Cost-effectiveness analysis with informative missing data: tools and strategies

Baptiste Leurent

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the University of London

April 2019

Department of Medical Statistics
Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine

Funded by the National Institute for Health Research (Grant number: DRF-12437)

Declaration

I, Baptiste Leurent, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Cost effectiveness analysis (CEA) of randomised trials are an important source of evidence for informing policy makers on how to best allocate limited resources. Missing data are a common issue in trial-based CEA, and methods such as multiple imputation are now commonly used to account for the missing values, assuming the data are 'missing at random' (MAR). This implies that the reasons for the missing data can be explained by the observed data. However, the missingness is often related to unobserved values, that is data are 'missing not at random' (MNAR, or 'informative'). For example, patients whose health status is relatively poor may be less likely to return health questionnaires, even conditional on their observed characteristics. In these settings, methodological guidance recommends assessing whether conclusions are sensitive to departures from the MAR assumption. Sensitivity analysis strategies for handling MNAR is an area of rapid development in medical statistics, but this form of uncertainty has not yet been appropriately addressed in health economics.

This PhD thesis aims to develop practical, accessible sensitivity analysis strategies and software tools to handle MNAR data in trial-based CEA.

The thesis critically assessed the statistical methods for handling MNAR data in CEA practice, and identified barriers to more widespread use of these methods, via a systematic review and stakeholder focus groups. The research then focused on two strategies to conduct sensitivity analysis under MNAR assumptions: pattern-mixture models, which involve imputing the data assuming MAR, then modifying the imputed values to reflect possible departures from that assumption; and reference-based imputation, where the data are imputed assuming a distribution borrowed from a 'reference group'. These approaches were illustrated in CEAs of 10TT and CoBalT trials, which evaluated weight loss and depression interventions. Software code and practical guidance are provided to facilitate implementation in practice.

A Margherita, Matteo, et Sergio

Acknowledgements

This report is independent research arising from a Doctoral Research Fellowship supported by the National Institute for Health Research. The views expressed in this publication are those of the author and not necessarily those of the NHS, the National Institute for Health Research (NIHR) or the Department of Health.

I am particularly grateful to my supervisors, Prof. James Carpenter and Dr. Manuel Gomes, for their precious advice and active support during the past three years.

I would like to thank the members of my advisory committee, Dr. Jonathan Bartlett, Prof. Richard Grieve, Prof. Steve Morris, and Prof. Ian White, the attendees of the Stakeholders' meetings, and Nicky Bailey, Anna Marriner, Richard Duncker, and Alan Berry, for helping steer this project in the right direction. I would also like to thank collaborators who shared their study data (and time) to support this work, including the 10TT, CoBalT, and SADD trials teams. I am also immensely grateful to my sponsor, NIHR, for giving me the opportunity to conduct this PhD.

This thesis would not have been what it is without a number of other people, but they probably would not fit all within the words limit. All those who helped me think about this project, or who helped me to manage it in any sense, including numerous colleagues in the Medical Statistics Department, at LSHTM, and beyond. You know who you are (or at least I do).

Finally, I would like to thank those who have supported me in my personal life, giving me the strength to start and go until the end of this thesis. My parents and Paolo, but also particularly Margherita, and my two little superheroes, who arrived and grew up with this project (promise, I will catch up on every minute I should have spent with you!).

Contents

Abbreviations

10TT	Ten Top Tips
BMCF	Baseline Mean Carried Forward
CBT	Cognitive Behavioural Therapy
CCA	Complete Case Analysis
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curve
CEP	Cost-Effectiveness Plane
CI	Confidence Interval
CIR	Copy Increments in Reference
HTA	Heath Technology Assessment
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
J2R	Jump to Reference
LMCF	Last Mean Carried Forward
LOCF	Last Observation Carried Forward
MI	Multiple Imputation
MVN	Multivariate Normal
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NHS	National Health Service
MAR	Missing At Random
MCAR	Missing Completely At Random
MNAR	Missing Not At Random

MNAR SA	Missing Not At Random Sensitivity Analysis
PMM	Pattern-Mixture Model
PMM-MI	Pattern-Mixture Model after Multiple Imputation
QALY	Quality-Adjusted Life Year
QoL	Quality of life
RCT	Randomised Controlled Trial
SD	Standard Deviation
SE	Standard Error
WHO	World Health Organisation
WTP	Willingness To Pay

Chapter 1

Introduction

1.1 Motivation

Health policy-makers worldwide aim to maximise population health given resource constraints. Cost effectiveness analyses of randomised trials are an important source of information to help decide which health care programmes to provide. Such studies can only provide a sound basis for policy-making if they use appropriate analytical methods. An outstanding concern is that there may be missing data, for example, because patients are lost to follow-up or fail to respond or fully complete quality-of life (QoL) or resource use questionnaires.

Any analysis of trials with missing data relies on untestable assumptions, and there is now greater awareness that naive analysis under a single simplistic assumption has the potential to bias statistical inference [1–3]. The main challenge is that — even given the information in the observed data — the reason for data being missing may well depend on the underlying, unobserved data values. In this case the data are said to be informatively missing or Missing Not At Random (MNAR). The consequence is that the statistical distribution of the missing effectiveness and/or cost data is systematically different from that of the observed data. The importance of this issue has gained wider recognition in recent statistical guidelines [4–6].

However, in trial based cost effectiveness evaluation (CEA), it is still very common to make the simple, and often naive, assumption that the data are Missing Completely At Random (MCAR) (i.e. that the reason for data being missing does not depend on the underlying values) [7]. Some studies investigate the implications of assuming MCAR by making a more plausible Missing At Random (MAR) assumption in a sensitivity analysis, typically using multiple imputation [8]. The MAR assumption states that, given the observed data, the reason for the missing data does not depend

on the underlying unseen values. However, this is an untestable assumption, and cost-effectiveness studies rarely investigate any potential departures from it. While statistical methods for dealing with informative missing data are available, they have yet to permeate economic evaluation studies. The reasons for this are unclear, but may be associated with a lack of adaptation of the biostatistics methods to the cost-effectiveness analysis context, or a lack of practical guidance or tools to implement them.

The particular issues raised by informative missingness are illustrated in the 10 Top Tips (10TT) [9] and CoBalT [10] trials. 10TT evaluated a brief weight-loss intervention for obese patients, and CoBalT Cognitive Behaviour Therapy (CBT) for patients with treatment-resistant depression, both in primary care setting. In both trials, not all randomised participants completed the entire study follow-up, and it is quite possible that data are informatively missing. For example, it is well-recognised in weight loss trials that completion could be associated with outcomes [11]. A key question is the extent to which the effectiveness and cost-effectiveness of the intervention is sensitive to the assumptions made about the missing data. Simply performing one analysis runs a high risk of placing unwarranted confidence in results which are, in fact, biased.

We need a framework for expressing the assumptions in an accessible way to the analysts and other Stakeholders, and appropriate statistical methods for valid inference under specific assumptions.

1.2 Aim and objectives

The overall aim of this thesis is to develop accessible tools and recommendations to perform cost-effectiveness analysis of randomised trials when data may be informatively missing.

To achieve this, the specific objectives are to:

1. review how missing data are currently addressed in trial-based CEA;
2. identify the needs and barriers of analysts and other Stakeholders in relation to implementation of CEA under MNAR assumptions;
3. provide tools and practical guidance to implement pattern-mixture model after multiple imputation in CEA; and
4. resolve methodological issues and provide tools to implement reference-based imputation in CEA.

Specifically, the plan for the thesis is as follows. Chapter 2 introduces the background and concepts relevant to the thesis. Then Chapter 3 reviews current missing data practice in CEA. Chapter 4 reports on discussions held with Stakeholders on the issue of informative missing data. In Chapters 5 and 6 we present two approaches for conducting sensitivity analysis in CEA with informatively missing data. We finish by a discussion of the findings and their implications for current practice and future research.

Chapter 2

Background

2.1 Cost-effectiveness analyses

2.1.1 Cost-effectiveness analysis alongside randomised trial

Health economics is a key discipline underpinning health systems functioning, and strongly influences how healthcare is funded and provided [12]. Economic evaluation is a branch of health economics, characterised by the comparison of different health 'options' (health technologies, public health interventions, etc.) in terms both of their costs and their consequences [13]. CEA is the main type of economic evaluation, where 'consequences' are measured in health units, such as number of malaria cases averted, or quality-adjusted life years gained. It is used by policy makers worldwide (e.g. NICE [14] and the World Health Organisation [15]) with the aim of maximising health benefits within a constrained budget. When comparing two (or more) alternative health interventions, a CEA typically aims to estimate the difference in costs relative to the difference in health units. For example, implementing a malaria intervention could cost on average £200 for each malaria case averted. This information is used by policy makers as an indicator of whether an intervention provides good value for money, and to identify the most 'cost-effective' choice among different options. To decide whether an intervention provides good value for money, policy makers typically have to attribute a monetary value they consider affordable to each 'unit' of health benefit. This is referred to as the willingness-to-pay (WTP) threshold.

Policy-makers often have to make decisions across different disease areas, for instance to decide whether it is more cost-effective to fund a malaria intervention or a new cancer treatment. Quality-adjusted life years (QALYs) — a composite measure of gain of life-expectancy and health-related quality-of-life [12, 16] — aim to facilitate such comparison across disease areas, and are therefore

commonly used as a generic measure of health benefit. In the UK, it is recognised that NICE typically uses a WTP threshold of around £20,000 or £30,000 per QALY gain to assess whether an intervention is cost-effective [17]. Given its relevance to decision makers, this thesis focus mainly on QALYs as effectiveness measure.

Randomised controlled trials provide an ideal vehicle to collect data on costs and effectiveness of new interventions [18, 19]. When a CEA is conducted based on randomised trial data, this is commonly referred to as a 'within trial' or 'trial-based' CEA. Collection of economic data as part of clinical trials is now increasingly common [20], and research funders such as NIHR HTA require the trials they fund to include an economic evaluation [21]. Randomised trials provide a key source of information for decision-making, as they allow unbiased estimation of the difference in costs and in effectiveness between interventions compared [18, 20]. It is therefore critical that methodological issues commonly encountered in RCTs, such as missing data, are appropriately taken into account in the CEA; this is the focus of this thesis.

2.1.2 Statistical analysis of cost-effectiveness data

Consider the main estimands of interest in CEA. Denote by $t = A, B$ the two interventions compared (this can be extended to any additional pairwise comparisons). CEA is principally concerned with estimating two quantities: $\Delta^C = \mu_B^C - \mu_A^C$ and $\Delta^E = \mu_B^E - \mu_A^E$, where μ_t^C and μ_t^E represent the expected (mean) cost and effectiveness of the intervention t. From these quantities, an incremental cost-effectiveness ratio (ICER) can be derived, defined as ICER= Δ^C/Δ^E, capturing the expected difference in costs per unit of effectiveness. The interpretation of the ICER is not always straightforward, as both the numerator and the denominator can be null or negative (the same ICER can represent two opposite results). The mean differences in cost and in effect are therefore also commonly represented graphically on a cost-effectiveness plane (CEP), with the difference in effect on the horizontal axis and in cost on the vertical axis (Figure 2.1). An intervention is said to 'dominate' its comparator if (Δ^C, Δ^E) is in the South-East quadrant (more effective and less costly), and to be 'dominated' if it sits in the North-West quadrant (less effective and costlier). When the intervention sits in the South-West, or more commonly, in the North-East quadrant (more effective but costlier), deciding whether an intervention represents good value for money will depend on the WTP threshold. To facilitate interpretation, a straight line defined by $\Delta^C = \lambda.\Delta^E$ can be plotted, where λ is the WTP threshold. An intervention will be considered cost effective when (Δ^C, Δ^E) sits below the line. At a given WTP, the interventions can also be compared in terms of Incremental Net Monetary Benefit (INMB). That is INMB= $\lambda.\Delta^E - \Delta^C$, representing

the average benefit of intervention B compared to A, in monetary unit. A positive INMB indicates that B is more cost-effective than A.

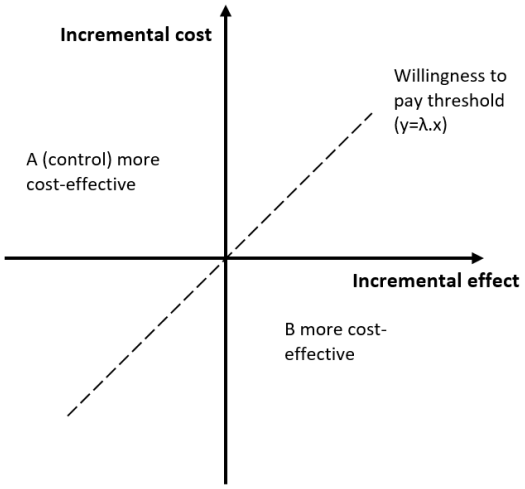


Figure 2.1: Cost-effectiveness plane

Cost-effectiveness data exhibit several challenging features compared to clinical outcomes, and the analytical methods differ, to a certain extent, from those typically used in the analysis of clinical effectiveness.

First, cost and effectiveness data are usually non-normally distributed. For example, the costs are typically right skewed, with a peak close to zero for participants using few health care resources, and a long right tail reflecting a minority of patients with very high costs [22]. This may explain why non-parametric bootstrapping [23] is commonly used. It is worth noting that CEA is concerned with inference about means (i.e. average cost of the intervention in the population), and some approaches for handling non-normal data, such as transforming the data, or comparing medians, are therefore not recommended [22].

Second, the cost and effectiveness endpoints are usually correlated. This needs to be taken into account when capturing the uncertainty in cost-effectiveness ratio. A popular method to express the joint uncertainty in Δ^C and Δ^E is to represent it graphically on the CEP. Non-parametric bootstrapping can be used, and the replicates of Δ^C and Δ^E plotted to represent the posterior joint distribution [24]. Parametric alternatives include using seemingly-unrelated regression which models cost and effect in separate models related though correlated error terms [25]; essentially a bivariate joint model.

Third, an important difference is that CEA and primary clinical trial analysis actually have different objectives. The latter typically aims to demonstrate the efficacy of a novel intervention, and is primarily interested in hypothesis testing. The former aims to inform decision-making and to

recommend the best decision based on the cost-effectiveness information available, recognising that both implementing or not implementing an intervention will have consequences in terms of cost and health. The CEA therefore aims to inform a choice, and reflect the uncertainty of that choice, not to provide statistical evidence for or against a specific hypothesis. A popular approach to address this question is to derive the probability for the intervention to be more cost-effective than the control, at a given WTP. This can be estimated non-parametrically, based on bootstrap replicates, or parametrically, based on a regression of the INMB, or seemingly-unrelated regression of the cost and effectiveness endpoints.

Fourth, whether an intervention is considered cost-effective will depend on the policy-makers' WTP. Analysts therefore needs to reports results across a range of WTP. A popular approach for this is to plot the probability for the intervention to be cost-effective over a range of WTP. This is referred to as the 'Cost-effectiveness Acceptability Curve' (CEAC) [26], which can be used by decision makers to identify the most cost-effective option, and the underlying uncertainty in that choice, depending on their WTP.

Finally, CEA are often based on self-reported measures, collected on multiple occasions, over a relatively long period of time. This has implications for the completeness of the data, and the choice of method to handle them, as discussed in Section 2.3.

2.2 Missing data

2.2.1 Missing data mechanisms

An important concept to discuss missing data is the taxonomy of the missing data mechanism, first developed by Rubin [27], then refined in various works, notably by Little and Rubin [28]. In the context of a randomised clinical trial with partially observed outcomes, let us denote by:

- X a set of fully observed variables, including the randomisation arm and baseline variables;
- $Y = (Y_{.1}, Y_{.2}, \cdots, Y_{.J})$ a set of J variables of interest which are partially observed;
- Y_{ij} the value of the variable j for the patient i;
- $M = (M_{.1}, M_{.2}, \cdots M_{.J})$ the matrix of indicators of whether Y is missing (for each patient i and variable j, $M_{ij} = 1$ if Y_{ij} is missing and 0 otherwise), also referred to as the 'missingness pattern'; and
- Y_{obs} the observed entries of Y, and Y_{miss} its missing components.

Let also $[X]$ denote the probability distribution of a random variable X, and $[X|Y]$ the conditional distribution of X given Y.

The missing data are said to be:

- Missing Completely At Random (MCAR) if the missingness is independent of X or Y, that is if $[M|X, Y_{obs}, Y_{miss}] = [M]$.
- Missing At Random (MAR) if the missingness depends of the observed variables, but is independent of the unobserved values, that is if $[M|X, Y_{obs}, Y_{miss}] = [M|X, Y_{obs}]$.
- Missing Not At Random (MNAR, also called 'informative', or 'non-ignorable' missingness) when the data are neither MCAR nor MAR, that is if the distribution of M is dependent of Y_{miss}, even after conditioning on X and Y_{obs}, or if $[M|X, Y_{obs}, Y_{miss}] \neq [M|X, Y_{obs}]$.

Based on the observed data, it is not possible to distinguish between MAR and MNAR mechanisms. MAR is an attractive assumption, as valid inference can be drawn based on the observed data. For example in a longitudinal study, it assumes the distribution of later follow-up data, given the baseline and earlier data, is the same whether or not the later follow-up data are observed. While MAR assumption is often questionable, MNAR is substantially more challenging as it implies some of the information needed to draw valid inference are not observed, and that further assumptions are needed. This probably explains why the MAR assumption is typically chosen for the primary analysis, while MNAR assumption is more often considered in the context of sensitivity analyses [28, 29].

For this thesis, we focus on missing outcome (response) variables as most relevant to the MNAR issue. For a discussion on missing baseline variables in randomised trials see for example White and Thompson [30] and Carpenter and Kenward [31].

2.2.2 Statistical models for informative missingness

Informative missing data implies that inference requires Y and M to be modelled jointly, but that part of this joint distribution cannot be identified from the data at hand. To conduct sensitivity analyses under MNAR assumptions (MNAR SA), two factorisation of the joint distribution are commonly used:

- Selection models:

  $$[M, Y|X] = [Y|X][M|X, Y]$$

  Where $[Y|X]$ represents the distribution of Y (conditionally on X) in the population, and $[M|Y, X]$ how the probability of the missingness pattern depends on Y and X. $[M|Y, X]$ is not estimable based on the observed data, but making different assumptions about its

form (*i.e.* how the chance of being missing varies with Y) allows sensitivity analyses to be performed (see below and Chapter 5).

- Pattern-mixture models (PMM):

$$[M, Y|X] = [M|X][Y|X, M]$$

Where $[M|X]$ represents the probability of each missing data pattern in the population, and $[Y|M, X]$ the distribution of Y within each of these pattern. $[Y|M, X]$ is unknown, but making different assumptions about its distribution (*i.e.* how the distribution of Y varies across pattern of observed and missing data) allows sensitivity analyses to be performed (see below and Chapter 5 and 6).

As described in Chapter 5, identification in selection models requires elicitation of how the chance of being observed is related to the outcome. For example, hypothesising that the chance of being missing doubles for each 0.5 QALY increase. Selection models have been commonly used in early work on informative missing data, as it could seem somehow more natural to analysts to express the missingness as a function of the outcome. They are also common in econometrics, where Heckmnan's selection model [32] has been commonly used to account for selection bias. The missing data model can be directly incorporated into the analysis model, for example using an inverse probability weighting approach [33, 34] or numerical integration [35]. However, selection models make untestable assumptions about the conditional distribution of the unobserved data, and results can be very sensitive to departure from these assumptions [28]. Another key disadvantage is that selection models formulate sensitivity analysis in a way that is not readily interpretable. A typical sensitivity parameter is the (log-)odds ratio of how a unit change in the partially observed outcome affects the chances of observing the data. This specification makes the elicitation of such parameters and the communication of the sensitivity analysis results challenging, particularly to non-statisticians.

The analysis of pattern-mixture models is covered in more details in Section 2.2.3. They are typically parametrised by the difference between observed and missing values. For example, assuming that participants with missing data have a 0.1 lower QALY than those observed. They have received more attention recently, particularly in clinical trials [31,36]. Their key advantage is that the model parameters are more readily interpretable, and can therefore be informed and reviewed by a broader audience, such as clinicians and decision-makers [33]. Another advantage is that they can easily be implemented with standard missing data methods, such as MI, as we will see in Section 2.2.3.

A third less common formulation of the MNAR problem is the shared parameter model, where Y and M are linked through a set of latent (unobserved) random variables. The joint probability of Y and M can then be decomposed as follows:

$$[M, Y, b|X] = [Y|X, b][M|X, Y, b][b] = [Y|X, b][M|X, b][b]$$

where b represents the set of latent variables, after conditioning on which M and Y are independent [36, 37]. Early references include works by Wu and Carroll [38] and Wu and Bailey [39], and a more recent review was conducted by Albert and Fullman [40]. Shared-parameter models have some appeal to address missing data with complex data structure. However, they raise additional assumptions (e.g. about the distribution of b) and computational issues, mainly related to the integration over the latent variables. They have been generally advised against by regulatory bodies due to "the many layers of assumptions" [41] and will not be considered further in this thesis as they do not appear particularly accessible to non-statisticians.

2.2.3 Methods for informative missingness sensitivity analyses in clinical trials

Statistical books on missing data commonly have a section covering the issue of informative missing data [28,29,31,36,37,41–43], introducing the concepts and challenges of informative missing data, and discussing different analytical approaches possible.

Journal articles on methods for MNAR sensitivity analysis (MNAR SA) in clinical trials are also numerous (see for example [34, 35, 44–54]). While there are some examples of selection model applications [49, 55–57], the focus has often been on pattern-mixture models, possibly due to their more accessible formulation [36] or their compatibility with popular missing data methods such as multiple imputation [58].

An approach that has been recurrently suggested — under various forms — is to perform a pattern-mixture model with a simple parameter capturing how the conditional distribution of the missing values $[Y_{miss}|X]$ may differ from the observed distribution $[Y_{obs}|X]$. This can be done by a shift parameter, usually noted δ, so that $[Y_{miss}|X] = [Y_{obs}|X] + \delta$. Or less commonly — but which may have relevance in CEA — using a scale parameter c so that $[Y_{miss}|X] = [Y_{obs}|X] \times c$. Note that the parameters δ and c are not quantities estimated from the data, but are capturing an assumption made about the missing data mechanism.

If relevant, more complex forms can be applied, for instance δ or c can vary depending on the reason for being missing or on the values of a covariate (e.g. treatment arm), or with the time since dropping out in longitudinal trials. Or both δ and c could be used simultaneously, or follow

a probabilistic distribution (as opposed to a fixed value). Typically sensitivity analyses will be performed over a range of values of δ or c, to represent a range of plausible scenarios.

To proceed with the analysis under the assumed pattern-mixture model, several statistical approaches are possible. Pattern-mixture models can naturally be fitted in a Bayesian framework [59]. They can also be estimated using a regression-based approach, using delta method [60] or bootstrapping to estimate the standard errors [5]. Multiple imputation is also a particularly convenient framework to implement PMM [42].

In simple situations (e.g. analysing an outcome at a single time-point), a rough approximation can by obtained by calculating the average of the observed and the transformed outcomes weighted by the proportion of missing data. The standard errors based on the observed data can then be used as a guide for inference.

Another approach particularly relevant in longitudinal trials, is a particular kind of pattern-mixture model, where the distribution within each pattern is defined by reference to another group of patients. This is termed a 'reference-based' approach, and is commonly implemented in a multiple imputation framework [48]. It is part of a broader family of 'controlled imputation' methods, which are gaining popularity in clinical trials sensitivity analyses. For recent reviews of controlled-imputation methods, see O'Kelly and Ratitch [36], Kenward [61], Molenberghs et al. [37] and Carpenter and Kenward [42]. The approach is described in detail in Chapter 6, but briefly, instead of drawing the missing values from an estimate of $[Y_{obs}|X]$ as would be done in a MAR imputation, reference-based imputation draws missing values from another 'reference' distribution, for example from the distribution of Y in the control group. This choice would be suitable if one believes that patients with missing data have stopped their treatment and behave like patients in the control group. The choice of the reference group, and how it is mimicked, will depend on the clinical context. One key advantage of reference-based imputation over selection and pattern-mixture models is that it does not require a quantitative elicitation of a sensitivity parameter, but it relies on a qualitative assumption which can be readily understood. One limitation of reference-based imputation is that it is not yet implemented in standard statistical software and it relies on user-written packages [36].

2.2.4 Uptake of informative missingness sensitivity analyses in clinical trials

The statistical literature on MNAR methods is extensive, but little attention has been paid to translation of these methods in practice. It is currently unclear to what extent MNAR sensitivity analyses have really permeated practice in the primary reporting of clinical trial results. Hayati et al. reviewed studies (not necessarily trials) in which MI was used, published in the *Lancet* and *New*

England Journal of Medicine between 2008 and 2013. Out of 103 studies, only three conducted sensitivity analyses departing from the MAR assumption, and only one reported clearly the method used. Bell et al. [62] conducted a review of 77 trials published in leading medical journals in 2013, and found that none reported sensitivity analyses under MNAR assumptions. However, leading medical journals often have limited word counts, and do not necessarily encourage the reporting of comprehensive sensitivity analyses. It is also possible that this has changed since. An updated review looking a broader journals would be of interest.

Both reviews identified the lack of robust sensitivity analyses as a key issue in adressing missing data, and discussed possible explanations. Hayati et al. mentioned the lack of explicit guidelines to conduct sensitivity analysis, while Bell et al. hypothesised that it could be due to a lack of knowledge or experience from the researchers, or a time-lag between methods development and software to implement them [62]. These assumptions may also apply to the CEA setting, as we will explore in Chapters 3 and 4.

2.3 Missing data in cost-effectiveness analyses

Missing data are a particularly relevant issue in trial-based CEA, which often relies on collecting rich data over a long-period of time. It is common for some of these data to be incomplete [7], for instance this could be because a participant dropped-out from the trial, missed a follow-up visit, or was not able to answer part of a questionnaire. Most trial-based CEA analysts are therefore confronted with missing data issues, and have to decide on the strategy to adopt and appropriate statistical methods to implement that strategy.

In comparison to the broader medical statistics literature, the issue of missing data historically received less attention in CEA. A review from 1999 shows that the issue of missing data was not mentioned in most trial-based CEA [63], and a paper in 2009 discussed how guidelines for CEA did not cover the issue of missing data [64]. The issue has however gained recognition since. Several publications raised the importance of the issue and methodological developments have been undertaken to address it. Articles published on the topic include general guidance on missing data in CEA, such as Faria et al. [65] and Manca et al. [66]. Other articles looked specifically at multiple imputation, and explain the benefits of the approach, illustrated by trial examples [64, 67]. Diaz et al. [68, 69] and Gomes et al. [70] have looked specifically at the issue of missing data in cluster-randomised trials. Several articles discussed missing data for quality of life data [16, 71–73] while others have looked specifically at missing cost data [67, 74]. More recent handbooks on methods for trial-based CEA generally discuss the implications and possible approaches to missing data

[75–78]. The 2015 ISPOR (International Society for Pharmacoeconomics and Outcomes Research) task force for good research practices in trial-based CEA [79] discuss the importance of appropriate data collection methods, the limitations of complete-case analysis and the advantages of multiple imputation (see below). The report recognises the plausibility of non-informative censoring and recommends using external information to try to address it.

Multiple imputation (MI) [6,8] has been the most widely discussed method for addressing missing data in trial-based CEA, and is the approach suggested in most of the recent literature [64–66, 75, 79–81]. The idea of multiple imputation builds on regression imputation (using the observed data to predict the missing values) by properly reflecting the uncertainty in the imputed values. This is achieved by conducting multiple rounds of imputations, each draws the missing values from the conditional predictive distribution of the missing data given the observed data. The substantive analysis model is then fitted to each of the imputed datasets separately, and the results combined for final inference using a specific formula (Rubin's rules). When models are correctly specified, multiple imputation gives valid inference under the MAR assumption. MI is not strictly necessary for the analysis of clinical effectiveness in RCTs, where it is common to have a model with a single outcome (with possibly repeated-measurements), and fully observed baseline variables. In this case valid estimates of treatment effect under MAR can be obtained by standard maximum likelihood methods conditioning on these variables [6].

The context of CEA raises important challenges to missing data methods, and by separating the imputation and analysis models MI provides practical advantages particularly relevant to that setting. First, it allows the inclusion in the imputation model of variables that are not in the analysis model, such as post-randomisation costs to estimate incremental effectiveness, or baseline variables to estimate mean costs or effectiveness. Second, cost and utilities are often based on multiple items, and MI is a convenient way to preserve the information from the observed items, before summing them up for the analysis. Similarly, utility and costs are typically measured over multiple assessments, but analysed in aggregated form, as opposed to repeated-measure (longitudinal) models more commonly used for clinical outcomes. This implies that if one assessment is missing the overall outcome is missing and the participant would therefore be excluded from a complete-case analysis. MI is a convenient way to preserve the information from the observed assessments, and impute the missing ones, before deriving the total cost or QALYs as usual. Finally, MI can also address the common features of CEA data, such as the correlation between endpoints, and the non-normality of the data [65, 82, 83].

In 2009, Noble et al. [7] conducted a review of missing data in trial-based CEA. They found that the reporting was very poor, with the extent of missing data and the approach used often unclear. They also found that complete-case analysis was the most commonly used approach (one third of the articles reviewed). MI was used in one fifth of the study, and only one fourth of the studies conducted sensitivity analyses (performing more than one approach to missing data). Since 2009, there has been wider recognition of the missing data problem and more methodological development. One of the first objectives of this PhD is therefore to assess the extent to which appropriate methods has permeated CEA practices.

While most of the recent developments focussed on the issue of missing data under MAR, the issue of informative missing data has received little attention. MNAR analyses are of particular relevance in that setting, as CEA often relies on self-reported measures and the dependence between missingness and outcomes (e.g. quality-of-life or resource-use) is a source of concern. In fact, most of the recent CEA guidance on missing data recognise the plausibility of informative missingness, and the importance of conducting sensitivity analyses [16, 65, 66, 68, 70, 75, 79, 80]. However, only one article provided practical advice on how to perform such analysis [65], while the others recognised the need for further development in that area.

This thesis therefore aims to address this issue, by developing accessible tools and recommendations to conduct sensitivity analyses for informatively missing data in trial-based CEA.

Chapter 3

Missing data in trial-based cost-effectiveness analysis: an incomplete journey

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Baptiste Leurent
Principal Supervisor	Prof. James Carpenter
Thesis Title	Cost-effectiveness analysis with informative missing data: tools and strategies

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Health Economics		
When was the work published?	24 March 2018		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	BL took responsibility for the article overall. Identified the studies, extracted the information (with the help of MG and JC) and conducted the analysis. Wrote the first draft, and final version of the article.

Student Signature: [redacted] **Date:** 05/09/18.

Supervisor Signature: [redacted] **Date:** 5th Sept 2018

Received: 3 February 2017 | Revised: 7 December 2017 | Accepted: 14 February 2018

DOI: 10.1002/hec.3654

HEALTH ECONOMICS LETTER

WILEY Health Economics

Missing data in trial-based cost-effectiveness analysis: An incomplete journey

Baptiste Leurent[1] | Manuel Gomes[2] | James R. Carpenter[1,3]

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

[2]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

[3]MRC Clinical Trials Unit, University College London, London, UK

Correspondence
Baptiste Leurent, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
Email: baptiste.leurent@lshtm.ac.uk

Funding information
Research Trainees Coordinating Centre, Grant/Award Number: DRF-12437; Medical Research Council, Grant/Award Number: MC_UU_12023/21

SUMMARY

Cost-effectiveness analyses (CEA) conducted alongside randomised trials provide key evidence for informing healthcare decision making, but missing data pose substantive challenges. Recently, there have been a number of developments in methods and guidelines addressing missing data in trials. However, it is unclear whether these developments have permeated CEA practice. This paper critically reviews the extent of and methods used to address missing data in recently published trial-based CEA.

Issues of the Health Technology Assessment journal from 2013 to 2015 were searched. Fifty-two eligible studies were identified. Missing data were very common; the median proportion of trial participants with complete cost-effectiveness data was 63% (interquartile range: 47%–81%). The most common approach for the primary analysis was to restrict analysis to those with complete data (43%), followed by multiple imputation (30%). Half of the studies conducted some sort of sensitivity analyses, but only 2 (4%) considered possible departures from the missing-at-random assumption.

Further improvements are needed to address missing data in cost-effectiveness analyses conducted alongside randomised trials. These should focus on limiting the extent of missing data, choosing an appropriate method for the primary analysis that is valid under contextually plausible assumptions, and conducting sensitivity analyses to departures from the missing-at-random assumption.

KEYWORDS

cost-effectiveness analysis, missing data, multiple imputation, randomised controlled trials, sensitivity analysis

This report is independent research supported by the National Institute for Health Research Trainees Coordinating Centre (Doctoral Research Fellowship, B.L., DRF-12437). The views expressed in this publication are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health. J.C. is supported by a Medical Research Council grant (MC_UU_12023/21).

1 | INTRODUCTION

Cost-effectiveness analyses (CEA) conducted alongside randomised controlled trials are an important source of information for health commissioners and decision makers. However, clinical trials rarely succeed in collecting all the intended information (Bell, Fiero, Horton, & Hsu, 2014), and inappropriate handling of the resulting missing data can lead to misleading inferences (Little et al., 2012). This issue is particularly pronounced in CEA because these usually rely on collecting rich, longitudinal information from participants, such as their use of healthcare services (e.g., Client Service Receipt Inventory; Beecham & Knapp, 2001) and their health-related quality of life (e.g., EQ-5D-3L; Brooks, 1996).

Several guidelines have been published in recent years on the issue of missing data in clinical trials (National Research Council, 2010; Committee for Medicinal Products for Human Use (CHMP), 2011; Burzykowski et al., 2010; Carpenter & Kenward, 2007) and for CEA in particular (Briggs, Clark, Wolstenholme, & Clarke, 2003; Burton, Billingham, & Bryan, 2007; Faria, Gomes, Epstein, & White, 2014; Manca & Palmer, 2005; Marshall, Billingham, & Bryan, 2009). Key recommendations include:

- taking practical steps to limit the number of missing observations;
- avoiding methods whose validity rests on contextually implausible assumptions, and using methods that incorporate all available information under reasonable assumptions; and
- assessing the sensitivity of the results to departures from these assumptions.

In particular, following Rubin's taxonomy of missing data mechanisms (Little & Rubin, 2002), methods valid under a missing-at-random (MAR) assumption (i.e., when, given the observed data, missingness does not depend on the unseen values) appear more plausible than the more restrictive assumption of missing completely at random, where missingness is assumed to be entirely independent of the variables of interest. Because we cannot exclude the possibility that the missingness may depend on unobserved values (missing not at random [MNAR]), an assessment of the robustness of the conclusions to alternative missing data assumptions should also be undertaken.

Noble and colleagues (Noble, Hollingworth, & Tilling, 2012) have previously reviewed how missing resource use data were addressed in trial-based CEA. They found that practice fell markedly short of recommendations in several aspects. In particular, that reporting was usually poor and that complete-case analysis was the most common approach. However, missing data research is a rapidly evolving area, and several of the key guidelines were published after that review. We therefore aimed to review how missing cost-effectiveness data were addressed in recent trial-based CEA.

We reviewed studies published in the National Institute for Health Research Health Technology Assessment (HTA) journal, as it provides an ideal source for assessing whether recommendations have permeated CEA practice. These reports give substantially more information than a typical medical journal article, allowing authors the space to clearly describe the issues raised by missing data in their study and the methods they used to address these. Our primary objectives were to determine the extent of missing data, how these were addressed in the analysis, and whether sensitivity analyses to different missing data assumptions were performed. We also provide a critical review of our findings and recommendations to improve practice.

2 | METHODS

The PubMed database was used to identify all trial-based CEA published in HTA between the January 1, 2013, and December 31, 2015. We combined search terms such as “randomised,” “trial,” “cost,” or “economic” to capture relevant articles (see Appendix A.1 for details of the search strategy). The full reports of these articles were downloaded then screened for eligibility by excluding all studies that were pilot or feasibility studies; reported costs and effects separately (e.g., cost-consequence analysis); or did not report a within-trial CEA.

For each included study, we extracted key information about the study and the analysis to answer our primary research questions. A detailed definition of each indicator extracted is provided in Appendix B. In a second stage, we drew on published guidelines and our experience to derive a list of recommendations to address missing data, and then re-reviewed the studies to assess to which extent they followed these recommendations (see Appendix B for further details).

Data analysis was conducted with Stata version 15 (StataCorp, 2017). The data from this review are available on request (Leurent, Gomes, & Carpenter, 2017).

3 | RESULTS

3.1 | Included studies

Sixty-five articles were identified in our search (Figure 1), and 52 eligible studies were included in the review (listed in Appendix A.2). The median time frame for the CEA was over 12 months, and the majority of trials (71%, $n = 37$) conducted a follow-up with repeated assessments over time (median of 2; Table 1). The most common effectiveness measure was the quality-adjusted life year (81%, $n = 42$). Other outcomes included score on clinical measures, or dichotomous outcomes such as "smoking status".

3.2 | Extent of missing data

Missing data was an issue in almost all studies, with only five studies (10%) having less than 5% of participants with missing data. The median proportion of complete cases was 63% (interquartile range, 47–81%; Figure 2). Missing data arose mostly from patient-reported (e.g., resource use and quality of life) questionnaires. The extent of missing data was generally similar for cost and effectiveness data, but 10 (19%) studies had more missing data in the latter (Table 1). The proportion of complete cases reduced, as the number of follow-up assessments increased (Spearman's rank correlation coefficient $\rho = -0.59$, p value $< .001$) and as the study duration increased ($\rho = -0.29$, $p = .04$).

3.3 | Approach to missing data

In the remaining assessments, we excluded the five studies with over 95% of complete cases. Three main approaches to missing data were used: complete-case analysis (CCA; Faria et al., 2014), reported in 66% of studies ($n = 31$), multiple imputation (MI; Rubin, 1987; 49%, $n = 23$), and ad hoc hybrid methods (17%, $n = 8$). For the primary analysis, CCA

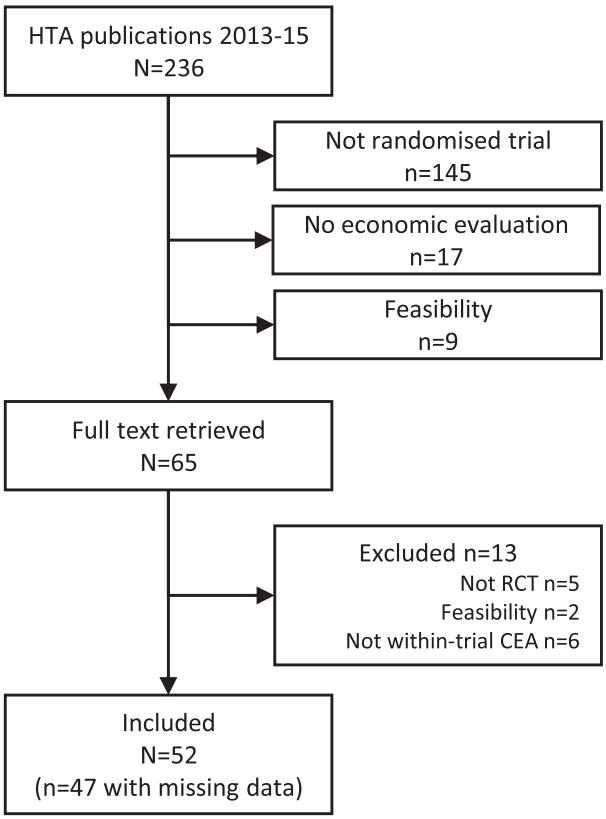


FIGURE 1 Studies selection flow diagram. CEA = cost-effectiveness analyses; HTA = health technology assessment; RCT = randomised controlled trial

TABLE 1 Characteristics of included studies ($n = 52$)

	n	%
	Median	**(IQR)**
General characteristics		
Publication year		
2013	14	27
2014	15	29
2015	23	44
CEA time frame		
0–11 months	22	42
12 months	19	37
≥24 months	11	21
Follow-up design		
Continuous (time to event)	4	8
One follow-up assessment	11	21
Repeated assessments	37	71
Effectiveness measure		
QALY	42	81
Binary	6	12
Clinical scale score	3	6
Time to recovery	1	2
Missing data		
Report exact number of complete cases	20	38
Proportion of complete cases[a]	0.63	(0.47–0.81)
Proportion complete effectiveness data ($n = 47$)	0.73	(0.55–0.86)
Proportion complete cost data ($n = 40$)	0.79	(0.67–0.92)
Differs between costs and effectiveness[b]		
Yes, more cost data missing	3	6
Yes, more effect data missing	10	19
No	22	42
No missing (<5%)	5	10
Unclear	12	23
Differs between arms[c]		
Yes	10	19
No	32	62
No missing (<5%)	5	10
Unclear	5	10

Note. IQR = interquartile range; QALY = quality-adjusted life year.

[a]Proportion of trial participants with complete cost-effectiveness data. An upper bound was used if exact number not reported.

[b]More than 5% difference in the proportion of participants with complete cost or effectiveness data.

[c]More than 5% difference in the proportion of complete cases between arms.

was the most commonly used method (43%, $n = 20$), followed by MI (30%, $n = 14$; Table 2). MI was more common when the proportion of missing data was high and when there were multiple follow-up assessments (see Table 3).

3.4 | Sensitivity analyses

Over half of the studies (53%, $n = 25$) did not conduct any sensitivity analysis around missing data, with 21% ($n = 10$) reporting CCA results alone and 11% ($n = 5$) MI results under MAR alone (Table 4). The remaining studies ($n = 22$, 47%) assessed the sensitivity of their primary analysis results to other approaches for the missing data. This was usually performing either MI under MAR, or CCA, when the other approach was used in the primary analysis. Other sensitivity analyses included using last observation carried forward or regression imputation.

Only two studies (4%) conducted sensitivity analyses, assuming data could be MNAR. In both studies, values imputed under a standard MI were modified to incorporate possible departures from the MAR assumption for both

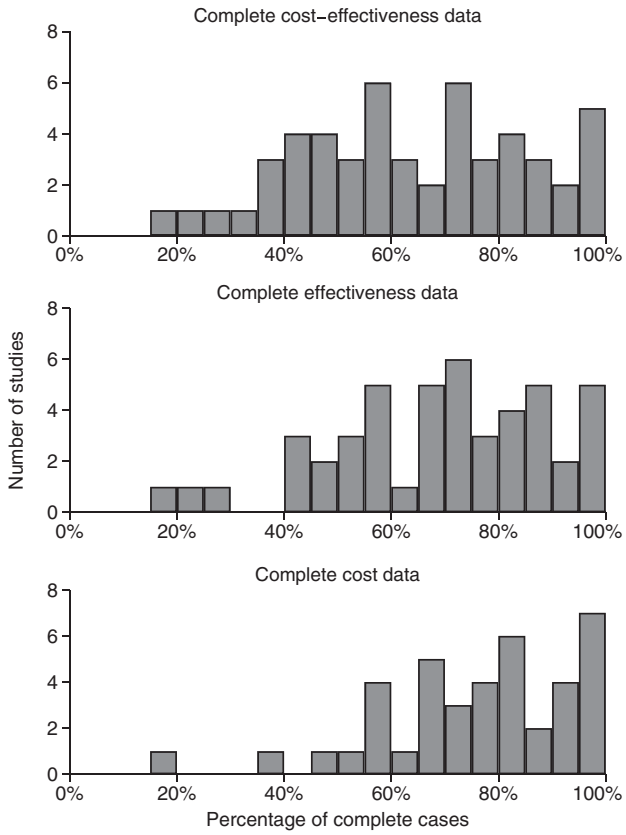


FIGURE 2 Proportion of trial participants with complete data for the primary cost-effectiveness analysis. Shown for cost-effectiveness (n = 52), effectiveness (n = 47, unclear in 5 studies), and cost data (n = 40, unclear in 12 studies)

TABLE 2 Methods for handling missing data in primary analysis ($n = 47$)

Primary analysis method	*n*	%
Complete-case analysis	20	43
Multiple imputation	14	30
Other—single methods		
Inverse probability weighting	1	2
Bayesian model, missing data as unknown parameter	1	2
Other—ad hoc hybrid methods[a]	8	17
Using a combination of		
Mean imputation[b]	6	
Regression imputation[c]	3	
Inverse probability weighting[d]	2	
Assuming failure when outcome missing	2	
Multiple imputation	1	
Last observation carried forward	1	
Unclear	3	6

[a]Ad hoc hybrid method = several approaches to missing data combined, for example, using mean imputation for missing individual resource use items and multiple imputation for fully incomplete observations.

[b]Mean imputation = replacing missing values by the average across other participants.

[c]Regression imputation = replace missing values by predicted value based on observed variables.

[d]Inverse probability weighting = analysing complete data, weighted according to their modelled probability of being observed. These methods are presented in more details in other references (Baio & Leurent, 2016; Faria et al., 2014).

TABLE 3 Approaches to missing data, by year, number of follow-ups, and extent of missing data ($n = 47$)

	Primary analysis method						Reported a sensitivity analysis			
	CCA		MI		Other		Yes		No	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Publication year										
2013 ($n = 13$)	6	46	3	23	4	31	5	38	8	62
2014 ($n = 15$)	9	60	1	7	5	33	6	40	9	60
2015 ($n = 19$)	5	26	10	53	4	21	11	58	8	42
Number of follow-up assessments[a]										
1 ($n = 10$)	7	70	1	10	2	20	3	30	7	70
≥2 ($n = 36$)	13	36	13	36	10	28	18	50	18	50
Proportion of complete cases[b]										
<50% ($n = 15$)	4	27	6	40	5	33	8	53	7	47
50–75% ($n = 18$)	10	56	4	22	4	22	9	50	9	50
75%–95% (n = 14)	6	43	4	29	4	29	5	36	9	64
Information missing[c]										
Similar ($n = 22$)	13	59	6	27	3	14	10	45	12	55
More cost missing ($n = 3$)	1	33	2	67	0	0	2	67	1	33
More effect missing ($n = 10$)	4	40	2	20	4	40	6	60	4	40

Note. % = row percentages. CCA = complete-case analysis; MI = multiple imputation.

[a]Excluding one study with continuous follow-up ($n = 46$).

[b]For the five studies with less than 5% of incomplete cases, four used CCA and one an ad hoc hybrid method for their primary analysis. One of the five studies conducted a sensitivity analysis to missing data.

[c]Excluding 12 studies where this was unclear ($n = 35$).

TABLE 4 Sensitivity analysis, overall, and by primary analysis method ($n = 47$)

	Sensitivity analysis method									
	None		CCA		MI (MAR)		MNAR		Other[a]	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Overall										
Total ($n = 47$)	25	53	11	23	9	19	2	4	5	11
By primary analysis										
CCA (n = 20)	10	50	0	0	8	40	0	0	2	10
MI (n = 14)	5	36	9	64	0	0	2	14	2	14
Other (n = 13)	10	77	2	15	1	8	0	0	1	8

Note. % = row percentages; CCA = complete-case analysis; MAR = assuming data missing at random; MI = multiple imputation; MNAR = assuming data missing not at random. Total may be more than 100% as some studies conducted more than one sensitivity analysis.

[a]Other methods used for sensitivity analysis include last observation carried forward ($n = 1$), regression imputation ($n = 1$), adjusting for baseline predictors of missingness ($n = 1$), imputing by average of observed values for that patient ($n = 1$), and an ad hoc hybrid method using multiple and mean imputation ($n = 1$).

the cost and effectiveness data using a simplified pattern-mixture model approach (Faria et al., 2014; Leurent et al., 2018). The studies then discussed the plausibility of these departures from MAR and their implications for the cost-effectiveness inferences.

3.5 | Recommendations criteria

Table 5 reports the number of studies that reported evidence of following the recommendations from Figure 3 (see Section 4). Most studies reported being aware of the risk of missing data, for example, by taking active steps to reduce them ($n = 35$, 74%). In addition, almost two-thirds of the studies ($n = 29$, 62%) reported the breakdown of missing data by arm, time point, and endpoint. Only about one-third of the studies have clearly reported the reasons for the missing

TABLE 5 Review of indicators based on recommendations criteria ($n = 47$)

Criterion[a]	Met[b]		Not met		Unclear	
	n	%	*n*	%	*n*	%
Prevent						
A1. Maximise response rate	35	74	12	26	0	0
A2. Alternative data sources	10	21	37	79	0	0
A3. Monitor completeness	17	36	30	64	0	0
Primary						
B1. Assumption for primary analysis	17	36	27	57	3	6
B2. Appropriate primary method	17	36	27	57	3	6
Sensitivity						
C1. Discuss departures from the primary assumption	0	0	47	100	0	0
C2. Consider broad range of assumptions	2	4	45	96	0	0
C3. Method valid under these assumptions	2	4	45	96	0	0
Report						
D1. Missing data by endpoint, arm, and time point	29	62	18	38	0	0
D2. Discuss reasons for missing data	16	34	31	66	0	0
D3. Describe methods used and assumptions	17	36	30	64	0	0
D4. Conclusions in light of missing data	1	2	46	98	0	0

[a]See Figure 3 and Appendix B for definition of each criterion.

[b]Report demonstrates evidence of having followed this recommendation. *Not met* if the recommendation was not followed or not mentioned. *Unclear* if some suggestions the criteria may have been met but information not clear enough. See Appendix B for detailed definitions and methodology used.

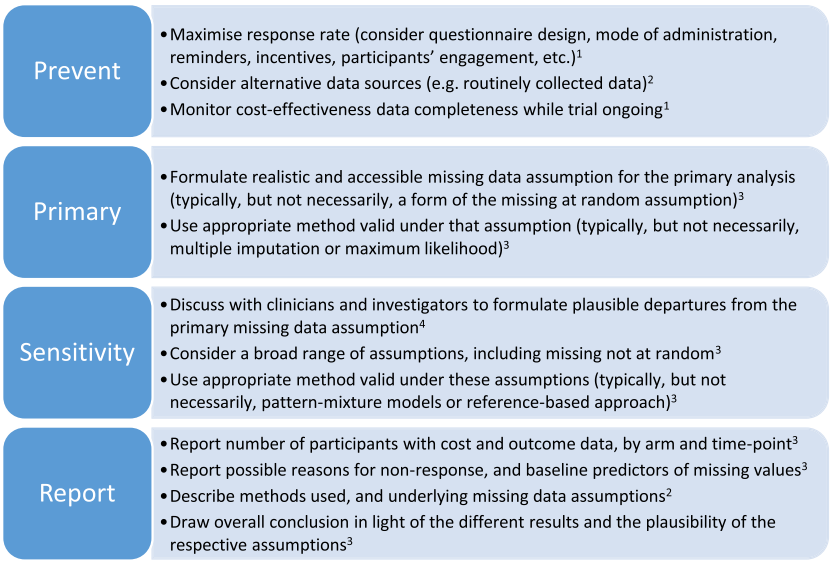


FIGURE 3 Recommendations for improving handling of missing data in trial-based cost-effectiveness analysis. References: 1, Little et al., 2012; 2, Noble et al., 2012; 3, Faria et al., 2014; and 4, Carpenter and Kenward 2007 [Colour figure can be viewed at wileyonlinelibrary.com]

data ($n = 16$, 34%) and the approach used for handling the missing data and its underlying assumptions ($n = 17$, 36%). Only one study (2%) appropriately discussed the implications of missing data in their cost-effectiveness conclusions.

4 | DISCUSSION

4.1 | Summary of findings

Missing data remain ubiquitous in trial-based CEA. The median proportion of participants with complete cost-effectiveness data was only 63%. This reflects the typical challenges faced by CEA of randomised controlled trials, which often rely on patient questionnaires to collect key resource use and health outcome data. Despite best efforts to ensure

completeness, a significant proportion of nonresponse is likely. This is consistent with other reviews, which also found no reduction of the extent of missing data in trials over time (Bell et al., 2014).

CCA remains the most commonly used approach for handling missing data in trial-based CEA, in contrast to recommendations. This approach makes the restrictive assumption that, given the variables in the analysis model, the distributions of the outcome data are the same, whether or not those outcome data are observed. This approach is also problematic because it can result in a loss in precision, as it discards participants who have partially complete data postrandomisation and who can provide important information to the analysis. Other unsatisfactory approaches based on unrealistic assumptions, such as last observation carried forward and single imputation, are also occasionally used.

MI (Rubin, 1987) assuming MAR has been widely recommended for CEA (Briggs et al., 2003; Burton et al., 2007; Faria et al., 2014; Marshall et al., 2009), allowing for baseline variables and postrandomisation data not in the primary analysis to be used for the imputation. It seems to be now more commonly used, with around half of the studies using MI for at least one of their analyses (up to 74% in 2015). Around one-third of the studies used MI for their primary CEA, which is higher than seen in primary clinical outcome analyses (8%; Bell et al., 2014).

On the other hand, sensitivity analyses to missing data remain clearly insufficient. Only two studies (4%) conducted comprehensive sensitivity analyses and assessed whether the study's conclusions were sensitive to departures from the MAR assumption (i.e., possible MNAR mechanisms). Half of the studies did not conduct any sensitivity analysis regarding the missing data. The remaining studies performed some sort of sensitivity analyses, but usually consisting of simple variations from the primary analysis, such as reporting CCA results in addition to MI. This may be more for completeness than proper missing data sensitivity analyses. For example, if MI is used for the primary analysis (having assumed that MAR is the realistic primary missing data assumption), a sensitivity analysis that involves CCA will make stronger missing data assumptions.

4.2 | Strengths and limitations

Our review follows naturally from the review of Noble et al. (2012) and gives an update of the state of play after the publication of several key guidelines. Our review, however, differs in scope and methods and cannot be directly compared with the results of Noble et al. One of the key strengths of this review is that HTA comprehensive reports allowed us to obtain a more complete picture of the missing data and the methods used to tackle it. HTA monographs are published alongside more succinct peer-reviewed papers in specialist medical journals, and they are often seen as the "gold-standard" for trial-based CEA in the UK. It seems therefore reasonable to assume that these are representative of typical practice in CEA. This review is, to our knowledge, the first to look at completeness of both cost and effectiveness data. A limitation is the use of a single-indicator "proportion of complete cases" to capture the extent of the missing data issue. This is however a clearly defined indicator and allows comparison with other reviews. The "recommendations indicators" also focused on the information reported in the study, not necessarily what might have been done in practice.

4.3 | Recommendations

A list of recommendations to address missing data in trial-based CEA is presented in Figure 3. Trial-based CEA are prone to missing data, and it is important that analysts take active steps at the design and data-collection stages to limit their extent (Bernhard et al., 2006; Brueton et al., 2013; National Research Council, 2010). Resource use questionnaires should be designed in a user-friendly way, and their completion encouraged during follow-up visits, possibly supported by a researcher (Mercieca-Bebber et al., 2016; National Research Council, 2010). Alternative sources should also be considered to minimise missing information, for example, administrative data or electronic health records (Franklin & Thorn, 2018; Noble et al., 2012).

For any study with missing data, clear reporting of the issue is required. Ideally, the study should report details of the pattern of missing data (Faria et al., 2014), possibly as an appendix. At a minimum, CEA studies should report for each analysis the number of participants included by trial arm, as recommended in the Consolidated Standards of Reporting Trials guidelines (Noble et al., 2012; Schulz et al., 2010).

Although CCA may be justifiable in some circumstances, the choice of CCA for the primary analysis approach appears difficult to justify in the presence of repeated measurements, because the loss of power (by discarding all patients with any missing values) across the different time points tends to be large. Other approaches valid under more plausible

MAR assumptions and making use of all the observed data, such as MI (Rubin, 1987); likelihood-based repeated measures models (Faria et al., 2014; Verbeke, Fieuws, Molenberghs, & Davidian, 2014); or Bayesian models (Ades et al., 2006), should be considered. In particular, MI has been increasingly used in CEA, and further guidance to support an appropriate use in this context is warranted.

An area with clear room for improvement is the conduct of sensitivity analyses. This review found that many studies used CCA for the primary analysis and MI as a sensitivity analysis, or vice-versa, and concluded that the results were robust to missing data. This is misleading because both of these methods rely on the assumption that the missingness is independent of the unobserved data. Although the MAR assumption provides a sensible starting point, it is not possible to determine the true missing-data mechanism from the observed data. Studies should therefore assess whether their conclusions are sensitive to possible departures from that assumption (National Research Council, 2010; Committee for Medicinal Products for Human Use (CHMP), 2011; Faria et al., 2014). Several approaches have been suggested to conduct analyses under MNAR assumptions. Selection models express how the probability of being missing is related to the value itself. Pattern-mixture models, on the other hand, capture how missing data could differ from the observed (Molenberghs et al., 2014; Ratitch, O'Kelly, & Tosiello, 2013). Pattern-mixture models appear attractive because they frame the departure from MAR in a way that can be more readily understood by clinical experts and decision makers and can be used with standard analysis methods such as MI (Carpenter & Kenward, 2012; Ratitch et al., 2013). MNAR modelling can be challenging, but accessible approaches have also been proposed (Faria et al., 2014; Leurent et al., 2018). Further developments are still needed to use these methods in the CEA context and to provide the analytical tools and practical guidance to implement them in practice.

5 | CONCLUSION

Missing data can be an important source of bias and uncertainty, and it is imperative that this issue is appropriately recognised and addressed to help ensure that CEA studies provide sound evidence for healthcare decision making. Over the last decade, there have been some welcome improvements in handling missing data in trial-based CEA. In particular, more attention has been devoted to assessing the reasons for the missing data and adopting methods (e.g., MI) that can incorporate those in the analysis. However, there is substantial room for improvement. Firstly, more efforts are needed to reduce missing data. Secondly, the extent and patterns of missing data should be more clearly reported. Thirdly, the primary analysis should consider methods that make contextually plausible assumptions rather than resort automatically to CCA. Lastly, sensitivity analyses to assess the robustness of the study's results to potential MNAR mechanisms should be conducted.

CONFLICT OF INTEREST

The authors have no conflict of interest.

ORCID

Baptiste Leurent http://orcid.org/0000-0001-6420-6567
Manuel Gomes http://orcid.org/0000-0002-1428-1232

REFERENCES

Ades, A. E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N., & Lu, G. (2006). Bayesian methods for evidence synthesis in cost-effectiveness analysis. *PharmacoEconomics*, *24*(1), 1–19. Available at: http://link.springer.com/10.2165/00019053-200624010-00001. Accessed May 25, 2017

Baio, G., & Leurent, B. (2016). An Introduction to Handling Missing Data in Health Economic Evaluations. In *Care at the End of Life* (pp. 73–85. Available at: http://link.springer.com/10.1007/978-3-319-28267-1_6). Cham: Springer International Publishing. Accessed September 2, 2016

Beecham, J. & Knapp, M., (2001). Costing psychiatric interventions, Gaskell London.

Bell, M. L., Fiero, M., Horton, N. J., & Hsu, C. H. (2014). Handling missing data in RCTs; a review of the top medical journals. *BMC Medical Research Methodology*, *14*(1), 118.

Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., ... Hürny, C. (2006). Missing quality of life data in cancer clinical trials: Serious problems and challenges. *Statistics in Medicine*, *17*(5–7), 517–532. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9549801

Briggs, A., Clark, T., Wolstenholme, J., & Clarke, P. (2003). Missing....presumed at random: Cost-analysis of incomplete data. *Health Economics*, *12*(5), 377–392.

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, *37*(1), 53–72.

Brueton, V. C., Tierney, J., Stenning, S., Harding, S., Meredith, S., Nazareth, I., & Rait, G. (2013). Strategies to improve retention in randomised trials. *The Cochrane Database of Systematic Reviews*, *12*(12). p.MR000032

Burton, A., Billingham, L. J., & Bryan, S. (2007). Cost-effectiveness in clinical trials: Using multiple imputation to deal with incomplete cost data. *Clinical Trials*, *4*(2), 154–161.

Burzykowski, T., Carpenter, J., Coens, C., Evans, D., France, L., Kenward, M., ... of the PSI Missing Data Expert Group (2010). Missing data: Discussion points from the PSI missing data expert group. *Pharmaceutical Statistics*, *9*(4), 288–297.

Carpenter, J., & Kenward, M. (2012). *Multiple imputation and its application*. Chichester: John Wiley & Sons.

Carpenter, J. R., & Kenward, M. G. (2007). *Missing data in randomised controlled trials-a practical guide*. Birmingham: National Institute for Health Research.

Committee for Medicinal Products for Human Use (CHMP) (2011). Guideline on missing data in confirmatory clinical trials. *London: European Medicines ...*, *44*(July 2010), 1–12.

Faria, R., Gomes, M., Epstein, D., & White, I. R. (2014). A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *PharmacoEconomics*, *32*(12), 1157–1170.

Franklin, M., & Thorn, J. (2018). Self-reported and routinely collected electronic data for trial-based evaluations: the current state of play in England and considerations for the future. *BMC Medical Research Methodology*, Submitted.

Leurent, B., Gomes, M. & Carpenter, J., (2017). Missing data in trial-based cost-effectiveness analysis review dataset. Available at: https://doi.org/10.17037/DATA.272.

Leurent, B., Gomes, M., Faria, R., Morris, S., Grieve, R., & Carpenter, J. R. (2018). Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *PharmacoEconomics*, Submitted.

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., ... Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, *367*(14), 1355–1360.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley & Sons.

Manca, A., & Palmer, S. (2005). Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. *Applied health economics and health policy*, *4*(2), 65–75.

Marshall, A., Billingham, L. J., & Bryan, S. (2009). Can we afford to ignore missing data in cost-effectiveness analyses? *The European Journal of Health Economics*, *10*(1), 1–3.

Mercieca-Bebber, R., Palmer, M. J., Brundage, M., Calvert, M., Stockler, M. R., & King, M. T. (2016). Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: A systematic review. *BMJ Open*, *6*(6), e010938.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2014). *Handbook of missing data methodology*. Boca Raton, FL: CRC Press.

National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: The National Academies Press. Available at: http://www.nap.edu/catalog/12955/

Noble, S. M., Hollingworth, W., & Tilling, K. (2012). Missing data in trial-based cost-effectiveness analysis: The current state of play. *Health Economics*, *21*(2), 187–200.

Ratitch, B., O'Kelly, M., & Tosiello, R. (2013). Missing data in clinical trials: From clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics*, *12*(6), 337–347.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. (pp. 15–19). Harvard University: John Wiley & Sons.

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine*, *152*(11), 726–732.

StataCorp, (2017). Stata Statistical Software: Release 15.

Verbeke, G., Fieuws, S., Molenberghs, G., & Davidian, M. (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, *23*(1), 42–59.

How to cite this article: Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. *Health Economics*. 2018;27:1024–1040. https://doi.org/10.1002/hec.3654

APPENDIX A

DETAILS OF STUDIES SELECTION

A.1 | PubMed search criteria and results

Search criteria:

("Health Technol Assess"[Journal])
AND ("2015/01/01"[Date - Publication]: "2015/12/31"[Date - Publication])
AND ("randomised"[Title] OR "randomized"[Title] OR "trial"[Title])
AND ("economic"[Title/Abstract] OR "cost*"[Title/Abstract])
NOT ("pilot"[Title] OR "feasibility"[Title])

Number of studies:

Search	Query	Items found
4	Search ("health Technol assess"[journal]) AND ("2013/01/01"[date - publication] : "2015/12/31"[date - publication]) AND ("randomised"[title] OR "randomised"[title] OR "trial"[title]) AND ("economic"[title/abstract] OR "cost*"[title/abstract]) NOT ("pilot"[title] OR "feasibility"[title])	65
3	Search ("Health Technol Assess"[Journal]) AND ("2013/01/01"[Date - Publication] : "2015/12/31"[Date - Publication]) AND ("randomised"[title] OR "randomized"[Title] OR "trial"[Title]) AND ("economic"[Title/Abstract] OR "cost*"[Title/Abstract])	74
2	Search ("Health Technol Assess"[Journal]) AND ("2013/01/01"[Date - Publication] : "2015/12/31"[Date - Publication]) AND ("randomised"[Title] OR "randomized"[Title] OR "trial"[Title])	91
1	Search ("Health Technol Assess"[Journal]) AND ("2013/01/01"[Date - Publication] : "2015/12/31"[Date - Publication])	236

A.2 | Included studies

Allen, S., Wareham, K., Wang, D., Bradley, C., Sewell, B., Hutchings, H., ... Phillips, C. (2013). A high-dose preparation of lactobacilli and bifidobacteria in the prevention of antibiotic-associated and Clostridium difficile diarrhoea in older people admitted to hospital: A multicentre, randomised, double-blind, placebo-controlled, parallel arm trial. Health Technology Assessment, 17(57), 1–140. http://doi.org/10.3310/hta17570

Ashby, R. L., Gabe, R., Ali, S., Saramago, P., Chuang, L.-H., Adderley, U., ... Torgerson, D. J. (2014). VenUS IV (Venous leg Ulcer Study IV) – compression hosiery compared with compression bandaging in the treatment of venous leg ulcers: A randomised controlled trial, mixed-treatment comparison and decision-analytic model. Health Technology Assessment, 18(57), 1–294. http://doi.org/10.3310/hta18570

Banerjee, S., Hellier, J., Romeo, R., Dewey, M., Knapp, M., Ballard, C., ... Burns, A. (2013). Study of the use of antidepressants for depression in dementia: the HTA-SADD trial—A multicentre, randomised, double-blind, placebo-controlled trial of the clinical effectiveness and cost-effectiveness of sertraline and mirtazapine. Health Technology Assessment (Winchester, England), 17(7), 1–166. http://doi.org/10.3310/hta17070

Bedson, E., Bell, D., Carr, D., Carter, B., Hughes, D., Jorgensen, A., ... Williams, N. (2014). Folate Augmentation of Treatment – Evaluation for Depression (FolATED): Randomised trial and economic evaluation. Health Technology Assessment, 18(48), 1–160. http://doi.org/10.3310/hta18480

Blyth, A., Maskrey, V., Notley, C., Barton, G. R., Brown, T. J., Aveyard, P., ... Song, F. (2015). Effectiveness and economic evaluation of self-help educational materials for the prevention of smoking relapse: Randomised controlled trial. Health Technology Assessment (Winchester, England), 19(59), 1–70, v–vi. http://doi.org/10.3310/hta19590

Brittenden, J., Cotton, S. C., Elders, A., Tassie, E., Scotland, G., Ramsay, C. R., ... Campbell, M. K. (2015). Clinical effectiveness and cost-effectiveness of foam sclerotherapy, endovenous laser ablation and surgery for varicose veins: Results from the Comparison of LAser, Surgery and foam Sclerotherapy (CLASS) randomised controlled trial. Health Technology Assessment (Winchester, England), 19(27), 1–342. http://doi.org/10.3310/hta19270

Carr, A. J., Cooper, C. D., Campbell, M. K., Rees, J. L., Moser, J., Beard, D. J., ... Ramsay, C. R. (2015). Clinical effectiveness and cost-effectiveness of open and arthroscopic rotator cuff repair [the UK Rotator Cuff Surgery (UKUFF) randomised trial]. Health Technology Assessment (Winchester, England), 19(80), 1–218. http://doi.org/10.3310/hta19800

Chakravarthy, U., Harding, S. P., Rogers, C. A., Downes, S., Lotery, A. J., Dakin, H. A., ... Reeves, B. C. (2015). A randomised controlled trial to assess the clinical effectiveness and cost-effectiveness of alternative treatments to Inhibit VEGF in Age-related choroidal Neovascularisation (IVAN). Health Technology Assessment, 19(78), 1–298. http://doi.org/10.3310/hta19780

Clark, T. J., Middleton, L. J., Cooper, N. A., Diwakar, L., Denny, E., Smith, P., ... Daniels, J. P. (2015). A randomised controlled trial of Outpatient versus inpatient Polyp Treatment (OPT) for abnormal uterine bleeding. Health Technology Assessment, 19(61), 1–194. http://doi.org/10.3310/hta19610

Cooper, S., Lewis, S., Thornton, J. G., Marlow, N., Watts, K., Britton, J., ... Coleman, T. (2014). The SNAP trial: a randomised placebo-controlled trial of nicotine replacement therapy in pregnancy – clinical effectiveness and safety until 2 years after delivery, with economic evaluation. Health Technology Assessment, 18(54), 1–128. http://doi.org/10.3310/hta18540

Costa, M. L., Achten, J., Plant, C., Parsons, N. R., Rangan, A., Tubeuf, S., ... Lamb, S. E. (2015). UK DRAFFT: A randomised controlled trial of percutaneous fixation with Kirschner wires versus volar locking-plate fixation in the treatment of adult patients with a dorsally displaced fracture of the distal radius. Health Technology Assessment, 19(17), 1–124. http://doi.org/10.3310/hta19170

Crawford, M. J., Sanatinia, R., Barrett, B., Byford, S., Dean, M., Green, J., ... Ward, H. (2014). The clinical effectiveness and cost-effectiveness of brief intervention for excessive alcohol consumption among people attending sexual health clinics: A randomised controlled trial (SHEAR). Health Technology Assessment, 18(30), 1–48. http://doi.org/10.3310/hta18300

Creswell, C., Cruddace, S., Gerry, S., Gitau, R., McIntosh, E., Mollison, J., ... Cooper, P. J. (2015). Treatment of childhood anxiety disorder in the context of maternal anxiety disorder: A randomised controlled trial and economic analysis. Health Technology Assessment, 19(38), 1–184. http://doi.org/10.3310/hta19380

Cunningham, S., Rodriguez, A., Boyd, K. A., McIntosh, E., & Lewis, S. C. (2015). Bronchiolitis of Infancy Discharge Study (BIDS): A multicentre, parallel-group, double-blind, randomised controlled, equivalence trial with economic evaluation. Health Technology Assessment, 19(71), 1–172. http://doi.org/10.3310/hta19710

Dennis, M., Sandercock, P., Graham, C., & Forbes, J. (2015). The Clots in Legs Or sTockings after Stroke (CLOTS) 3 trial: A randomised controlled trial to determine whether or not intermittent pneumatic compression reduces the risk of post-stroke deep vein thrombosis and to estimate its cost-effectiveness. Health Technology Assessment, 19(76), 1–90. http://doi.org/10.3310/hta19760

Everard, M. L., Hind, D., Ugonna, K., Freeman, J., Bradburn, M., Dixon, S., ... Cross, E. (2015). Saline in Acute Bronchiolitis RCT and Economic evaluation: Hypertonic saline in acute bronchiolitis – randomised controlled trial and systematic review. Health Technology Assessment, 19(66), 1–130. http://doi.org/10.3310/hta19660

Forster, A., Dickerson, J., Young, J., Patel, A., Kalra, L., Nixon, J., ... Farrin, A. (2013). A cluster randomised controlled trial and economic evaluation of a structured training programme for caregivers of inpatients after stroke: The TRACS trial. Health Technology Assessment, 17(46), 1–216. http://doi.org/10.3310/hta17460

Gates, S., Perkins, G., Lamb, S., Kelly, C., Thickett, D., Young, J., ... Gao Smith, F. (2013). Beta-Agonist Lung injury TrIal-2 (BALTI-2): A multicentre, randomised, double-blind, placebo-controlled trial and economic evaluation of intravenous infusion of salbutamol versus placebo in patients with acute respiratory distress syndrome. Health Technology Assessment, 17(38), v–vi, 1–87. http://doi.org/10.3310/hta17380

Goyder, E., Hind, D., Breckon, J., Dimairo, M., Minton, J., Everson-Hock, E., ... Cooper, C. (2014). A randomised controlled trial and cost-effectiveness evaluation of "booster" interventions to sustain increases in physical activity in middle-aged adults in deprived urban neighbourhoods. Health Technology Assessment, 18(13), 1–210. http://doi.org/10.3310/hta18130

Grant, A., Boachie, C., Cotton, S., Faria, R., Bojke, L., Epstein, D., ... Campbell, M. (2013). Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial). Health Technology Assessment, 17(22), 1–167. http://doi.org/10.3310/hta17220

Gupta, J. K., Daniels, J. P., Middleton, L. J., Pattison, H. M., Prileszky, G., Roberts, T. E., ... Kai, J. (2015). A randomised controlled trial of the clinical effectiveness and cost-effectiveness of the levonorgestrel-releasing intrauterine

system in primary care against standard treatment for menorrhagia: The ECLIPSE trial. Health Technology Assessment, 19(88), 1–118. http://doi.org/10.3310/hta19880

Halligan, S., Dadswell, E., Wooldrage, K., Wardle, J., von Wagner, C., Lilford, R., ... Atkin, W. (2015). Computed tomographic colonography compared with colonoscopy or barium enema for diagnosis of colorectal cancer in older symptomatic patients: Two multicentre randomised trials with economic evaluation (the SIGGAR trials). Health Technology Assessment, 19(54), 1–134. http://doi.org/10.3310/hta19540

Handoll, H., Brealey, S., Rangan, A., Keding, A., Corbacho, B., Jefferson, L., ... Torgerson, D. (2015). The ProFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation) trial – a pragmatic multicentre randomised controlled trial evaluating the clinical effectiveness and cost-effectiveness of surgical compared with non-surgical treatment for proxi. Health Technology Assessment, 19(24), 1–280. http://doi.org/10.3310/hta19240

Iliffe, S., Kendrick, D., Morris, R., Masud, T., Gage, H., Skelton, D., ... Belcher, C. (2014). Multicentre cluster randomised trial comparing a community group exercise programme and home-based exercise with usual care for people aged 65 years and over in primary care. Health Technology Assessment, 18(49), 1–106. http://doi.org/10.3310/hta18490

Kuyken, W., Hayes, R., Barrett, B., Byng, R., Dalgleish, T., Kessler, D., ... Byford, S. (2015). The effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse/recurrence: Results of a randomised controlled trial (the PREVENT study). Health Technology Assessment, 19(73), 1–124. http://doi.org/10.3310/hta19730

Lall, R., Hamilton, P., Young, D., Hulme, C., Hall, P., Shah, S., ... Lamb, S. (2015). A randomised controlled trial and cost-effectiveness analysis of high-frequency oscillatory ventilation against conventional artificial ventilation for adults with acute respiratory distress syndrome. The OSCAR (OSCillation in ARDS) study. Health Technology Assessment, 19(23), 1–178. http://doi.org/10.3310/hta19230

Lenney, W., McKay, A. J., Tudur Smith, C., Williamson, P. R., James, M., Price, D., & MASCOT Study Group. (2013). Management of Asthma in School age Children On Therapy (MASCOT): A randomised, double-blind, placebo-controlled, parallel study of efficacy and safety. Health Technology Assessment (Winchester, England), 17(4), 1–218. http://doi.org/10.3310/hta17040

Little, P., Hobbs, F. R., Moore, M., Mant, D., Williamson, I., McNulty, C., ... Mullee, M. (2014). PRImary care Streptococcal Management (PRISM) study: in vitro study, diagnostic cohorts and a pragmatic adaptive randomised controlled trial with nested qualitative study and cost-effectiveness study. Health Technology Assessment, 18(6), vii–xxv, 1–101. http://doi.org/10.3310/hta18060

Littlewood, E., Duarte, A., Hewitt, C., Knowles, S., Palmer, S., Walker, S., ... Gilbody, S. (2015). A randomised controlled trial of computerised cognitive behaviour therapy for the treatment of depression in primary care: The Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) trial. Health Technology Assessment, 19(101), 1–174. http://doi.org/10.3310/hta191010

Livingston, G., Barber, J., Rapaport, P., Knapp, M., Griffin, M., Romeo, R., ... Cooper, C. (2014). START (STrAtegies for RelaTives) study: A pragmatic randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of a manual-based coping strategy programme in promoting the mental health of carers of people with dementia. Health Technology Assessment, 18(61), 1–242. http://doi.org/10.3310/hta18610

Logan, P. A., Armstrong, S., Avery, T. J., Barer, D., Barton, G. R., Darby, J., ... Leighton, M. P. (2014). Rehabilitation aimed at improving outdoor mobility for people after stroke: a multicentre randomised controlled study (the Getting out of the House Study). Health Technology Assessment, 18(29), vii–viii, 1–113. http://doi.org/10.3310/hta18290

McMillan, A., Bratton, D. J., Faria, R., Laskawiec-Szkonter, M., Griffin, S., Davies, R. J., ... Morrell, M. J. (2015). A multicentre randomised controlled trial and economic evaluation of continuous positive airway pressure for the treatment of obstructive sleep apnoea syndrome in older people: PREDICT. Health Technology Assessment, 19(40), 1–188. http://doi.org/10.3310/hta19400

Molassiotis, A., Russell, W., Hughes, J., Breckons, M., Lloyd-Williams, M., Richardson, J., ... Ryder, W. (2013). The effectiveness and cost-effectiveness of acupressure for the control and management of chemotherapy-related acute and delayed nausea: Assessment of Nausea in Chemotherapy Research (ANCHoR), a randomised controlled trial. Health Technology Assessment, 17(26), 1–114. http://doi.org/10.3310/hta17260

Morris, R., Malin, G., Quinlan-Jones, E., Middleton, L., Diwakar, L., Hemming, K., ... Kilby, M. (2013). The Percutaneous shunting in Lower Urinary Tract Obstruction (PLUTO) study and randomised controlled trial: evaluation of the effectiveness, cost-effectiveness and acceptability of percutaneous vesicoamniotic shunting for lower urinary tract obstruction. Health Technology Assessment, 17(59), 1–232. http://doi.org/10.3310/hta17590

Mouncey, P. R., Osborn, T. M., Power, G. S., Harrison, D. A., Sadique, M. Z., Grieve, R. D., ... Rowan, K. M. (2015). Protocolised Management In Sepsis (ProMISe): A multicentre randomised controlled trial of the clinical effectiveness and cost-effectiveness of early, goal-directed, protocolised resuscitation for emerging septic shock. Health Technology Assessment, 19(97), 1–150. http://doi.org/10.3310/hta19970

Murray, D. W., MacLennan, G. S., Breeman, S., Dakin, H. A., Johnston, L., Campbell, M. K., ... Grant, A. M. (2014). A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: The Knee Arthroplasty Trial (KAT). Health Technology Assessment, 18(19), 1–235, vii–viii. http://doi.org/10.3310/hta18190

Nicholson, K. G., Abrams, K. R., Batham, S., Medina, M. J., Warren, F. C., Barer, M., ... Zambon, M. (2014). Randomised controlled trial and health economic evaluation of the impact of diagnostic testing for influenza, respiratory syncytial virus and Streptococcus pneumoniae infection on the management of acute admissions in the elderly and high-risk 18- to 64-y. Health Technology Assessment, 18(36), 1–274, vii–viii. http://doi.org/10.3310/hta18360

Orgeta, V., Leung, P., Yates, L., Kang, S., Hoare, Z., Henderson, C., ... Orrell, M. (2015). Individual cognitive stimulation therapy for dementia: a clinical effectiveness and cost-effectiveness pragmatic, multicentre, randomised controlled trial. Health Technology Assessment, 19(64), 1–108. http://doi.org/10.3310/hta19640

Pickard, R., Starr, K., MacLennan, G., Kilonzo, M., Lam, T., Thomas, R., ... McClinton, S. (2015). Use of drug therapy in the management of symptomatic ureteric stones in hospitalised adults: A multicentre, placebo-controlled, randomised controlled trial and cost-effectiveness analysis of a calcium channel blocker (nifedipine) and an alpha-blocker (tam. Health Technology Assessment (Winchester, England)), 19(63), vii–viii, 1–171. http://doi.org/10.3310/hta19630

Powell, C., Kolamunnage-Dona, R., Lowe, J., Boland, A., Petrou, S., Doull, I., ... Williamson, P. (2013). MAGNEsium Trial In Children (MAGNETIC): A randomised, placebo-controlled trial and economic evaluation of nebulised magnesium sulphate in acute severe asthma in children. Health Technology Assessment, 17(45), v–vi, 1–216. http://doi.org/10.3310/hta17450

Russell, I., Edwards, R., Gliddon, A., Ingledew, D., Russell, D., Whitaker, R., ... Park, K. (2013). Cancer of Oesophagus or Gastricus – New Assessment of Technology of Endosonography (COGNATE): report of pragmatic randomised trial. Health Technology Assessment, 17(39), 1–170. http://doi.org/10.3310/hta17390

Salisbury, C., Foster, N., Hopper, C., Bishop, A., Hollinghurst, S., Coast, J., ... Montgomery, A. (2013). A pragmatic randomised controlled trial of the effectiveness and cost-effectiveness of "PhysioDirect" telephone assessment and advice services for physiotherapy. Health Technology Assessment, 17(02), 1–157, v–vi. http://doi.org/10.3310/hta17020

Scott, D. L., Ibrahim, F., Farewell, V., O'Keeffe, A. G., Ma, M., Walker, D., ... Kingsley, G. (2014). Randomised controlled trial of Tumour necrosis factor inhibitors Against Combination Intensive Therapy with conventional disease-modifying antirheumatic drugs in established rheumatoid arthritis: The TACIT trial and associated systematic reviews. Health Technology Assessment, 18(66), 1–164. http://doi.org/10.3310/hta18660

Sharples, L., Glover, M., Clutterbuck-James, A., Bennett, M., Jordan, J., Chadwick, R., ... Quinnell, T. (2014). Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea–hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive airwa. Health Technology Assessment, 18(67), 1–296. http://doi.org/10.3310/hta18670

Stallard, P., Phillips, R., Montgomery, A., Spears, M., Anderson, R., Taylor, J., ... Sayal, K. (2013). A cluster randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of classroom-based cognitive–behavioural therapy (CBT) in reducing symptoms of depression in high-risk adolescents. Health Technology Assessment, 17(47), vii–xvii, 1–109. http://doi.org/10.3310/hta17470

Thursz, M., Forrest, E., Roderick, P., Day, C., Austin, A., O'Grady, J., ... Ternent, L. (2015). The clinical effectiveness and cost-effectiveness of STeroids Or Pentoxifylline for Alcoholic Hepatitis (STOPAH): A 2×2 factorial randomised controlled trial. Health Technology Assessment, 19(102), 1–104. http://doi.org/10.3310/hta191020

Underwood, M., Lamb, S., Eldridge, S., Sheehan, B., Slowther, A., Spencer, A., ... Taylor, S. (2013). Exercise for depression in care home residents: a randomised controlled trial with cost-effectiveness analysis (OPERA). Health Technology Assessment, 17(18), 1–281. http://doi.org/10.3310/hta17180

Ussher, M., Lewis, S., Aveyard, P., Manyonda, I., West, R., Lewis, B., ... Coleman, T. (2015). The London Exercise And Pregnant smokers (LEAP) trial: A randomised controlled trial of physical activity for smoking cessation in pregnancy with an economic evaluation. Health Technology Assessment, 19(84), 1–136. http://doi.org/10.3310/hta19840

Watson, J., Crosby, H., Dale, V., Tober, G., Wu, Q., Lang, J., ... Coulton, S. (2013). AESOPS: A randomised controlled trial of the clinical effectiveness and cost-effectiveness of opportunistic screening and stepped care interventions for older hazardous alcohol users in primary care. Health Technology Assessment, 17(25), 1–158. http://doi.org/10.3310/hta17250

Wiles, N., Thomas, L., Abel, A., Barnes, M., Carroll, F., Ridgway, N., ... Lewis, G. (2014). Clinical effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: The CoBalT randomised controlled trial. Health Technology Assessment, 18(31), 1–167, vii–viii. http://doi.org/10.3310/hta18310

Williams, M. A., Williamson, E. M., Heine, P. J., Nichols, V., Glover, M. J., Dritsaki, M., ... Lamb, S. E. (2015). Strengthening And stretching for Rheumatoid Arthritis of the Hand (SARAH). A randomised controlled trial and economic evaluation. Health Technology Assessment, 19(19), 1–222. http://doi.org/10.3310/hta19190

Wolf, A., McKay, A., Spowart, C., Granville, H., Boland, A., Petrou, S., ... Gamble, C. (2014). Prospective multicentre randomised, double-blind, equivalence study comparing clonidine and midazolam as intravenous sedative agents in critically ill children: The SLEEPS (Safety profiLe, Efficacy and Equivalence in Paediatric intensive care Sedation) st. Health Technology Assessment, 18(71), 1–212. http://doi.org/10.3310/hta18710

APPENDIX B

INDICATORS DEFINITION

B.1 | Primary indicators

Indicator	Definition	Notes
Proportion of complete cases	Proportion of randomised participants for whom all data were available for the primary cost-effectiveness analysis	If the number of complete-cases was not clearly reported, we estimated an "upper bound," from information, such as the proportion of participants with complete cost, or effect, data. See definition of primary analysis below.
Proportion complete effectiveness data	Proportion of randomised participants for whom all effectiveness data were Available for the primary cost-effectiveness analysis	Same as above
Proportion complete cost data	Proportion of randomised participants for whom all cost data were available for the primary cost-effectiveness analysis	Same as above
Report exact number of complete cases	Whether the number of participants with complete cost and effectiveness data was clearly reported.	
More missing costs or effectiveness	Whether the proportion of complete cases differ between cost and effectiveness variable.	Considered "similar" when the proportion of complete cases was within 5% of each other.
Primary analysis method	Methods used to address missing data in the primary (base case) cost-effectiveness analysis	When multiple effectiveness measures, time-frames, or cost perspectives were reported, without a base-case clearly defined, we considered the analysis based on quality-adjusted life years (QALYs) over the longest within-trial follow-up period, from the NHS and social services cost perceptive.
Conducted a sensitivity analysis to missing data	Report results under more than one approach for addressing missing data	

B.2 | Secondary indicators: Derived from the recommendations list

B.2.1 | Methods

Because these aspects could have been mentioned in multiple parts in the monograph, we used a systematic approach, looking for keywords and checking the most relevant paragraphs in the full report.

i. Search in PDF: "Missing"; "Participation"; "Completion"; "Incomplete"; "Response"; "Non-response"; "Monitor"; "MCAR"; "MAR"; "MNAR."
ii. If did not find "steps to reduce missing data," also check "reminder," "incentive," "telephone," and "contact."
iii. Then, check relevant paragraphs manually: data source for cost-effectiveness data; beginning of CEA results; and CEA conclusions.

B.2.2 | Answers

"Yes": The recommendation was clearly mentioned, and the criteria therefore met.

"No": The recommendation was not clearly mentioned or found. The recommendation may still have been followed but not reported (or at least not found with the above strategy).

"Unclear": There was some suggestions the criteria may have been met but not enough information to be sure.

Recommendation	Indicator definition	Examples "yes"	Examples "no"	Notes
A1. Maximise response rate (consider questionnaire design, mode of administration, reminders, incentives, participants' engagement, etc.)	Mention taking steps to maximise response rate	Reminder, incentives, home/hospital visit, multiple attempts,	Mention response was maximised for clinical outcome but not reported for cost-effectiveness endpoints	Can be for overall trial data if implicit includes cost or effect data. Except if steps are clearly for non-CE variables only (e.g., primary outcome only).
A2. Consider alternative data sources (e.g., routinely collected data)	Mention that considered missing data issues when choosing appropriate source, OR mention more than one source used for a CE data.	Use of electronic health records or administrative data, e.g., hospital episode statistics were used to supplement trial's data, for example, about hospital admissions post-randomisation (which might be otherwise missing).	Using routine data as a primary source: e.g., resource use taken primarily from administrative/hospital records.	
A3. Monitor cost-effectiveness data completeness while trial ongoing	Mentioned monitoring data completeness while trial ongoing.	Data managers checked inconsistent and missing data (if not clear "while trial ongoing" but mention monitoring probably fine). Mention taking new steps to reduce MD (e.g., incentive) as realised lots of MD after trial started.	Mention data checks for inconsistencies, but no mention of checking missing data.	Can be for overall trial data. Except if monitoring clearly for non-CE variables only (e.g., primary outcome only).
B1. Formulate realistic and accessible missing data assumption for the primary analysis (typically, but not necessarily, a form of the missing at random assumption)	Primary (base-case) CEA based on reasonable missing data assumptions. (likely MAR, or alternative if well justified).	– Used MI for primary analysis - well justified and clear alternative	– Hybrid method, except if clearly explain and justify underlying assumptions	

(Continues)

(Continued)

Recommendation	Indicator definition	Examples "yes"	Examples "no"	Notes
B2. Use appropriate method valid under that assumption (typically, but not necessarily, multiple imputation or maximum likelihood)	Use appropriate analysis method.	– MI for primary analysis - Bayesian under MAR - well justified and clear alternative	– Use unadjusted CCA when reporting data are MAR.	
C1. Discuss with clinicians and investigators to formulate plausible departures from the primary missing data assumption		Conducted MNAR SA + mention elicitation.	Did not conduct MNAR SA	
C2. Consider a broad range of assumptions, including missing not at random mechanisms		Conducted MNAR SA	Did not conduct MNAR SA	
C3. Use appropriate method valid under these assumptions (typically, but not necessarily, pattern-mixture models or reference-based approach)		Conducted MNAR SA, and used an appropriate method (PMM, etc.).	Did not conduct MNAR SA	
D1. Report number of participants with cost and outcome data, by arm and time-point	Report number (or %) of complete or missing data. Split at least by effectiveness vs. cost, time point (when applicable), and arm		Reported missing data by endpoint and arm, but not by time point.	Do not have to be all at the same time (split by endpoint + time + arm), can be three separate table/texts.
D2. Report possible reasons for non-response, and baseline predictors of missing values	Mention something about main reason for the missing data, OR Explore factors associated with it.	Comment on why missing data (e.g., "because patients were too ill"). Or explore baseline factors associated with missingness	No mention of reasons for MD in the CE section.	Have to be specific to the CE missing data, or clearly mentioning something like "reasons for MD are discussed in clinical analysis section ..."
D3. Describe methods used, and underlying missing data assumptions	Clearly state the method used to address missing data, AND the underlying assumption.		No report of missing data assumption or method used	
Draw overall conclusion in light of the different results and the plausibility of the respective assumptions	Conduct sensitivity analyses, and interpret results appropriately.	Did MNAR SA and appropriate conclusion.	– Did not conduct sensitivity analyses – Conducted sensitivity analyses, but no comment/conclusion – Did MI and CC and only say "results did not change/robust to missing data"	

Chapter 4

Stakeholders' seminars

4.1 Preamble

Chapter 3 confirmed the importance of the missing data issue in trial-based CEA, and that MNAR sensitivity analyses were rarely conducted. We now need to better understand the reasons behind this gap between recommendations and practice. In addition, this thesis aims to develop relevant and accessible tools and recommendations, and it is therefore critical to understand the viewpoints and expectations of those conducting or using CEA. The best way to gather this information appeared to hold interactive seminars, encouraging discussion around MNAR issues between participants using semi-structured questionnaires.

4.2 Aims

To deepen our understanding of practitioners' views on MNAR missing data in CEA, we organised two interactive seminars with a number of Stakeholders (i.e. those conducting or using CEA, including health economists, decision modellers and members of decision committees). The objectives of the seminars were to:

- understand the current awareness of, and experience with, MNAR missing data in CEA;
- explore awareness about the potential for, and implications of, conducting MNAR sensitivity analyses in CEA and using these to inform decisions on resource allocation;
- identify current barriers to more widespread use of MNAR sensitivity analyses; and
- identify participants' views most relevant ways to conduct and report these sensitivity analyses.

4.3 Methods

4.3.1 Seminars' organisation and data collection

Interactive seminars were used to collect this information.

Two of these interactive seminars took place, one in the University of York Centre for Health Economics, in April 2016, and the other at the London School of Hygiene and Tropical Medicine (LSHTM) in June 2016.

The participants were recruited by advertising the seminars using the relevant mailing lists. They were informed of the seminars' structure, a presentation followed by discussion groups, and that participation in either part was entirely voluntary. A consent form (Appendix B) was circulated to inform of the data collection, and to authorise the use of their anonymised responses for research purpose.

Each seminar started with a 45 minutes presentation (see Section 4.3.2), introducing the problem of informative missing data in CEA, and outlining different approaches to conduct sensitivity analyses, illustrated with examples. The presentation was then followed by 30 minutes small group discussions. The attendees were split into groups of two or three, and asked to:

- complete a consent form (Appendix B);
- complete a short survey, about their role and experience with missing data (Appendix C);
- discuss one or more 'theme' questionnaires within their group (Appendix D).

Discussion questionnaires were organised around four themes:

- methodological approaches to deal with MNAR data;
- presentation of sensitivity analysis results;
- elicitation of expert opinion to inform sensitivity analyses; and
- barriers to MNAR methods implementation.

Each theme questionnaire included a list of questions to be discussed within group, and participants were asked to report their answers on the paper questionnaires, which were then collected at the end. The participants' answers to the questionnaires, and other comments discussed during the seminars, were then compiled in a Word document. See Appendices C and D, with the participants's answers in blue. Personal annotations were then added (in green), highlighting the key findings and implications for the remaining of the thesis.

4.3.2 Presentation

The seminars started with a 45 minutes presentation, introducing the problem of informative missing data in CEA, and presenting possible approaches to conduct and report MNAR sensitivity analyses (see slides in Appendix A). To exemplify the various approaches, we used data from the 10TT weight loss trial, described in more details in Chapter 5 (see also Beeken et al. [9] and Patel et al. [84]).

The presentation focussed on key approaches which appeared potentially relevant in the CEA context, based on their ability to address trial-based CEA complexities (see Section 2.1.2), their availability (i.e. methodological developments and software), and the accessibility of their assumptions to the different stakeholders. Three of them were based on pattern-mixture models with a δ parameter capturing how the missing-data distribution differed from the MAR distribution (see 2.2.3). The first approach 'back-of-the-envelope' or 'simple approximation' approach arithmetically calculated the MNAR estimates, using the MAR estimates and the departure parameter δ, weighted by the proportion of missing data (see Section 2.2.3). The second, 'MI + MNAR delta', referred to shifting MAR-MI imputed data by δ, as described in detail in Chapter 5. The third one was similar, but instead of shifting the data by a fixed value, a stochastic distribution was used. Finally, we discussed 'reference-based' imputation, introduced in Section 2.2.3 and covered in detail in Chapter 6. Because this approach had not yet been implemented in CEA, we only introduced the principles for the clinical effectiveness outcome.

We illustrated the different approaches, and discussed alternative ways to report the sensitivity analyses results. This presentation aimed to provide sufficient background knowledge and information to facilitate the ensuing small group discussions.

4.4 Results

4.4.1 Number of participants and response rate

Twenty participants attended the seminar and participated in the small group discussion, seven in York and thirteen at LSHTM. Not all participants completed the short survey, with nearly half of the participants at LSHTM (6/13) not completing (or returning) the survey form. The reason is not clear, but participants may have directly started their small group discussions, or completed a single survey for their group. Each of the four themes were discussed by three to seven different groups of participants, with 18 theme questionnaires completed in total (see Appendix D).

4.4.2 Participants' survey

Participants' characteristics based on the survey can be seen in Appendix C. Most considered themselves as 'modellers' and a minority as 'trial-based analysts'. Five sat on, or reviewed evidence for, decision-making committees (e.g. NICE Technology Appraisal Committee).

Three quarters (10/14) reported facing missing data issues 'regularly' or 'very often'. Every participant (14/14) reported having already thought about the issue of informative missing data, and a majority (8/13) reported having at least once faced a situation where they thought something more could have been done in terms of missing data sensitivity analysis. These findings support that missing data, and MNAR SA, are perceived as relevant issues by health economists.

4.4.3 MNAR approaches

The participant's answers to the different themes questionnaires are reported in Appendix D. The first theme was about the different methods to implement MNAR sensitivity analyses.

None of the sensitivity analysis methods discussed particularly stood out as most appealing. Participants recognised the pros and cons in all of the approaches presented. Technically sound methods appeared more appealing, but they needed to remain accessible for the analysts, and for the decision-makers. Participants listed perceived advantages and limitations for the different methods discussed. For instance, reference-based imputation was well received, as more intuitive and not making assumptions about δ, but it was recognised as more complex to implement. None of the methods appeared clearly irrelevant, but it seemed there was greater potential in focussing on the more 'advanced' methods (e.g. 'MI + MNAR delta' and 'reference-based') than the 'simple approximation' approach.

Questions about familiarity with the different statistical frameworks indicated that multiple imputation appeared now relatively well established, at least among analysts. By contrast, both analysts and users appeared less familiar with Bayesian methods.

4.4.4 Reporting

Opinion differed on the most appropriate way to report sensitivity analyses results. For example, some participants really liked the 'contour plot' (reporting results across a range of sensitivity parameters values, see for example Chapter 5, Figure 6), but others found it confusing. What appeared the most intuitive and accepted approach generally was the reporting of different MNAR scenarios on a CEAC (see, for example, Chapter 5, Figure 4). An interesting suggestion was to present the INB for different values of δ, which was developed and included in the next Chapter.

Overall it seems the appeal of the different reporting approaches depends on the context and preferences, but that the overlaying CEAC may be the most intuitive generally. In the following chapters we will therefore focus on CEAC, but also provide other presentation options.

4.4.5 Elicitation

This questionnaire responses highlighted the challenges of elicitation. In particular it appears challenging for clinicians to give informed views about parameters on a utility scale. Reference-based ideas appeared more intuitive, but were also mentioned as potentially difficult and not always implementable without a suitable control group. The list of 'experts' that should be involved in the elicitation included clinicians, patients representatives, health economists and statisticians.

4.4.6 Barriers

In this themed discussion participants reflected on what they perceived to be the main reasons for sensitivity analyses not being more commonly conducted. As with the other themes, the answers varied between participants. The three key factors that emerged were: "would not know how to do"; "do not have the software"; and "would take too much time". The first two items indicate the need for guidance and software code to conduct this analysis, as we will try to address in Chapters 5 and 6. The third item indicates the importance of accessible methods, that are quick and easy to implement. Again, this was kept in mind in Chapters 5 and 6.

Interestingly, participants did not seem to think MNAR sensitivity analyses were not conducted because they were perceived to be irrelevant, but more because of the challenges in conducting these analyses. They also mentioned additional reasons, such as investigators not being very keen as they can increase the uncertainty, or because it was not common practice.

The main practical challenges to implement the analysis for the respondents appeared to be the lack of software, and the elicitation of the sensitivity parameters (δ).

The implications of (MNAR) missing data in model-based CEA was also discussed. It seems that the issue of missing data is commonly ignored, or that when analysts have access to the trial data, multiple-imputation is often used (under MAR). However, modellers appeared potentially interested in conducting MNAR sensitivity analyses. The issue of MNAR sensitivity analyses in model-based CEA is discussed further in Chapter 7 (sections 7.3.1 and 7.7.3).

We also asked about the popularity of different statistical software, and Stata was the most commonly used (followed by R and Excel). Our review of the HTA trials (Chapter 3) also indicated that Stata

was the dominant software, used in 75% of the CEA (data not shown). The tools in Chapters 5 and 6 were therefore developed in Stata.

4.5 Conclusions

Participants engaged with the topic and enjoyed the interactive discussion groups. The summaries of the discussions provided very valuable information for the PhD.

First, it confirmed the importance of the missing data issue in the CEA context. Second, it raised numerous points which have helped shape this thesis. Most usefully, it gave me further information on the CEA context and the viewpoints of analysts, which will help ensure the recommendations and tools developed are relevant in practice.

While it was 'reassuring' to find that MNAR sensitivity analysis was indeed a highly relevant issue, progress appear hampered by a lack of knowledge and software. This is in line with the hypothesis advanced by Bell and al. [62] in the clinical effectiveness context. Further, opinions differed between participants, suggesting scope for different analysis and reporting approaches, depending on the context and preferences. This highlights the difficulty of suggesting a one-size-fits-all approach for MNAR sensitivity analyses in CEA. In the remaining chapters, we will therefore focus on two different approaches, but we will also indicate variations around these approaches, e.g. in terms of reporting, or of choosing the sensitivity analysis scenarios.

The main limitation was that these 'focus group' discussions were held in leading CEA research centres, with a limited sample of participants. These may not be representative of the CEA analysts and users more generally. For example, they may be more aware of missing data issues and methods. Nevertheless, these views both complement and reinforce the information from Chapter 3, and are extremely valuable in informing the research agenda for this thesis.

Based on these discussions, and on Chapter 3, the approach that appeared to have the most potential, in terms of relevance, accessibility, and flexibility, was the 'MI + MNAR delta' approach. This will therefore be the focus of Chapter 5.

The reference-based was also very well received. There was wide agreement it seemed more intuitive than eliciting quantitative parameters from experts, and easier to understand for decision makers. Indeed, on of the key challenge with elicitation is to know whether the elicited information is clinically and empirically relevant. The implementation appeared the main drawback of this method, which we will aim to address in Chapter 6.

Chapter 5

Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial

5.1 Preamble

The Stakeholders' seminars strongly suggested that key barriers to more widespread conduct of MNAR sensitivity analyses are the lack of knowledge on how to do in practice and the lack of software code. This chapter is an article published in PharmacoEconomics (Springer) which aims to address these key issues.

After explaining the importance of the MNAR issue, and different approaches to conduct sensitivity analyses under MNAR, the article focuses on one approach that, based on Chapters 2 to 4, appeared particularly accessible and relevant in the CEA context. This approach, pattern-mixture models after multiple-imputation, is based on modifying data multiply-imputed under MAR to reflect plausible MNAR mechanisms. Widespread use of multiple-imputation in trial-based CEA, as identified in Chapters 3 and 4, makes it an ideal vehicle to implement these sensitivity analyses.

The article uses the 10TT weight-loss trial as an example to illustrate the approach. It starts by providing a relatively simple example of MNAR sensitivity analysis, showing the analysis step by step and the corresponding Stata code. It then covers further key topics for practitioners, including how to decide on the relevant sensitivity parameters, and how to report the sensitivity analyses results.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Baptiste Leurent
Principal Supervisor	Prof. James Carpenter
Thesis Title	Cost-effectiveness analysis with informative missing data: tools and strategies

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	PharmacoEconomics		
When was the work published?	20 April 2018		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	BL took responsibility for the article overall. Conducted the statistical analysis and developed the implementation code. Wrote the first draft, and final version of the article.

Student Signature: [redacted] **Date:** 05/09/18

Supervisor Signature: [redacted] **Date:** 5 Sept 2018

PharmacoEconomics (2018) 36:889–901
https://doi.org/10.1007/s40273-018-0650-5

PRACTICAL APPLICATION

Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis: A Tutorial

Baptiste Leurent[1] · Manuel Gomes[2] · Rita Faria[3] · Stephen Morris[2] · Richard Grieve[4] · James R. Carpenter[1,5]

Published online: 20 April 2018

Abstract Cost-effectiveness analyses (CEA) of randomised controlled trials are a key source of information for health care decision makers. Missing data are, however, a common issue that can seriously undermine their validity. A major concern is that the chance of data being missing may be directly linked to the unobserved value itself [missing not at random (MNAR)]. For example, patients with poorer health may be less likely to complete quality-of-life questionnaires. However, the extent to which this occurs cannot be ascertained from the data at hand. Guidelines recommend conducting sensitivity analyses to assess the robustness of conclusions to plausible MNAR assumptions, but this is rarely done in practice, possibly because of a lack of practical guidance. This tutorial aims to address this by presenting an accessible framework and practical guidance for conducting sensitivity analysis for MNAR data in trial-based CEA. We review some of the methods for conducting sensitivity analysis, but focus on one particularly accessible approach, where the data are multiply-imputed and then modified to reflect plausible MNAR scenarios. We illustrate the implementation of this approach on a weight-loss trial, providing the software code. We then explore further issues around its use in practice.

Key Points for Decision Makers

Cost-effectiveness analysis of randomised trials with missing data should assess the robustness of their findings to possible departures from the missing at random assumption.

Multiple imputation provides a flexible and accessible framework to conduct these sensitivity analyses.

Sensitivity analysis results should be reported in a transparent way, allowing decision-makers to assess the plausibility of their respective assumptions.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40273-018-0650-5) contains supplementary material, which is available to authorized users.

✉ Baptiste Leurent
baptiste.leurent@lshtm.ac.uk

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

2 Department of Applied Health Research, University College London, London, UK

3 Centre for Health Economics, University of York, York, UK

4 Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

5 MRC Clinical Trials Unit at University College London, London, UK

1 Introduction

Cost-effectiveness analyses (CEA) of randomised trials are an important source of information to help decide which health care programmes to provide. A common issue is that there may be missing data, for example, because patients withdrew from the trials or failed to respond to study questionnaires, and this could result in biased findings and, ultimately, wrong decisions being taken.

There is now greater awareness that simple approaches, such as discarding the participants with missing data, are generally unsatisfactory [1–5]. The benefits of methods that make use of all the available data and offer valid inference under 'missing at random' (MAR) assumptions are now well recognised, and recent years have seen an increase in the use of such methods in CEA, in particular multiple imputation (MI) [6, 7].

A key concern, however, is that conditional on the observed data, the probability of cost-effectiveness data being missing may still depend on the underlying unobserved values, i.e. data may be 'missing not at random' (MNAR). For example, after adjusting for observed prognostic factors, the chances of completing quality-of-life questionnaires may depend on the patient's (unobserved) quality-of-life status. This raises particular challenges to cost-effectiveness inferences because the analyst cannot formally choose between MAR and MNAR given the data at hand. Therefore, conducting sensitivity analyses to assess whether conclusions are robust to plausible departures from MAR is widely recommended [1, 2, 8–10], and these are particularly relevant for CEA which usually rely on patient-reported outcomes. However, a recent review has found that, in practice, cost-effectiveness studies rarely conduct such a sensitivity analysis [7]. We discussed this issue with stakeholders (academics from the University of York and the London School of Hygiene and Tropical Medicine analysing or reviewing cost-effectiveness evidence for health care decision making), and an important barrier that was identified was the lack of software tools and guidance to conduct these analyses.

This tutorial paper aims to address this gap by presenting an accessible framework and practical guidance to conduct sensitivity analysis for trial-based CEA with missing data. This builds on previous guidance on missing data in CEA [1, 3, 4], by focusing on sensitivity analysis approaches to address MNAR. This paper introduces different approaches to MNAR analyses, but focuses particularly on the implementation of pattern-mixture models using MI [11] as it was highlighted as the most accessible and flexible approach during our discussions with stakeholders. This tutorial assumes familiarity with the conduct of MI (under the MAR assumption), which has been covered elsewhere [3, 4, 12, 13].

The remaining sections of this paper are organised as follows. Section 2 provides a brief overview of the different approaches for MNAR analysis. Section 3 illustrates a framework for MNAR sensitivity analysis, based on a weight-loss trial, the Ten Top Tips (10TT) study. Section 4 discusses possible extensions to the proposed approach and further considerations for implementing it in practice.

2 Overview of Missing Not at Random (MNAR) Analysis Methods

2.1 Missing Data Mechanisms

The classification of the missing data mechanisms proposed by Little and Rubin [14] provides a useful context. Data are said to be missing 'completely at random' (MCAR), when missingness occurs for reasons unrelated to the analysis question, and hence independent of the variables of interest. In this case, the observed data are representative of the overall data and analysing the participants with complete data will give valid results. A less restrictive assumption is that the data are 'missing at random' (MAR), so that the probability of a value being missing may be dependent on observed data (e.g. intervention group, or participants' age), but—given the observed data—independent of the underlying value itself. Finally, if, after taking into account the observed variables, the chance of observing the data is still associated with its value (for example, if, after controlling for preceding data, a patient is less likely to complete a health questionnaire when in poorer health), the data are said to be 'missing not at random' (MNAR, also called 'informative', or 'non-ignorable' missingness).

When missing data are MAR, valid conclusions can be drawn from the data available using an appropriate approach, such as MI [15]. MI has been widely recommended as a flexible, practical approach to handle missing data in CEA studies [1, 3–5, 12], and its uptake has been steadily increasing [6, 7]. The idea of MI follows from regression imputation (using the observed data to predict the missing values), but appropriately takes into account the uncertainty in the imputed values. To achieve this, missing observations are replaced by plausible values drawn from an appropriate predictive distribution of the missing values given the observed data. To reflect the fact that imputed values are estimated rather than known, and hence uncertain, this process is repeated several times to create several complete datasets. The analysis model is then fitted to each 'complete' dataset, and the results are combined for inference using Rubin's MI rules [15], which recognise the uncertainty both within imputations (sampling uncertainty) and between imputations (uncertainty due to missing data).

Analysis under MNAR is more challenging, as it implies some relevant information is unobserved, and it requires additional untestable assumptions to proceed with the analysis. This naturally makes the MAR assumption the typical starting point for the primary analysis of clinical trials [16, 17]. However, because we cannot determine the true missing data mechanism, sensitivity analyses should

be conducted in order to assess whether conclusions are robust to plausible departures from the MAR assumption [1, 2, 8–10].

2.2 MNAR Modelling Frameworks

Various approaches have been proposed in the statistical literature to conduct analysis under MNAR. These vary according to how they formulate the MNAR model, how they fit this model, and how the unobserved parameters are informed and results reported as part of a sensitivity analysis strategy. Here, we briefly review some of the main MNAR modelling frameworks; for a more comprehensive description, see Molenberghs et al. [11]. There are two main ways to model possible departure from MAR: selection models and pattern-mixture models.

Selection models specify the mechanism by which the data are observed (or 'selected') as a function of the underlying data values [15, 18]. For example, 'for each decrease of 0.1 in quality of life, the chance of being missing doubles' formulates the MNAR problem in selection model terms. Selection models were commonly used in early work on informative missing data; an example in econometrics is Heckman's selection model [19], which is used to address selection bias. They have the attractive feature that the missing data model can be directly incorporated into the analysis model, for example, using an inverse probability weighting approach [18, 20] or numerical integration [21]. However, selection models make untestable assumptions about the conditional distribution of the unobserved data, and results can be very sensitive to departure from these assumptions, as has been shown elsewhere [14, 22–24]. This limitation is particularly relevant for CEA studies, as the cost and effectiveness endpoints tend to be difficult to parametrise. Another disadvantage is that selection models formulate sensitivity analysis in a way that is not readily interpretable. For example, a typical sensitivity parameter is the (log-)odds ratio of how a unit change in the partially observed outcome affects the chances of observing the data. This specification makes the elicitation of such parameters challenging, as well as the interpretation and communication of the sensitivity analysis results.

Pattern-mixture models, on the other hand, formulate the MNAR problem in terms of the different distributions between the missing and observed data. The overall distribution of a variable is seen as a mixture of the distribution of the observed and the distribution of the missing values ('pattern-mixture') [18, 25]. For example, 'participants with missing data have a 0.1 lower quality of life than those observed' corresponds to a pattern-mixture formulation. Pattern-mixture models have received increasing attention over time [26], a key advantage being that they rely on more easily interpretable parameters [3, 18, 27–29]—such as the mean difference between missing and observed data—and have therefore been favoured in the context of clinical trial sensitivity analysis [30, 31]. Different approaches can be used to formulate and analyse pattern-mixture models, as we will see in the next section.

Other forms of MNAR modelling have also been proposed, but these can be seen as special cases of selection or pattern-mixture models. In shared-parameter models, the outcome and the missingness are linked through a latent (unobserved) variable [32]. They have been particularly used in the context of structural equation modelling. Another approach which is gaining interest for use in longitudinal trials is 'reference-based' or 'controlled' imputation, where missing data are assumed to follow a distribution borrowed from another trial arm [33]. This approach is yet to be explored in the CEA setting.

While any of the methods above would allow an appropriate assessment of departures from MAR, we will focus on the pattern-mixture approach in the remainder of this paper because (1) it allows for more interpretable parameters, hence making this approach more accessible and transparent; (2) it seems to be the main approach currently used in clinical trial sensitivity analysis [7, 34]; (3) our discussion with stakeholders confirmed this approach was also appealing in the CEA context; and (4) pattern-mixture models can be easily implemented using standard missing data methods, such as MI, and build naturally on the MAR analysis, as we will see below.

2.3 Sensitivity Analysis with Pattern-Mixture Models

An approach for MNAR sensitivity analysis that has often been suggested—under various forms—is to perform a pattern-mixture model with a parameter capturing how the distribution of the missing values Y_{miss} could differ from the conditional distribution based on the observed data Y_{obs} [15, 18, 30, 35]. This can be done, for example, by using an 'offset' parameter δ (delta) representing the average difference between the missing and observed values ($Y_{\text{miss}} = Y_{\text{obs}} + \delta$). An alternative modification is to use a multiplicative 'scale' parameter c, so that $Y_{\text{miss}} = Y_{\text{obs}} \times c$. For example, missing values could be assumed to be 10% lower than those observed, or $c = 0.9$. Figure 1 illustrates an example of such modelling with a rescaling parameter. In that example, a participant who drops out from the trial is assumed to have on average a 10% lower quality of life compared to a participant with similar characteristics who remained in the trial. Note that this parameter is not derived from the data, but is used to express one possible assumption about the (unknown) missing data mechanism.

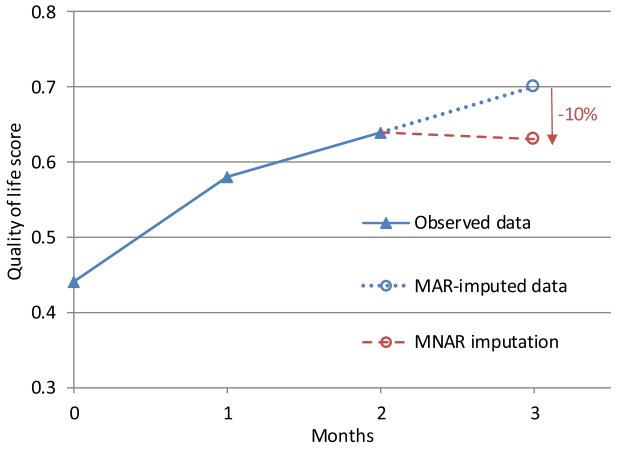


Fig. 1 Example of pattern-mixture assumptions with rescaling. Quality-of-life score over time for a trial participant, where missing data are assumed to be 10% lower ($c = 0.9$) than would have been imputed under a missing at random assumption. *MAR* missing at random, *MNAR* missing not at random

Sensitivity analyses are then typically conducted over a range of plausible values for this parameter, assessing how different assumptions could result in different findings. Several approaches can be used to inform the values of the parameter in practice, and these are discussed further in Sect. 4.3. We also discuss in Sect. 4.2 alternative parametrisations that can be used to capture how missing and observed data might differ.

Several approaches have been proposed to fit pattern-mixture models, for example, within a Bayesian framework [28, 36] or as an arithmetic function of the observed estimates and using bootstrap or sandwich estimators to derive the standard errors [18, 28]. But a particularly convenient and flexible framework to fit these models is MI [11, 15, 26, 37]. An approach commonly adopted in practice consists of simply modifying multiply-imputed data to reflect possible departures from the MAR assumption [3, 7, 16, 38]. It involves the following steps:

1. Use MI to impute the missing values under an MAR assumption.
2. Modify the MAR-imputed data to reflect a range of plausible MNAR scenarios, for example, by multiplying the imputed values by c, or by adding δ.
3. Analyse the resulting dataset as one would a usual multiply-imputed dataset, fitting the analysis model to each imputed dataset and combining the results using Rubin's rules.

This approach is straightforward to implement in any statistical software, and allows the effect of different MNAR mechanisms on the conclusion to be easily assessed, as we will illustrate in the next section.

3 Illustrative Application

3.1 The Ten Top Tips (10TT) Trial

3.1.1 Overview of the Trial and Cost-Effectiveness Analysis

The 10TT trial was a two-arm, individually randomised, controlled trial of a weight-loss intervention for obese adults attending general practices in the UK [39]. The intervention comprised self-help material delivered by a practice nurse, providing the patients with a set of ten simple weight-control behaviours, with strategies to make them habitual. The participants randomised to the control arm received care as usual from their general practices.

The primary trial outcome was weight loss at 3 months, but participants were followed for 2 years to assess longer-term outcomes and cost-effectiveness. Health-related quality of life (HRQoL) was measured by EQ-5D-3L questionnaires [40, 41] completed during study visits at baseline and 3, 6, 12, 18 and 24 months, and quality-adjusted life years (QALYs) were derived by the 'area under the curve', combining both time and utilities [10]. Total costs were measured from the National Health Service (NHS) perspective over the 2-year study period and based on the intervention costs and the health resource use data collected from the practice records at the end of the trial. More details on the trial and CEA can be found in the respective publications [39, 42, 43].

3.1.2 Missing Data

The trial recruited 537 participants, but only 313 (58%) completed the last follow-up at 2 years. Missing data were a major challenge for the CEA because only 31% of randomised participants had complete HRQoL and cost data. Missing data were mostly driven by missing EQ-5D data, from participants who had either withdrawn from the trial (76% of the missing HRQoL) or missed a follow-up appointment (24%). Resource use data were derived from the general practitioner records and were complete for 73% of the participants (all the health care data were missing for the remaining 27%). Details of the missing data by arm are shown in Fig. 2. Although non-significant, missing data appeared to be more common in the intervention arm (27 vs 34% of complete cases, p value = 0.075).

The primary CEA of the trial [43] was conducted under the MAR assumption, using MI to impute the missing cost and HRQoL values. It is, however, recognised in weight-loss trials that participants who drop out could be those with poorer outcomes [44]. This means that the chance of

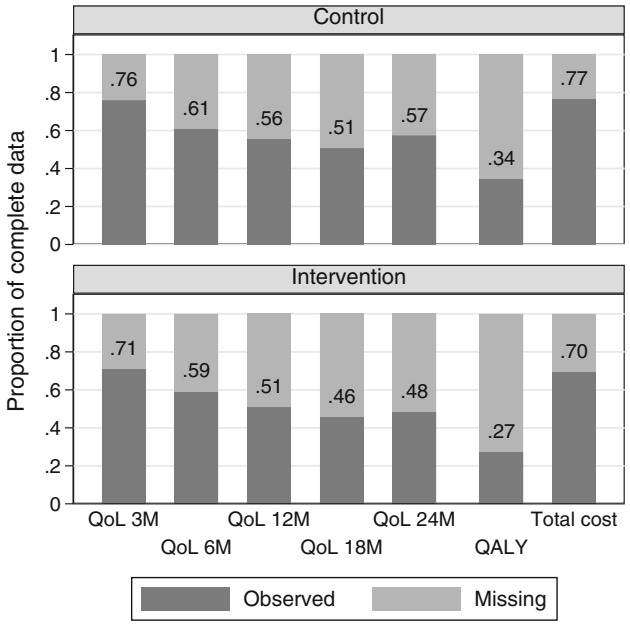


Fig. 2 Proportion of complete cost-effectiveness outcomes in 10TT trial, by arm: QoL at each time point (3, 6, 12, 18, and 24 months), overall QALY, and total cost. $N = 270$ in control arm and 267 in intervention arm. *QALY* quality-adjusted life year, *QoL* quality of life, *10TT* Ten Top Tips

observing endpoints such as weight loss or HRQoL could be dependent on their actual value, i.e. data are likely to be MNAR. It is therefore important to assess the cost-effectiveness results under different assumptions regarding the missing data, including plausible MNAR mechanisms, as we will illustrate in Sect. 3.2.

3.1.3 Cost-Effectiveness Analysis Methods

The CEA conducted in this tutorial follows the main characteristics of the methods used for the trial's primary CEA [43], with some simplifications made to allow a clear focus on the sensitivity analysis. Details of the analysis variables are presented in Online Appendix 1 [see the electronic supplementary material (ESM)]. Effectiveness was measured in QALYs, and costs were captured by the total health care use over the trial period (Sect. 3.1.1), as derived for the primary analysis [43]. A discount rate of 3.5% per year was applied to both cost and effect.

Results are presented in terms of incremental cost, incremental QALYs and incremental net monetary benefit (INMB) at a cost-effectiveness threshold of £20,000 per QALY. These were estimated alongside their 95% confidence intervals (CIs) using non-adjusted linear regression, comparing the 10TT arm to the control arm. Non-parametric bootstrap [45] was also used to produce the cost-effectiveness plane [46], representing the uncertainty in incremental cost and effect estimates, and the cost-effectiveness acceptability curve (CEAC) [47], representing the probability of 10TT being cost-effective at different thresholds. We focus on INMB rather than the incremental cost-effectiveness ratio (ICER) as the intervention was cost-saving. All the analyses were conducted in Stata version 15 [48].

3.2 Sensitivity Analysis Example

In this section, we use the 10TT trial to illustrate MNAR sensitivity analyses using a pattern-mixture approach following MI, as described at the end of Sect. 2.3.

3.2.1 MNAR Scenarios Explored

Several approaches can be used to decide on the relevant MNAR scenarios for the sensitivity analyses, and this is discussed further in Sect. 4.3. In this example, we considered that the missing HRQoL data may be MNAR, while the MAR assumption is likely to hold for the missing cost data (MNAR costs are discussed in Sect. 4.1). It was postulated that patients who failed to complete an EQ-5D questionnaire at a specific follow-up assessment were likely to have been in relatively poorer health (Sect. 3.1.2). More specifically, we assumed patients' HRQoL could be up to 10% lower ($c = 0.9$), compared to the MAR setting ($c = 1$). This sensitivity parameter c was allowed to differ by arm, with up to a 5% difference between the two arms (this reflects that the missing data mechanism may not be the same in the two arms, but that it is unlikely to be perfect MAR in one arm and strong MNAR in the other). This resulted in seven different MNAR scenarios, with $c = 1.0$, 0.95, or 0.9 for either arm (Table 1).

3.2.2 Implementation of the Analysis in Stata

The annotated Stata code to conduct the analysis is provided in Online Appendix 2 (see the ESM), and the dataset is described in Online Appendix 1.

Step 1. Performing Multiple Imputation

The first step of the analysis is to conduct standard MI (under an MAR assumption), to 'fill in' the variables with missing data. The missing HRQoL at each time point and total costs were imputed stratified by arm, using a linear model based on each other, and baseline characteristics (age, sex, study centre, weight, body mass index and baseline HRQoL). We conducted MI by chained equations, using predictive-mean matching, and created 50 imputations. Note that alternative MI approaches, for example, linear regression, would not affect the proposed sensitivity

Table 1 Cost-effectiveness of 10TT under different MNAR assumptions for missing quality-of-life data

Scenario number	MNAR rescaling parameters[a]		Incremental cost[b] (£) [95% CI]	Incremental QALYs [95% CI]	INMB[c] (£) [95% CI]	Probability cost-effective[c] (%)
	$c_{control}$	c_{10TT}				
1 (MAR)	1	1	− 35 [− 504 to 434]	− 0.004 [− 0.074 to 0.066]	− 49 [− 1632 to 1534]	48
2	1	0.95	− 35 [− 504 to 434]	− 0.037 [− 0.107 to 0.032]	− 713 [− 2280 to 853]	19
3	0.95	1	− 35 [− 504 to 434]	0.026 [− 0.044 to 0.095]	550 [− 1022 to 2121]	75
4	0.95	0.95	− 35 [− 504 to 434]	−0.008 [− 0.076 to 0.061]	− 115 [− 1670 to 1440]	44
5	0.95	0.90	− 35 [− 504 to 434]	− 0.041 [− 0.109 to 0.027]	− 780 [− 2321 to 762]	16
6	0.90	0.95	− 35 [− 504 to 434]	0.022 [− 0.046 to 0.091]	484 [− 1063 to 2030]	73
7	0.90	0.90	− 35 [− 504 to 434]	− 0.011 [− 0.078 to 0.057]	− 181 [− 1714 to 1352]	41

All results are based on imputed data and comparing the 10TT arm to the control arm (n = 537). For participants with complete cost and effectiveness data (n = 166; 31%), the observed incremental cost was − £65 [95% CI − 924 to 794], incremental QALYs was − 0.040 [− 0.169 to 0.088], INMB was − £741 [− 3645 to 2163], and probability cost-effective was 31%

CI confidence interval, *INMB* incremental net monetary benefit, *MAR* missing at random, *MNAR* missing not at random, *QALY* quality-adjusted life year, *10TT* Ten Top Tips

[a]How missing quality-of-life data are assumed to differ from the MAR-imputed values. $c_{control}$ = 0.9 means that all imputed quality-of-life values in the control arm have been reduced by 10%

[b]Missing costs assumed to be MAR in all scenarios

[c]At a cost-effectiveness threshold of £20,000/QALY

analysis strategy. More detailed guidance on conducting MI in Stata is provided elsewhere [3, 13, 49].

Step 2. Modifying Imputed Data

To obtain the imputed data under MNAR, we simply need to multiply each MAR-imputed value by c. For example:

```
replace qol_3=qol_3*0.9 if miss_qol_3==1 & arm==0
```

will multiply the imputed values of qol_3 in the control arm by 0.9.

Different versions of the modification could be implemented at this stage (see Sect. 4.2), for example, by alternatively considering an 'offset' additive parameter d:

```
replace qol_3=qol_3 + d if miss_qol_3==1 & arm==0
```

This can be done in turn for each of the scenarios, or storing each of the scenario parameters in a table (matrix) allows Stata to execute this in one step, using a loop. The modified data can then be saved in a single large dataset to facilitate the remaining steps.

Step 3. Analysing the MNAR Dataset

The CEA analysis is then applied as usual to each of the MNAR multiply-imputed datasets. To estimate the incremental costs, QALYs and net monetary benefit and their 95% CIs, we have used the 'mi estimate' command, which fits the analysis model on each of the imputed datasets, then combines the results using Rubin's rules [15]. We have also used a non-parametric bootstrap approach to produce the cost-effectiveness plane and the CEAC, with the implementation described in Online Appendix 2 (see the ESM). Further guidance on the analysis of multiply-imputed cost-effectiveness data can be found elsewhere [1, 3, 4, 12].

Step 4. Reporting

Clear reporting of the sensitivity analysis results is key to ensure their implications are well understood. We recommend a table which presents the summary findings for each scenario (Table 1). Figure 3, which plots the cost-effectiveness plane for the different MNAR scenarios is also useful to understand the effect of each MNAR assumption, as discussed in the next section. Our discussions with stakeholders indicated that the most intuitive way to summarise the findings was probably overlaying CEACs, showing the probability of the intervention being cost-effective at different thresholds, for each MNAR scenario (Fig. 4). Alternative presentations of the sensitivity analysis results are discussed in Sect. 4.5.

3.2.3 Results

The 10TT CEA results under the different missing data scenarios are reported in Table 1, Figs. 3 and 4. In particular, the CEAC (Fig. 4) shows that the probability of 10TT being cost-effective remains relatively stable when

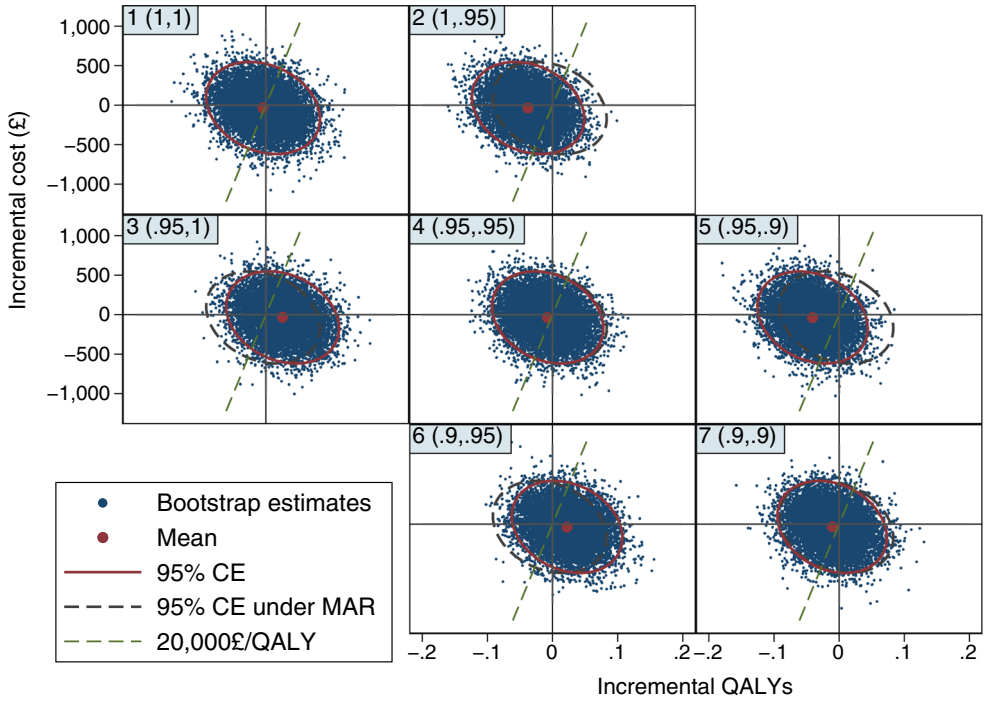


Fig. 3 Cost-effectiveness planes under different MNAR assumptions. Headings in top-left corner indicate the scenario number and the MNAR rescaling parameters (c_{control}, $c_{10\text{TT}}$). For example, (0.9, 0.9): imputed quality-of-life values have been reduced by 10% in both arms. Each plane is based on 10,000 bootstrap replicates, from 50 imputed datasets. 95% CEs are shown (solid ellipse), alongside the 95% CEs under MAR (scenario 1) as a reference (dashed ellipse). Dashed lines indicate the cost-effectiveness threshold of £20,000 per QALY. *CE* confidence ellipse, *MAR* missing at random, *MNAR* missing not at random, *QALY* quality-adjusted life year, *10TT* Ten Top Tips

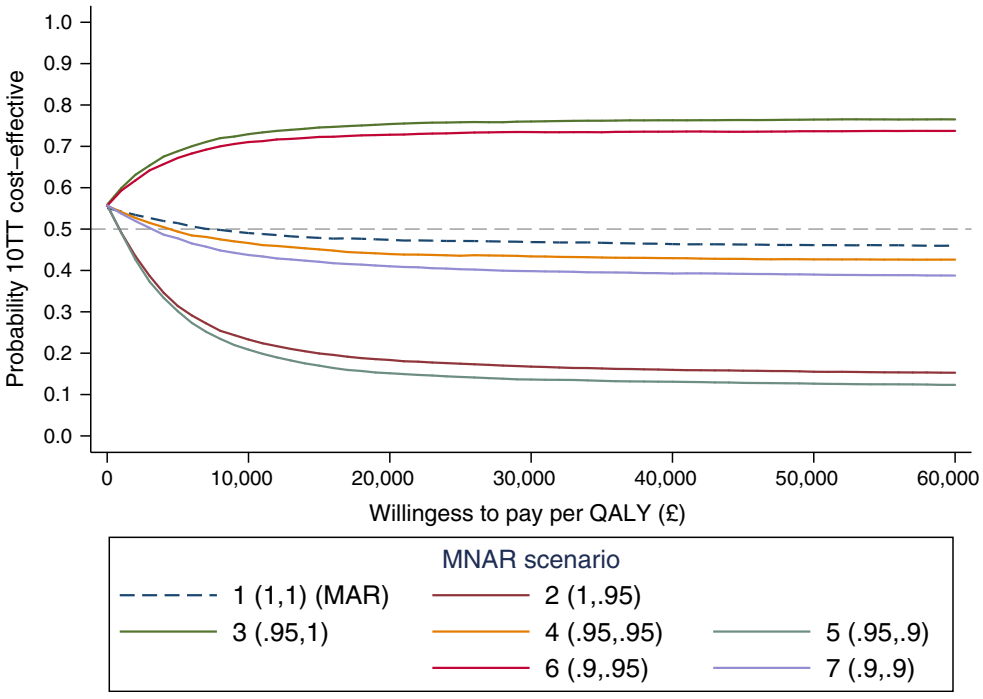


Fig. 4 Cost-effectiveness acceptability curves under different MNAR assumptions. Legend indicates the scenario number and the MNAR rescaling parameters (c_{control}, $c_{10\text{TT}}$). For example, (0.9,0.9): imputed quality-of-life values have been reduced by 10% in both arms. *MAR* missing at random, *MNAR* missing not at random, *QALY* quality-adjusted life year, *10TT* Ten Top Tips

MAR departures are assumed to be the same across randomised arms (scenarios 1, 4 and 7). This is also seen in Table 1, where the alternative departures from MAR had little effect on the incremental QALYs in these scenarios. This will usually be the case when the missing data pattern is broadly similar across treatment arms, as the MNAR bias applies roughly equally to each arm and cancels out in the treatment comparison.

As we move through the other scenarios, however, 10TT alternates between being cost-effective and not depending on which arm is assumed to have a stronger MNAR mechanism. For example, 10TT appear unlikely to be cost-effective when we assumed stronger MNAR (lower c) for the treatment arm, with a probability of being cost-effective around 0.2 at £20,000 per QALY. Table 1 also shows how the incremental QALYs vary across the different scenarios, while the width of the 95% CI remains relatively similar. Since the magnitude of the incremental QALYs was relatively small, different missing data mechanisms across arms led to substantially different incremental QALYs estimates.

The impact of the different MNAR assumptions can also be readily described in the cost-effectiveness plane (Fig. 3). On the diagonal, where the MAR departures are assumed to be the same in both arms, the joint distribution of incremental QALYs and cost remains relatively unchanged. However, differential changes of the sensitivity parameter (c) between arms lead to a shift in the distribution of incremental QALYs to the right (10TT more cost-effective) or left (10TT less cost-effective). These shifts essentially reflect the impact of the MAR departures on the incremental QALYs seen in Table 1. For example, for scenarios where c is lower (stronger departure from MAR) in the treatment arm (upper-right off-diagonal plots), the joint distribution is shifted to the left and the proportion of points below the cost-effectiveness threshold (£20,000 per QALY) is lower (10TT less likely to be cost-effective).

4 Extensions

Section 3 provided a relatively simple example of a sensitivity analysis. In this section, we discuss possible extensions and further issues around their implementation in practice.

4.1 Missing Cost

In our base-case example, we considered departures from the MAR assumption for the effectiveness endpoint (HRQoL) only. However, it is possible to consider MNAR sensitivity analysis for the cost data as well, following a similar approach.

Table 2 presents the results of a sensitivity analysis for 10TT when both the missing cost and HRQoL data were considered to be MNAR. This involves four parameters, capturing the MAR departure in total costs and HRQoL, in each arm. The missing costs were assumed to be somewhere between MAR and up to 10% higher than observed (i.e. participants who dropped out may have higher health care use). Table 2 suggests that the departures from MAR for the cost endpoint would only have a marginal effect on the overall results, while departures for the HRQoL endpoint can strongly affect the conclusions, particularly if the missing data mechanisms differ between arms. More details on the analysis and the Stata code are provided in Online Appendix 3 (see the ESM).

As the number of variables increase, so does the number of sensitivity parameters, whose values we have to specify. The number of plausible combinations of these parameters can quickly become overwhelming, and it may be best to focus on a limited number of scenarios, or on the parameters that affect the results the most, to allow for a meaningful interpretation.

4.2 Alternative MNAR Parametrisation

In our example, we have rescaled the MAR-imputed HRQoL by a multiplicative factor. As discussed in Sect. 2.3, another popular pattern-mixture approach is to 'offset' the data by an additive factor. This is commonly used for continuous outcomes measured on a readily interpretable scale, such as EQ-5D, which is anchored at 0 (death) and 1 (full health). However, for cost data, a multiplicative reduction may be more intuitive; for example, a '10% reduction' may be more readily understood than a '£200 reduction' as the latter is context specific. A multiplicative transformation may therefore be more appealing in the CEA context.

The values of the MNAR parameters could also be varied according to other factors. With longitudinal data, the departure from MAR can be assumed constant over time—as was considered here—or changing over time, for example, with the parameter increasing with time since withdrawal [31, 37]. The parameter can also be applied at different levels of data aggregation, for example, assuming only one of the resource use components is likely to be MNAR. Different parameters could also be used according to the reasons for discontinuing the trial.

In principle, pattern-mixture models are very flexible and the distribution of unobserved data could take any shape or form. While it can be tempting to consider more complex models (e.g. additional parameters), it can make elicitation and interpretation challenging. In our view, simple offsets or rescaling of the MAR distribution (allowed to differ by arm) should usually provide sufficient

Table 2 Cost-effectiveness of 10TT under different MNAR assumptions for missing cost and effectiveness quality-of-life data

Scenario description	Incremental cost (£) [95% CI]	Incremental QALYs [95% CI]	INMB[a] (£) [95% CI]	Probability cost-effective[a] (%)
MAR	− 35 [− 504 to 434]	− 0.004 [− 0.074 to 0.066]	− 49 [− 1632 to 1534]	48
Same MNAR parameters[b] in the two arms				
− 10% QoL in both arms	− 35 [− 504 to 434]	− 0.011 [− 0.078 to 0.057]	− 181 [− 1714 to 1352]	41
+ 10% cost in both arms	− 25 [− 512 to 462]	− 0.004 [− 0.074 to 0.066]	− 59 [− 1650 to 1532]	47
− 10% QoL and + 10% cost	− 25 [− 512 to 462]	− 0.011 [− 0.078 to 0.057]	− 191 [− 1733 to 1350]	40
Different MNAR parameters[b] in the two arms				
− 10% QoL in intervention arm	− 35 [− 504 to 434]	− 0.071 [− 0.139 to − 0.002]	− 1378 [− 2932 to 176]	4
− 10% QoL in control arm	− 35 [− 504 to 434]	0.056 [− 0.014 to 0.125]	1148 [− 415 to 2711]	93
+ 10% cost in intervention arm	20 [− 459 to 499]	− 0.004 [− 0.074 to 0.066]	− 104 [− 1691 to 1483]	45
+ 10% cost in control arm	− 80 [− 558 to 398]	− 0.004 [− 0.074 to 0.066]	− 4 [− 1591 to 1583]	50

All results are based on imputed data and comparing the 10TT arm to the control arm (n = 537)

CI confidence interval, *INMB* incremental net monetary benefit, *MAR* missing at random, *MNAR* missing not at random, *QALY* quality-adjusted life year, *QoL* quality of life, *10TT* Ten Top Tips

[a]At a cost-effectiveness threshold of £20,000/QALY

[b]How missing cost and QoL data are assumed to differ from MAR-imputed values

span for a comprehensive sensitivity analysis, while remaining sufficiently transparent.

4.3 Choosing the MNAR Parameters

One of the main concerns about conducting an MNAR analysis is how to choose plausible sensitivity parameter values. Several approaches and sources of information can be used for this purpose. One potential approach is to formally elicit 'experts' beliefs on the missing data distribution [28]. These 'experts' can be anyone who can contribute knowledge in understanding the missing data, such as trial investigators, clinicians, or patients. Mason et al. have developed a useful framework for eliciting expert opinion about MNAR mechanisms in CEA [36]. The experts' beliefs, capturing the most likely value for the MNAR parameters, and the uncertainty in that value, can then be incorporated into the analysis model (see Sect. 4.4).

Alternatively, one could simply use a 'tipping point' or threshold analysis approach. This involves changing the MNAR parameter until a different conclusion is reached (for example, being or not being cost-effective). The analyst can then discuss with the relevant experts the plausibility of this value. This approach is appealing because it is more readily implemented and less time-consuming than formal elicitation, and may provide sufficient information for the decision problem at hand, especially when results are robust to a wide range of assumptions. However, what constitutes a 'change of conclusion' may not be uniquely defined, and it may be difficult to implement with multiple sensitivity parameters.

An intermediate approach would be to agree on plausible sensitivity scenarios with those involved in the trial or regulators, for example, at a steering committee meeting. A 'most likely' scenario and several 'most extreme' scenarios could be agreed on, without formally eliciting the uncertainty in the parameters. The scenarios should cover all plausible situations, so that readers can be confident that missing data are unlikely to affect the CEA conclusions beyond what is reported in the sensitivity analysis.

Analysts should also consider how missing data are addressed in the trial primary (clinical) analysis, and the elicitation could be done jointly when suitable. The elicitation should ideally be conducted around the final stages of data collection and be 'pre-specified' before the trial results are known.

Overall, a clear understanding of the reasons for missing data in the specific trial context, discussions with relevant 'experts', and insights drawn from the literature are key to inform the choice of sensitivity parameters.

4.4 Probabilistic Parameters

An alternative to reporting results for specific sensitivity parameters values is to incorporate the uncertainty around the parameters into the analysis model. This is a natural approach when a formal elicitation of the parameter's value and its uncertainty has been conducted (Sect. 4.3). While the analysis can be conducted using a Bayesian framework [36], it can also be implemented using MI [28, 37]. To do so, instead of rescaling all the imputed dataset by a fixed value, a random parameter value is drawn from the elicited

distribution for each of the imputed datasets. An example is provided in Online Appendix 4 (see the ESM).

This probabilistic approach is particularly appealing as it incorporates the uncertainty related to MNAR into the analytical model, providing a 'single' answer. It can be particularly relevant, for example, if the result is to be incorporated in a larger decision model.

However, some stakeholders found this approach less comprehensive than the reporting under different MNAR scenarios. Indeed, this approach also relies on making a single assumption (that the uncertainty was captured appropriately), whereas a range of plausible scenarios may be more readily interpretable in showing how different missing data mechanisms could result in different conclusions.

4.5 Presentation of Results

We have shown how to report the results for different MNAR scenarios by displaying the resulting CEACs. This was flagged by stakeholders as an accessible way to report the results, but they have also recognised that alternative graphical representations may be preferred depending on the decision problem at hand. In this section, we illustrate some of these graphical tools (Stata code provided in Online Appendix 5; see the ESM).

For example, Fig. 5 shows the INMB (and CIs) for values of the *c* parameter, ranging from 0.8 to 1. The parameter is applied to both arms simultaneously, or only one of the arms.

Alternatively, a more comprehensive description of possible combinations of the sensitivity parameters across treatment arms is plotted in Fig. 6. This 'colour-coded graph' (or contour plot) provides a useful tool to interpret

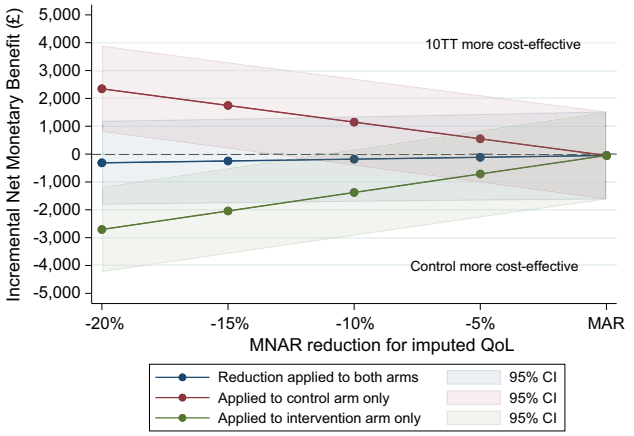


Fig. 5 Alternative presentation: incremental net monetary benefit of 10TT compared to control arm (at £20,000/QALY), for different values of the MNAR rescaling parameter. *CI* confidence interval, *MAR* missing at random, *MNAR* missing not at random, *QoL* quality of life, *10TT* Ten Top Tips

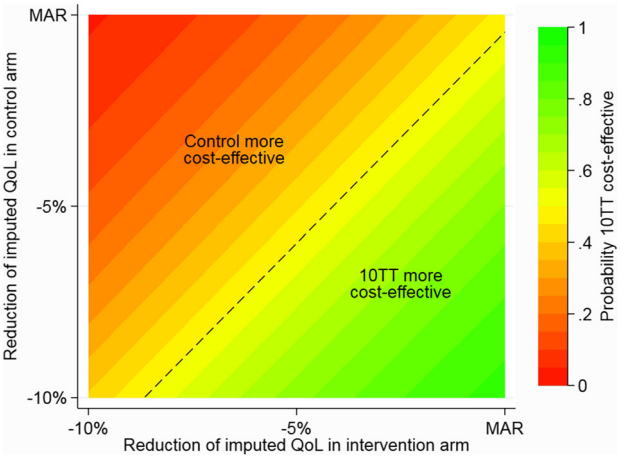


Fig. 6 Alternative presentation: contour plot of the probability of 10TT being more cost-effective than control (at £20,000/QALY), for different values of MNAR rescaling parameters in the control and intervention arms. *MAR* missing at random, *MNAR* missing not at random, *QALY* quality-adjusted life year, *QoL* quality of life, *10TT* Ten Top Tips

the implications of different departures from MAR on the overall decision. For example, it illustrates that for lower values of *c* (stronger departure from MAR) in the intervention arm compared to the control group, the 10TT intervention is unlikely to be cost-effective (red/orange area).

5 Discussion

In this tutorial, we have outlined different approaches for conducting sensitivity analysis for missing data in CEA. We focused on one particularly accessible approach, based on pattern-mixture modelling with MI, and illustrated how it can be implemented in practice. While this is not, in any sense, the final word, we believe that more widespread use of the approach described here would represent a substantial step towards realising the regulatory call for sensitivity analysis.

As Sect. 2 highlights, numerous approaches to MNAR analyses are possible, and there is a large literature on this topic [11, 18, 37]. However, we believe the approach illustrated here has the key advantages of accessibility, flexibility, and transparency. Transparency is indeed the principal requirement for these sensitivity analyses to serve their purpose, as the plausibility of their underlying assumptions needs to be clearly understood and critically assessed by a broad readership [2, 16, 31]. The straightforward implementation of the analysis within an MI framework makes it accessible to the increasing number of analysts who are routinely using MI. It can also be readily implemented within any statistical software with MI (Stata, R, SAS, SPSS, etc.).

Ready implementation allows the focus to be on identifying relevant MNAR scenarios and assessing their plausibility. We discussed here several approaches that can be used in practice, whose suitability will depend on each situation. Some approaches are more rigorous, but more time-consuming, while others are cruder, but still informative. Deciding on the relevant scenarios is likely to involve discussion with other collaborators, and the analysts should be able to explain the different assumptions in non-technical language. Another challenge is the reporting of the results: how can the analyst ensure that the sensitivity analysis is comprehensive, without being overwhelming for the readers? We have suggested a framework where the analysis is conducted under a limited number of plausible scenarios, and the results reported in a table and on a combined CEAC, but also discussed alternative presentations.

The proposed framework is not without some limitations, however. First, every trial raises different issues, and it is not possible to recommend a universal framework for MNAR sensitivity analyses. The framework suggested here is nevertheless relatively flexible, and should be suitable in a wide range of settings, including longitudinal and cluster-randomised trials. Secondly, an assumption such as 'the missing HRQoL are 10% lower' could be too simplistic to capture the varied reasons behind missing data. However, it is important to consider this in light of several aspects. We are primarily interested here, as is usually the case in randomised trials, in estimating mean differences between groups. To obtain valid conclusions, it is therefore not necessary to predict accurately each missing value, but only the *average* difference between observed and missing data. Also, the true missing data mechanism is always unknown, and the aim of the sensitivity analysis is not to provide a definitive answer, but to indicate how conclusions could differ under different missing data assumptions. Finally, the framework proposed here was for continuous outcomes such as cost and quality of life. While the main ideas of the framework are relevant for other outcomes (e.g. binary or survival), they do raise additional challenges, especially around model compatibility and elicitation [37]. For example, differences between observed and missing data in terms of 'odds ratios' may be more difficult to elicit and interpret.

While this tutorial focuses on within-trial CEA, a similar sensitivity analysis approach could possibly be used in observational settings, for example, when analysing routinely collected data, where the issue of informative missing data may arguably be even more important.

This tutorial highlights several areas where further research could improve the value of CEA for decision making in the presence of missing data. A particularly interesting alternative MNAR approach is 'reference-based' or 'controlled' imputation, where the missing data are assumed to follow a distribution that is 'borrowed' from another group. For example, in a trial comparing a drug to placebo, it could be assumed that patients dropping out from the experimental arm have stopped taking their treatment, and therefore follow a similar pattern to that seen in the control arm [33]. This approach is appealing as it sidesteps the elicitation of quantitative parameters required for selection or pattern-mixture models, and instead formulates the MNAR assumption in a qualitative way. It was well received when discussed with stakeholders, but, to our knowledge, has not yet been used in the CEA context. Relevant areas for further research also include incorporating the sensitivity analysis results into broader decision models and, related to this, conducting sensitivity analysis without patient-level data. One possibility could be to approximate the MNAR bias based on the proportion of missing data, and to retain the analysis standard errors as a measure of sampling uncertainty. Further guidance on how to best address missing binary and survival endpoints is still needed. While we propose some routes for eliciting sensitivity parameters, this critical aspect deserves further attention, and is likely to evolve as MNAR analyses become more routinely performed.

In summary, CEA based on incomplete data should routinely assess whether the study's conclusions are robust to potential departures from the standard MAR assumption. This paper described some approaches to conducting these sensitivity analyses, and illustrated the application of a practical, accessible framework using pattern-mixture models with MI. This approach builds on the increasing use of MI in CEA and should provide an important step towards improving practice in trial-based CEA.

Acknowledgements The authors are particularly grateful to the Ten Top Tips trial team for sharing the trial data, in particular, Nishma Patel, Rebecca Beeken and Victoria Vickerstaff, for their help in accessing and using the data. We would also like to thank the attendees of the stakeholders' seminars at the University of York Centre for Health Economics and at the London School of Hygiene and Tropical Medicine, and the patients' representatives, for their invaluable insight. Finally, thank you to Katy Morgan for revising the manuscript and improving its English.

Author Contributions BL, MG and JC had the original idea of the tutorial. BL conducted the statistical analysis, wrote the first draft of the manuscript and developed the implementation code in Stata. All authors critically reviewed the manuscript and approved its final content. BL acts as the overall guarantor.

Compliance with Ethical Standards

Funding This report is independent research supported by the National Institute for Health Research (NIHR) (Doctoral Research Fellowship, B. L., DRF-12437, Senior Research Fellowship, R. G., SRF-2013-06-016). S. M. was in part supported by the NIHR Collaboration for Leadership in Applied Health Research and Care

(CLAHRC) North Thames at Bart's Health NHS Trust. J. C. is supported by a Medical Research Council (MRC) Grant MC_UU_12023/21. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflict of interest J. C. is funded by the MRC for his work at the MRC Clinical Trials Unit at University College London. He receives royalties for his book *Multiple Imputation and its Application* (Wiley, 2013). He has received payment for consultancy and courses relating to missing data from GSK and Novartis, amongst others. He is a member of a Data and Safety Monitoring Board for a Pfizer trial. B. L., R. F., S. M. and R. G. declare that they have no conflict of interest. J. C. and M. G. have received payments for missing data courses in the UK and overseas.

References

1. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing….presumed at random: cost-analysis of incomplete data. Health Econ. 2003;12:377–92.
2. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012;367:1355–60.
3. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics. 2014;32:1157–70.
4. Manca A, Palmer S. Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. Appl Health Econ Health Policy. 2005;4:65–75.
5. Marshall A, Billingham LJ, Bryan S. Can we afford to ignore missing data in cost-effectiveness analyses? Eur J Health Econ. 2009;10:1–3.
6. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. Health Econ. 2012;21:187–200.
7. Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: an incomplete journey. Health Econ. 2018. https://doi.org/10.1002/hec.3654.
8. Burzykowski T, Carpenter J, Coens C, Evans D, France L, Kenward M, et al. Missing data: discussion points from the PSI missing data expert group. Pharm Stat. 2010;9:288–97.
9. Committee for Medicinal Products for Human Use (CHMP). Guideline on missing data in confirmatory clinical trials. London Eur Med Agency. 2010.
10. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015.
11. Molenberghs G, Fitzmaurice G, Kenward MG, Tsiatis A, Verbeke G. Handbook of missing data methodology. Boca Raton: CRC Press; 2014.
12. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clin Trials. 2007;4:154–61.
13. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
14. Little RJA, Rubin DB. Statistical analysis with missing data. Hoboken: Wiley; 2002.
15. Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken: Wiley; 1987.
16. Carpenter JR, Kenward MG. Missing data in randomised controlled trials-a practical guide. Birmingham: National Institute for Health Research; 2007.
17. Bell ML, Fiero M, Horton NJ, Hsu C-H, Gravel J, Opatrny L, et al. Handling missing data in RCTs; a review of the top medical journals. BMC Med Res Methodol. 2014;14:118.
18. National Research Council. Principles and methods of sensitivity analyses. Washington: The National Academies Press; 2010. p. 83–106.
19. Heckman JJ. Sample selection bias as a specification error. Econometrica. 1979;47:153–61.
20. Carpenter JR, Kenward MG, White IR. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. Stat Methods Med Res. 2007;16:259–75.
21. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Appl Stat. 1994;43:49.
22. Glynn RJ, Laird NM, Rubin DB. Selection modeling versus mixture modeling with nonignorable nonresponse. In: Draw inferences from self-selected samples. New York: Springer; 1986. p. 115–42.
23. Little RJA. A note about models for selectivity bias. Econometrica. 1985;53:1469–74.
24. Kenward MG. Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. Stat Med. 1998;17:2723–32.
25. Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc. 1993;88:125–34.
26. Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-mixture models. Biostatistics. 2002;3:245–65.
27. Mallinckrodt CH. Preventing and treating missing data in longitudinal clinical trials: a practical guide. Cambridge: Cambridge University Press; 2013.
28. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomised controlled trials. Clin Trials. 2007;4:125–39.
29. Daniels MJ, Hogan JW. Missing data in longitudinal studies: strategies for Bayesian modeling and sensitivity analysis. Boca Raton: Chapman & Hall; 2008.
30. Permutt T. Sensitivity analysis for missing data in regulatory submissions. Stat Med. 2016;35:2876–9.
31. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharm Stat. 2013;12:337–47.
32. Creemers A, Hens N, Aerts M, Molenberghs G, Verbeke G, Kenward MG. A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. Biometric J. 2010;52:111–25.
33. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23:1352–71.
34. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. BMC Med Res Methodol. 2015;15:30.
35. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med. 1999;18:681–94.

36. Mason AJ, Gomes M, Grieve R, Ulug P, Powell JT, Carpenter J. Development of a practical approach to expert elicitation for randomised controlled trials with missing health outcomes: application to the IMPROVE trial. Clin Trials. 2017;14:357–67.
37. Carpenter JR, Kenward MG. Sensitivity analysis: MI unleashed. In: Multiple imputation and its application. Wiley Online Library; 2013. p. 229–68.
38. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ. 2011;342:1–4.
39. Beeken RJ, Leurent B, Vickerstaff V, Wilson R, Croker H, Morris S, et al. A brief intervention for weight control based on habit-formation theory delivered through primary care: results from a randomised controlled trial. Int J Obes. 2017;41:246–54.
40. EuroQol Group. EuroQol–a new facility for the measurement of health-related quality of life. Health Policy. 1990;16:199–208.
41. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35:1095–108.
42. Beeken RJ, Croker H, Morris S, Leurent B, Omar R, Nazareth I, et al. Study protocol for the 10 Top Tips (10TT) trial: randomised controlled trial of habit-based advice for weight control in general practice. BMC Public Health. 2012;12:667.
43. Patel N, Beeken RJ, Leurent B, Omar RZ, Nazareth I, Morris S. Cost-effectiveness of habit-based advice for weight control versus usual care in general practice in the Ten Top Tips (10TT) trial: economic evaluation based on a randomised controlled trial. BMJ Open. Submitted.
44. Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med. 2003;348:2136–7.
45. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Stat Med. 2000;19:3219–36.
46. Black WC. The CE plane: a graphic representation of cost-effectiveness. Med Decis Mak. 1990;10:212–4.
47. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves - facts, fallacies and frequently asked questions. Health Econ. 2004;13:405–15.
48. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
49. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30:377–99.

Chapter 6

Reference-based multiple imputation

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Baptiste Leurent
Principal Supervisor	Prof. James Carpenter
Thesis Title	Cost-effectiveness analysis with informative missing data: tools and strategies

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Health Economics
Please list the paper's authors in the intended authorship order:	B Leurent, M Gomes, S Cro, N Wiles, J Carpenter
Stage of publication	To be submitted to Health Economics

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	BL took responsibility for the article overall. He extented the existing framework and Stata code to cost-effectiveness data, and conducted the analysis. He wrote the first draft and final version of the paper.

Student Signature: [redacted] **Date:** 26/09/18

Supervisor Signature: [redacted] **Date:** 28 Sept 2018

Reference-based multiple imputation for missing data sensitivity analyses in trial-based cost-effectiveness analysis

Baptiste Leurent[1], Manuel Gomes[2], Suzie Cro[3], Nicola Wiles[4], James R. Carpenter[1,5]

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

[2]Department of Applied Health Research, University College London, London, UK

[3]Imperial Clinical Trials Unit, School of Public Health, Imperial College London, London, UK

[4]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[5]MRC Clinical Trials Unit, University College London, London, UK

ABSTRACT

Missing data are a common issue in cost-effectiveness analysis (CEA) alongside randomised trials, and are often addressed assuming the data are "missing at random" (MAR). However, this assumption is often questionable and sensitivity analyses are required to assess the implications of departures from MAR. Reference-based multiple imputation provides an attractive approach for conducting such sensitivity analyses, because missing data assumptions are framed in an intuitive way by making reference to other trial arms. For example, a plausible not-at-random mechanism in a placebo-controlled trial would be to assume that participants in the experimental arm who drop out stop taking their treatment, and have similar outcomes to those in the placebo arm.

Drawing on the increasing use of this approach in other areas, this paper aims to extend and illustrate the reference-based multiple imputation approach in CEA. It introduces the principles of reference-based imputation, and proposes an extension to the CEA context. The method is illustrated in the CEA of the CoBalT trial evaluating cognitive behavioural therapy for treatment resistant depression. Stata code is provided. We find that reference-based multiple imputation provides a relevant and accessible framework for assessing the robustness of CEA conclusions to different missing data assumptions.

1 | INTRODUCTION

Cost-effectiveness analyses (CEA) of randomised trials provide an important source of information for decision-making, but are often limited by incomplete data collection. For example, participants may withdraw before the end of the study or fail to complete a questionnaire. This is particularly common in longitudinal studies, where data are collected at multiple follow up points, as is often the case in CEA. There has been substantial progress in methods for handling missing data in CEA (Briggs, Clark, Wolstenholme, & Clarke, 2003; Burton, Billingham, & Bryan, 2007; Faria, Gomes, Epstein, & White, 2014; Manca & Palmer, 2005; Oostenbrink & Al, 2005), particularly those that allow valid inferences under the assumption that data are missing at random (MAR) (R. J. A. Little & Rubin, 2002), and in recent years there has been an increase in the uptake of methods such as multiple imputation (Gabrio, Mason, & Baio, 2017; Leurent, Gomes, & Carpenter, 2018; Noble, Hollingworth, & Tilling, 2012). The MAR assumption often provides a desirable starting point for missing data analyses as it implies that any differences between individuals with missing and complete information can be explained by differences in the observed data. However, this assumption may not always hold, as the missingness could depend on unobserved values, i.e. data are missing not at random (MNAR). For example, participants in poorer health may be less likely to complete health-related quality of life questionnaires, conditional on their observed characteristics.

Because the true missing data mechanism is unknown, methodological guidelines recommend conducting sensitivity analyses to departures from the MAR assumption, considering alternative, plausible MNAR mechanisms (Burzykowski et al., 2010; Committee for Medicinal Products for Human Use (CHMP), 2010; Faria et al., 2014; R. J. Little et al., 2012). However, these sensitivity analyses are not routinely conducted (Bell et al., 2014; Gabrio et al., 2017; Leurent, Gomes, & Carpenter, 2018), perhaps due to the lack of accessible methods, or because of the challenges of formulating relevant missing data assumptions beyond MAR. One approach that is receiving increasing attention in clinical trials is reference-based multiple imputation (Carpenter, Roger, & Kenward, 2013; Keene, Roger, Hartley, & Kenward, 2014; Kenward, 2015; R. Little & Yau, 1996). This approach recognises that individuals with missing data could differ from those who complete the study, and – reflecting this – the data are imputed using a different distribution. For example, in a placebo-controlled drug trial, participants in the experimental arm who drop out may stop taking their treatment, and be expected to have similar outcomes to those in the placebo arm. A key advantage of this approach over other methods that have been proposed (Gabrio, Daniels, & Baio, 2018; Leurent, Gomes, Faria, et al., 2018; Mason, Gomes, Grieve, & Carpenter, 2018) is that the

departure from MAR is captured in a qualitative way, making the formulation of the problem more intuitive and accessible to a broader audience, including clinicians and decision makers.

Drawing on recent work (Carpenter et al., 2013), this paper extends and illustrates the reference-based multiple imputation approach to address MNAR data in trial-based CEA. In particular, we focus on adapting the approach to jointly model costs and effectiveness, and allow for different patterns of missingness on cost and effectiveness endpoints over time.

This paper is organised as follows: Section 2 introduces the CoBalT trial, which is used as a motivating example to illustrate the methods. Section 3 introduces the reference-based multiple imputation approach, its extension to the CEA framework, and its implementation in Stata (StataCorp., 2017). Section 4 illustrates the methods, applied to the CoBalT trial. The paper finishes with a discussion of the proposed methods.

2 | CASE-STUDY

2.1 Overview of the CoBalT trial

CoBalT was a two-arm individually randomised controlled trial of Cognitive Behavioural Therapy (CBT) as an adjunct to pharmacotherapy for treatment resistant depression (Wiles et al., 2014, 2013). Patients with treatment-resistant depression were recruited from UK primary care practices between 2008 and 2010, and randomised to either usual care for depression (including pharmacotherapy), or to CBT in addition to usual care. CBT consisted of 12 to 18 sessions delivered by a trained therapist at the general practice or a nearby location, and followed standard CBT manuals (Thomas et al., 2012). The trial's primary outcome was clinical response, defined as a 50% reduction in depressive symptoms (Beck Depression Inventory-II (BDI) (Beck, Steer, & Brown, 1996)) at six months compared with baseline. The trial had an originally planned sample size of 472 participants recruited, to provide 90% power to detect an odds ratio of 2.0 (or an absolute difference of 16%) in clinical response, at the 5% level.

2.2 Cost-effectiveness analysis

A one-year CEA was conducted alongside the trial to assess the cost-effectiveness of CBT in addition to usual care and has been reported in detail elsewhere (Hollinghurst et al., 2010; Wiles et al., 2014). For the purpose of this article, we follow broadly the CEA methods described in Hollinghurst et al. (Hollinghurst et al., 2010) with some simplifications made to allow a clearer focus

on the relevant methodology (e.g. focussing only on unadjusted cost-utility analysis and total costs from the National Health Service (NHS) and Personal Social Service (PSS) perspective). Briefly, health-related quality-of-life (QoL) was measured by the EQ-5D-3L (EuroQol Group, 1990) at baseline, 6 and 12 months and converted into utility scores using a standard set of UK valuations (Dolan, 1997). Quality-adjusted life-years (QALYs) were derived by the 'area under the curve' approach, combining both time and utility scores (Drummond, Sculpher, Claxton, Stoddart, & Torrance, 2015). Costs were measured from the NHS and PSS perspective over a 12 months' period, using resource-use data from the general practice records and patient-reported health service use. Missing QoL data at 6 and 12 months, and total costs were imputed under different assumptions using the referenced-based multiple imputation approach described in Section 3. Participants' baseline QoL, age, sex and BDI were used as covariates in the imputation model, and a set of 100 imputations were performed. The resulting multiply-imputed datasets were analysed using Rubin's rules (Rubin, 1987). Mean differences between arms in QALYs and costs (and 95%CIs) were estimated using unadjusted linear regression, and divided to obtain the incremental cost-effectiveness ratio (ICER) of CBT compared with usual care. The probability of CBT being cost-effective at different willingness-to-pay thresholds (and the resulting cost-effectiveness acceptability curve (CEAC) (Fenwick, O'Brien, & Briggs, 2004)) were derived using unadjusted 'seemingly unrelated regressions' (Willan, Briggs, & Hoch, 2004) for QALYs and for costs. All analyses were performed in Stata version 15 (StataCorp., 2017).

2.3. Missing data pattern and descriptive results

The trial enrolled 469 participants, and 101 (22%) had some cost or effectiveness data missing. Table 1 shows the frequency of each missing data pattern for the cost and effectiveness variables. The cost endpoint had slightly more missing data than the effectiveness endpoint, and the missing data were mostly monotone (when QoL was unobserved at 6 months, QoL at 12 months and total costs tended to be missing as well) but there were also some participants with interim missing data (QoL missing at 6 months, but observed at 12). There was no important difference between arms, with 77% (182/235) of participants providing complete data in the usual care arm and 79% (186/234) in the CBT arm. Missing data were mostly due to participants withdrawing from the study or being lost to follow-up during the trial, and were more common in men and younger participants (Wiles et al., 2014) .

Table 2 reports the observed mean and SD for the cost-effectiveness variables. The mean QoL over time is also shown by missing data pattern in Figure 1. In the participants with complete

effectiveness data, the QoL tended to increase over time, with a greater improvement in the CBT arm than the usual care arm, particularly between baseline and 6 months. Participants with missing data tended to have a lower QoL at baseline.

The primary CEA (Hollinghurst et al., 2010) was conducted using multiple imputation, assuming missing data were MAR. However, due to the nature of the illness, it was argued that those with poorer outcomes could have been more likely to drop out of the trial (i.e. data may be MNAR). To ensure the study provides sound evidence it is therefore important to assess the extent to which the cost-effectiveness inferences are robust to departures from the primary MAR assumption. Reference-based imputation provides a particularly appealing framework to conduct these sensitivity analyses under varying missing data assumptions, as we will see in the following sections.

TABLE 1 Missing data patterns of CoBalT cost and effectiveness variables

Missing data pattern				Usual care (N=235)		CBT (N=234)		Total (N=469)	
QoL baseline	QoL 6 months	QoL 12 months	Total cost	n	%	n	%	n	%
✓	✓	✓	✓	182	77.4	186	79.5	368	78.5
✓	✓	✓	×	13	5.5	6	2.6	19	4.1
✓	×	✓	✓	0	0.0	2	0.9	2	0.4
✓	✓	×	×	18	7.7	14	6.0	32	6.8
✓	×	✓	×	3	1.3	3	1.3	6	1.3
✓	×	×	×	19	8.1	23	9.8	42	9.0

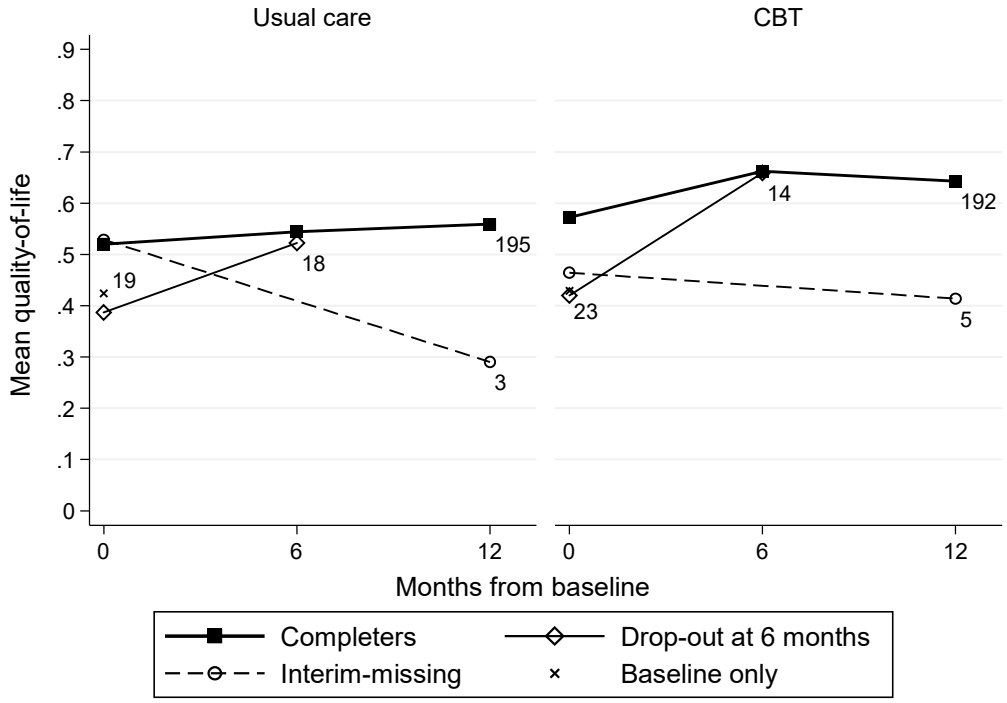
Note. QoL: health-related quality-of-life measured by the EQ-5D-3L; CBT: cognitive behavioural therapy. Ticks indicate observed data, crosses indicate missing data.

TABLE 2 Summary statistics of CoBalT cost and effectiveness variables

	Usual care (N=235)			CBT (N=234)		
Variable	n	Mean	SD	n	Mean	SD
QoL baseline	235[†]	.502	.311	234	.547	.315
QoL 6 months	213	.542	.329	206	.662	.303
QoL 12 months	198	.555	.358	197	.637	.338
QALYs	195	.542	.292	192	.635	.279
Total cost (£)	182	799	725	188	1,803	1,115

Note. QoL: health-related quality-of-life measured by EQ-5D-3L; QALYs: quality-adjusted life years; CBT: cognitive behavioural therapy; SD: standard deviation.
[†]One missing baseline QoL was mean-imputed.

FIGURE 1 CoBalT mean quality-of-life scores, by treatment arm and missing data pattern



Note. CBT: cognitive behavioural therapy.
The number of participants in each pattern is indicated next to the last observation. Linear change is assumed between time-points.

3 | REFERENCE-BASED MULTIPLE IMPUTATION

This section starts by introducing the basic principles of the reference-based multiple imputation approach drawing on recent work by Carpenter et al. (Carpenter et al., 2013). We then provide some technical details, and describe how the approach can be extended for the CEA setting and implemented in standard statistical software.

3.1 Introduction

Reference-based multiple imputation is part of the reference-based (or 'controlled' or 'placebo-based') approaches to handling missing data (Ayele, Lipkovich, Molenberghs, & Mallinckrodt, 2014; Carpenter et al., 2013; Keene et al., 2014; Kenward, 2015; R. Little & Yau, 1996; Lu, 2014; Tang, 2018) which belong to a broader class of 'pattern-mixture models' to model MNAR data (R. J. A. Little, 1993; Ratitch, O'Kelly, & Tosiello, 2013). It can be seen as an extension of ad hoc single imputation MNAR methods, such as assuming "missing=still smoking" commonly used in smoking

cessation trials (West, Hajek, Stead, & Stapleton, 2005), but appropriately capturing random variations and imputation uncertainty in a multiple imputation framework. Instead of a single imputation, an appropriate distribution is used to draw multiple plausible values. This distribution can come from any 'reference' group, but a typical choice in randomised trials would be to use the control arm information. For example, in a placebo-controlled trial, we may wish to use the distribution from the placebo arm to impute outcomes of active-arm individuals who dropped-out (assuming these have stopped taking their treatment). Multiple imputation provides a convenient framework to implement this approach, because it naturally builds on the MAR elements (Carpenter et al., 2013). Once a multivariate model has been fitted assuming MAR, the different elements of the model can be used as 'building blocks' to construct the desired distribution under MNAR. We describe this more formally in the next section.

3.2 Generic algorithm

Consider a randomised controlled trial, where an outcome (say QoL) is measured at multiple time-points. Let i=1,…,N index the N participants randomised in the trial, and T_i indicate the randomisation arm of participant i. Let j index the time-points, j=0,…,J , with j=0 the baseline measurement. Y_{ij} denotes the value of the outcome for participant i at time j. Let Y_{Oi} and Y_{Mi} denote the vectors of observed and missing variables for participant i. For now, let us also assume that all the missing data are due to drop out, so that for a participant i, data are all observed until time-point $D_i \in \{0,\ldots,J\}$, and missing thereafter. So $Y_{Oi}=(Y_{i0},\ldots,Y_{iD_i})^T$, and $Y_{Mi}=(Y_{iD_{i+1}},\ldots,Y_{iJ})^T$.

To impute the missing values, we need to define a distribution for the missing data Y_{Mi}, given the treatment arm and observed data, that is $(Y_{Mi} \mid Y_{Oi}, D_i, T_i)$. Under MAR, this distribution is independent of D_i, and is $(Y_{Mi} \mid Y_{Oi}, T_i)$. Under MNAR assumptions, however, it depends on D_i and we need to define the distribution according to some plausible assumption. A practical option is to make statements about the unobserved data by reference to other groups of participants in the trial (typically participants in different treatment arms).

Reference-based multiple imputation involves the following steps (Carpenter et al., 2013; Cro, Morris, Kenward, & Carpenter, 2016):

1. For each treatment arm separately, fit a multivariate normal (MVN) model for Y_{ij} using the observed data (assuming MAR).
2. Draw a mean vector and a covariance matrix from the posterior distribution of the MVN model parameters.

3. For each participant with missing data, use the draw from step 2 to form the joint distribution of Y_{Oi} and Y_{Mi}. Different assumptions can be used to construct this joint distribution (see Section 3.3.).

4. For each participant, use the joint distribution to construct the conditional distribution of Y_{Mi} given Y_{Oi}, and draw random values from that conditional distribution to impute the missing data.

5. Repeat steps 2-4 *m* times to construct *m* imputed datasets.

The analysis can then be conducted as with standard multiply-imputed datasets. That is the parameters of interest and their variances are estimated by fitting the model of interest to each dataset, and combined using Rubin's rules (Rubin, 1987). Guidance for the analysis of multiply-imputed cost-effectiveness data is provided elsewhere (Briggs et al., 2003; Burton et al., 2007; Faria et al., 2014; Manca & Palmer, 2005)

3.3 Constructing the joint distribution

Several options to construct the joint distribution of the observed and unobserved data have been proposed, each reflecting a different MNAR mechanism (Carpenter et al., 2013). The appropriate choice will be context-specific, but here we describe some options that may be of particular relevance to trial-based CEA. Each of these options is illustrated in Figure 2.

Randomised-arm MAR. The distributions of the missing and observed values, conditionally on the observed variables, are assumed to be the same. The joint distribution follows a multivariate normal with mean and covariance corresponding to the participant's randomised arm estimates. It corresponds to the default assumption with the standard multiple imputation approach. This is the natural choice when missingness is assumed independent of the outcome, or to estimate a 'de jure' (per protocol) estimand, censoring after any protocol deviation.

Jump to reference (J2R). After drop-out, the participant's conditional outcomes are assumed to 'jump' to those of the reference group (typically the control arm). The joint-distribution is a MVN model with mean parameters from the randomised arm until D_i, and from the reference arm afterward. The covariance matrix corresponds to the parameters from the randomised arm until D_i, and to the reference arm for the conditional components of the post drop-out variables, given the pre drop-out measurements. It corresponds to assuming that, after dropping-out,

participants from the active arm have the same outcomes as the reference-arm individuals. This is a plausible choice when the treatment effect is lost after the individual leaves the study.

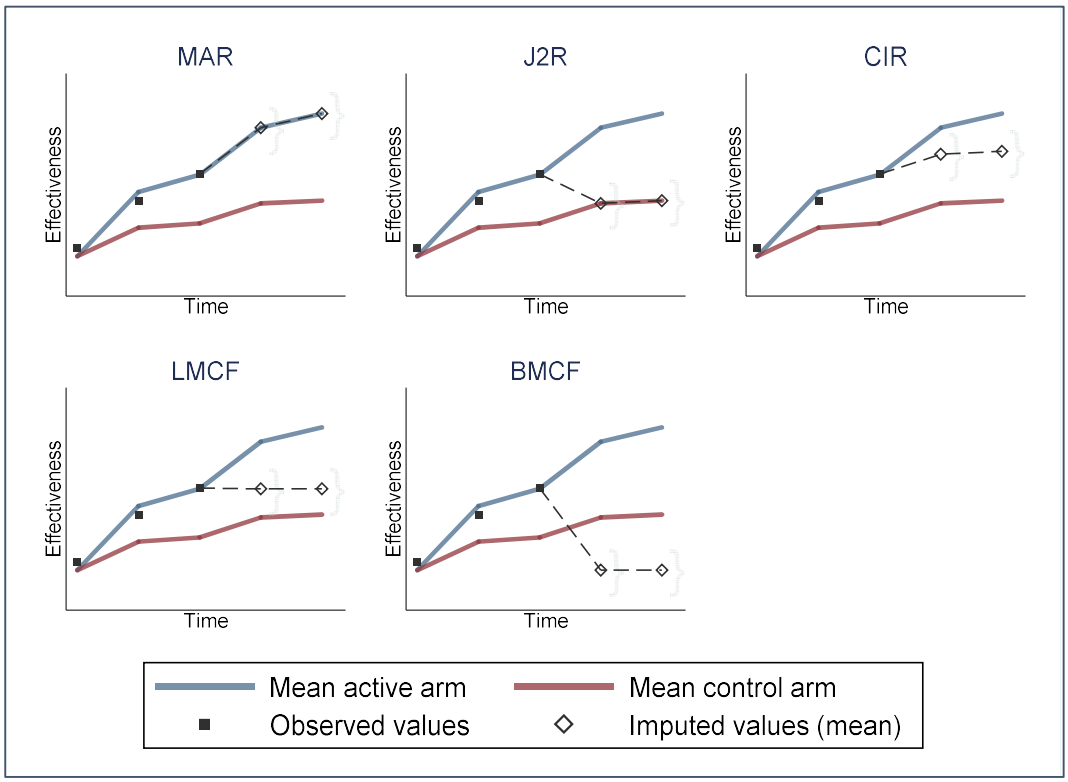
Copy increments in reference (CIR). After drop-out, the participant's conditional outcomes mimic (parallel) the gradient from a reference group. The joint distribution mean parameters follow those from the randomised arm until drop-out, and the increment in mean from the reference group thereafter. The covariance matrix is formed as under J2R. This may be plausible when participants are expected to maintain the treatment benefits accrued until drop-out, then follow (parallel) the outcome trajectory from the reference group after that.

Last mean carried forward (LMCF). After drop-out, the participant's conditional outcomes remain stable, around the mean at that last time-point (from their treatment arm). The joint distribution is a MVN with mean parameters from the randomised arm until drop-out, and the mean parameter from their randomised arm at time D_i for all the following time-points. The covariance parameters follow those of the randomised arm. This is an appropriate choice when the outcomes are likely to remain stable, on average, after drop-out. Note that this is distinct from the ad-hoc "last observation carried forward" approach (Molenberghs et al., 2004), as values are drawn from a well-defined posterior distribution.

Baseline mean carried forward (BMCF). After drop-out, the participant's conditional outcomes are assumed to 'jump' back to the baseline mean level. The joint distribution is a MVN with mean parameters from the randomised arm until drop-out, and the mean parameter from their randomised arm at baseline for the following time-points. The covariance parameters follow those of the randomised arm. This assumption may be plausible when participants are anticipated to lose treatment benefits and return to their baseline values. Although this option was not considered in Carpenter et al. (Carpenter et al., 2013), it seemed relevant in our motivating example.

Note that for J2R and CIR, we need to specify a reference group (typically the control arm). For a participant already in the reference group, the distribution will be the same as under MAR.

FIGURE 2 Illustration of reference-based imputation options



Note. MAR: missing at random; J2R: jump to reference; CIR: copy increments in reference; LMCF: last mean carried forward; BMCF: baseline mean carried forward.
Black squares are observed values for a participant in the active arm dropping-out after the third time-point. Hollow diamonds are the mean of the imputed values for that participant, under the different assumptions. The curly brackets represent the imputation uncertainty around that mean. The reference group (for J2R and CIR) is the control arm.
Note that for clarity the participant is assumed to follow closely the mean of its arm before drop-out. The imputed values will actually depend of the observed data, and for, example, a participant with higher values before drop-out will tend to have higher imputed values.

3.4 Extension to cost-effectiveness data

In this section, we build on the original framework described above to handle key features commonly encountered in CEA.

3.4.1 Handling cost and effectiveness endpoints

The MVN framework in which the algorithm is implemented can be extended naturally to accommodate additional endpoints. While in the original model the size of the Y_{ij} vector was defined by the number of repeated measures, it can be extended to a vector of size $J = J_e+J_c$, where J_e and J_c capture the number of repeated measures of effectiveness and costs, respectively. This results in a MVN model defined, for each treatment arm, by a mean vector of size J, and a variance-covariance

matrix of size JxJ, and therefore $J(J+3)/2$ parameters. The joint-distribution options described above in Section 3.3 can be logically extended for two distinct endpoints, which do not necessarily have to follow the same follow up measurements schedule ($J_e \neq J_c$), or missing data pattern. For example, with J2R, the distribution of the unobserved cost and effectiveness variables, conditional on the observed variables, can be set to follow the conditional distributions for the corresponding variables from the reference arm. Or with LMCF, it is simply the mean for the corresponding endpoint (cost or effectiveness) that is carried forward to the following time-points.

3.4.2 Allowing for differential missing data assumptions between endpoints

An important feature of cost-effectiveness data is that the mechanism that gives rise to missing costs may differ from that of missing effectiveness data. For example, the data may come from different sources (e.g. case-report forms versus patient-reported questionnaires), or be collected at different time points. We may want to assume that only one of the endpoints is MNAR, and that the other may be MAR. To allow for this, for each participant i, Y_{Mi} can be split in two vectors: Y_{MARi}, consisting of the MAR-missing variables, and Y_{MNARi}, of the MNAR-missing variables. The conditional distribution of Y_{MNARi} given Y_{Oi} and Y_{MARi} can then be defined following the options described above. The mean parameters are straightforward to derive, following the principles described in Section 3.3, with the mean parameters from the MAR-missing variables corresponding to those from the randomised-arm. For MAR, LMCF, and BMCF, the covariance matrix will be that of the randomised arm. For J2R and CIR, the covariance matrix requires some further derivation and the technical details are reported in Appendix F. Once the joint distribution has been defined, the remaining steps of the algorithm (see Section 3.2) can be followed, drawing values for (Y_{MARi}, Y_{MNARi}) conditionally on Y_{Oi} for each participant.

3.4.3 Interim missing data

So far it was assumed missing data were monotone within each endpoint (cost or effectiveness), so that all data were missing after a given point in time. It is however common for trial-based CEA to have interim-missing data (an endpoint measure is missing at a particular time point, but observed at a subsequent follow-up point). We have extended the reference-based framework to accommodate this. If the interim and drop-out missing data mechanisms are the same, the joint distribution can be naturally defined. For example, with J2R, we can assume that for each individual, the missing (interim or drop-out) variables follow the distribution from the reference group

conditionally on the observed data. Similarly, for LMCF, the mean carried forward can be drawn from the last observation before the missing time-point. However, the reasons for the interim missing data may differ from those of the drop-out, and it will sometime seem sensible to assume only drop-out missing data are MNAR (while interim-missing are more likely to be MAR). In this case, the joint-distribution can be built following the approach described in Section 3.4.2, with the interim-missing data added to the vector of Y_{MARi} variables. The MNAR endpoints would then follow the specified distribution, conditionally on the observed and the interim-missing variables.

3.5 Implementation in Stata

Drawing on the `mimix` Stata command (Cro et al., 2016) we developed `CEmimix`, a Stata do-file to implement reference based multiple imputation for cost-effectiveness data. The code is reported in Appendix G, and instructions for using `CEmimix` are provided in Appendix H. In brief, the user needs to specify the list of effectiveness and cost variables, the treatment arm variable, any additional imputation covariates, and the choice of imputation methods for the effectiveness and cost endpoints. The program then follows the algorithm described in Section 3.2, and returns the corresponding multiply-imputed datasets which can be analysed using the `mi estimate` command in Stata. Optionally, it allows the user to specify different imputation methods for the interim-missing data, and to restrict the multiple imputation to a subset of participants. Further technical details are provided in the code file (Appendix G) and in Cro et al. (Cro et al., 2016).

4 | RESULTS

In this section we illustrate the reference-based multiple imputation approach for assessing the sensitivity of the CoBalT cost-effectiveness results to different missing data assumptions.

4.1 MAR analysis

For the base-case analysis, we assumed missingness was independent of the unobserved outcome values given the observed data (MAR). It is not possible to test whether this assumption holds based on the observed data, but it often constitutes a sensible starting point. Results are reported in Table 3 and Figure 3. Under MAR, participants in the CBT arm had significantly higher QALYs (0.088, 95%CI: [0.035 to 0.142]) and costs (£996, 95%CI: [802 to 1,190]) than the usual care arm. This resulted in an

ICER of £11,260 per QALY, and a 90.8% probability of CBT being cost-effective at a willingness-to-pay threshold of £20,000 per QALY.

TABLE 3 CoBalT reference-based imputation results under MAR and J2R assumptions

	Usual care (N=235) **Mean [95%CI]**	**CBT** (N=234) **Mean [95%CI]**	**Difference** (N=469) **Mean [95%CI]**
MAR assumption			
QoL at 6 months	0.537 [0.494 to 0.581]	0.653 [0.611 to 0.694]	0.115 [0.055 to 0.175]
QoL at 12 months	0.547 [0.498 to 0.595]	0.625 [0.579 to 0.671]	0.079 [0.012 to 0.145]
QALYs	0.531 [0.492 to 0.569]	0.619 [0.582 to 0.657]	0.088 [0.035 to 0.142]
Costs (£)	803 [694 to 912]	1,798 [1,641 to 1,956]	996 [802 to 1,190]
ICER (£/QALY)			11,260
Probability cost-effective[†]			90.8%
J2R assumption[‡]			
QoL at 6 months	0.537 [0.494 to 0.581]	0.640 [0.597 to 0.683]	0.103 [0.042 to 0.164]
QoL at 12 months	0.547 [0.498 to 0.595]	0.614 [0.566 to 0.661]	0.067 [0.000 to 0.134]
QALYs	0.531 [0.492 to 0.569]	0.610 [0.572 to 0.649]	0.079 [0.025 to 0.134]
Costs (£)	803 [694 to 912]	1,615 [1,464 to 1,767]	813 [630 to 996]
ICER (£/QALY)			10,244
Probability cost-effective[†]			90.8%

Note. MAR: missing at random; J2R: jump to reference; CBT: cognitive-behavioural therapy; QoL: quality-of-life; QALYs: quality-adjusted life-years; ICER: incremental cost-effectiveness ratio.
Based on *m* =100 imputations.
[†]at £20,000/QALY.
[‡] J2R assumption: assuming QoL and costs for drop-out participants in CBT arm jump to usual-care values. Interim-missing QoL assumed to be MAR.

4.2 MNAR sensitivity analyses

We then conducted sensitivity analysis under different MNAR assumptions. First it was assumed that participants dropping out from the CBT arm stopped engaging with the intervention, and that

their QoL and costs followed ('jumped to') that of the control group from that point onwards (J2R). Interim missing data were assumed to be MAR. We can see in Table 3, how this assumption affected the mean QoL and cost estimates at the different time points. The QoL estimates in the CBT arm reduced towards the values of the usual care arm, resulting in a smaller difference in overall QALYs (0.079, 95%CI: [0.025 to 0.134]). Similarly, for the cost, the CBT arm costs were lower than under MAR, resulting in a smaller difference between arms (813, 95%CI: [630 to 996]). Overall, under this assumption the ICER of CBT was slightly lower than under MAR (£10,244 per QALY), but the probability of being cost-effective at £20,000 per QALY was unaffected (90.8%). We can see on the CEAC (Fig. 3) that MAR and J2R results were relatively similar across different willingness to pay thresholds, with a probability of CBT being cost-effective above 90% for any willingness to pay threshold above £20,000 per QALY.

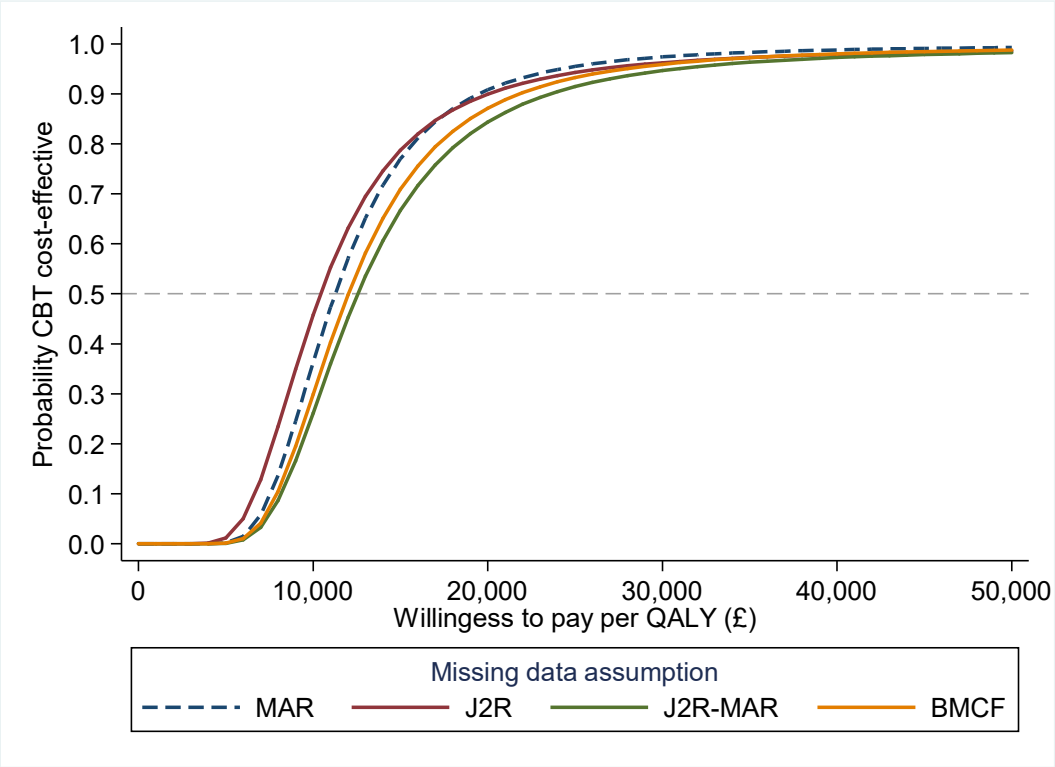
We then explored the impact of further missing data assumptions, for which results are summarised in Table 4. We first conducted the same sensitivity analysis, but assumed that interim-missing QoL data also 'jumped to reference' (J2R-interim). This had little impact on the results (Table 4). We then assumed that only QoL were J2R, and that costs were MAR (J2R-MAR). This was to represent a conservative scenario (for CBT cost-effectiveness), assuming that participants dropping-out from the CBT arm 'jumped to' the QoL from the usual care group, but that costs would still be similar to completers in the CBT arm. Finally, we conducted a more extreme scenario where we assumed QoL of drop-out participants went back to baseline values (BMCF). Note that this is likely conservative in terms of within-arm QALYs, but not necessarily in term of difference between arms. While the exact estimates varied slightly under these different missing data assumptions, none significantly affected the CEA conclusions, with an ICER ranging from £10,244 to £12,552 per QALY, and a probability of being cost effective between 84.4% and 90.8% at £20,000 per QALY (Table 4 and Fig. 3). Overall, these results suggest that for any willingness to pay above £20,000 per QALY, CBT is likely to provide good value for money, and the trial CEA conclusions appear robust to various missing data mechanisms.

TABLE 4 Summary of cost-effectiveness results under different missing data assumptions

Missing data assumption	Difference in QALYs Mean [95%CI]	Difference in costs (£) Mean [95%CI]	ICER (£/QALY)	Probability cost-effective[†]
MAR	0.088 [0.035 to 0.142]	996 [802 to 1,190]	11,260	90.8%
J2R[‡]	0.079 [0.025 to 0.134]	813 [630 to 996]	10,244	90.8%
J2R interim[§]	0.078 [0.024 to 0.132]	813 [630 to 996]	10,423	90.0%
J2R-MAR [¶]	0.079 [0.025 to 0.134]	997 [801 to 1,192]	12,552	84.4%
BMCF [#]	0.083 [0.029 to 0.137]	996 [802 to 1,190]	12,016	87.2%

Note. MAR: missing at random; J2R: jump to reference; BMCF: baseline mean carried forward; CBT: cognitive-behavioural therapy; QoL: quality-of-life; QALYs: quality-adjusted life-years; ICER: incremental cost-effectiveness ratio; CI: confidence interval.
Based on N=469 participants and *m* =100 imputations.
Note that results on the 368 participants with complete cost and effectiveness data were incremental QALYS= 0.091 (95%CI [0.032 to 0.149]) and incremental costs= £1,011 (95%CI [817 to 1,204]).
[†]at £20,000/QALY.
[‡] J2R assumption: assuming QoL and costs for drop-out participants in CBT arm jump to usual-care values. Interim-missing QoL assumed to be MAR.
[§] Same as [‡], but interim-missing QoL were assumed J2R.
[¶] Same as [‡], but missing costs were assumed MAR.
[#] Assuming QoL for drop-out participants goes back to baseline values. Missing costs and interim-missing QoL were assumed MAR.

FIGURE 3 Cost-effectiveness acceptability curve under different missing data assumptions



Note. MAR: missing at random; J2R: jump to reference; BMCF: baseline mean carried forward; CBT: cognitive-behavioural therapy; QALYs: quality-adjusted life-years.
J2R-interim not shown, similar to J2R curve.

5 | DISCUSSION

This study proposes a sensitivity analysis framework for addressing MNAR cost-effectiveness data using the reference-based multiple imputation approach. Drawing on recent work proposed to address missing clinical outcomes in longitudinal trials (Carpenter et al., 2013), our paper extends reference-based multiple imputation to jointly handle missing cost and effectiveness endpoints, and allows for features commonly seen in CEA, such as interim missing data. We illustrated the approach

in the CoBalT trial, evaluating the cost-effectiveness of CBT as an adjunct to pharmacotherapy for primary care patients with treatment-resistant depression (Wiles et al., 2014). We formulated contextually plausible departures from the MAR assumption, and found that the trial cost-effectiveness conclusions were robust to varied missing data assumptions. The software code was also provided, with instructions, to facilitate the implementation of the methods.

The development of sensitivity analysis strategies for addressing potential departures from the MAR assumption in trial-based CEA is an active area of research (Faria et al., 2014; Gabrio et al., 2018; Leurent, Gomes, Faria, et al., 2018; Mason et al., 2018). One of the key strengths of reference-based imputation compared to other sensitivity analysis approaches is the intuitive formulation of the missing data assumptions. This matters because the main challenge when conducting missing-data sensitivity analysis is to formulate assumptions which are contextually relevant and accessible to a broad audience. While similar claims have been made using other pattern mixture models (Faria et al., 2014; Leurent, Gomes, Faria, et al., 2018; Mason et al., 2018), these typically formulate departures from MAR in terms of quantitative differences between observed and missing data, which are less straightforward to interpret. Another strength of this approach is that it can be conveniently implemented after MAR multiple imputation, which is increasingly used to address missing data in trial-based CEA (Gabrio et al., 2017; Leurent, Gomes, & Carpenter, 2018). While the potential of reference-based imputation is more obvious in longitudinal trials, it is also relevant with single follow-up trials, and provides a convenient way to conduct 'worst-case'-type scenarios while appropriately preserving the variance and imputation uncertainty.

A potential limitation of the proposed approach is that its current implementation relies on a MVN model, while QALYs and costs are likely to be non-normally distributed. MVN multiple imputation is, however, recognised as robust to non-normal data, as long as the estimators of interest are normally distributed (Lee & Carlin, 2017; Schafer, 1997) . This is expected to be the case for most trial-based CEA but could be an issue in small trials. For validation, we compared the CoBalT CEmimix results under MAR to multiple imputation by chained equations using predictive mean-matching – which has been recommended to handle non-normal data (White, Royston, & Wood, 2011) – using the `mi impute chained` command in Stata. We obtained very similar results, for example, the mean difference between arms in QALYs under MAR was 0.088 (95%CI [0.034 to 0.141]), with chained equations imputation, compared to 0.088 (95%CI [0.035 to 0.142]) with CEmimix MVN imputation (Table 3).

The estimation of the variance parameters in reference-based approaches has been a source of discussion (Ayele et al., 2014; Gao et al., 2017; Lu, 2014; Seaman, White, & Leacy, 2014). In particular, the model used for the imputation step differs from the one used for the analysis, an issue referred to as 'incongeniality'. While the definition of what should be the appropriate variance estimator when making assumptions about unobserved data is still an area for debate, a recent study showed that the use of Rubin's rules with reference-based multiple imputation has a desirable "information-anchored" property, in the sense that the amount of information lost by the missing data under MNAR is similar to the information loss caused by the missing data under MAR (Cro, Carpenter, & Kenward, 2018). It is worth noting that while multiple imputation provides a particularly convenient framework for implementation, the principles of reference-based are not necessarily tied with those of multiple imputation, and alternative frameworks have been proposed (Lu, 2014).

One key challenge of MNAR sensitivity analyses concerns the choice of plausible missing data assumptions in practice (Faria et al., 2014; Leurent, Gomes, Faria, et al., 2018). While such assumptions are made more transparent in the proposed method, these still need to be informed by subject-matter knowledge and discussion with relevant 'experts' (e.g. trial investigators and practitioners, clinical experts, and patient representatives). The plausibility of each assumption is likely to be a matter of debate, but it is important to keep in mind that the true missing data mechanism is always unknown, and that the aim of the sensitivity analyses is to indicate how results could differ under a range of plausible assumptions (Morris, Kahan, & White, 2014). If sensitivity analyses results differ importantly, investigators should draw conclusions in light of the different results and the plausibility of the respective assumptions (Leurent, Gomes, & Carpenter, 2018). An additional complexity in CEA is to formulate relevant assumptions for each endpoint in light of their differential nature. For example, the J2R assumption is generally seen as conservative for the effectiveness, as it assumes no treatment effect in those with missing data. This may not be the case for the cost endpoint as the difference is typically expected in the opposite direction (new treatment more expensive) and a J2R assumption then becomes liberal.

Reference-based methods are still relatively novel, and there is scope for further research. While the normality assumption has been found reasonable for multiple imputation under MAR, assessing the robustness of the proposed approach to non-normal data in realistic settings is warranted. This paper considered scenarios where one endpoint was assumed to be MAR and the other MNAR, but did not allow for multiple MNAR mechanisms simultaneously (e.g. assuming the effectiveness follows J2R and the cost LMCF). Another development would be to allow for different mechanisms

for different components of the endpoint. For example, assuming that self-reported resource use items are MNAR, while other costs items based on medical records are MAR. Finally, this paper focused on addressing continuous outcomes, as these are most common in CEA (Leurent, Gomes, & Carpenter, 2018), but extending to other types of effectiveness measure (e.g. binary or time-to-event) would provide a valuable contribution.

In conclusion, this study directly addresses the lack of accessible methods for handling MNAR data in trial-based CEA. Reference-based multiple imputation is relatively straightforward to implement and facilitates the formulation of relevant, accessible assumptions. We hope this approach will help future CEA based on incomplete trial data to routinely conduct sensitivity analyses departing from the MAR assumption.

REFERENCES

Ayele, B. T., Lipkovich, I., Molenberghs, G., & Mallinckrodt, C. H. (2014). A Multiple-Imputation-Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials. *Journal of Biopharmaceutical Statistics*, *24*(2), 211–228.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II, Beck Depression Inventory: Manual*. Psychological Corporation.

Bell, M. L., Fiero, M., Horton, N. J., Hsu, C.-H., Gravel, J., Opatrny, L., … Fergusson, D. (2014). Handling missing data in RCTs; a review of the top medical journals. *BMC Medical Research Methodology*, *14*(1), 118.

Briggs, A., Clark, T., Wolstenholme, J., & Clarke, P. (2003). Missing....presumed at random: Cost-analysis of incomplete data. *Health Economics*, *12*(5), 377–392.

Burton, A., Billingham, L. J., & Bryan, S. (2007). Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clinical Trials*, *4*(2), 154–161.

Burzykowski, T., Carpenter, J., Coens, C., Evans, D., France, L., Kenward, M., … Yu, L. M. (2010). Missing data: Discussion points from the PSI missing data expert group. *Pharmaceutical Statistics*, *9*(4), 288–297.

Carpenter, J. R., Roger, J. H., & Kenward, M. G. (2013). Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics*, *23*(3), 1352–1371.

Committee for Medicinal Products for Human Use (CHMP). (2010). *Guideline on missing data in confirmatory clinical trials*. *London: European Medicines Agency* (Vol. 44).

Cro, S., Carpenter, J. R., & Kenward, M. G. (2018). Information-Anchored Sensitivity Analysis: Theory and Application. Retrieved from https://arxiv.org/abs/1805.05795

Cro, S., Morris, T. P., Kenward, M. G., & Carpenter, J. R. (2016). Reference-based sensitivity analysis via multiple imputation for longitudinal trials with protocol deviation. *The Stata Journal*, *16*(2), 443–463.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, *35*(11), 1095–1108.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press.

EuroQol Group. (1990). EuroQol--a new facility for the measurement of health-related quality of life.

Health Policy (Amsterdam, Netherlands), *16*(3), 199–208.
Faria, R., Gomes, M., Epstein, D., & White, I. R. (2014). A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. *PharmacoEconomics*, *32*(12), 1157–1170.
Fenwick, E., O'Brien, B. J., & Briggs, A. (2004). Cost-effectiveness acceptability curves - Facts, fallacies and frequently asked questions. *Health Economics*, *13*(5), 405–415.
Gabrio, A., Daniels, M. J., & Baio, G. (2018). A Bayesian Parametric Approach to Handle Missing Longitudinal Outcome Data in Trial-Based Health Economic Evaluations. Retrieved from http://arxiv.org/abs/1805.07147
Gabrio, A., Mason, A. J., & Baio, G. (2017). Handling Missing Data in Within-Trial Cost-Effectiveness Analysis: A Review with Future Recommendations. *PharmacoEconomics - Open*, *1*(2), 79–97.
Gao, F., Liu, G., Zeng, D., Diao, G., Heyse, J. F., & Ibrahim, J. G. (2017). On inference of control-based imputation for analysis of repeated binary outcomes with missing data. *Journal of Biopharmaceutical Statistics*, *27*(3), 358–372.
Hollinghurst, S., Peters, T. J., Kaur, S., Wiles, N., Lewis, G., & Kessler, D. (2010). Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: Randomised controlled trial. *British Journal of Psychiatry*, *197*(4), 297–304.
Keene, O. N., Roger, J. H., Hartley, B. F., & Kenward, M. G. (2014). Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharmaceutical Statistics*, *13*(4), 258–264.
Kenward, M. G. (2015). Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. *Clinical Investigation*, *5*(3), 311–320.
Lee, K. J., & Carlin, J. B. (2017). Multiple imputation in the presence of non-normal data. *Statistics in Medicine*, *36*(4), 606–617.
Leurent, B., Gomes, M., & Carpenter, J. R. (2018). Missing data in trial-based cost-effectiveness analysis: An incomplete journey. *Health Economics*, *27*(6), 1024–1040.
Leurent, B., Gomes, M., Faria, R., Morris, S., Grieve, R., & Carpenter, J. R. (2018). Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis: A Tutorial. *PharmacoEconomics*, *36*(8), 889–901.
Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, *88*(421), 125–134.
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley.
Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., … others. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, *367*(14), 1355–1360.
Little, R., & Yau, L. (1996). Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs. *Biometrics*, *52*(4), 1324.
Lu, K. (2014). An analytic method for the placebo-based pattern-mixture model. *Statistics in Medicine*, *33*(7), 1134–1145.
Manca, A., & Palmer, S. (2005). Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. *Appl Health Econ Health Policy*, *4*(2), 65–75.
Mason, A. J., Gomes, M., Grieve, R., & Carpenter, J. R. (2018). A Bayesian framework for health economic evaluation in studies with missing data. *Health Economics*.
Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C., & Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, *5*(3), 445–464.
Morris, T. P., Kahan, B. C., & White, I. R. (2014). Choosing sensitivity analyses for randomised trials: principles. *BMC Medical Research Methodology*, *14*(1), 11.
Noble, S. M., Hollingworth, W., & Tilling, K. (2012). Missing data in trial-based cost-effectiveness analysis: the current state of play. *Health Economics*, *21*(2), 187–200.
Oostenbrink, J. B., & Al, M. J. (2005). The analysis of incomplete cost data due to dropout. *Health Economics*, *14*(8), 763–776.

Ratitch, B., O'Kelly, M., & Tosiello, R. (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics*, *12*(6), 337–347.
Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys, (JOHN WILEY & SONS).
Schafer, J. (1997). *Analysis of Incomplete Multivariate Data* (Vol. 72). Chapman & Hall.
Seaman, S. R., White, I. R., & Leacy, F. P. (2014). Comment on "Analysis of Longitudinal Trials With Protocol Deviations: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation," by Carpenter, Roger, and Kenward. *Journal of Biopharmaceutical Statistics*, *24*(6), 1358–1362.
StataCorp. (2017). Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.
Tang, Y. (2018). Controlled pattern imputation for sensitivity analysis of longitudinal binary and ordinal outcomes with nonignorable dropout. *Statistics in Medicine*, *37*(9), 1467–1481.
Thomas, L. J., Abel, A., Ridgway, N., Peters, T., Kessler, D., Hollinghurst, S., … Wiles, N. (2012). Cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment resistant depression in primary care: The CoBalT randomised controlled trial protocol. *Contemporary Clinical Trials*, *33*(2), 312–319.
West, R., Hajek, P., Stead, L., & Stapleton, J. (2005). Outcome criteria in smoking cessation trials: proposal for a common standard. *Addiction*, *100*(3), 299–303.
White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399.
Wiles, N., Thomas, L., Abel, A., Barnes, M., Carroll, F., Ridgway, N., … Lewis, G. (2014). Clinical effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: the CoBalT randomised controlled trial. *Health Technology Assessment*, *18*(31), 1–167, vii–viii.
Wiles, N., Thomas, L., Abel, A., Ridgway, N., Turner, N., Campbell, J., … Lewis, G. (2013). Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *The Lancet*, *381*(9864), 375–384.
Willan, A. R., Briggs, A. H., & Hoch, J. S. (2004). Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics*, *13*(5), 461–475.

Chapter 7

Discussion

7.1 Summary

Missing data are a common issue in cost-effectiveness analysis of randomised trials, and it is widely recognised that sensitivity analyses should be conducted to assess the robustness of conclusions to varied assumptions about the unknown distribution of the missing values. This is, however, rarely done in practice. This thesis aimed to address this issue by developing practical, accessible sensitivity analysis strategies and software tools to handle MNAR data in trial-based CEA. Thus we reviewed current practice in the literature and engaged with analysts and other researchers. Then, informed by this, we proposed two accessible frameworks to conduct these sensitivity analyses.

A review of recent trial-based CEA confirmed that missing data was indeed a ubiquitous issue, and that MNAR sensitivity analyses were rarely conducted. Discussions with academics involved in analysing or reviewing cost-effectiveness evidence, however, indicated that this was broadly acknowledged to be unsatisfactory, and that there was corresponding interest in addressing this. The main barriers appeared to be a lack of time combined with lack of practical methods and guidance to conduct such sensitivity analysis.

The relevance of the issue was also highlighted in two trials, 10 Top Tips, evaluating a brief intervention for weight loss, and CoBalT, evaluating CBT for treatment-resistant depression. In both trials, it seemed possible that completion of the trial could be dependent on health outcomes, and therefore the robustness of CEA conclusions to different MNAR assumptions should be assessed.

In the following two chapters, we therefore proposed frameworks for conducting MNAR sensitivity analyses. The first one was based on a relatively straightforward method, consisting of directly modifying multiply imputed data to reflect different pattern-mixture assumptions. The main appeal

of the method is its simplicity and flexibility. We saw in the review that MI was commonly used to address missing data, and this approach provides a particularly accessible and rapid way to conduct sensitivity analysis after MI (under MAR) has been conducted. One of the key challenge remains the elicitation of the sensitivity parameters. An alternative approach capturing the departure in a more readily interpretable qualitative way, is reference-based imputation. While this approach has gained popularity for the analysis of clinical outcomes [36, 61], it had not yet been implemented in the CEA context. In Chapter 6, we developed a theoretical framework and associated software to apply this approach in a CEA context, illustrating it on the CoBalT trial, and providing the software code to conduct the analysis. The intuitive formulation of the missing data assumption and ease of implementation (once the imputation code has been written) is likely to make this method particularly appealing in CEA.

In both of our motivating examples, conducting the sensitivity analyses provided quantifiable insight on the possible consequences of different missing data mechanisms on the cost-effectiveness estimates. In one trial, the estimates varied widely, reinforcing the initial conclusions that the cost-effectiveness of the technology was highly uncertain. In the second trial, cost-effectiveness estimates were robust to different missing data assumptions, and unlikely to affect decision regarding adoption of the technology in practice.

7.2 Specific contributions

A first specific contribution from the thesis was to provide an update on the current state of play regarding missing data in CEA, following on from the Noble review in 2012 [7]. Missing data are a critical issue in trial-based CEA and it is key to remain aware of current practice. We also conducted a critical review of current practice, and derived a checklist of key recommendations. This checklist should provide analysts with simple points to consider when designing or analysing CEA, and, if followed, could significantly reduce the impact of missing data in future trials.

This review, and the discussion with Stakeholders (Chapter 4), offered important insights into the current CEA context, and perceptions of missing data issues. It provides a point of reference and motivation for developing relevant and accessible methods in this context. The innovative structure of informal discussion, organised around a seminar presentation, could be of interest to other methodological researchers, when a gap between methods and practice has been identified.

We published an article providing an overview of different approaches for MNAR analysis, and suggesting a practical framework to address this issue (Chapter 5). This simple guidance was

clearly missing, and should be an important contribution to raise awareness of the issue, and provide cost-effectiveness analysts with a starting-point to approach the issue.

The most important contribution of the thesis is probably the development of two accessible tools to conduct CEA under MNAR assumptions. In Chapter 5, we focussed on one particularly accessible and flexible method, PMM-MI. We made practical for performing the analysis, such as how to decide the MNAR parameters, and how to report the results. In Chapter 6 we extended an existing method, reference-based imputation, to the CEA context. We developed a theoretical framework, addressing key issues for the approach to be applicable to CEA data. One important contribution may be the more formal way of addressing data that could be MAR-missing within that framework.

In order to facilitate the uptake of these methods, Chapters 5 and 6 provided software to implement these sensitivity analysis approaches. In particular, as part of Chapter 5, we uploaded some data examples to an online repository to enable users to replicate the sensitivity analysis conducted in this paper [85].

7.3 Other general contributions emerging from the thesis

7.3.1 The National Institute for Health and Care Excellence

During my PhD, I have explored the potential impact of using appropriate methods for handling missing data, in particular sensitivity analyses, on the decision-making process of agencies such as the UK National Institute for Health and Care Excellence (NICE). This research sought to better understand NICE's methods guidance, whether there were any recommendations about handling missing data, and how alternative MNAR SA approaches could help improve the decision-making process. For this, I met and corresponded with various staff involved with NICE, attended a technology appraisal committee meeting, and reviewed several technology appraisal guidance and clinical guidelines.

Although this remained an informal assessment, it became clear that decisions for technology appraisal were commonly made on cost-effectiveness models typically trying to model the cost and health outcomes of patients over a life-time, rather than individual trials cost-effectiveness conclusions. Randomised trials seem mostly to be used as one of the sources of evidence to populate the model parameters, sometimes from single trials, or from systematic reviews combining multiple trials. Modellers are typically faced with a high number of uncertainties, and seemed usually more concerned by the generalisability of the parameters (to different population, over a lifetime, etc.) rather than the internal validity of randomised trials. Typically, if missing data was an

important issue in the studies used to inform the model, this would be recognised as a limitation in the overall reliability, but the impact would not be quantified systematically. In the NICE guidance for technology appraisal [86] and for clinical guidelines [87], missing data is noted as one example of area of uncertainty that could require sensitivity analyses, but no further recommendations are provided.

This suggest that randomised trials are a natural starting point to improve more general practice for missing data sensitivity analysis — hence the focus of this thesis. Better sensitivity analysis in trials should raise awareness of the issue, and encourage the uptake of appropriate methods in other study designs. The thesis also focussed on developing accessible strategies which can be readily interpreted by those involved in the decision making process; this is key to their adoption.

7.3.2 Multiple imputation

Multiple imputation had a prominent place in this thesis, and provided a practical, flexible framework for estimation and inference for the two main approaches we focused on. MI has become the dominant method to address missing data in health research in recent years, particularly in CEA, as we saw in our review. While its suitability to address MAR missing data in a variety of settings is widely recognised, this thesis harnessed the MI framework for addressing potential departures from the missing data assumption. This was clearly illustrated in the two approaches proposed in this thesis, where MI provided an accessible, user-friendly approach to sensitivity analysis. The first involves modifying the multiply-imputed data to reflect a given MNAR pattern-mixture assumption. The second is also based on a MI framework, but this time it is the MI procedure itself that is modified, altering the conditional predictive distribution to reflect the MNAR assumption before drawing the individual values.

The omnipresence of MI in trial-based CEA 8under MAR) was observed in our review, and is likely to continue to increase in the future. It is therefore critical that analysts are familiar with the approach and avoid its common pitfalls. In our review, the use of MI did not appear always optimal. The reporting of the exact imputation model and of the results was not always clear, the number of imputations was typically small (e.g. 5), and the use of imputation covariates often limited.

7.3.3 MI and bootstrap

While MI has been commonly used with the non-parametric bootstrap, there is little guidance on how to combine the non-parametric bootstrap with the MI procedure. Alternative approaches have been proposed, which often involve conducting first MI then bootstrapping the imputed data, or first

bootstrapping the data and conducting MI within each bootstrap sample. Some of these approaches were evaluated in an article published during the course of this PhD [88], but not specifically to address non-normal outcomes. I supervised a MSc student to explore the relative merits of different approaches in CEA with missing data, particularly in the presence of skewed data such as costs [89]. Briefly, we found that the different approaches we compared to combine MI and BS were unbiased and that the confidence interval coverages seemed appropriate with large samples, but problematic with small sample and strong skewness. It was not clear whether bootstrapping improved inference with skewed outcomes. Given the popularity of non-parametric bootstrap in CEA, further evaluation and understanding of relative performance compared to alternative approaches would be of interest.

7.3.4 Bayesian analysis

A fully Bayesian framework is in many ways a natural framework for conducting analysis under MNAR assumptions. Indeed, MNAR implies that some external information is needed to proceed with the estimation, which is not very natural in a frequentist framework, but much more so in Bayesian. Bayesian models can naturally include additional parameters on which there may be little or no information in the data, so naturally allowing for a given selection or pattern-mixture model. Reference-based imputation can also be implemented in a Bayesian framework [90]. It was one of the approaches considered at the beginning of the PhD, but the review (Chapter 3), and discussion with Stakeholders (Chapter 4) indicated that the framework was still relatively unfamiliar to many analysts and decision-makers. In the review, only one trial CEA was conducted in a Bayesian framework. Our aim was to develop methods with as broad a reach as possible, and MI seemed a more suitable candidate for this purpose, as discussed above. However, while this PhD was ongoing, several articles were published on Bayesian methods to address missing data in CEA, in particular for MNAR analysis [91, 92]. While the analytical techniques are different, the principles and questions addressed tend to be very similar. For example, as discussed in Chapter 6, although the tools to conduct expert elicitation are generally developed with a Bayesian analysis in mind, they can also be used in a frequentist framework. These parallel developments in Bayesian methods for MNAR missing data in CEA are indicative of the current interest in the topic, and should provide additional tools to conduct sensitivity analyses, particularly relevant for those familiar with Bayesian methods.

7.3.5 Objective choice of MNAR scenarios and parameter values

MNAR sensitivity analysis typically involve external judgement and reporting of multiple results, implying a higher risk of subjective analysis and interpretation. This has several implications.

First, it is important to define clearly a single primary (base-case) missing data assumption. The MAR assumption seems a natural choice for the base-case assumption in most situations. Even when some departure from MAR sounds highly plausible, as in our two examples, it can be difficult to assess whether any specific MNAR assumption is any closer to the true missing data mechanism. Formulation of the MNAR model is challenging, and there is a risk that minor misspecification could result in important departure from the 'true' result. Some may disagree and support that a properly elicited MNAR assumption should be more reliable than simply using the MAR assumption by default. Nevertheless, the challenges of eliciting something that is actually unknown has been recognised [93].

For the sensitivity analyses, when several scenarios are considered, these scenarios should ideally be decided a priori and by independent experts (it is relatively easy to try to favour one arm over another if the MNAR approach used allows for different parametrisation by arm). The relative plausibility of each assumption should also be agreed, giving an indication of what appears to be the most plausible assumption(s), and the ones that are more to test results under 'best/worst case'-type scenarios. This would limit selective reporting when discussing the findings, for example focussing on the only scenario where the intervention appears cost-effective.

However, there is a tension in such recommendations. First, while we advise choosing the scenarios a priori, the data collected may also be useful to decide the scenarios or parameters. For example, knowing more about the missing data or the difference between arms may be of relevance to decide the sensitivity analyses of interest. Secondly, while we recommend independent decisions, the 'experts' with more knowledge of the missing data mechanism are probably those that have been closely involved in the trial. But of course those may well have some bias, possibly unconscious, toward favouring one or the other of the treatment arm.

Above all, this highlights the critical importance of having MNAR methods based on transparent and accessible assumptions. Readers (of CEA results) need to be able to understand the assumptions behind each analysis, and judge of their plausibility. If assumptions are unclear, readers will (and should) be suspicious of the results provided. Having fully transparent assumptions is therefore probably the most efficient way to ensure that analysts are sensible in their conduct — and interpretation — of MNAR sensitivity analyses.

7.4 Limitations

Inevitably, the research reported in this thesis has several limitations. First, all the tools and recommendations were based on findings from our review of the HTA studies and discussions with Stakeholders. It is therefore dominated by a UK academic perspective, and it is possible that other countries or settings face different issues, for example QALYs or MI may be less established.

Secondly, it focuses on trial-based individual-patient level analysis. Some of the tools and recommendations appear relatively transferable to other settings, such as PMM-MI with observational studies. However, we did not provide recommendations for more distinct settings such as decision modelling (see above), meta-analysis, or expected value of perfect information.

Another limitation is that we did not perform a systematic assessment of the different sensitivity analysis methods. No head to head comparison of different methods was performed, or simulations conducted to assess the relative performance under different situations. We did not fully consider alternative approaches which could be of relevance for MNAR analysis, such as selection models or Bayesian methods. The relative advantages and disadvantages of the different approaches were discussed, but more in terms of accessibility and applicability than statistical validation (Chapters 4 and 5).

We developed all our software code in Stata, as it was the dominating software identified in the review (75%) and Stakeholders' discussion, but implementation in other software, such as R, could increase accessibility. While the adaptation is straightforward for PMM-MI, reference-based imputation would require further work. Another limitation is that we focus mainly on continuous outcomes, as, again, QALYs was seen as the dominating measure of effectiveness (81%) in the HTA review. Addressing MNAR SA for different type of outcomes (e.g. binary or survival), raises additional challenges such as the modelling technique and parameters elicitation, and are topics for future work.

7.5 Implications for analysts

Drawing on the findings from the literature review and Stakeholders meetings, there are a number of implications for practice:

- Sensitivity analyses for informative-missing data should be routinely conducted. Chapters 5 and 6 propose accessible, off-the-shelf methods for conducting these.
- The missing data issue and sensitivity analyses should be considered early. Analysts should decide which approach appears more suitable, and what information will need to be collected.

- Analysts should engage effectively with trial investigators and clinical experts. They should be able to communicate clearly the implications of MNAR data and possible ways to address it, ensuring that the choice of sensitivity analysis is contextually plausible.
- Analysts should report MNAR methods and results in a clear and transparent way. This is to ensure readers can understand the methods used, and judge the plausibility of the missing data assumption(s) made.
- More generally, analysts should keep in mind that the best way to address the issue is to minimise the occurrence of missing data. Chapter 3 suggests some active steps to consider at the design stage in order to minimise the scope for missing data.

7.6 Implications for health-care decision making

Whenever using information based on incomplete data, it is important that those involved in the decision-making process understand the extent and impact that missing data may have on cost-effectiveness inferences, and ultimately on decisions about resource allocation. Currently, it seems the issue of missing data is usually recognised as a limitation, but its possible implications on decisions are rarely or inappropriately quantified.

In Chapters 3 and 5 we have described the key concepts and statistical principles for addressing issues raised by missing data, to help analysts assess how realistic missing data assumptions behind an analysis appear, and the appropriateness of the methods considered for handling the missing data. As highlighted is this thesis, a fundamental aspect of appropriately handling the missing data is about reporting the results in a way that they are readily interpretable by the different Stakeholders. This will enable those involved in the decision process to recognise the full extent of the uncertainty to the decision at hand.

An increased awareness of the issue and the recognition that data are not systematically MCAR or MAR is likely to result in additional uncertainty in decision models. But in the end, it will give more confidence that this uncertainty has been appropriately been taken into account in the final decision, and so improve the allocation of limited resources.

Decision-making authorities such as NICE should recognise more clearly the possible implications of missing data, and provide further guidance on how they should be addressed, both at the study design and analysis stage.

7.7 Directions for future research

This thesis identified some areas for further methodological research.

7.7.1 Non-normal distributions

A key feature of cost-effectiveness data is that they are typically non-normally distributed (e.g. strongly skewed with a peak at zero for costs). While the methods proposed should be robust to non-normality in practice, simulations studies would be of interest to understand when the validity of the approaches may be questionable. Clear recommendations for addressing imputation of non-normally distributed data (specifically in the CEA context), are still lacking. This could be looked at from the perspective of the MNAR SA methods suggested here.

7.7.2 Simulations studies

As mentioned above, further simulation studies would help understanding when the methods suggested may or may not be valid. In addition to departures from normality, it would be of interest to assess how each method performs if the missing data mechanism is misspecified. One challenge with validation of missing data methods is that typically if the simulated missing-data mechanism corresponds to the one assumed in the model, then results will be valid, and if they differ then they will not. Thus validity of results typically depends on how the missing data mechanism assumed differ from the true mechanism, but this is unknown in practice. Nevertheless, having an idea of how far results may get when misspecified may be of interest, for example, by generating missing data in a selection model framework, and analysing with one of the pattern-mixture models presented here.

7.7.3 Cost-effectiveness modelling

As discussed in Section 7.3.1, cost-effectiveness models have a key role in the decision-making process, and further research on MNAR SA in this context would be welcome. For example, how should analysts account of the MNAR possibility when using estimates from a trial with incomplete data? If sensitivity analyses are reported, they could probably be used directly in the model, using deterministic or probabilistic sensitivity analysis.

If no sensitivity analyses have been conducted, one approach would be to approximate the MNAR SA estimates arithmetically. For example, making assumptions similar to the PMM-MI approach and assuming that, conditionally on the analysis variables, the missing values of an outcome are on average δ higher than the observed, then the MNAR mean can be obtained by simply adding

$\delta.\pi$ to the estimated mean (where π is the proportion of missing data) [60]. This approach has not been, to our knowledge, extended to repeated measures, but if sufficient information is reported, it should be possible to derive arithmetically the MNAR cost-effectiveness parameters estimates in a similar way. It would be interesting to compare the performance of such approximation to the more formal PMM-MI approach implemented in Chapter 5.

7.8 Concluding remarks

Informative missing data in cost-effectiveness analysis of randomised trials is an important issue which has been neglected and often ignored. Based on a survey of the literature, and structured discussion with analysts, we proposed and developed accessible tools and strategies to address the issue. While these are in no sense the final word, this work both raises awareness of the issue and provide practical solutions. We also hope our papers, software, and examples, will encourage cost-effectiveness analysts to routinely conduct sensitivity analyses to assess robustness of their conclusions to possible departures from the MAR assumption. Over the longer term, this should ensure that missing data uncertainty is appropriately incorporated in the decision making process. The proposed methods should help future studies provide sounder evidence about the effectiveness and cost-effectiveness of health interventions, which will contribute to better decisions about resource allocation and improving population health.

Bibliography

[1] Committee for proprietary medicinal products (CPCM). Points to consider on missing data.; 2001. November.

[2] Burzykowski T, Carpenter J, Coens C, Evans D, France L, Kenward M, et al. Missing data: Discussion points from the PSI missing data expert group. Pharmaceutical Statistics. 2010;9(4):288–297.

[3] Lewis JA. Statistical principles for clinical trials (ICH E9): An introductory note on an international guideline. Statistics in Medicine. 1999 aug;18(15):1903–1904.

[4] Committee for Medicinal Products for Human Use (CHMP). Guideline on missing data in confirmatory clinical trials; 2010.

[5] Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. New England Journal of Medicine. 2012;367(14):1355–1360.

[6] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338(jun29 1):b2393.

[7] Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. Health economics. 2012 feb;21(2):187–200.

[8] Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons; 1987.

[9] Beeken RJ, Leurent B, Vickerstaff V, Wilson R, Croker H, Morris S, et al. A brief intervention for weight control based on habit-formation theory delivered through primary care: results from a randomised controlled trial. International Journal of Obesity. 2017 feb;41(2):246–254.

[10] Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment

resistant depression: results of the CoBalT randomised controlled trial. The Lancet. 2013 feb;381(9864):375–384.

[11] Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med. 2003;348(21):2136–2137.

[12] Morris S, Devlin N, Parkin D. Economic Analysis in Health Care. John Wiley & Sons; 2007.

[13] Drummond MF, Sculpher MJ, Torrance GW, O'Brien, Stoddart BJ, L G. Methods for the economic evaluation of health care programmes. vol. 3; 2005.

[14] NICE. NICE | The National Institute for Health and Care Excellence;. Available from: https://www.nice.org.uk/.

[15] WHO. WHO | World Health Organization;. Available from: https://www.who.int/.

[16] Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: A review of methodology and transparency. Health Economics. 2004;13(12):1203–1210.

[17] McCabe C, Claxton K, Culyer AJ. The NICE Cost-Effectiveness Threshold. PharmacoEconomics. 2008;26(9):733–744.

[18] Gray AM. Cost-effectiveness analyses alongside randomised clinical trials. Clinical Trials: Journal of the Society for Clinical Trials. 2006 dec;3(6):538–542.

[19] Gray A, Clarke PM, Wolstenholme JL, Wordsworth S. Applied methods of cost-effectiveness analysis in health care. Oxford University Press; 2011.

[20] O'Sullivan AK, Thompson D, Drummond MF. Collection of Health-Economic Data Alongside Clinical Trials: Is There a Future for Piggyback Evaluations? Value in Health. 2005 jan;8(1):67–79.

[21] HTA. Health Technology Assessment Programme;. Available from: https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technology-assessment/.

[22] Barber JA, Thompson SG. Analysis of cost data in randomized trials: An application of the non-parametric bootstrap. Statistics in Medicine. 2000;19(23):3219–3236.

[23] Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Statistical Science. 1986 feb;1(1):54–75.

[24] Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. Health Economics. 1997 jul;6(4):327–340.

[25] Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. Health Economics. 2004 may;13(5):461–475.

[26] Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves - Facts, fallacies and frequently asked questions. Health Economics. 2004;13(5):405–415.

[27] Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581–592.

[28] Little RJA, Rubin DB. Statistical analysis with missing data. Wiley; 2002.

[29] Mallinckrodt CH. Preventing and treating missing data in longitudinal clinical trials: a practical guide. Cambridge University Press; 2013.

[30] White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Statistics in Medicine. 2005 apr;24(7):993–1007.

[31] Carpenter JR, Kenward MG. Missing data in randomised controlled trials — a practical guide; 2007. Available from: `http://researchonline.lshtm.ac.uk/4018500/1/rm04{_}jh17{_}mk.pdf`.

[32] Heckman JJ. Sample Selection Bias as a Specification Error. Econometrica. 1979;47(1):153–161.

[33] National Research Council. Principles and Methods of Sensitivity Analyses. In: The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press; 2010. p. 83–106.

[34] Carpenter JR, Kenward MG, White IR. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. Statistical Methods in Medical Research. 2007;16(3):259–275.

[35] Diggle P, Kenward MG. Informative Drop-Out in Longitudinal Data Analysis. Applied Statistics. 1994;43(1):49.

[36] O'Kelly M, Ratitch B. Clinical trials with missing data : a guide for practitioners. Wiley-Blackwell, editor; 2014.

[37] Molenberghs G, Fitzmaurice G, Kenward MG, Tsiatis A, Verbeke G. Handbook of Missing Data Methodology. CRC Press; 2014.

[38] Wu MC, Carroll RJ. Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process. Biometrics. 1988 mar;44(1):175.

[39] Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. Biometrics. 1989 sep;45(3):939–55.

[40] Albert PS, Follmann DA. Shared-parameter models. In: Longitudinal data analysis. Chapman & Hall/CRC: Boca Raton; 2009. p. 433–452.

[41] National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press; 2010. Available from: `http://www.nap.edu/catalog/12955/`.

[42] Carpenter JR, Kenward MG. Multiple imputation and its application. John Wiley & Sons; 2012.

[43] Molenberghs G, Kenward M. Missing data in clinical studies. vol. 61. John Wiley & Sons; 2007.

[44] Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Statistics in Medicine. 2014;33(24):4170–4185.

[45] Magder LS. Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios, and risk differences. Controlled Clinical Trials. 2003;24(4):411–421.

[46] Permutt T. Sensitivity analysis for missing data in regulatory submissions. Statistics in Medicine. 2016;35(17):2876–2879.

[47] Hollis S. A graphical sensitivity analysis for clinical trials with non-ignorable missing binary outcome. Statistics in Medicine. 2002;21(24):3823–3834.

[48] Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. Journal of Biopharmaceutical Statistics. 2013;23(3):1352–71.

[49] Galimard JE, Chevret S, Protopopescu C, Resche-Rigon M. A multiple imputation approach for MNAR mechanisms compatible with Heckman's model. Statistics in Medicine. 2016;35(17):2907–2920.

[50] Kaciroti NA, Raghunathan T. Bayesian sensitivity analysis of incomplete data: Bridging pattern-mixture and selection models. Statistics in Medicine. 2014;33(27):4841–4857.

[51] Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, et al. Analyzing incomplete longitudinal clinical trial data. Biostatistics. 2004;5(3):445–464.

[52] Moreno-Betancur M, Chavance M, Lawson AB, Lee D, MacNab Y. Sensitivity analysis of incomplete longitudinal data departing from the missing at random assumption: Methodology and application in a clinical trial with drop-outs. Statistical Methods in Medical Research. 2016;25(4):1471–1489.

[53] White IR, Kalaitzaki E, Thompson SG. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. Statistics in Medicine. 2011;30(27):3192–3207.

[54] Jackson D, White IR, Seaman S, Evans H, Baisley K, Carpenter J. Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation. Statistics in Medicine. 2014;33(27):4681–4694.

[55] DeSouza CM, Legedza ATR, Sankoh AJ. An Overview of Practical Approaches for Handling Missing Data in Clinical Trials. Journal of Biopharmaceutical Statistics. 2009 oct;19(6):1055–1073.

[56] Mallinckrodt CH, Lin Q, Molenberghs M. A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. Pharmaceutical Statistics. 2013 jan;12(1):1–6.

[57] Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, et al. Analyzing incomplete longitudinal clinical trial data. Biostatistics. 2004 jul;5(3):445–464.

[58] Carpenter JR, Kenward MG. Sensitivity analysis: MI unleashed. In: Multiple Imputation and its Application. Wiley Online Library; 2013. p. 229–268.

[59] Daniels MJ, Hogan JW. Missing data in longitudinal studies : strategies for Bayesian modeling and sensitivity analysis. Boca Raton: Chapman & Hall/CRC; 2008.

[60] White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomised controlled trials. Clinical Trials. 2007;4:125–139.

[61] Kenward MG. Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. Clinical Investigation. 2015;5(3):311–320.

[62] Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. BMC Medical Research Methodology. 2014 dec;14(1):118.

[63] Johnston K, Buxton M, Jones D, Fitzpatrick R. Assessing the costs of healthcare technologies in clinical trials HTA Health Technology Assessment NHS RD HTA Programme. Health Technology Assessment. 1999;3(6).

[64] Marshall A, Billingham LJ, Bryan S. Can we afford to ignore missing data in cost-effectiveness analyses? European Journal of Health Economics. 2009 feb;10(1):1–3.

[65] Faria R, Gomes M, Epstein D, White IR. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. PharmacoEconomics. 2014;32(12):1157–1170.

[66] Manca A, Palmer S. Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. Appl Health Econ Health Policy. 2005;4(2):65–75.

[67] Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clinical Trials. 2007;4(2):154–161.

[68] Diaz-Ordaz K, Kenward MG, Grieve R. Handling missing values in cost effectiveness analyses that use data from cluster randomized trials. Journal of the Royal Statistical Society Series A: Statistics in Society. 2014;177(2):457–474.

[69] DiazOrdaz K, Kenward MG, Gomes M, Grieve R. Multiple imputation methods for bivariate outcomes in cluster randomised trials. Statistics in Medicine. 2016;35(20):3482–3496.

[70] Gomes M, Díaz-Ordaz K, Grieve R, Kenward M. Multiple imputation methods for handling missing data in cost-effectiveness analyses that use data from hierarchical studies: an application to cluster randomized trials. Medical decision making. 2013;33(8):1051–1063.

[71] Morris S, Baio G, Freemantle N, Round J, Hunter RM, Butt T. An Educational Review of the Statistical Issues in Analysing Utility Data for Cost-Utility Analysis. PharmacoEconomics. 2015;33(4):355–366.

[72] Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? Quality of Life Research. 2015;24(4):805–815.

[73] Curran D, Bacchi M, Schmitz SFHSU, Molenberghs G, Sylvester RJ. Identifying the Types of Missingness in Quality of Life Data From Clinical Trials. Statistics. 1998;17:739–756.

[74] Briggs A, Clark T, Wolstenholme J, Clarke P. Missing....presumed at random: Cost-analysis of incomplete data. Health Economics. 2003;12(5):377–392.

[75] Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015.

[76] Willan AR, Briggs AH. Statistical analysis of cost-effectiveness data. vol. 37. John Wiley & Sons; 2006.

[77] Round J. Care at the End of Life: An Economic Perspective. Springer; 2016.

[78] Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluation in clinical trials. OUP Oxford; 2014.

[79] Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-Effectiveness Analysis Alongside Clinical Trials II—An {ISPOR} Good Research Practices Task Force Report. Value in Health. 2015;18(2):161–172.

[80] Baio G, Leurent B. An Introduction to Handling Missing Data in Health Economic Evaluations. In: Care at the End of Life. Cham: Springer International Publishing; 2016. p. 73–85. Available from: http://researchonline.lshtm.ac.uk/2535953/.

[81] Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. Bmj. 2011;342(apr07 2):d1548–d1548.

[82] Lee KJ, Carlin JB. Multiple imputation in the presence of non-normal data. Statistics in Medicine. 2017;36(4):606–617.

[83] Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Medical Research Methodology. 2014 dec;14(1):75.

[84] Patel N, Beeken RJ, Leurent B, Omar RZ, Nazareth I, Morris S. Cost-effectiveness of habit-based advice for weight control versus usual care in general practice in the Ten Top Tips (10TT) Trial: economic evaluation based on a randomised controlled trial. BMJ Open;Submitted.

[85] Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for missing data in cost-effectiveness analysis: Stata code. Figshare; 2018. Available from: https://dx.doi.org/10.1038/ijo.2016.206.

[86] National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 | Guidance and guidelines | NICE; 2013. Available from: https://www.nice.org.uk/process/pmg9/.

[87] National Institute for Health and Care Excellence. Developing NICE guidelines: the manual | Guidance and guidelines | NICE; 2014. Available from: https://www.nice.org.uk/process/pmg20/.

[88] Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. Statistics in Medicine. 2018 jun;37(14):2252–2266.

[89] Carroll O. Combining Multiple Imputation and Non-parametric Bootstrapping to Analyse Skew Distributed Cost Data; 2017. Available from: https://www.researchgate.net/publication/329269791.

[90] Liu GF, Pang L. On analysis of longitudinal clinical trials with missing data using reference-based imputation. Journal of Biopharmaceutical Statistics. 2016 sep;26(5):924–936.

[91] Mason AJ, Gomes M, Grieve R, Carpenter JR. A Bayesian framework for health economic evaluation in studies with missing data. Health Economics. 2018 jul;.

[92] Mason AJ, Gomes M, Grieve R, Ulug P, Powell JT, Carpenter J. Development of a practical approach to expert elicitation for randomised controlled trials with missing health outcomes: Application to the IMPROVE trial. Clinical Trials. 2017;14(4):357–367.

[93] Heitjan DF. Commentary on Mason et al. Clinical Trials: Journal of the Society for Clinical Trials. 2017 aug;14(4):368–369.

Appendices

Appendix A

Stakeholders' seminar - Slides

Cost-effectiveness analysis with informative missing data

Baptiste Leurent
Dr Manuel Gomes
Prof. James Carpenter

Health Economics seminar, LSHTM
29th June 2016

NHS
National Institute for Health Research

Seminar aims

- Not a standard seminar where we tell you about our wonderful work
- Instead, we want to:
  - raise the issue of data missing not at random in trials, which is typically overlooked when the trial contributes to the decision making process
  - discuss some of the assumptions, and corresponding methods that can be used when data are MNAR
  - Understand your thinking on these issues, e.g.
    - Not an issue
    - An issue, but don't understand methods
    - Understand the methods, but believe they are the wrong ones
    - Would like to use the methods, but no code/time
    - …..

Outline

- Introduction → MG
- Examples → BL
  - The 10TT trial
  - MCAR/MAR methods:
    - Complete case results
    - Multiple imputation
  - MNAR methods
    - I. Simple approximation
    - II. MNAR MI
    - III. Including uncertainty in delta
    - IV. Reference-based method → JC
- General comments
- Conclusion

Missing data and CEA

- Missing data are common in RCTs
  - Loss of power
  - Risk of bias
- Particularly features in CEA
  - composite outcomes (cost, QALY)
  - More complex mechanism
  - Intermittent patterns
- Typical assumptions about the reasons for the missing data:
  - MCAR: e.g. lost questionnaires
  - MAR: prob missingness unrelated to unobserved values given the observed; e.g. older people less likely to return QOL questionnaires
- Missing data receiving increasing attention
  - MI more commonly used, still active area

Informative missing data

- **MNAR** (or 'informative', or 'non-ignorable') is particularly problematic
  - Missing data is related to unobserved values, conditional on the observed
  - We cannot test for MNAR, given the data at hand
- But often plausible
  - Patients in poor health may be less likely to return their EQ-5D (conditional on X) questionnaires because they are depressed.
- MNAR methods have been commonly applied in other areas, but have not permeated CEA
  - Ongoing review HTA publications: 1 MNAR SA / 26 CEA in 2015
- How to appropriately translated these to CEA?

Informative missing data

- **Good starting point**: sensitivity analysis
  - Are the CEA results sensitive to departures from MAR?

Sensitivity analyses are often formulated:

- In terms of the missing data (selection) mechanism
  - how do the probability of missingness relates to the unobserved value?

  $[Y,M|X] = [Y|X]\,\mathbf{[M|Y,X]}$ → Selection models
- According to differences between the observed and unobserved data

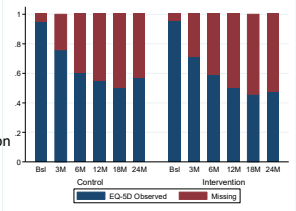
  $[Y,M|X] = [M|X]\,\mathbf{[Y|M,X]}$ → Pattern-mixture models
- Choice of method often related to the level of confidence about the likely values of the sensitivity parameters

The '10 Top Tips' trial

- The trial:
  - **Design:** 2-arms randomised controlled trial
  - **Population:** Obese patients in 14 UK primary care practices
  - **Intervention:** Leaflet with 10 simple tips for weight loss, given by a practice nurse. Compared to usual care.
  - **Follow-up:** 2 years (3, 6, 12, 18, 24 months)
- CEA methods:
  - Cost: 2 years, NHS perspective, GP records, no missing.
  - QALY: 2 years, EQ-5D at each visit. Area under the curve
  - Incremental Net Monetary Benefit, Cost-Effectiveness Acceptability Curve
  - Non-parametric bootstrapping
  - *Only to illustrate MNAR methods*

Missing data in 10TT

- 537 patients randomised
- Around 60% FU at 24M
- But complete case (all EQ-5D available) = 30% !
- Slightly more missing in intervention arm (e.g. CC = 27% vs. 34%)
- Cost fully observed



- MAR or MNAR??
  Weight loss trials → Less successful more likely to drop out[1]

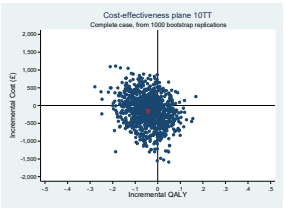
[1] Whare 2010

Methods presented

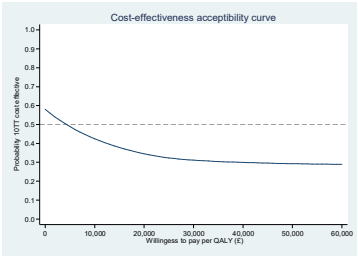
- MCAR/MAR methods:
  - Complete case results
  - Multiple imputation
- MNAR methods
  I. Simple approximation
  II. MNAR MI
  III. Including uncertainty in delta
  IV. Reference-based imputation

Complete case results

	Control arm (n=91)	10TT (n=72)	Difference (n=163)
Cost (£)*	2,138 (1,490; 2,786)	1,985 (1,430; 2,540)	-153 (-1,027; 721)
QALY	1.518 (1.437; 1.599)	1.472 (1.366; 1.578)	-0.046 (-0.177; 0.084;)



Complete case results

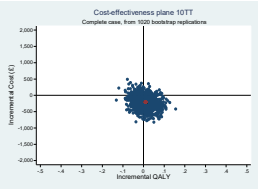


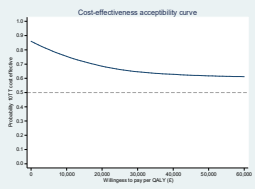
CC → Assume MAR | baseline variables adjusted for
→ Inefficient, drop participants any missing EQ-5D or cost.

Multiple imputation

- MI : Replace each missing value by a set of plausible values given the observed data. Analyse each dataset, then combine results.
- MI commonly used in CEA
  - Can include variables post-randomisation in imputation model → MAR more likely
  - Participants with partially missing data still contribute.
    - Either cost or effect available
    - Cost and QALY = multiple components → missing if any missing
- 10TT MI methods:
  - 30 imputations
  - Variables in imputation model: age, sex, GP, total cost, EQ-5D scores
  - Chained equations, Stata
  - Predictive-mean matching
  - Stratified by arm

MI – 10TT results





→ MI better than CC
- Valid under less stringent missing data assumptions
- Make use of more data → reduce uncertainty

→ 10TT looks possibly cost-effective

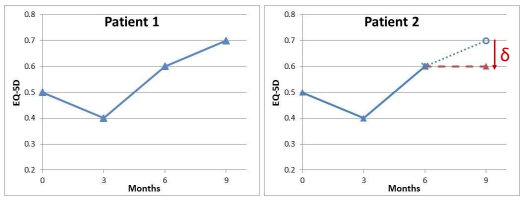
Still assume MAR (observed values are representative of the missing EQ-5D). **What if was not the case??**

Methods presented

- MCAR/MAR methods:
  - Complete case results
  - Multiple imputation
- MNAR methods
  I. Simple approximation
  II. MNAR MI
  III. Including uncertainty in delta
  IV. Reference-based imputation

MNAR sensitivity analyses – PMM Framework

- All MNAR methods presented here will be in a pattern-mixture framework
- In all (except last method) departure from MAR is captured by a parameter "delta":
  → $E(Y_{miss}|X) = E(Y_{obs}|X) + \boldsymbol{\delta}$
- **δ** = "Average difference in Y for missing vs. observed values (conditional on X)"
- If δ = 0 → MAR
- Allowed to vary by arm: δ_0, δ_1 → Important as results sensitive
- PMM could take any form in principle. Shift in mean probably most intuitive/simple.



I) Simple approximation

- Assume we have no access to the data. Only complete-case results reported previously.
- For arm *i* : π_i = proportion of observed data
  μ_i = observed mean
  δ_i = MNAR parameter
  → $\mu_{i,MNAR} = \pi_i \mu_i + (1-\pi_i)(\mu_i + \delta_i)$

	Control arm (n=91)	10TT (n=72)	Difference (n=163)
Cost (£)*	2,138 (1,490; 2,786)	1,985 (1,430; 2,540)	-153 (-1,027; 721)
QALY	1.518 (1.437; 1.599)	1.472 (1.366; 1.578)	-0.046 (-0.177; 0.084)

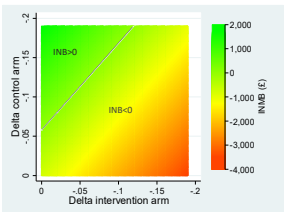
- **$IncQALY_{MNAR}$** = $\pi_1\mu_1 + (1-\pi_1)(\mu_1+\delta_1) - \pi_0\mu_0 + (1-\pi_0)(\mu_0+\delta_0)$
  = **$IncQALY_{CC} + (1-\pi_1)\delta_1 - (1-\pi_0)\delta_0$**

- What about sampling uncertainty?
  → Observed standard error good approximation
  (correcting for bias, not regaining information lost)

Simple approximation - 10TT example

- Complete case: INMB at £20,000/QALY :
  £-777, 95%CI: [-3,711 ; 2,155]
- What if those who dropped out from the study had a 0.1 lower QALY than completers?
  $INMB_{MNAR(-0.1,-0.1)}$ = £-901, 95% CI: [-3,835 ; 2,030]

- Can show results across a range of δ_0 , δ_1:



Simple approximation - Comments

- Not ideal here: based on CC = not full use of data
- Can be applied to adjusted results using same formula[1]
- To MI results? What would be an appropriate π ?

- But can still give an idea: are the results sensitive to missing data?
- Interesting when
  - No better alternative (=no IPD?)
  - Modelling?

[1] White et al. 2007

Formal MNAR modelling

- If have access to IPD, can do things better.
- Let's assume $E[Y_{miss} | X] = E[Y_{obs} | X] + \delta$

  → We can have valid estimates of E[Y] and standard errors (so CI, hypothesis testing, CEAC, etc.) under this assumption, *even if Y not always observed*

- Not directly done in software, but can be solved relatively easily:
  - Maximum likelihood
  - Bootstrapping
  - Bayesian
  - MI
- Here we will show in a MI framework

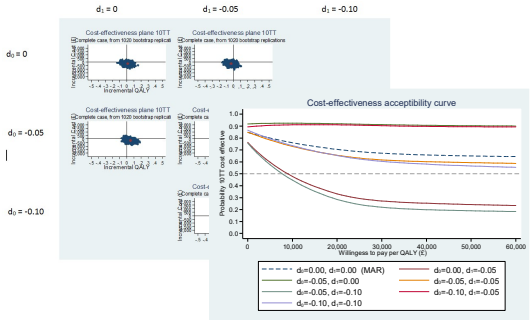
Methods presented

- MCAR/MAR methods:
  - Complete case results
  - Multiple imputation
- MNAR methods
  - I. Simple approximation
  - II. MNAR MI
  - III. Including uncertainty in delta
  - IV. Reference-based imputation

II) MNAR MI

- Again, we assume $E[Y_{miss} | X] = E[Y_{obs} | X] + \delta$
- Straightforward implementation in MI
  - i) Do MAR multiple imputation
  - ii) Shift imputed data of δ_i
  - iii) Analyse MI data as usual
- 10TT example:
  - δ = shift in individual EQ-5D scores
  - δ = { 0, -0.05, -0.10 } (= *Drop out probably worst health, somewhere between 0 to 0.1 lower QoL*)
  - δ varies between arms. δ_0 and δ_1 more likely to be relatively close to each other.
  - → 7 different 'scenarios' for δ

MNAR MI – 10TT example



Methods presented

- MCAR/MAR methods:
  - Complete case results
  - Multiple imputation
- MNAR methods
  - I. Simple approximation
  - II. MNAR MI
  - III. Including uncertainty in delta
  - IV. Reference-based imputation

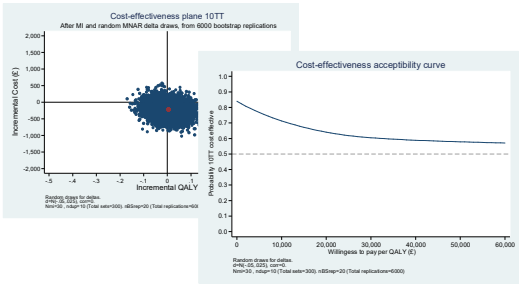
III) Incorporating uncertainty in delta

- So far, presented results with different possible fixed values of deltas
- Another approach is to determine a possible distribution for delta and incorporate it in the analysis model.
- → Will get a unique answer
- How to get distribution?
  - Expert opinion
  - See White et al. (2007), Mason et al. (in preparation)
- Bayesian concepts, but not necessarily analysis

- 10TT example:
  - $\begin{pmatrix}\delta_0\\ \delta_1\end{pmatrix} \sim N\left(\begin{pmatrix}-0.05\\ -0.05\end{pmatrix}, 0.025^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$
  - ρ can be difficult to elicit. Can do with (0,0.5,1), usually little difference
  - Here shown for $\rho = 0$.
  - In MI framework: MI + random draws for δ, added to the imputed EQ-5D

25

Uncertainty in delta - 10TT example



26

Methods presented

- MCAR/MAR methods:
  - Complete case results
  - Multiple imputation
- MNAR methods
  - I. Simple approximation
  - II. MNAR MI
  - III. Including uncertainty in delta
  - IV. Reference-based imputation

27

IV) Reference based MI

- Reference based MNAR analysis seeks to avoid specifying the parameters describing the difference between MAR and MNAR data (e.g. delta above)
- Instead, we make a qualitative statement about the behaviour of patients who deviate. For example, after deviation they may
  - Jump to the distribution (behaviour) of patients in the control arm, *or*
  - Track (parallel) the distribution (behaviour) of patients in the control arm, *or*
  - …
- Such qualitative statements *implicitly* specify parameters describing the differences between MAR and MNAR.
- However, framing it this way may be more accessible.

28

Reference based MI

- Computationally, it is convenient to implement reference based methods using multiple imputation
- This allows us to explore how inferences from our primary analysis model varies across a range of scenarios.
- Relative to the primary analysis (under MAR) the resulting sensitivity analysis are *information preserving*.
- Software is available for SAS (The 5 Macros) and Stata (mimics) from www.missingdata.org.uk
- Further details in Carpenter & Kenward (2013) *Multiple Imputation and its Application* (Wiley)
- Only used for continuous data, effectiveness (licencing) thus far…

29

Comments (1/2)

- Examples were illustrative: simple methods, arbitrary delta, ...
- MNAR → some information not in data → need external info
- Probably as sensitivity analysis. MAR as primary?
- Many other approaches possible
  - PMM ↔ Selection model, shared parameters, etc.
  - MI ↔ Bayesian, likelihood, bootstrap, etc.
    - → Which framework most promising for CEA?
- Missing cost ?
- Presentation of results
  - 3 main possible approaches illustrated:
    - i. Tipping point/contour plot
    - ii. Range of specific scenarios
    - iii. Combined uncertainty
  - Challenges: many scenarios possible (+ if cost)
  - How to make it comprehensive but informative?

30

Comments (2/2)

- Elicitation
  - How to decide relevant MNAR scenarios
  - Particularly challenging *and* critical for "combined analysis"
  - Typically little/no data available
  - Ideally : *a priori*
  - Reference-based: information from the data only
- Modelling
  - MNAR methods more `natural' in IPD. What place in modelling?
- Observational studies
- Outcomes other than QALY (binary, survival)
- Unlikely 'one size fits all'. Can recommend principles. Can make methods more accessible.

Conclusion

- No magic, avoiding missing data best solution!
- Missing data → Make assumptions → What if do not hold?
- Results sensitive, especially to different mechanisms between arms
- **Should be done more often**
- Particularly when:
  i. High missingness
  ii. MNAR plausible
  iii. Different drop out behaviours between arm
- This PhD aims to facilitate the use of these methods. Need to understand needs of those conducting, and using, CEA → next part of seminar
- Any comments? Examples? baptiste.leurent@lshtm.ac.uk

References

Books on missing data, with chapters on sensitivity analyses:

Carpenter, J. & Kenward, M., 2007. *Missing data in randomised controlled trials-a practical guide*, NIHR. Available at http://www.missingdata.org.uk/

National Research Council, 2010. *The Prevention and Treatment of Missing Data in Clinical Trials*, The National Academies Press. Available at http://www.nap.edu/

Carpenter, J. & Kenward, M., 2012. *Multiple imputation and its application*, John Wiley & Sons.

Little, R. & Rubin, D., 2014. *Statistical analysis with missing data*, John Wiley & Sons.

Molenberghs, G. et al., 2014. *Handbook of Missing Data Methodology*, CRC Press.

MNAR SA articles mentioned:

Faria, R. et al., 2014. *A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials*. PharmacoEconomics, 32(12), pp.1157–1170.

White, I.R. et al., 2007. *Eliciting and using expert opinions about dropout bias in randomised controlled trials. Clinical Trials*, 4, pp.125–139.

Mason, A. et al. (in preparation). *Development of Prior Elicitation for Missing Health Outcomes: application to the IMPROVE trial.*

(many other works not mentioned here)

Other ref:

Ware, J.H., 2003. *Interpreting incomplete data in studies of diet and weight loss*. N Engl J Med, 348(21), pp.2136–2137.

Beeken, R.J. et al., 2012. *Study protocol for the 10 Top Tips (10TT) trial: randomised controlled trial of habit-based advice for weight control in general practice*. BMC Public Health, 12, p.667.

Discussion groups

- **We would like to hear your views!**
- Four themes:
  1) Methodological approaches
  2) Presentation of results
  3) Elicitation
  4) Barriers
- On your tables:
  - Consent form
  - Survey + theme questionnaires
- Tasks:
  1) Sign consent form - Keep one copy, leave one
  2) Complete survey
  3) Group discussion :
     - ❑ Split in group of 2-3
     - ❑ Pick one theme (if possible start by the highlighted ones)
     - ❑ Take notes on sheet (or blank sheet at end of questionnaire) → Will collect them
     - ❑ If time, can discuss other themes.

Appendix B

Stakeholders' seminar - Consent form

Cost effectiveness analysis with informative missing data: tools and strategies

LSHTM seminar discussion – 29th June 2016

Informed Consent Form

The purpose of this meeting is to discuss the direction of the PhD research, and to advise on the development of methods for cost effectiveness analysis in presence of informative missing data. In particular, we would like to have your views on how to design, conduct, and report appropriate analyses for dealing with informative missing data in economic evaluations.

Although the meeting will not be recorded, the content of the discussions may be used in future presentations, PhD thesis or other publications. This constitutes primary data collection, and we therefore need your formal consent for using this information.

The individual contributions to the meeting will be anonymised. You can tell us anytime if you do not wish some specific content of the meeting to be reported.

If you have any questions or concerns, please contact us at baptiste.leurent@lshtm.ac.uk

By signing below, I acknowledge that I have read and understood the above information.
I agree to participate, and understand that I can discontinue my participation at any time.

_____	_____	_____
Name of Participant	Signature	Date

Please sign and return one copy, and keep the other copy for your records.

Appendix C

Stakeholders' seminar - Survey

Stakeholders meeting

Participants' survey

N=14. 7 from York, 7 from LSHTM (6 LSHTM group discussion participants did not complete survey)

1. Your role

ID	Trial-based	Modeller	Sitting on decision panel	Reviewing evidence for panel	Other
01		1			
02		1			
03		1		1	
04		1			Econometrist
05	1	1		1	
06					PhD
07	1	1	1	1	
21	1				
22		1	1		
23		1			
24	1	1			
25	1	1			
26		1			
27		1			+ Observational

➔ Mostly modellers. 5 working on trial-based. 4 sitting/reviewing for panels

2. As an analyst, or person using results from CEA, how often do you face the issue of missing data?

ID	1-Never	2- Rarely	3- Occasionally	4- Regularly	5- Very often
01			1		
02				1	
03			1		
04			1		
05				1	
06				1	
07					1
21				1	
22		1			
23				1	
24				1	
25					1
26					1
27					1

➔ Missing data is a common problem. 10/14 (71%) regularly face missing data issues.

3. Before today, did you ever thought about the issue of "Missing Not At Random"?
4. If yes, did you ever face a situation where you thought something more could be done but did not for some reasons (not sure how to do, no time, etc.)?

	Ever thought of MNAR?		**Ever wanted to do more?**	
	No	**Yes**	No	**Yes**
01		1	1	
02		1		1
03		1	.	.
04		1		1
05		1		1
06		1		1
07		1	1*	
21		1		1
22		1†	1	
23		1		1
24		1	1	
25		1		1
26		1		1
27		1	1	

* "Unclear when MNAR is relevant and significant"

† "Not much though"

➔ Everyone thought about MNAR. Majority (8/13, 61%) ever wanted/thought about doing more.

Appendix D

Stakeholders' seminar - Discussion groups

Stakeholders meeting

Answers to discussion groups, and other notes made during seminar

Combining notes from York (28th April) and LSHTM (29th June) discussions.

Content:

Answers from questionnaires
- Theme 1- Methodological approaches
- Theme 2- Presentation of results
- Theme 3- Elicitation
- Theme 4- Barriers

Notes of questions during presentation and discussion

Note:
In blue are participants' answers/comments
In red are my comments/thoughts

Number of respondents to each themes:
Theme 1 - 7
Theme 2 - 4
Theme 3 - 3
Theme 4 - 4
Total = 18

THEME 1 – METHODOLOGICAL APPROACHES

Q1. How familiar/comfortable do you think analysts and "reader" of CEA are with the following approaches? Please mark from 1(few familiar with it) to 5(most familiar with it)

	CEA analysist					Average	CEA users					Average
	1	2	3	4	5		1	2	3	4	5	
Multiple imputation		2		4	1	**3.6**	3	2	1	1		**2.0**
Bayesian	2	1	3		1	**2.6**	6		1			**1.3**
Non-parametric bootstrapping	1		3	1	2	**3.4**	5	1	1			**1.4**
Regression-based*	1		1	2	3	**3.9**	2	2	2		1	**2.4**

*Regression model appropriate for cost-effect, such as seemingly unrelated regression, or linear regression on net monetary benefit.

➔ Regression-based and MI most used/understood approach. Seems like developing methods in MI context is the most promising. Probably not go into the Bayesian route.

Q2. MNAR SA is often a balancing act between making things simple enough to be done and understood, and more complex and accurate methods. Do you think it is more important to focus on methods that are easy to conduct, or one that are technically sound?

Depends on what is the appropriate methodology: irrespective of whether the methods are complex, you would hope that they would be accepted by researchers following dissemination.

Technically sound.

Like the approach MAR as base-case, and MNAR as SA.
Would be good to show difference in results + value foregone based on different methods (EVPI-type analysis).

Methods technically sound, but need to be explained in a non-technical way.

Technically sound if have to pick one.

To make more analysts adopt some approach to conduct MNAR analyses→ most simple.
To provide decisions makers with best evidence → most accurate.

➔ Prefers "best methods". Even if more complex will be adopted if is the appropriate way. (still think there is space for "Back of the envelope" when don't have access to IPD / for modellers)

Q3. Which of the two common framework for MNAR analysis do you think is more appealing for CEA?

- Pattern mixture, where as discussed in the seminar, the sensitivity parameter describes how different the observed and missing data are;
- Selection model, where the sensitivity parameter relates the chance of observing the data to the underlying (but potentially unobserved) value.

The SM seems more intuitive and like the fact that it explicitly recognises the potential situation where we never observe the values.

Pattern-mixture

As analyst, using a selection model (how the prob of missingness related to unobserved values) have more face-value.
Has to do also with how intuitive each model is and how much judgement (subjective) they are.

Pattern-mixture, easier to understand and implement.

Pattern mixture potentially sound simpler.

Reference-based imputation appealing (unclear to me which category this fits in)

Mixed views. Little time to explain well the difference between the 2 approaches during presentation, some liked SM as well.
+ see question on elicitation, reported SM could be more intuitive for clinician.

Q4: What do you see as the pros and cons of each approaches shown during the seminar?

	Advantages	Limitations
Simple approximation	Straightforward Relatively easy to understand and implement Simple, don't need access to IPD	If there are lots of different patterns? No estimates of uncertainty
MI + MNAR delta	A lot more sophisticated. Easier/less complex than MI+delta distribution	Assumptions on delta Need IPD Constant delta too restrictive
MI + delta distribution	This one was my favourites of those seen. Seemed more amenable for estimating the opportunity cost of missing data. More comprehensive.	Assumptions on delta

	Build on the above Seems more natural that delta is a distribution	Need to specify distribution
Referenced-based	No explicit assumptions on delta Sounds in tune with common economic eval modelling Intuitive for experts/ decision makers	Seems not easy to implement (More?) complex to implement Do we (trust?) the experts?
Bayesian*	Framework of combining data + priors	May not be accessible for the analyst

- One liked the MI + delta distribution.
- One mentioned that only ref-based don't need to elicit delta, but may be more complex to implement.
- One liked simple approx. as other seems more complex to implement (I guess when not familiar with MI)

Not really sure what new all this tells us.
Ref-based seems to have some appeal, and probably worth developing.
From the discussion in York people seems to like the delta-distribution approach (for EVPI, or as a formal way to get `final` decision uncertainty, taking into account the possibility of MNAR). In LSHTM preferred more the MI+ fixed values of deltas instead of trying to incorporate all uncertainty in one analysis.

➔ Seems there is still scope for most methods! Depending on context, different methods may be more/less suitable. See summary comment at the end of the doc.

THEME 2 – PRESENTATION OF RESULTS

You have seen earlier three type of presenting results:

i) Individual results for a range of specific scenarios.

ii) A "tipping-point" approach,
(contour plot example, showing results as a continuum on a wide range of values and letting reader decide whether decision would be affected in any plausible range.)

iii) A combined analysis, considering the uncertainty in departure from MAR into the uncertainty modelling.

Q1. Were these approaches clear to you when presented? Any was particularly clear/ unclear?

"Tipping-point" not so clear. The 2 other approaches OK.

Yes, clear

i) was most clear, but CEAC always tricky for non-experts. ii) and iii) (somewhat?) unclear.

Yes they were all clear to us. The delta-graph (ii) was the clearest.

Q2. Which of the 3 type presentational approaches do you think is most relevant for people using cost-effectiveness analysis?
Do you think the presentation should depend of the audience? If so, how?
i) (CEAC with different deltas) most relevant and can be used with all audience (CEA analysts or users).

i) quite flexible and allows more engagement from ? [Stake (= stakeholders) / State ?]

i)
Not ideally, but for pragmatic purposes might be necessary.
CEAC don't go down well with clinical audience.

We think method ii) is the clearest, but needs a lot of explanation to non-experts. It could have the complete case added as a comparator too. + shows the multiple thresholds.
Yes, more details may be required for non-experts.

- ➔ No agreement. After York was starting to think contour plot maybe too confusing, but actually some liked it at LSHTM.
- ➔ Some don't like CEAC as may not be clear for non health-economist. I think still have to go with it as one way to present, recommended by NICE and commonly used, see HTA review. Keep in mind the audience not familiar with CEAC, probably need to show how to calculate other result such as ICER and INB as well.

➔ Mixed view on “combined”, some don’t like to combine the uncertainty in one analysis. Some think that is the best/most realistic/comprehensive.

Q3. Any other way to present the results you thought of?
Could present INB versus delta
Contingent on λ but `accepted’ cut: 20K or 30K

Could add contours to (iii) like in (ii) to show uncertainty.
Could do (ii) for costs and effect separately

➔ Need to try INB with delta graph.

Q4. For most approaches, the way to present the results was dependent on the number of sensitivity parameters. In the seminar presentation, there were two varying parameters (delta for EQ-5D in each arm), but what about if there were more parameters (such as delta for costs varying as well, so having four parameters): would that affect any of your comments above?

No, this would simply add more graphs in approach i).

The case for presenting one result which internalise the joint uncertainty is better.

Harder to process but still prefer ii)

➔ Not so much on this. Except that the “joint uncertainty” may be better when too complex. (But needs more elicitation of correlations or do not matter so much?)

Any other thoughts/comments?

- ?? about observational data
- What about when have immature survival data?
- Selection according to centre/patient prognosis
- Allow for delta to differ by subgroup

➔ In the end think contour plot may not be so clear/obvious, or at least need a couple of minutes to understand well, which people may not do if screening a paper quickly. CEAC with difference scenarios may look less scary, familiar with graph, should understand each curve correspond to a different scenario and just a matter of understanding what these different scenarios mean.

➔ Preference for CEAC with different scenarios. CEAC commonly used in HTA. this was confirmed by answers to Q2 as well.
Drawback compared to tipping point is that needs some sort of elicitation to decide scenarios [tipping point more flexible, shows all “possible” results and let the reader decide]. I think OK, no way around this, seems easier to convince that needs some sort of elicitation, that to push to use a graph that no one understands readily.

➔ Also have to try suggestion to present INB.

- ➔ Some audience not comfortable with CEAC, do no focus only on this graph but also how to calculate INB etc. (but in a similar way, with fixed scenarios more than combining distribution)

Summary:

- ➔ Tipping point may not be as obvious at initially thought
- ➔ Preferred the CEAC with the individual scenarios.
- ➔ If many scenarios/parameters, the “internalised uncertainty” could make presentation easier.

THEME 3 – ELICITATION

Q1. The required elicitation will depend of the approach used. Are there some approaches you have seen today where you think the elicitation will be particularly straightforward or difficult?

Reminder of four main approaches for elicitation:

- Elicitation of delta in a pattern mixture framework
  - Specifying a set of deltas of interest
  - Specifying a distribution for the deltas
- Elicitation in a selection model* framework, where we seek to elicit how the probability of observing a value relates to that value.
- Using a referenced-based approach

[PMM:] Difficult to elicit delta
[Reference-based:] Care needed as subjective judgement

In my experience it can be difficult to give clinicians a feel for what the utility scale means, so a selection model can be a more intuitive alternative.
In the area I work (mental health economics), reference-based approach seems potentially difficult to implement.

→ Utility not so familiar for clinician. But not sure why SM easier, still relies on understanding utility (maybe even more?).
→ Reference based may not be suitable in all situation. Need a suitable control.

Q2. In your opinion, who should be involved in eliciting MNAR scenarios?

Clinical collaborators
Patients representative
Health economists

Clinicians
Patient groups
Statisticians on the clinical trial

Q3. Post-hoc MNAR SA are probably not ideal, but are they acceptable in some situations, or Should they always be pre-specified?

Not conducting a post-hoc MNAR SA entails on implicit assumptions too, so if a case when MNAR appears plausible a post-hoc appears appropriate.

It's best so specify in study proposal but I suppose the details and extent of methos may be specified post-hoc.

Q4. What will be the main challenges of eliciting missingness mechanisms for both cost and effect? How would you address these?

Making assumptions about missing service use (or cost) may not be as easy as with effects

Q5. What do you think about the idea of having a standard criterion for delta in the pattern-mixture approach, such as setting delta to be +/- 1 standard deviation of the variable?

→ No answer to this question. Had some discussions on this, see notes at the end.

THEME 4 – BARRIERS

Q1. Here is a list of possible reasons why MNAR may not be more commonly conducted currently. How important do you think each reason is?

	1 Not a reason why	2 Probably one of the reasons	3 important reason	4 Major reason	**Average**
Not relevant	2	2			1.5
Interesting, but many other possible sensitivity analyses to perform.	1	2		1	2.25
Never thought about it.		3		1	2.5
Would not know how to do			3	1	**3.25**
Do not have the software / technical code to perform it			3	1	**3.25**
Would not know how to decide the relevant scenarios	1	3			1.75
Would take too much time		1	2	1	**3**
No asked for by the funders/investigators	2			2	2.5
It would not be understood		2	2		2.5
Not comfortable in using info not from the data	3	1			1.25
Not commonly done	2		2		2
Not done for the clinical outcome		3		1	2.5

Note : two questionnaires seems to have been completed as a group, and reported exact same answers.

→ Not so much agreement!
Some findings:

- "**Would not know how to do**" and "**Do not have the software**" identified as important reason by all.
  → Good to work on these two aspects during PhD
- "**Would take too much time"**
  → Scope for simple methods: BoE, using standard values for deltas.
- **"Not asked by the investigator"** identified as key reasons by the 2 respondents in York, but not at all by LSHTM.
  → Difficult to affect this directly. But if methods become more common, then may be asked more often. Shows importance of dissemination. Journal reviewers may be easier to reach than "funders/investigators" initially.
- "Not relevant" was not important → demonstrate people think the issue *is* relevant

Other reasons?
Might not give a clear answer → creates more uncertainty. (→ less publishable?)

Accepted practice (challenged at review?)

→ Fair point! Investigators/analysts may not have interest, prefer to assume MAR and put uncertainty of missing data "under the carpet" (+ save time). May need incentive (request by reviewers?) → importance of dissemination.

Q2. Suppose you want to conduct a MNAR SA for a trial soon. What do you anticipate the problems will be? What would you like to see/have to help you performing it?

Identification of delta.
Interpretation

More practical guidance and code on how to do it, including how to carry it to do the probabilistic analysis and code for graphs.
Often the clinical report on the trials does not report it → There should be a need for MNAR analysis in the clinical report.

Coding→ Stata commands!
Fixing delta. Specifically $\delta_T \neq \delta_C$ (evidence?!)
→ precendence for delta

Code! Stata

(Precedence? Meant if publishes a few delta in other papers can use same? Or want more research to get info on possible delta?)

→ Deciding relevant delta was a recurrent issue.
→ Want code + paper explaining methods in practice.
→ (Influence performing MNAR in clinical effectiveness? Probably beyond scope of my PhD. If done more in CEA, may be more common in clinical as well)

Q3. Suppose you are a modeller and want to use data from a trial where missing data is a key issue (let us say 50% missing at the time-point of interest).

a. What would be the common way to approach this currently?

Attempt to perform MAR method? Such as MI

The common way is to use it and not to think about it. This is not the most appropriate way. There is probably scope for a BMJ reporting paper on missing data to explain the issue in simple terms to clinicians and policy maker.

Access to IPD-> MI.
Otherwise ignore!
Not top priority unless treatment effect

b. Do you think modellers would use MNAR results if those were reported? How would they approach this if more than one result was reported?

If key drivers of decision uncertainty, some assessment of plausibility of delta scenarios

I think they would use it if they were reported, but in the sensitivity analyses.

Possibly → SA.

Possibly

→ Seems likely to start with MAR, then may possibly check if affect results when use results from MNAR scenarios.
→ May try to see if delta is a "key driver of decision uncertainty". (How done? Fits with the EVPI idea? Think more: matters for PhD? Can be tackled or focus on the rest first?)

c. If the only results reported are complete-case, do you think modellers would be interested in conducting their own `ad hoc' MNAR analysis?

I think so. Probably first conducting MAR analysis given greater precedence.

I think yes but uptake may be slow at start.

Yes, possibly

Yes, possibly.

→ Potential interest by modellers.

Q6. If codes are developed to conduct MNAR SA, which software do you think are the most relevant? (How familiar are analysts who may want to conduct MNAR SA with the following software?)

	1 Few analysts use it	2 Some analyst use it	3 Often used	4 Most/all analysts familiar with it.	**Average**
Stata			1	3	3.75
R			4		3
SAS		4			2
WinBugs / Other Bayesian		4			2
Excel			3	1	3.25

Twice same answers, worked as group.

→ Stata (and R) main software for IPD.

Any other thoughts?

Mean delta – focus seems sensible as first step, but perhaps distribution idea seems interesting.

Appendix E

Tutorial - Supplementary material

Online Appendix 1 – Ten Top Tips data description

Variable	Description	N	Missing (n)	Missing (%)	Min	Max	Mean	Standard Deviation
Baseline data								
`pid`	Participant code	537	0	0	1	537	-	-
`site_code`	Practice code	537	0	0	1	14	-	-
`arm`	Trial arm	537	0	0	0	1	-	-
`sex`	Gender	537	0	0	1	2	-	-
`age`	Age (years)	537	0	0	18	83	56.80	12.73
`weight_0`	Weight at baseline (kg)	537	0	0	70	177	100.83	17.20
`bmi_0`	BMI at baseline (kg/m^2)	537	0	0	30	61	36.38	5.10
`qol_0`	HRQoL at baseline	537	0	0	-.18	1	0.75	0.25
Quality of life								
`qol_3`	HRQoL – 3 months	395	142	26.4	-.18	1	0.77	0.25
`qol_6`	HRQoL – 6 months	322	215	40	-.07	1	0.77	0.25
`qol_12`	HRQoL – 12 months	286	251	46.7	-.18	1	0.77	0.24
`qol_18`	HRQoL – 18 months	259	278	51.8	-.26	1	0.75	0.25
`qol_24`	HRQoL – 24 months	284	253	47.1	-.14	1	0.77	0.24
`qaly`	QALYs	166	371	69.1	-.1	2	1.50	0.42
Cost								
`totalcost`	Total costs 0-24 months (£)	393	144	26.8	23	27578	1997.81	2546.65

HRQoL= Health-related quality of life (derived from EQ-5D questionnaire), QALYs = Quality Adjusted Life Years, BMI = Body mass index

Notes: missing data at baseline were mean-imputed. Data stored in a 'wide' format, with one record per participant.

Online Appendix 2 – MNAR sensitivity analysis of the 10TT trial – Stata code

Step 1 - Perform multiple imputation under MAR

```
   use "10TT_CEA_tutorial.dta", clear

   *Generate missing data indicators
      //Needed for step 2
      misstable sum  qol_3-qol_24 totalcost qaly ,  gen(miss_)

   *Set imputation
      mi set flong
      mi register imputed qol_3-qol_24 totalcost
      mi register passive qaly

   *Perform imputation
      mi impute chained (pmm, knn(10)) qol_3-qol_24 totalcost = i.sex age i.site_code weight_0
bmi_0 qol_0, add(50) by(arm) rseed(123456)
      //Multiple imputation by chained equations, using predictive mean matching.
      //Imputing the Qol at each follow up and the total cost, using baseline variables.
      //Performing 50 imputations, stratified by arm.
      //rseed()is for reproducibility.
      //Alternatively could use "mi impute chained (regress)" or "mi impute mvn"

   *Calculate QALY
      mi passive: replace qaly=0.125*qol_0 + 0.25*qol_3 + 0.375*qol_6 + (0.25*1.965)*qol_12 +
(0.5*0.965)*qol_18 + (0.25*0.965)*qol_24
      //Area under the curve of the individual QoL (discounted by 3.5% the 2nd year)
      //'mi passive' replace QALY only in imputed datasets.

   *Save imputed dataset
      save "10TT_MI.dta", replace
```

Step 2 - Modifying imputed data

```
   ** Define MNAR scenarios of interest
      //MNAR parameters values are stored in a matrix
      matrix mnar_param = ( 1.0,1.0 \ 1.0,0.95 \ 0.95,1.0 \ 0.95,0.95 \ 0.95,0.9 \ 0.9,0.95 \
0.9,0.9 )
      matrix colnames mnar_param = C_ctr C_int
      matrix list mnar_param
      global nscen = rowsof(mnar_param)  // Saving number of scenarios in global macro

   ** Modify MI data
      clear
      save "10TT_MI_MNAR.dta", replace emptyok //Empty dataset to start with
      forvalues s = 1/$nscen {    //Loop over each MNAR scenarios
            use "10TT_MI.dta" , clear

            *Save scenario info
               gen scenid=`s'
               local q0 = mnar_param[`s',1]
               local q1 = mnar_param[`s',2]
               gen scenario= "`s' (`q0',`q1') "

            *Modify QoL values
               foreach var of varlist qol_3 qol_6 qol_12 qol_18 qol_24 {
                  replace `var'=`var'*mnar_param[`s',1] if miss_`var'==1 & arm==0
                  replace `var'=`var'*mnar_param[`s',2] if miss_`var'==1 & arm==1
                  }

            *Calculate modified QALY
               mi passive: replace qaly=0.125*qol_0 + 0.25*qol_3 + 0.375*qol_6 +
(0.25*1.965)*qol_12 + (0.5*0.965)*qol_18 + (0.25*0.965)*qol_24

            *Append and save
               // The results for all MNAR scenarios are appended in a large dataset to
facilitate remaining steps.
               append using "10TT_MI_MNAR.dta"
               save  "10TT_MI_MNAR.dta", replace
            }
```

Step 3a - Analysis: Incremental cost, effect, ICER and INMB, using Rubin's rules.

```
*** Incremental cost and effect
   *Cost:
      //Have not been MNAR-modified, same for all scenarios.
      use "10TT_MI_MNAR.dta" if scenid==1, clear
      mi estimate: mean totalcost if arm==0
      mi estimate: mean totalcost if arm==1
      mi estimate: regress totalcost arm

   *QALY:
      forvalues s = 1/$nscen {
         use "10TT_MI_MNAR.dta" if scenid==`s', clear
         list scenario in 1
         mi estimate: mean qaly if arm==0
         mi estimate: mean qaly if arm==1
         mi estimate: regress qaly arm
         }

*** ICER
   forvalues s = 1/$nscen {
      use "10TT_MI_MNAR.dta" if scenid==`s', clear
      list scenario in 1
      *Incremental cost
         mi estimate: regress totalcost arm
         local incc= el(e(b_mi),1,1)
      *QALY
         mi estimate: regress qaly arm
         local incq= el(e(b_mi),1,1)
      *Display ICER
         list scenario in 1
         display "MNAR scenario `s' : ICER = "  `incc'/`incq'
      }

*** NMB
   forvalues s = 1/$nscen {
      use "10TT_MI_MNAR.dta" if scenid==`s', clear
      list scenario in 1
      gen inb20=qaly*20000-totalcost
      mi estimate: regress inb20 arm
      }

*** Probability cost effective
   forvalues s = 1/$nscen {
      use "10TT_MI_MNAR.dta" if scenid==`s', clear
      list scenario in 1
      gen inb20=qaly*20000-totalcost
      mi estimate: regress inb20 arm
      local pce = normal(el(r(table),1,1)/el(r(table),2,1) )
      display "Probabiliy cost effective = " `pce'
      }
```

Step 3b - Analysis: CEP and CEAC plots, using non-parametric bootstrap

```
*** Bootstrap
   //Conduct bootstrap re-sampling on imputed dataset
   //Note that alternatives to bootstrap could have been considered here (cf. Faria et al.)

   ** Set up
      *Program returning incremental cost and effect
         capture program drop ceestim
         program define ceestim , rclass
            regress qaly arm
            return scalar inc_qaly = _b[arm]
            regress totalcost arm
            return scalar inc_cost = _b[arm]
            end
      *Dataset to store BS estimates
         clear
         save "bootstrap_mnar.dta", replace emptyok //Empty dataset

   ** Run bootstrap
      //Note: different approaches have been suggested to combine MI and BS
      //(cf. Schomaker and Heumann, arXiv:1602.07933)
```

```
        //Here we are using one possible approach: drawing bootstrap samples from each of the
imputed dataset separately, then pooling the estimates.
        forvalues s = 1/$nscen {
           forvalues m = 1/50 {
              use "10TT_MI_MNAR.dta" if scenid==`s' & _mi_m==`m', clear
              //Open one MI dataset, for one MNAR scenario.
              bootstrap inc_cost=r(inc_cost) inc_qaly=r(inc_qaly), ///
                      reps(200) strata(arm) saving("bsres.dta", replace) : ceestim
                    // Bootstrapping the incremental cost and effect.
                    // 200 BS replications for each imputed dataset, stratified by arm.
              *Pool all estimates
                 use "bsres.dta", clear
                 gen _mi_m=`m'
                 gen scenid=`s'
                 append using "bootstrap_mnar.dta"
                 save "bootstrap_mnar.dta", replace
              }
           }

     ** Clean bootstrap dataset
        *Add scenario label (used for graphs)
           use "10TT_MI_MNAR.dta", clear
           keep scenid scenario
           duplicates drop
           merge 1:m scenid using "bootstrap_mnar.dta", nogenerate
        *Sort and save
           sort scenid _mi_m
           compress
           save "bootstrap_mnar.dta", replace


  *** Cost-effectiveness plane
     use "bootstrap_mnar.dta" , clear
     bysort scenario: egen meanc=mean(inc_cost)
     bysort scenario: egen meanq=mean(inc_qaly)

     *Graph
        graph twoway scatter inc_cost inc_qaly, msize(*0.1) || scatter meanc meanq, ///
                    by(scenario, holes(3 7) compact leg(off)) ///
                    xlab(-0.2(0.1)0.2) xtitle("Incremental QALY") ///
                    ylab(-1000(500)1000,  nogrid angle(horizontal)) ///
                    ytitle("Incremental Cost (£)") ///
                    yli(0,lc(black) lw(thin))  xli(0,lc(black) lw(thin)) ///
                    name(CEP, replace)


  *** Cost-Effectiveness Acceptability Curve

     *Calculate probability cost effective at different threshold.
        postfile ceac scenid str12 scenario wtp proba using "ceac.dta", replace
        // Set-up 'postfile' to store results
        forvalues s = 1/$nscen {
           use "bootstrap_mnar.dta" if scenid==`s', clear
           forvalues wtp = 0(1000)60000 {
              qui: count if (inc_qaly*`wtp'-inc_cost)>0
              local p = `r(N)' / _N  //Proportion of cost-effective BS replicates
              post ceac (scenid[1]) (scenario[1])  (`wtp') (`p')
              }
           }
        postclose ceac //Closing postfile

     *Graph
        use "ceac.dta", clear
        separate proba, by(scenario) veryshortlabel gen(proba_)
           //Create a new variable for each MNAR scenario (needed to show on same graph)
        graph twoway line proba_* wtp, ///
               xlab(0(10000)60000, format(%9.0fc)) ///
               xtitle("Willingess to pay per QALY (£)") ///
               yscale(range(0 1)) ylab(0(0.1)1, nogrid angle(horizontal) format(%2.1f)) ///
               ytitle("Probability 10TT cost effective") ///
               yline(0.5,lc(gs10) lw(thin) lpattern(dash)) ///
               legend(label(1 "1 (1,1) (MAR)") title("MNAR scenario")) rows(3) hole(3 7)) ///
               name(CEAC, replace)
```

Online Appendix 3 – MNAR sensitivity analysis for missing cost and effectiveness

We reanalysed the 10TT trial data, this time considering the missing cost data could also be MNAR.

1) Scenarios of interest

The MNAR scenarios were defined by four parameters:

- c_{Q0} : the MNAR rescaling factor for the imputed quality-of-life score (QoL) in the control group. For example, c_{Q0} =0.9 correspond to reducing MAR-imputed QoL values in the control group by 10%.
- c_{Q1} : rescaling factor for HRQoL in the intervention group
- c_{c0} : rescaling factor for total cost in the intervention group
- c_{c1} : rescaling factor for total cost in the intervention group

We considered eight scenarios, covering a range of MNAR variation for cost and QoL. We considered the missing QoL more likely to be lower, and the cost higher, than under MAR.

Scenario description	**MNAR rescaling parameters**			
	QoL in control group	**QoL in intervention group**	**Cost in control group**	**Cost in intervention group**
1. (MAR)	1	1	1	1
Same parameters in both arms				
2. -10% QoL in both arms	-10%	-10%	1	1
3. +10% cost in both arms	1	1	+10%	+10%
4. -10% QoL and +10% cost	-10%	-10%	+10%	+10%
Different parameters by arm				
5. -10% QoL in intervention arm	1	-10%	1	1
6. -10% QoL in control arm	-10%	1	1	1
7. +10% cost in intervention arm	1	1	1	+10%
8. +10% cost in control arm	1	1	+10%	1

2) Stata code to transform QoL and cost

```
   ** Define scenarios
      matrix mnar_param = (1.0,1.0,1.0,1.0 \  0.9,0.9,1.0,1.0 \ 1.0,1.0,1.1,1.1 \
0.9,0.9,1.1,1.1 \ 1.0,0.9,1.0,1.0 \ 0.9,1.0,1.0,1.0\ 1.0,1.0,1.0,1.1\ 1.0,1.0,1.1,1.0)
      matrix colnames mnar_param = Q0 Q1 C0 C1
      matrix list mnar_param
      global nscen = rowsof(mnar_param)  // Global macro, number of scenarios.

   ** Modify MI data
      clear
      save "10TT_MI_MNAR_cost.dta", replace emptyok
      forvalues s = 1/$nscen {
            use "10TT_MI.dta" , clear
            *MNAR parameters variable
               gen scenid=`s'
               local q0 = mnar_param[`s',1]
               local q1 = mnar_param[`s',2]
               local c0 = mnar_param[`s',3]
               local c1 = mnar_param[`s',4]
               gen scenario= "`s' (`q0',`q1',`c0',`c1') "
            *Modify QoL values
               foreach var of varlist qol_3 qol_6 qol_12 qol_18 qol_24 {
                  replace `var'=`var'*mnar_param[`s',1] if miss_`var'==1 & _mi_m>0 & arm==0
                  replace `var'=`var'*mnar_param[`s',2] if miss_`var'==1 & _mi_m>0 & arm==1
                  }
            *Modify cost
```

```
            replace totalcost=totalcost*mnar_param[`s',3] if miss_totalcost==1 & _mi_m>0 &
arm==0
            replace totalcost=totalcost*mnar_param[`s',4] if miss_totalcost==1 & _mi_m>0 &
arm==1
         *Calculate new QALY
            mi passive: replace qaly=0.125*qol_0+0.25*qol_3+0.375*qol_6+(0.25*1.965)*qol_12
+(0.5*0.965)*qol_18+(0.25*0.965)*qol_24
         *Append and save
            append using "10TT_MI_MNAR_cost.dta"
            save "10TT_MI_MNAR_cost.dta", replace
      }
```

3) Results

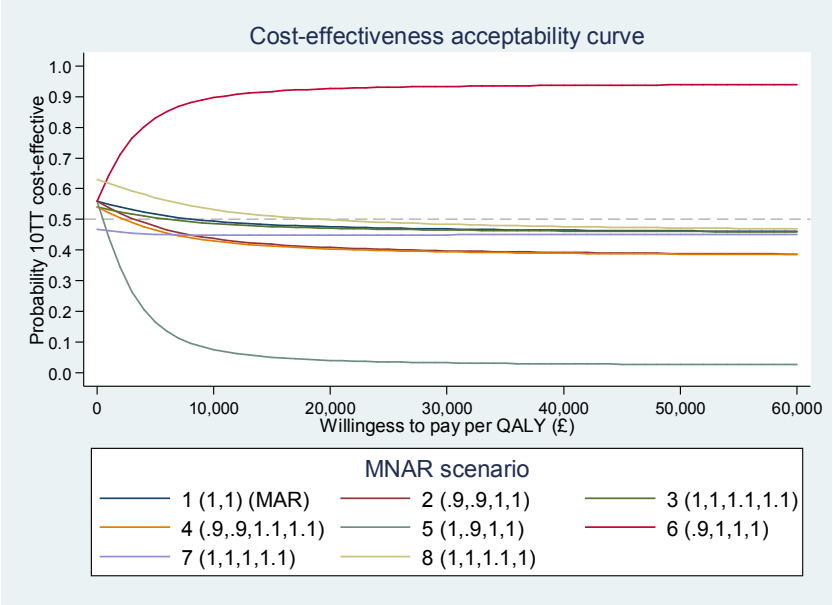
Scenario description	Incremental cost (£) [95% CI]	Incremental QALYs [95% CI]	INMB[a] (£) [95% CI]	Probability cost-effective[a]
1. MAR	-35 [-504 to 434]	-.004 [-.074 to .066]	-49 [-1,632 to 1,534]	48%
Same MNAR parameters[b] in the two arms				
2. -10% QoL in both arms	-35 [-504 to 434]	-.011 [-.078 to .057]	-181 [-1,714 to 1,352]	41%
3. +10% cost in both arms	-25 [-512 to 462]	-.004 [-.074 to .066]	-59 [-1,650 to 1,532]	47%
4. -10% QoL and +10% cost	-25 [-512 to 462]	-.011 [-.078 to .057]	-191 [-1,733 to 1,350]	40%
Different MNAR parameters[b] in the two arms				
5. -10% QoL in intervention arm	-35 [-504 to 434]	-.071 [-.139 to -.002]	-1378 [-2,932 to 176]	4%
6. -10% QoL in control arm	-35 [-504 to 434]	.056 [-.014 to .125]	1148 [-415 to 2,711]	93%
7. +10% cost in intervention arm	20 [-459 to 499]	-.004 [-.074 to .066]	-104 [-1,691 to 1,483]	45%
8. +10% cost in control arm	-80 [-558 to 398]	-.004 [-.074 to .066]	-4 [-1,591 to 1,583]	50%

MAR missing at random, MNAR missing not at random, QALY quality-adjusted life year, INMB incremental net monetary benefit, QoL quality-of-life, 10TT Ten Top Tips, CI confidence interval

[a] At a cost-effectiveness threshold of £20,000/QALY.

[b] How missing cost and QoL data are assumed to differ from MAR-imputed values.

4) Cost-effectiveness acceptability curve



We can see here that a departure from the MAR assumption for the costs is unlikely to affect significantly the findings, even if the missing costs are assumed 10% higher than under MAR only in the intervention arm.
However, departure from the MAR assumption for QoL could importantly affect the conclusions, particularly if the MNAR mechanism is not the same in each arm. The results for varying MNAR parameters for QoL, as reported in the Section 3 of the tutorial, is probably of primary interest in this case.

Online Appendix 4 - Probabilistic MNAR parameters

1) Distribution of the parameters

In this example, let us assume we believe the rescaling parameter *c* to be around 0.95, with a standard deviation of 0.025 (this standard deviation corresponds to being 95% certain that the true parameter value is somewhere between 1 and 0.90). We want to draw two correlated values from that distribution (c_{control} , $c_{10\text{TT}}$).

A correlation would capture how the values of c_{control} and $c_{10\text{TT}}$ are related, for example if the departure from MAR is strong in one arm, it could be more likely to also be the case in the other arm (positive correlation). The difficulty of eliciting the correlation parameter has been discussed elsewhere[1,2]. One solution is to simply assume independence. Indeed, this should result in a slightly conservative estimate for the difference between arms (assuming the correlation is usually positive), and the difference will typically be negligible[1,2] (to confirm this, the analysis could also be repeated with different correlations).

We will therefore draw two parameters from the following distribution:

$$\begin{pmatrix} c_{Control} \\ c_{10TT} \end{pmatrix} \sim N\left(\begin{pmatrix} 0.95 \\ 0.95 \end{pmatrix}, 0.025^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

2) Analysis implementation

We incorporated the random draw as part of the multiple imputation procedure, by drawing a different set of (c_{control} , $c_{10\text{TT}}$) for each of the imputed dataset, and rescaling each dataset accordingly. The standard analysis of the multiply imputed datasets (i.e. Rubin's rules) should then take into account of both the imputation and the MNAR uncertainty, and approximate a fully Bayesian analysis[1]. Note that a sufficient number of imputations are needed to perform multiple draws of (c_{control} , $c_{10\text{TT}}$), to obtain sufficiently stable results (negligible Monte Carlo error).

3) Stata code

```
** Define parameters distribution
   matrix C = (1,0.0,1) //Uncorrelated draw
   global mu = 0.95
   global sd = 0.025

** Modify MI data
   //Each MI dataset is MNAR-modified according to parameters drawn from the distribution
   use "10TT_MI.dta", clear

   *Draw random parameters
      set seed 1234  //seed for reproducibility
      drawnorm c0 c1, corr(C) cs(lower) //Draw values from 2 correlated normal distribution
      replace c0 = $mu + $sd *c0  // Transform to wanted mean and SD
      replace c1 = $mu + $sd *c1
      bysort _mi_m: replace c0 = c0[1]  //Same parameter value for each imputed dataset
      bysort _mi_m: replace c1 = c1[1]

   *Modify QoL values
      foreach var of varlist qol_3 qol_6 qol_12 qol_18 qol_24 {
         replace `var'=`var'*c0 if miss_`var'==1 & arm==0
         replace `var'=`var'*c1 if miss_`var'==1 & arm==1
         }
```

[1] White I.R., et al. "Eliciting and using expert opinions about dropout bias in randomized controlled trials." *Clinical Trials* 4.2 (2007): 125-139.

[2] Carpenter J., and Kenward M., "Multiple imputation and its application." John Wiley & Sons, 2012.

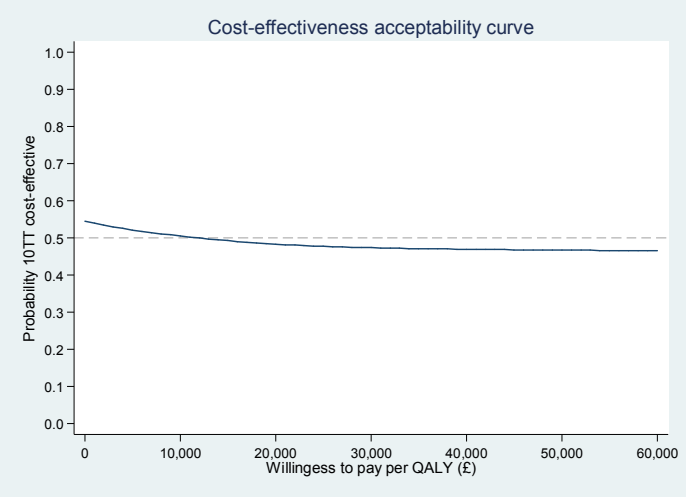
```
   *Calculate modified QALY
      mi passive: replace qaly=0.125*qol_0 + 0.25*qol_3 + 0.375*qol_6 + (0.25*1.965)*qol_12 +
(0.5*0.965)*qol_18 + (0.25*0.965)*qol_24

   *Save
      save "10TT_MI_MNAR_probabilistic.dta", replace
```

4) Results

Scenario description	Incremental cost (£) [95% CI]	Incremental QALYs [95% CI]	INMB[a] (£) [95% CI]	Probability cost-effective[a]
Probabilistic MNAR parameters	-35 [-504 to 434]	-.004 [-.085 to .076]	-50 [-1816 to 1716]	47.8%

QALY quality-adjusted life-years, INMB incremental net monetary benefit

[a] At a cost-effectiveness threshold of £20,000/QALY.



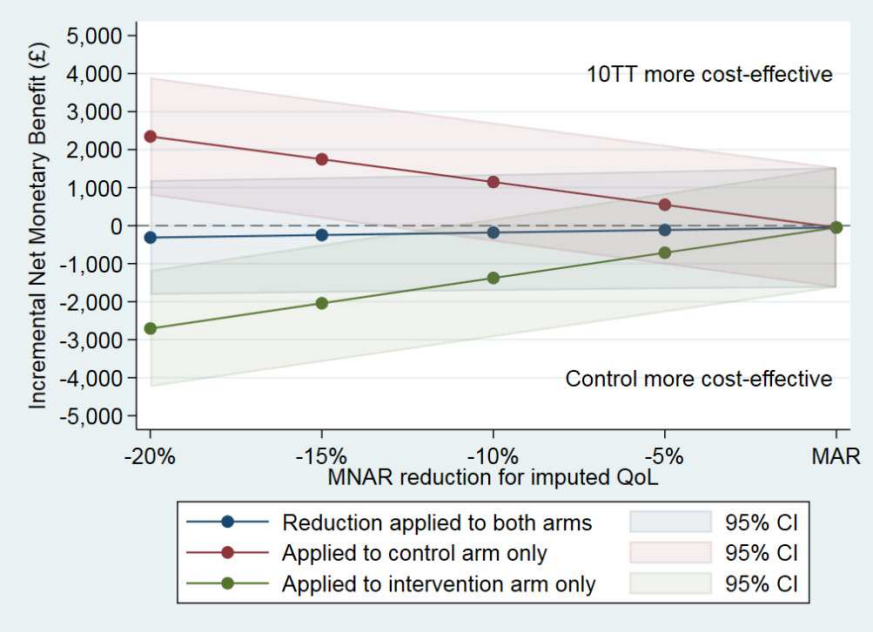
We can see that the resulting incremental estimates are very close to the MAR results. This is likely to be the case when i) the proportion of missing data is relatively similar between arms, and ii) the MNAR parameters have the same mean value in each arm.

However, the probabilistic analysis increases the uncertainty (wider confidence intervals, less steep CEAC), although this is not strongly seen here.

Note that the *within* arm QALYs estimates would differ from those under MAR (not shown here). The rescaling parameter is assumed to be around 0.95, meaning missing QoL are expected to be somewhat lower than under MAR, resulting in lower QALYs (in both arms).

Online Appendix 5 – Presentation of results

1) Net Monetary Benefit over range of MNAR parameter values



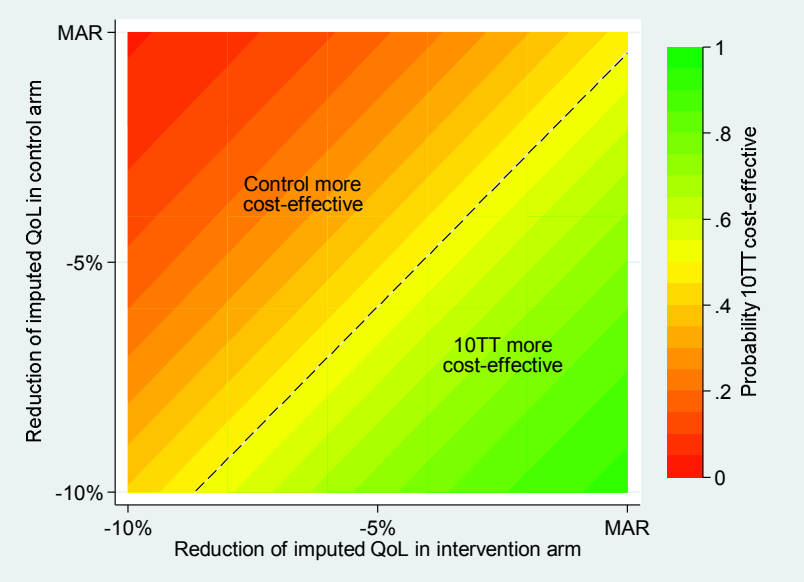
Note: example shown here with INMB, but could also display other results such as ICER or probability of being cost-effective.

Stata code:

```
** Calculate results for different MNAR parameters

postfile inbgraph c c0 c1 armscat inb ll ul using "inbgraph.dta", replace  //Set up 'postfile'
to save results

forvalues c=0.8 0.85 to 1.0 { // Range of MNAR parameter values
forvalues armscat=1/3 {       // 3 series: applying MNAR parameter to either arm, or both
   local c0=cond(`armscat'==1 | `armscat'==2,`c',1)
   local c1=cond(`armscat'==1 | `armscat'==3,`c',1)
   display "MNAR param = ( " `c0' " ; " `c1'" )"
   *Modify QoL values
      use "10TT_MI.dta" , clear
      foreach var of varlist qol_3 qol_6 qol_12 qol_18 qol_24 {
         replace `var'=`var'*`c0' if miss_`var'==1 & _mi_m>0 & arm==0
         replace `var'=`var'*`c1' if miss_`var'==1 & _mi_m>0 & arm==1
         }
   *Calculate modified QALY and INB
      mi passive: replace qaly=0.125*qol_0 + 0.25*qol_3 + 0.375*qol_6 + (0.25*1.965)*qol_12 +
(0.5*0.965)*qol_18 + (0.25*0.965)*qol_24
      mi passive: generate inb=qaly*20000-totalcost
   *Do MI analysis and save results
         mi estimate: regress inb arm
         matrix res=r(table)
         post inbgraph (`c') (`c0') (`c1') (`armscat') (res[1,1]) (res[5,1]) (res[6,1])
   }
   }
postclose inbgraph

** Graph set up
   use "inbgraph.dta", clear
   *Reshape to have one row by parameter (and 3 series, by 'armscat')
      drop c0 c1
      tab armscat
      reshape wide inb ll ul, i(c) j(armscat)
      list, noobs
   *Convert to "-%" for presentation
      gen cperc=string(-(1-c), "%8.2f")+"%"
```

```
** Graph
   graph twoway (connected  inb1 c) (connected inb2 c) (connected inb3 c) ///
              (rarea ll1 ul1 c, color(navy%10)) ///
              (rarea ll2 ul2 c, color(maroon%10)) ///
              (rarea ll3 ul3 c, color(forest_green%10)), ///
                yline(0.5,lc(gs5) lw(thin) lpattern(dash)) ///
                text(-4000 1 "Control more cost-effective", placement(w)) text(4000 1 "10TT
more cost-effective", placement(w)) ///
                xtitle("MNAR reduction for imputed QoL")  xlabel(0.80 "-20%" 0.85 "-15%" 0.90
"-10%" 0.95 "-5%"  1.0 "MAR") ///
                ytitle("Incremental Net Monetary Benefit (£)") ylabel(-5000(1000)5000, angle(0)
format(%9.0fc)) ///
                legend(order(1 4 2 5 3 6) label(1 "Reduction applied to both arms") label(2
"Applied to control arm only") label(3 "Applied to intervention arm only") label(4 "95% CI")
label(5 "95% CI") label(6 "95% CI"))
```

2) Contour plot for the probability of 10TT being cost-effective by MNAR parameter



Note: example shown here with probability of being cost-effective, but could also display other results such as INMB or ICER.

Stata code:

```
** Calculate results for different MNAR parameters

postfile contour c0 c1 inmb ll ul pce using "contour_graph.dta", replace  //Set up 'postfile'

forvalues c0=0.9 0.92 to 1.0 {
   forvalues c1=0.9 0.92 to 1.0 {
      *Modify QoL values
         use "10TT_MI.dta" , clear
         foreach var of varlist qol_3 qol_6 qol_12 qol_18 qol_24 {
            replace `var'=`var'*`c0' if miss_`var'==1 & _mi_m>0 & arm==0
            replace `var'=`var'*`c1' if miss_`var'==1 & _mi_m>0 & arm==1
            }
      *Calculate modified QALY and INB
         mi passive: replace qaly=0.125*qol_0 + 0.25*qol_3 + 0.375*qol_6 + (0.25*1.965)*qol_12
+ (0.5*0.965)*qol_18 + (0.25*0.965)*qol_24
         mi passive: generate inb=qaly*20000-totalcost
      *Do MI analysis and save results
         mi estimate: regress inb arm
         matrix res=r(table)
         local pce = normal(res[1,1]/res[2,1])
         post contour (`c0') (`c1') (res[1,1]) (res[5,1]) (res[6,1]) (`pce')
         //Saving INMB, 95%CI, and probability cost-effective (at £20,000/QALY)
      }
```

```
   }
postclose contour

** Graph
** Set up
   use "contour_graph.dta",clear
   label var c0 "MNAR parameter in control arm "
   label var c1 "MNAR parameter in intervention arm"
   label var pce "Probability 10TT cost-effective"

** Graph
   graph twoway (contour pce c0 c1, ccuts(0(0.05)1) zlabel(#6) scolor(red) ecolor(green)) ///

              (contourline pce c0 c1, ccuts(0.5) colorlines lpattern(dash)), ///
              ytitle("Reduction of imputed QoL in control arm") ylabel(0.90 "-10%" 0.95 "-5%"
1.0 "MAR", angle(0)) ///
              xtitle("Reduction of imputed QoL in intervention arm") xlabel(0.90 "-10%" 0.95 "-
5%" 1.0 "MAR")  ///
              text(0.965 0.935 "Control more" "cost-effective") text(0.93 0.975 "10TT more"
"cost-effective")
```

Appendix F

Reference-based - Covariance matrix derivation

Appendix - Covariance matrix derivation for J2R and CIR

Let μ_A, μ_R, Σ_A, and Σ_R denote the mean vectors and covariance matrices under MAR for the active and reference arm respectively. For any individual, these matrices can be partitioned according to whether the corresponding variables are observed (1), MAR-missing (2), or MNAR-missing (3).

$$\text{So that } \Sigma_A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,1} & A_{3,2} & A_{3,3} \end{bmatrix} \text{ and } \Sigma_R = \begin{bmatrix} R_{1,1} & R_{1,2} & R_{1,3} \\ R_{2,1} & R_{2,2} & R_{2,3} \\ R_{3,1} & R_{3,2} & R_{3,3} \end{bmatrix}$$

Let us denote

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \Sigma_{1,3} \\ \Sigma_{2,1} & \Sigma_{2,2} & \Sigma_{2,3} \\ \Sigma_{3,1} & \Sigma_{3,2} & \Sigma_{3,3} \end{bmatrix} \right)$$

the distribution for an individual in the active arm assumed to "jump-to-reference" for Y_3. That is we want the distribution of (Y_1, Y_2) to follow the distribution from the active arm, and the conditional distribution of Y_3 given (Y_1, Y_2) to follow the corresponding conditional distribution from the reference arm.

The distribution of (Y_1, Y_2) is straightforward:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_{A1} \\ \mu_{A2} \end{bmatrix}, \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \right)$$

Let $_{12}$ denotes the joined elements of the observed and MAR-missing variables. In the reference arm the conditional distribution $[Y_3|Y_{12} = y_{12}]$ is given by

$$N\Big(\Big[\mu_{A3} + (y_{12} - \mu_{A12})^T R_{12,12}^{-1} R_{12,3} \Big], \Big[R_{3,3} - R_{3,12} R_{12,12}^{-1} R_{12,3} \Big] \Big)$$

Following the same decomposition for the jump-to-reference distribution, we have the following constraints on the covariance parameters:

$$\Sigma_{12,12} = A_{12,12} \tag{1}$$

$$\Sigma_{12,12}^{-1} \Sigma_{12,3} = R_{12,12}^{-1} R_{12,3} \tag{2}$$

$$\Sigma_{3,3} - \Sigma_{3,12} \Sigma_{12,12}^{-1} \Sigma_{12,3} = R_{3,3} - R_{3,12} R_{12,12}^{-1} R_{12,3} \tag{3}$$

Which can be resolved by

$$\Sigma_{12,12} = A_{12,12} \tag{4}$$

$$\Sigma_{12,3} = A_{12,12} R_{12,12}^{-1} R_{12,3} \tag{5}$$

$$\Sigma_{3,3} = R_{3,3} - R_{3,12} R_{12,12}^{-1} (R_{12,12} - A_{12,12}) R_{12,12}^{-1} R_{12,3} \tag{6}$$

The J2R distribution is therefore:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_{A1} \\ \mu_{A2} \\ \mu_{R3} \end{bmatrix}, \begin{bmatrix} A_{1,1} & A_{1,2} & \Sigma_{1,3} \\ A_{2,1} & A_{2,2} & \Sigma_{2,3} \\ \Sigma_{3,1} & \Sigma_{3,2} & \Sigma_{3,3} \end{bmatrix} \right)$$

where $\Sigma_{12,3}$, $\Sigma_{3,12}$, and $\Sigma_{3,3}$ are defined in (5) and (6).

For CIR, the covariance matrix will be defined similarly, but the mean for Y_3 will be defined by the *increment* in mean from the reference arm. For MAR, LMCF, and BMCF the covariance matrix is that of the active arm.

Appendix G

Reference-based - CEmimix Stata code

```
set more off

version 15

cd "H:\_PhD\6_Reference-based\Stata\CEmimix"

capture: log close
local t = substr("$S_TIME",1,2)
log using "..\log\ce_mimix_`c(current_date)'_`t'.log", replace


*************************************************************
*** REFERENCE-BASED IMPUTATION OF COST-EFFECTIVENESS DATA ***
*************************************************************

*********************************************************************
//Program to conduct reference-based multiple imputation with cost-effectiveness data
//See acommpanying instructions on using the do file

* Version:
* Author: BL, based on mimix program originally developed by Cro et al. (Stata J. 2016 16(2):443-463)
* Date: 01 October 2018
* Stata version 15

** CONTENT:
  * I  -  SET-UP
  * II -  DEFINE ROUTINES
  * III - PREPARE DATA FOR IMPUTATION
  * IV -  RUN MVN
  * V  -  MNAR IMPUTATION, FOR EACH ARM AND PATTERN
  * VI -  SAVE AS MI DATASET

*********************************************************************



**********************
***  I - SET-UP     ***
**********************
//Define here program parameters (dataset,varaible names, imputation method, etc.)
//Please refer to the do-file instructions

  macro drop _all


** Parameters **

  global m = 2 //Number of imputations

  global emethod  J2R   //MAR J2R CIR LMCF BMCF
  global cmethod  MAR   //MAR J2R CIR LMCF BMCF
  global interimMAR effect cost   // effect, cost, or leave blank

  global restrictto  //Restrict MNAR imputation to a specific subgroup (e.g. arm==1). Leave blank otherwise.

  global seed   //Specify seed for reproducibility. Leave blank for random seed

  *Set-up for COBALT dataset:
    global data ..\COBALT\COBALT_short.dta
    global effectv  eq5d0 eq5d6 eq5d12
    global costv  tcost
    global covariates  age sex
    global idv   ptidno
    global treatv   arm
    global refgroup 0
    global saving "data/COBALT_imputed.dta"



** Check parameters
  //Basic error checks

  if !inlist("$emethod","MAR","J2R","CIR","LMCF","BMCF") | !inlist("$cmethod","MAR","J2R","CIR","LMCF","BMCF") {
    display as error "Please specify imputation method for effect and cost: MAR J2R CIR LMCF or BMCF"
    exit
    }
  if inlist("$emethod","J2R","CIR","LMCF","BMCF") & inlist("$cmethod","J2R","CIR","LMCF","BMCF") & "$emethod"!="$cmethod" {
    display as error "Different MNAR mechanisms for effect and costs not allowed"
    exit
    }
  if !strpos("$interimMAR", "effect") & !strpos("$interimMAR", "cost") & "$interimMAR"!="" {
    display as error " 'interimMAR' should be 'effect', 'cost', or nothing"
    exit
    }
  if (inlist("$emethod","J2R","CIR") | inlist("$cmethod","J2R","CIR")) & "$refgroup"=="" {
    display as error "Please specify reference group for CIR or J2R"
    exit
    }
  if "$idv"=="" | "$treatv"=="" {
    display as error "Please specify treatment arm and patient identifier variables"
    exit
    }
```

```
93
94
95
96  ********************************
97  ***  II - DEFINE ROUTINES     ***
98  ********************************
99    //Define Mata functions used in imputation step
100
101     mata: mata clear
102
103 ** Mata functions to manipulate list of variables
104   mata
105     // Common : Returns common elements between 2 vectors
106       real vector common(real vector V1, real vector V2)
107       {
108         st_local("v1",invtokens(strofreal(V1)))
109         st_local("v2",invtokens(strofreal(V2)))
110         stata("local l2: list v1 & v2")
111         res=strtoreal(tokens(st_local("l2")))
112         return(res)
113       }
114     // Join: Returns elements in either of 2 vectors
115       real vector join(real vector V1, real vector V2)
116       {
117         st_local("v1",invtokens(strofreal(V1)))
118         st_local("v2",invtokens(strofreal(V2)))
119         stata("local l2: list v1 | v2")
120         res=sort(strtoreal(tokens(st_local("l2")))',1)'
121         return(res)
122       }
123     // Exclude: Returns elements of V1, not contained in V2.
124       real vector exclude(real vector V1, real vector V2)
125       {
126         st_local("v1",invtokens(strofreal(V1)))
127         st_local("v2",invtokens(strofreal(V2)))
128         stata("local l2: list v1 - v2")
129         res=sort(strtoreal(tokens(st local("l2")))',1)'
130         return(res)
131       }
132   end
133
134 ** Mata function to build conditional covariance matrix
135   //Used for J2R and CIR imputation
136   //Build joint covariance matrix, so that MNAR-missing variables follow distribution from reference arm,
    conditionally on observed or MAR-missing variables.
137   //Parameters = covariance matrix in active arm; covariance in reference arm; indicator of observed or MAR
    variables; indicator of MNAR-missing varibales.
138   //See technical details in Appendix
139   mata
140     real matrix condcov(real matrix SigmaA, real matrix SigmaR, real vector vobsmar, real vector vmnar)
141       {
142       A11 = SigmaA[vobsmar,vobsmar] //Decompose var/covar in active and reference arm
143       R11 = SigmaR[vobsmar,vobsmar]
144       R12 = SigmaR[vobsmar,vmnar]
145       R22 = SigmaR[vmnar,vmnar]
146       J11=A11  //Solve contraints (see Appendix)
147       J12=A11*invsym(R11)*R12
148       J22=R22-(R12)'*invsym(R11)*(R11-A11)*invsym(R11)*R12
149       J = J(cols(SigmaA),cols(SigmaA),.)  //Build joint covariance matrix
150       J[vobsmar,vobsmar]=J11
151       J[vobsmar,vmnar]=J12
152       J[vmnar,vobsmar]=J12'
153       J[vmnar,vmnar]=J22
154     return(J)
155     }
156   end
157
158
159
160 ********************************************
161 *** III - PREPARE DATA FOR IMPUTATION    ***
162 ********************************************
163
164 *** Open original dataset
165   use "$data" , clear
166   describe
167   list in 1/5, noobs
168
169 *** Prepare macros and variables for program
170
171   **Global macros
172     *Seed
173       capture: set seed $seed
174         //Affect random draws in MVN, and imputation steps. Will obtain same draws with same sorted data (and do
    file)
175     *Outcomes list
176       global responses $effectv $costv
177     *Number of variables
178       global nresp: word count $responses
179       global ncov: word count $covariates
180       global nvar = $nresp + $ncov  //"nct" in mimix
181     *Number of treatment group
182       tab $treatv
```

```
183         global ntreat = r(r)
184       *First effectiveness and cost variable
185         //Note: in mata, variables order always $effectv $costv $covariates
186         global v1effect=1
187         local neff: word count $effectv
188         global v1cost= `neff' + 1  // Cost var = first after effectiveness vars.
189       *Interim-MAR option
190         if strpos("$interimMAR", "effect") global eintmeth iMAR
191           else global eintmeth $emethod
192         if strpos("$interimMAR", "cost") global cintmeth iMAR
193           else global cintmeth $cmethod
194       *Check
195         macro list
196
197     ** New variables
198       *Treatment arm variable
199         egen m_treat=group($treatv)  //Recoding = 1,2,..
200         tab m_treat
201       *New reference-group code
202         if "$refgroup"!="" sum m_treat if $treatv==$refgroup
203         global m_refer=r(max)
204         display "New reference arm code = $m_refer"
205       *Observation ID
206         if substr("`:type $idv'",1,3)=="str" encode($idv), generate(m_id)
207         else gen m_id = $idv
208         duplicates report m id
209       *Missing data pattern
210         qui: generate m_pattern = 0
211         local i=0
212         foreach var of varlist $responses  {
213           local k2 = 2^(`i++')  //Will assign a unique number by pattern, for any number of variables.
214           qui: replace m_pattern = m_pattern + `k2'  if `var'== .
215           }
216         tab m_pattern m_treat ,m
217       *MNAR subgroup
218         //Use specified method by default, except if not part of "restcitto" ...
219         gen m_allmar=0
220         if "$restrictto"!="" replace m_allmar= !($restrictto) //AllMAR=1 only if observation not in "restrictto"
    subgroup.
221         qui: count if m_allmar==0
222         if "$restrictto"!="" display "MNAR imputation restricted to `r(N)' observations out of " _N
223         if `r(N)'==0  display as error "No observations MNAR-imputed - Check 'restrictto' option"
224
225   *** Sort and save
226     *Save dataset
227       //Original dataset + system variables. Will be used to merge with imputed data at the end
228       sort m_id
229       compress
230       save "originalext.dta", replace
231
232     *Save reduced version for imputation
233       keep m_id m_treat $responses $covariates m_pattern m_allmar
234       order m_id m_treat $responses $covariates m_pattern m_allmar
235       sort m_treat m_pattern m_allmar m_id //Sort by treat arm, missing data pattern, then PID.
236       compress
237       save "m_d2.dta", replace
238
239
240   ************************
241   ***  IV -  RUN MVN    ***
242   ************************
243    // Fit a multivariate normal model to the observed data, for each arm
244    // Then draw mean/covariance parameters from their posterior distribution
245
246
247   ** Set-up MCMC burn-in parameters
248     local burnin = 100  //Number of iterations for the initial burn-in period
249     local burnbetween 100  //Number of iterations between imputation
250     local burninM = `burnin' + (($m-1)*`burnbetween')  //Total number of iterations
251
252   *** Run MVN for each treament arm, and save parameters
253      forvalues i = 1/$ntreat {
254       **Set-up
255         use "m_d2.dta" if m_treat == `i', clear
256         mi set wide  //Wide faster, can set to mlong if size error
257         qui: mi register imputed $responses $covariates
258
259       **MVN
260         display as text "Performing imputation procedure for arm " as result "`i'" as text " of " as result
    "$ntreat" as text "..."
261         mi impute mvn $responses $covariates , mcmconly burnin(`burninM')  prior(jeffreys) initmcmc(em, iter(1000))
    saveptrace(mimix_parms_a`i', replace)
262           //Note: Used only to fit MVN model and save trace, not doing imputation.
263
264       **Save parameters
265         //Using values from the MCMC trace. Saving every 'burnbetween' iteration is like doing random draws from
    from posterior distribution of the parameters
266
267         *Open trace
268           mi ptrace describe mimix_parms_a`i'
269           mi ptrace use mimix_parms_a`i', clear
270
271         *Save every 100 iterations:
```

```
272           local burn = `burnin' - 1
273           drop in 1/`burn'
274           keep if !mod(_n-1,`burnbetween')
275           generate m_treat = `i'
276           drop m iter
277           capture mata: mata drop mimix_all
278           mata: mimix_all= st_data( ., .) //Copy dataset (all params, m_treat) into mimix_all
279         *Save mean and covariance in matrices, for each m:
280           forvalues k=1/$m {
281             display _n " Draw for group `i', imputation `k' "
282             *Save mean matrice:
283               mata: mean_group`i'_imp`k' = mimix_all[`k',1..$nvar]
284                 *mata: mean_group`i'_imp`k'
285             *Save covariance matrices:
286               mata: mata VAR group`i' imp`k'=J($nvar,$nvar,0)
287               local step = $nvar+ 1
288               forvalues r = 1/$nvar {
289                 forvalues j = 1/$nvar{
290                   if `j' <= `r' {
291                     mata: mata_VAR_group`i'_imp`k'[`r', `j'] = mimix_all[`k', `step']
292                     local step = `step' + 1
293                   }
294                 }
295               }
296               mata: mata_VAR_group`i'_imp`k' = makesymmetric(mata_VAR_group`i'_imp`k')
297                 *mata: mata VAR group`i' imp`k'
298           } //End of saving mean and cov matrices
299
300     }  //End of MVN loop.
301
302
303
304   ****************************************************************
305   ***  MNAR IMPUTATION, FOR EACH ARM AND MISSIGN DATA PATTERN   ***
306   ****************************************************************
307
308   **** Set up
309
310     ** Describe data
311       use "m_d2.dta", clear
312       describe
313       tab m_pattern m_treat,m
314
315     ** Save characteristics of each arm+pattern group
316
317       *First and last observation
318         gen n=_n
319         bysort m_treat m_pattern m_allmar: gen nfirst=n[1]
320         bysort m_treat m_pattern m_allmar: gen nlast=n[_N]
321       *Number of missing var
322         egen nmiss=rowmiss($responses $covariates)
323       *Contract
324         contract m_treat m_pattern m_allmar nfirst nlast nmiss
325         rename _freq ncount
326         gen groupID=_n
327       *Order var and save in a matrix
328         mkmat m_treat m_pattern m_allmar ncount nfirst nlast nmiss groupID , mat(m_group)
329         matrix list m_group
330       *Save number of combinations/groups
331         global max_indicator=_N
332
333     ** Indicator of effect/cost/MAR/MNAR variables
334         mata: mata_responses=J(1,0,.)
335         mata: mata_eff=J(1,0,.)
336         mata: mata_cost=J(1,0,.)
337         mata: mata_meth_mar=J(1,0,.)
338         mata: mata_meth_mnar=J(1,0,.)
339         local j=0
340         foreach var in $responses $covariates {  //Note: Variables identified by their position, use always same
      order
341           local j=`j'+1
342           if strpos("$responses","`var'")  mata: mata_responses=(mata_responses,`j')
343           if strpos("$effectv","`var'")  mata: mata_eff=(mata_eff,`j')
344           if strpos("$costv","`var'")    mata: mata_cost=(mata_cost,`j')
345           if strpos("$effectv","`var'")*("$emethod"=="MAR") | strpos("$costv","`var'")*("$cmethod"=="MAR") {
346             mata: mata_meth_mar=(mata_meth_mar,`j')
347             }
348           if strpos("$effectv","`var'")*("$emethod"!="MAR") | strpos("$costv","`var'")*("$cmethod"!="MAR") {
349             mata: mata_meth_mnar=(mata_meth_mnar,`j')
350             }
351           }
352
353
354     ** Empty matrix to save imputed data
355       global new varlist m treat m $responses $covariates m id  //List of variables to be saved after each
      mata-imputation (used when converting back to Stata)
356       mata: mata_all_new=J(0,$nvar+3,.) // Size= nvar+3(treat,m,ID)
357
358
359   **** Begining of "for each imputation group" loop
360     //Split data in imputation groups ( = arm + missing data pattern).
361     //For each group do:  1) Build joint distribution from MAR parameters 2) Draw missing values from that
      distribution 3) Redo 1-2 M times.
```

```
362     //Note: large loop, encompasses "foreach imputation" loop, see below.
363
364     forvalues i= 1/$max_indicator {  //For each imputation group
365       *local i=11 //COBALT 10: last qol and cost missing, in active arm. 11: 2 qol + cost missing. 9 = interim +
      cost missing
366
367       display _n "--- Imputation for group `i' of $max_indicator ---"
368
369       ** Set up
370         //Group charateristics, before going into "for each m" loop.
371
372         *Save group characteristics
373           matrix list m_group
374           local trt_grp= m_group[`i',1]
375           local pattern = m_group[`i',2]
376           local allmar= m_group[`i',3]
377           local ncount= m_group[`i',4]
378           local nfirst= m_group[`i',5]
379           local nlast= m_group[`i',6]
380           local miss_count= m_group[`i',7]
381           local refer = $m_refer  //Note: reference arm currently same for everyone, but allow to change if needed.
382
383         *Indicator of complete/missing var
384           qui: use m_d2.dta, clear
385           mata: mata_miss = J(1,0,.)
386           mata: mata nonmiss = J(1,0,.)
387           local j=0
388           foreach var of varlist $responses  $covariates {
389             local j=`j'+1
390             if (`var'[`nfirst']>=.)  mata: mata_miss=(mata_miss,`j')
391               else  mata: mata_nonmiss=(mata_nonmiss,`j')
392             }
393
394         *Indicator of interim-MAR missing
395           *Last observed cost/effect:
396             mata: st_numscalar("lastobse",rowmax((common(mata_eff,mata_nonmiss),0))) //Adding a 0 so is "0" if
      empty matrix.
397             mata: st_numscalar("lastobsc",rowmax((common(mata_cost,mata_nonmiss),0))) //Adding a 0 so is "0" if
      empty matrix.
398           *Testing whether interim (+MAR option specified), for each missing variable:
399             mata: st_local("misslist",invtokens(strofreal(mata_miss)))
400             mata: mata_int_mar = J(1,0,.)
401             foreach v of local misslist {
402               if (`v'>=$v1effect & `v'<lastobse & "$eintmeth"=="iMAR" )  | (`v'>=$v1cost & `v'<lastobsc &
      "$cintmeth"=="iMAR"){
403                 mata: mata_int_mar=(mata_int_mar,`v')
404                 }
405               }
406           *Check
407             mata: mata_int_mar
408
409         *Indicator of forced-MAR variables
410           //If "restricto" specified, impute all var under MAR for observations not in that subgroup.
411           if `allmar'==1 mata: mata_allmar=mata_responses
412           else mata: mata_allmar=J(1,0,.)
413
414         *Identify MAR-missing variables
415           //Will be MAR if either i)Main imputation-method=MAR or ii) is interim-MAR or iii) participant not in
      "restrictto" subgroup
416           //Note: use mata "common" and "join" functions defined above
417           mata: mata mar2=join(mata meth mar,join(mata int mar,mata allmar))
418           mata: mata_marmiss=common(mata_mar2,mata_miss) // MAR and actually missing. Will be those MAR-imputed for
      that pattern.
419
420         *Identify MNAR-missing variables
421           // Main-method-MNAR, except for interim-MAR missing or if participant not in "restrictto" subgroup.
422           mata: mata_mnar2=exclude(mata_meth_mnar,join(mata_int_mar,mata_allmar)) // Main-method-MNAR, except for
      interim-MAR missing or "allmar"
423           mata: mata_mnarmiss=common(mata_mnar2,mata_miss) // MNAR and actually missing. Will be those MNAR-imputed
      for that pattern.
424
425         *Indicator of any MNAR missing variables:
426           mata: st_local("n_mnar_miss",strofreal(cols(mata_mnarmiss)))
427
428         *Check all indicators:
429           display as txt _n "Variables imputation status for group `i' (var numbered as: effect,cost,covariates)"
430           display as txt "Observed:"
431             mata: mata_nonmiss
432           display as txt "MAR-missing:"
433             mata: mata_marmiss
434           display as txt "MNAR-missing:"
435             mata: mata_mnarmiss
436
437         *Save observed data
438           //Save responses,covariates,ID in a mata matrix
439           qui: use m_d2.dta, clear
440           qui: keep in `nfirst'/`nlast'
441           keep $responses $covariates m_id
442           order $responses $covariates m_id
443           mata: mata_obs= st_data( . , .)
444
445
446       *** Begining "for each imputation" loop
```

```
447
448         forvalues imp = 1/$m {
449             display "." _cont
450
451           ** If no missing data, copy data directly
452             if `miss_count' == 0 {
453               if `imp'==1 dis "No missing"
454               *Copy observed data
455                 mata: mata_new = (J(`ncount',1,`trt_grp'), J(`ncount',1, `imp'), mata_obs)  //Dataset with Arm +
    imp_number + observed data
456               *Append to existing
457                 mata: mata_all_new = (mata_all_new \ mata_new)
458               }
459
460           ** If missing data, build the joint distribution (mean vector, and covariance matrix)
461             else {
462               *All MAR
463                 if `n_mnar_miss'==0  {  // No MNAR missing
464                   if `imp'==1 dis "Imputation (Method = MAR)"
465                   mata: mata_Meansv=mean_group`trt_grp'_imp`imp'
466                   mata: Sigma = mata_VAR_group`trt_grp'_imp`imp'
467                   }
468               *J2R
469                 if (`n_mnar_miss'>0) & ("$emethod" == "J2R" | "$cmethod" == "J2R")  {  //If any cost or effect is
    J2R.
470                   if `imp'==1 dis "Imputation (Method = J2R)"
471                   *Mean
472                     mata: mata_Meansv=mean_group`trt_grp'_imp`imp'
473                     mata: mata_Meansv[1,mata_mnarmiss]=mean_group`refer'_imp`imp'[1,mata_mnarmiss]  //Replacing
    Mean from reference group for MNAR variables
474                   *Covariance
475                     mata: mata_nonmiss_marmiss=join(mata_nonmiss,mata_marmiss)  //Observed or MAR-missing variables.
476                     mata: Sigma=condcov(mata_VAR_group`trt_grp'_imp`imp', mata_VAR_group`refer'_imp`imp',
    mata_nonmiss_marmiss,mata_mnarmiss)
477                   }
478               *CIR
479                 if (`n_mnar_miss'>0) & ("$emethod" == "CIR" | "$cmethod" == "CIR")  {  //If any cost or effect is CIR
480                   if `imp'==1 dis "Imputation (Method = CIR)"
481                     **Mean
482                       mata: mata_Meansv=mean_group`trt_grp'_imp`imp'
483                       mata: MeansC=mean_group`refer'_imp`imp'
484                       *Effect
485                         mata: mata_mnarmiss_e=common(mata_mnarmiss,mata_eff) // Effectiveness var MNAR-missing
486                         mata: st_local("vlist",invtokens(strofreal(mata_mnarmiss_e)))
487                         foreach v of local vlist {
488                           if `v'==$v1effect mata: mata_Meansv[1,`v'] = MeansC[1,`v']  //If first var missing, copy
    from reference arm
489                           else mata: mata_Meansv[1,`v'] = mata_Meansv[1,`v'-1] + (MeansC[1,`v']-MeansC[1,`v'-1])
    //Previous mean (in current arm) + increment in mean in refer group
490                         }
491                       *Cost
492                         mata: mata_mnarmiss_c=common(mata_mnarmiss,mata_cost)
493                         mata: st_local("vlist",invtokens(strofreal(mata_mnarmiss_c)))
494                         foreach v of local vlist {
495                           if `v'==$v1cost mata: mata_Meansv[1,`v'] =  MeansC[1,`v']  //If first var missing, copy
    from reference arm
496                           else mata: mata_Meansv[1,`v'] = mata_Meansv[1,`v'-1] + (MeansC[1,`v']-MeansC[1,`v'-1])
    //Previous mean (in current arm) + increment in mean in refer group.
497                         }
498                     **Covariance
499                       mata: mata nonmiss marmiss=join(mata nonmiss,mata marmiss)  //Observed or MAR-missing
    variables.
500                       mata: Sigma=condcov(mata_VAR_group`trt_grp'_imp`imp', mata_VAR_group`refer'_imp`imp',
    mata_nonmiss_marmiss,mata_mnarmiss)
501                   }
502               *LMCF
503                 if (`n_mnar_miss'>0) & ("$emethod" == "LMCF" | "$cmethod" == "LMCF") { //Either effect or cost is
    LMCF
504                   if `imp'==1 dis "Imputation (Method = LMCF)"
505                   *Mean
506                     mata: mata_Meansv=mean_group`trt_grp'_imp`imp'
507                     *Effect
508                       mata: mata_mnarmiss_e=common(mata_mnarmiss,mata_eff) // Effectiveness variables MNAR-missing
509                       mata: st local("vlist",invtokens(strofreal(mata mnarmiss e)))
510                       foreach v of local vlist {
511                         if `v'>$v1effect {  //Note: if first var missing, use the mean
512                           mata: mata_Meansv[1,`v'] = mata_Meansv[1,`v'-1]   // Copying previous mean
513                         }
514                       }
515                     *Cost
516                       mata: mata_mnarmiss_c=common(mata_mnarmiss,mata_cost)
517                       mata: st_local("vlist",invtokens(strofreal(mata_mnarmiss_c)))
518                       foreach v of local vlist {
519                         if `v'>$v1cost {
520                           mata: mata Meansv[1,`v'] = mata Meansv[1,`v'-1]
521                         }
522                       }
523                   *Covariance
524                     mata: Sigma = mata_VAR_group`trt_grp'_imp`imp'  //Using MAR covariance from that arm
525                   }
526               *BMCF
527                 if (`n_mnar_miss'>0) & ("$emethod" == "BMCF" | "$cmethod" == "BMCF") { //Either effect or cost is
    BMCF
```

```
528                 if `imp'==1 dis "Imputation (Method = BMCF)"
529                 *Mean
530                   mata: mata_Meansv=mean_group`trt_grp'_imp`imp'
531                   *Effect
532                     mata: mata_mnarmiss_e=common(mata_mnarmiss,mata_eff) // Effectiveness variables MNAR-missing
533                     mata: st_local("vlist",invtokens(strofreal(mata_mnarmiss_e)))
534                     foreach v of local vlist {
535                       mata: mata_Meansv[1,`v'] = mata_Meansv[1,$v1effect]  // Copying mean of first variable
536                     }
537                   *Cost
538                     mata: mata_mnarmiss_c=common(mata_mnarmiss,mata_cost)
539                     mata: st_local("vlist",invtokens(strofreal(mata_mnarmiss_c)))
540                     foreach v of local vlist {
541                         mata: mata_Meansv[1,`v'] = mata_Meansv[1,$v1cost ]
542                     }
543                 *Covariance
544                   mata: Sigma = mata_VAR_group`trt_grp'_imp`imp'  //Using MAR covariance from that arm
545                 }
546
547             **Check joint distribution
548               *mata: mata_Meansv
549               *mata: Sigma
550
551             ** Perform imputation
552               * Expand mean vector to n observations
553                 mata: mata Means=J(`ncount', 1, mata Meansv)
554               * Decompose the covariance matrix observed/missing
555                 mata: S11 = Sigma[mata_nonmiss, mata_nonmiss] //Covariance observed var.
556                 mata: S12 = Sigma[mata_nonmiss, mata_miss] //Covariance for observed(row)Xmissing(col) var
557                 mata: S22 = Sigma[mata_miss, mata_miss]  //Covariances missing var
558               *Draw missing values conditionally on observed
559                 mata: m1=mata_Means[., mata_nonmiss]  //Mean param for all observed var (n times)
560                 mata: m2=mata_Means[., mata_miss]  //Mean param for all missing var (n times)
561                 mata: raw1=mata_obs[., mata_nonmiss] //Observed values matrix.
562                 mata: meanval = m2 + (raw1 - m1)*invsym(S11)*S12 //Expectation given observed values.
563                 mata:conds=S22-S12'*invsym(S11)*S12
564                 mata: U = cholesky(conds)
565                 mata: Z = invnormal(uniform(`ncount',`miss_count'))  //Drawn n*nmiss standard normal
566                 mata: mata_y1 = meanval + Z*U'  //Draw n X nmiss following N((cond mean),Covar). = Imputed values.
567               *Merge all variables
568                 mata: mata_new =J(`ncount',$nvar,.)  //Empty mat n*nvar
569                 mata: mata_new[.,mata_nonmiss] = mata_obs[.,mata_nonmiss] //Add observed val
570                 mata: mata_new[.,mata_miss] = mata_y1[.,.] //Add imputed val
571                 mata: GI=J(`ncount',1,`trt_grp')  //Treatment group
572                 mata: II=J(`ncount',1,`imp')   //Imputation number
573                 mata: ID = mata_obs[.,cols(mata_obs)] //Last column of mata_obs = ID
574                 mata: mata_new=(GI, II, mata_new, ID)
575               *Append to existing data
576                 mata: mata_all_new = (mata_all_new \ mata_new)
577
578             } //End of "if missing" loop.
579
580           } //End of "for m" loop
581
582     } //End of "for each group" loop
583
584
585   *** Check data
586     clear
587     getmata($new_varlist)=mata_all_new
588     describe
589     list in 1/5, header noobs
590     count
591     dis _N/$m //Check: number of obs in original dataset?
592
593
594
595   *********************************
596   ***** SAVE AS MI DATASET   ****
597   *********************************
598
599     *Prepare imputed data
600       clear
601       getmata($new varlist)=mata all new  //Convert mata "all new" to Stata
602       keep $responses m m_id //Other var will be in original dataset, no need to keep them here.
603       sort m_id m
604       tempfile imputedv
605       save `imputedv', replace
606
607     *Add other variables from original dataset
608       use originalext.dta, clear
609       count
610       sort m_id
611       merge 1:m m_id using `imputedv', nogen update
612       count //OK, N*$m
613
614     *Add _m=0 (=observed data)
615       append using originalext.dta
616       replace m=0 if m==.
617
618     *Convert to MI format
619       mi import flong , m(m) id($idv) clear
620       mi register imputed $responses $covariates
```

```
621       mi describe
622       list in 1/5
623
624     *Clean and save
625       describe
626       drop m_treat-m  //Drop programming var
627       sort $idv _mi_m
628       list in 1/10, sepby($idv)
629       compress
630       label data "Reference-based imputed ($emethod-$cmethod) - `c(current_date)'"
631       save "$saving", replace
632         //! Will overwrite dataset if already exist.
633
634     ** Delete temporary datasets
635       //Temporary datasets created for the program
636       erase originalext.dta
637       erase m_d2.dta
638       forvalues i = 1/$ntreat {
639         erase mimix_parms_a`i'.stptrace
640         }
641
642
643
644
645   ***** END *****
646   capture: log close
647
```

Appendix H

Reference-based - CEmimix instructions

Supporting material 2

Using CEmimix.do

I) Stata version

The do-file needs Stata software[1] to run. It was developed with Stata version 15, and should work with subsequent versions (type "`version 15`" at beginning of execution). It may work with other recent versions but not earlier than 11, when `mi impute` was implemented.

II) Data input format

The input dataset needs to be store in Stata 'wide' format. That is, with a single record per participant. The effectiveness and cost data should be stored as separate variables for each time-point. It also needs to include a treatment arm and patient identifier variable.

III) Set up

The user will need to define the following global macros at the beginning of the do-file:

Required information:

m: The number of imputation

emethod: Imputation method for the effectiveness variable. Should be one of **MAR**, **J2R**, **CIR**, **LMCF**, or **BMCF**

cmethod: Imputation method for the cost variable. Should be one of **MAR**, **J2R**, **CIR**, **LMCF**, or **BMCF**

data: Name of dataset to be imputed. Can use path if dataset is not in working directory. E.g. `"Data\mydata.dta"`

effectv: Effectiveness variable(s), in chronological order[2]. E.g. `eq5d0 eq5d6 eq5d12`

costv: Cost variable(s), in chronological order. E.g. `tcost0 tcost6 tcost12`

idv : Unique patient identifier variable. Can be either numerical or string.

treatv : Trial treatment arm variable. Need to be numerical. Multiple arms are allowed.

Optional:

[1] StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC

[2] The temporal order of the variables is used for CIR, LMCF, BMCF, and for defining interim-missing. For BMCF, the first variable need to be the baseline measurement.

refgroup : Reference group number (as coded in *treatv*). Need to be defined for J2R and CIR, can be left blank otherwise.

interimMAR: Whether interim-missing data should be imputed under MAR. Can be **cost** and/or **effect**. If not specified, uses the imputation method specified above.

covariates: Additional variables used in the imputation model. Covariates need to be complete[3] and numerical. Categorical variables need to be split in dummy (binary) variables.

seed: Specifies the seed for the random number generator. If a seed is specified (and data are sorted in same order), the program will return the same results on separate runs. If nothing is specified, a random seed is used and different runs will result in a different set of imputed data.

restrictto: Specify a subgroup on which to restrict the reference-based imputation. Other individuals are imputed under MAR. For example, could restrict MNAR assumption to a single arm or to participants who dropped-out for a specific reason. E.g. `(arm==1 & reason=="withdraw")`.

using: Specify file name to save imputed dataset. Note that it will overwrite file with same name if already exist.

This do-file was primarily designed to be run on its' own once these macros have been defined. If needed, the remaining of the do-file can be modified to accommodate particular situations (e.g. other outcome types) but this is not covered here.

IV) Output

On completion, CEmimix returns a multiply-imputed dataset corresponding to the options specified. If 'saving' was specified, the dataset is also saved in corresponding folder. This dataset is in Stata `mi flong` format, and can be analysed with the `mi:` family of commands. Note that CEmimix only returns a dataset and does not conduct any analysis. An example of analysis can be seen below.

V) Example

Here is an example of CEmimix set up for CoBalT study, under J2R assumptions for effectiveness and MAR for costs. Interim missing are assumed MAR, and 100 imputations are performed.

```
global m  100          //Number of imputations
global emethod  J2R    //MAR J2R CIR LMCF BMCF
global cmethod  MAR    //MAR J2R CIR LMCF BMCF

global data  COBALT.dta
global effectv  eq5d0 eq5d6 eq5d12
global costv  tcost
global covariates  age sex
global idv  ptidno
global treatv  arm
global refgroup  0
```

[3] If a baseline covariate is not fully observed, it could be mean-imputed beforehand (White & Thompson, *Statistics in medicine*, *24*(7), pp.993-1007)

```
	global interimMAR  effect cost  // effect, cost, or leave blank
	global restrictto    //Restrict MNAR imputation to a specific subgroup
	global seed //Specify seed for reproducibility. Leave blank for random seed
	global saving  COBALT_imputed_J2R.dta
```

And an example of analysis of the resulting dataset:

```
use COBALT_imputed_J2R.dta, clear
*Calculate imputed QALYs
	mi passive: gen qaly=0.5*(eq5d0+eq5d6)/2 +0.5*(eq5d6+eq5d12)/2
*Incremental QALYs and costs
	mi estimate: regress qaly arm
	local incq=el(r(table),1,1)
	mi estimate: regress tcost arm
	local incc=el(r(table),1,1)
*ICER
	display "ICER = " `incc'/`incq'
*Probability cost effective
	gen inb20=qaly*20000-tcost
	mi estimate: regress inb20 arm
	local pce = normal(el(r(table),1,1)/el(r(table),2,1) )
	display "Probability cost effective = " round(`pce'*100,0.1) "%"
```