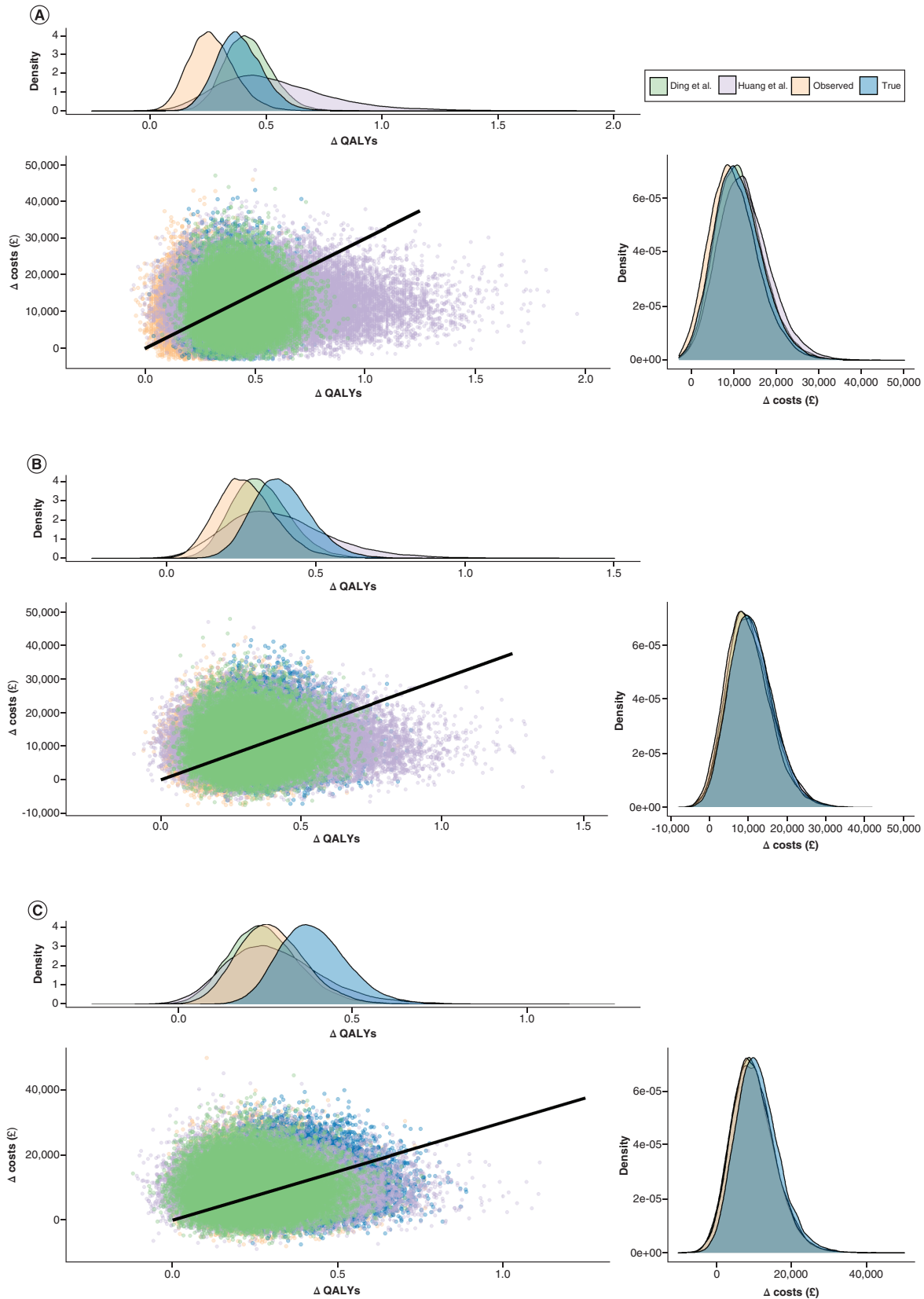


Figure 4. The cost-effectiveness plane (change in costs [Δ costs] versus change in QALYs [Δ QALYs]) generated by the probabilistic sensitivity analysis showing the true and observed relationship as well as that estimated by Huang *et al.* and Ding *et al.* methods. There was (A) good, (B) poor and (C) incorrect knowledge of the unmeasured confounder. Huang *et al.* [9] requires individual patient-level data, whereas Ding *et al.* [10] requires aggregate data. QALYs: quality-adjusted life-years.

this study was to apply QBA approaches to demonstrate a reduction in bias, in the HTA setting an understanding of the direction of bias is also potentially useful.

Health technology evaluation varies by jurisdiction. In many jurisdictions, a CE approach is followed where health benefits are compared with additional costs: the approach followed by NICE in England and Wales is one such example. The purpose of CEMs is to evaluate the change in cost and outcomes arising from an intervention. These models are developed in a systematic and transparent way to assist in decision making. Since most CEMs are informed by multiple data sources and assumptions and are subject to uncertainty, a crucial component in assessing the CE of an intervention is to conduct sensitivity analyses to characterise uncertainty around the input parameters [18]. We have proposed in this study that QBA methods assessing the sensitivity of the treatment effect to unmeasured confounding could also be incorporated into the existing framework of CEM with some natural extensions to the commonly used approaches to sensitivity analyses. By incorporating QBA methods into existing CEM frameworks, it will increase the likelihood of their use and reduce the burden of developing *de novo* approaches to presenting sensitivity analysis results.

Although QBA methods present a useful tool to assess the sensitivity of both clinical and CE results, there are several factors to consider when applying these methods. One key consideration is the parameterisation of the methods, such as defining the values of the sensitivity parameters. Ideally, data for this would be sourced from internal or high-quality external datasets, in which case one might expect to obtain relatively accurate estimates to parameterise the model. Where such data is not available, lower-quality external data or formal expert elicitation methods may be used; however, these would be expected to produce commensurately less accurate estimates of the sensitivity parameters. This study demonstrated that QBA methods reduced the bias of CE results where there was relatively good knowledge and, to an extent, poor knowledge regarding the unmeasured confounder. In the incorrect knowledge scenario, the simulation study used parameters that in no way reflected the true associations of the unmeasured confounder to highlight that reasonable parameterisation of these methods is key to obtaining plausible results. In fact, when there is very poor knowledge of an unmeasured confounder, the application of the QBA methods could inflate the existing bias. As such, the study supports the use of these methods in this setting when relatively good data is available to inform parameterisation of the QBA but raises questions regarding the application of these methods when data on the confounders is limited. In cases where estimates are expected to be inaccurate, one could account for this by widening the range of parameters to cover a broader range of possible scenarios, thereby adding this uncertainty into the CEM results. Further, an alternative implementation of QBA may be pursued in such cases, under which one implements a scenario analysis which identifies the amount of unmeasured confounding that would be required to alter the results of the CEM by a meaningful extent (e.g., move the ICER above a specified threshold) and then consider whether this level of confounding is plausible. For transparency and to avoid investigator bias, it is important to prespecify in a protocol and statistical analysis plan the QBA and parameterisation approach.

Other aspects of QBA, beyond the parameterisation of the methods, that should be considered include the availability of methods for the outcome effect measure of interest (e.g., odds ratio, HR, RR) and how the treatment effect evidence is synthesised. Further discussion of these considerations is given in a companion paper [2]. Notably, although this study explored the impact of unmeasured confounding in a CE setting and the use of QBA to adjust for such confounding, there are other biases, such as misclassification, that may impact the results of NRS and that should be considered. Additional sources of bias relevant to the CE setting are related to key CE inputs such as costs and utility values. However, the focus of this article is to highlight how the impact of bias on efficacy estimates affects the CE results and therefore to study the effect of this source of bias in isolation; other sources of bias were not considered in the CE analysis. In practice, one would ideally first identify all potential sources of bias and their assumed causal relationships and seek to address as many as possible through the choice of appropriate data source(s), study design(s) and analyses. One should then determine a suitable QBA approach to account for any remaining biases that might be expected to have a meaningful impact on the results [19]. Ideally, a holistic approach to incorporating all potential sources of bias, including both measured and unmeasured confounding, should be utilised with a CEM; however, exploration of such an approach was beyond the scope of this study.

This study utilised a three-state partitioned survival CEM which is commonly used in oncology and other disease areas. A limitation of PartSA models is that the modelled survival end points are structurally independent, with potential implications for the extrapolation of these outcomes. Therefore, the results of this study may not be generalisable to other modelling frameworks such as state transition models. Further work is required to assess the generalizability of the results of this study in other CEM frameworks, such as state transition models that

model explicitly the natural history of a disease and treatment effects, as well as the assumptions underpinning extrapolations of outcomes beyond the clinical study follow-up period. Additionally, the QBA methods employed in this study assumed a single binary unmeasured confounder; there are other approaches in the literature for nonbinary and multiple unmeasured confounders [2]. Further, the scenarios regarding the level of knowledge of the unmeasured confounder were selected for demonstration; in practice, one would not typically know how good one's assumptions are to inform the parameterization of the methods, and therefore careful consideration of the available data on a case-by-case basis is required. In addition, if the interest lies in outcomes other than survival outcomes, such as linear models with continuous outcomes, one could consider those methods described by Oster [20] and Cinelli and Hazlett [21].

Current guidance from HTA agencies regarding the use of QBA methods is limited, and further support from HTA agencies is needed to ensure alignment on best practices when utilising nonrandomised sources of data to inform the estimation of treatment effects. This includes guidance on the acceptable methods, their integration within CE models (where relevant) and the process of prespecifying the sensitivity parameters. NICE, for example, has developed the QuEENS checklist to assess the quality of effectiveness estimates generated from NRS [22]; however, guidance with respect to unmeasured confounding is limited. Development of such guidelines would aid users in the practical application and interpretation of these methods. Further methodological work will also help further this discussion, and although we have utilised simulated data here, future work might use real-world examples in which RCT data has been used in a CEM to inform a HTA decision and assess whether a similar conclusion could have been obtained using an external nonrandomised comparator arm, with and without the use of QBA.

Conclusion

This article illustrates how QBA methods can provide useful tools, when applied sensibly, to help improve the robustness of treatment effects derived from NRS. The usefulness of such an endeavour will crucially depend on the appropriateness and plausibility of the quality of the evidence used to inform the sensitivity parameters. The application of such approaches in the HTA setting can provide decision makers with a more transparent understanding of the impact of unmeasured confounding on the HTA decision. The inclusion of QBA methods within a CEM is a natural extension to the existing CEM framework for assessing the sensitivity of results through both DSA and PSA. The integration of QBA methods into CEM can improve the robustness and interpretation of the results from a CEM in light of likely unmeasured confounding when using RWD. Further methodological guidance QBA methods in HTA and their application to real-world HTA case studies are needed.

Summary points

- The application of quantitative bias analysis (QBA) methods to adjust the treatment effect estimate from nonrandomised studies can reduce the biases in cost–effectiveness results due to unmeasured confounding.
- The definition of the sensitivity parameters and the quality of the available information to inform these need to be considered prior to the application of QBA methods.
- Further guidance from health technology assessment agencies on the best practices to account for unmeasured confounding is needed.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/ce-2022-0030

Author contributions

All authors contributed to the study conception and design. Data simulation and methodology was conducted by TP Leahy. Analysis was conducted by TP Leahy and D Tzelis. The first draft of the manuscript was written by TP Leahy and D Tzelis, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Financial & competing interests disclosure

This study was funded by F. Hoffmann-La Roche AG. TP Leahy, C Sammon and D Tzelis are employees of PHMR Ltd. S Ramagopalan and J Ray are employees of F. Hoffmann-La Roche AG. The authors have no other relevant affiliations or financial involvement with

any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

1. National Institute for Health and Care Excellence. *Guide to the Processes of Technology Appraisal*. London, UK, (2018).
2. Leahy TP, Sammon C, Kent S, Ramagopalan S, Groenwold RH. Sensitivity analyses for unmeasured confounding in non-randomised studies: considerations for application in health technology assessment. *J. Comp. Eff. Res.* 11(12), 10.2217/cer-2022-0029 (2022).
3. 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).
4. Kent S, Salcher-Konrad M, Boccia S *et al.* The use of nonrandomized evidence to estimate treatment effects in health technology assessment. *J. Comp. Eff. Res.* 10(14), 1035–1043 (2021).
5. Thorlund K, Dron L, Park JJ, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. *Clin. Epidemiol.* 12, 457 (2020).
6. Alexander M, Wolfe R, Ball D *et al.* Lung cancer prognostic index: a risk score to predict overall survival after the diagnosis of non-small-cell lung cancer. *Br. J. Cancer* 117(5), 744–751 (2017).
7. Sehgal K, Gill RR, Widick P *et al.* Association of performance status with survival in patients with advanced non-small cell lung cancer treated with pembrolizumab monotherapy. *JAMA Netw. Open* 4(2), e2037120–e2037120 (2021).
8. Cox DR. Regression models and life-tables. *J. R. Stat. Soc. Series B Stat. Methodol.* 34(2), 187–202 (1972).
9. 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).
10. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology* 27(3), 368–377 (2016).
11. Carnegie NB, Harada M, Hill JL. Assessing sensitivity to unmeasured confounding using a simulated potential confounder. *J. Res. Educ. Eff.* 9(3), 395–420 (2016).
12. Vanderweele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* 167(4), 268–274 (2017).
13. Vanderweele T. On a square-root transformation of the odds ratio for a common outcome. *Epidemiology (Cambridge, Mass)* 28(6), e58 (2017).
14. Sjölander A. A note on a sensitivity analysis for unmeasured confounding, and the related E-value. *J. Causal Inference* 8(1), 229–248 (2020).
15. Greenland S. Basic methods for sensitivity analysis of biases. *Int. J. Epidemiol.* 25(6), 1107–1116 (1996).
16. Mok T, Camidge D, Gadgeel S *et al.* Updated overall survival and final progression-free survival data for patients with treatment-naive advanced ALK-positive non-small-cell lung cancer in the ALEX study. *Ann. Oncol.* 31(8), 1056–1064 (2020).
17. Ramalingam SS, Vansteenkiste J, Planchard D *et al.* Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC. *N. Engl. J. Med.* 382(1), 41–50 (2020).
18. Briggs AH, Weinstein MC, Fenwick EA *et al.* Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health* 15(6), 835–842 (2012).
19. Lash TL, Fox MP, Macle hose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int. J. Epidemiol.* 43(6), 1969–1985 (2014).
20. Oster E. Unobservable selection and coefficient stability: theory and evidence. *J. Bus. Econ. Stat.* 37(2), 187–204 (2019).
21. Cinelli C, Hazlett C. Making sense of sensitivity: extending omitted variable bias. *J. R. Stat. Soc. Series B Stat. Methodol.* 82(1), 39–67 (2020).
22. Faria R, Alava MH, Manca A, Wailoo AJ. National Institute for Health and Care Excellence. *The Use of Observational Data to Inform Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data; NICE DSU Technical Support Document*. London, UK (2015).