ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2020)

A Bayesian parametric approach to handle missing longitudinal outcome data in trial-based health economic evaluations

Andrea Gabrio,

University College London, UK

Michael J. Daniels

University of Florida, Gainesville, USA

and Gianluca Baio

University College London, UK

[Received August 2018. Final revision August 2019]

Summary. Trial-based economic evaluations are typically performed on cross-sectional variables, derived from the responses for only the completers in the study, using methods that ignore the complexities of utility and cost data (e.g. skewness and spikes). We present an alternative and more efficient Bayesian parametric approach to handle missing longitudinal outcomes in economic evaluations, while accounting for the complexities of the data. We specify a flexible parametric model for the observed data and partially identify the distribution of the missing data with partial identifying restrictions and sensitivity parameters. We explore alternative non-ignorable missingness scenarios through different priors for the sensitivity parameters, calibrated on the observed data. Our approach is motivated by, and applied to, data from a trial assessing the cost-effectiveness of a new treatment for intellectual disability and challenging behaviour.

Keywords: Bayesian statistics; Cost-effectiveness; Longitudinal data; Missing data; Sensitivity analysis

1. Introduction

Economic evaluation alongside randomized clinical trials (RCTs) is an important and increasingly popular component of the process of technology appraisal (National Institute for Health and Care Excellence, 2013). A typical analysis of individual level data involves the comparison of two interventions for which suitable measures of clinical benefits and costs are observed on each patient enrolled in the trial at different time points throughout the follow-up.

Typically, clinical benefits are measured through multiattribute utility instruments (e.g. EQ-5D-3L: `http://www.euroqol.org`), costs are obtained from clinic resource records and both are summarized in cross-sectional quantities, e.g. quality-adjusted life years (QALYs). The main objective of the economic analysis is

(a) to combine the population-average clinical benefits (or effectiveness) and costs to determine the most 'cost-effective' intervention, given current evidence, and

Address for correspondence: Andrea Gabrio, Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK.
E-mail: ucakgab@ucl.ac.uk

 0964–1998/20/183000

(b) to assess the effect of the uncertainty in the model inputs on the decision-making process (Claxton, 1999; Briggs, 2000; Spiegelhalter *et al.*, 2004; O'Hagan *et al.*, 2004; Sculpher *et al.*, 2005; Briggs *et al.*, 2006; Jackson *et al.*, 2009; Baio, 2012).

Individual level data from RCTs are almost invariably affected by missingness. The recorded outcome process is often incomplete because of individuals who drop out or are observed intermittently throughout the study, causing some observations to be missing. In most applications, the economic evaluation is performed on the cross-sectional variables, computed using only the data from the individuals who are observed at each time point in the trial (completers), with at most limited sensitivity analysis to missingness assumptions (Noble *et al.*, 2012; Gabrio *et al.*, 2017; Leurent *et al.*, 2018a). This, however, is an extremely inefficient approach as the information from the responses of all partially observed subjects is completely lost and it is also likely to be biased unless the completers are a random sample of the subjects on each arm (Little and Rubin, 2002).

When there are missing data, inferences about parameters of interest cannot be obtained without assumptions about the distribution of the missing responses that cannot be verified from the data. When the information from the available data cannot fully explain the systematic differences between observed and unobserved data, missingness can introduce bias in the analysis (Rubin, 1987; Little and Rubin, 2002). Dealing with informative missingness is not straightforward as inference can be drawn only under untestable assumptions about the unobserved data and is often sensitive to the particular assumptions that are made (Molenberghs *et al.*, 1997). It is therefore desirable to assess the robustness of the inference by varying these assumptions in a principled way (Scharfstein *et al.*, 1999; Vansteelandt *et al.*, 2006; Daniels and Hogan, 2008).

The problem of informative missingness is often embedded within a more complex framework, which makes the modelling task in economic evaluations particularly challenging. Specifically, the effectiveness and cost data typically present a series of complexities that need to be simultaneously addressed to avoid biased results. First, the presence of a bivariate outcome requires the use of appropriate methods that deal with correlation (O'Hagan and Stevens, 2001). Second, repeated utility and cost measurements are typically collected during the trial period for each individual. Under such designs, incomplete data are not unusual, as some of the subjects are not available to be measured at all time points, which presents a considerable modelling challenge for the statistician (Molenberghs *et al.*, 2015). Third, outcome data typically have empirical distributions that are highly skewed. The adoption of parametric distributions that can account for skewness (e.g. beta for the utilities and gamma or log-normal distributions for the costs) has been suggested to improve the fit (Nixon and Thompson, 2005; Thompson and Nixon, 2005). In addition, data may exhibit spikes at one or both of the boundaries of the range for the underlying distribution that may induce high skewness in the data that is difficult to capture by using standard parametric models (Cooper *et al.*, 2003). For example, some patients in a trial may not accrue any cost at all or some individuals may be associated with perfect health, i.e. unit utility. The use of more flexible formulations, known as *hurdle models*, explicitly accounts for these 'structural' values. Hurdle models are essentially a mixture between a point mass distribution (the spike) and a parametric model fit to the natural range of the relevant variable without the boundary values. Hurdle models have been applied in economic evaluations for handling either costs or QALYs (Baio, 2014; Gabrio *et al.*, 2019).

To our knowledge, there are no approaches in the literature that can jointly handle the complexities of the data in economic evaluations, while also providing a principled approach to explore non-ignorable missingness. Alternative approaches have been proposed in the literature

to handle separately some of the complexities that affect utility and cost data: correlation and skewness (Gomes *et al.*, 2012), structural values (Basu and Manca, 2012) and missingness under an ignorable missingness assumption (Diaz-Ordaz *et al.*, 2014; Ng *et al.*, 2016). However, these approaches do not simultaneously account for all the complexities of the data and therefore may lead to incorrect results and misleading cost-effectiveness conclusions. Recently, Leurent *et al.* (2018b) proposed a delta adjustment method within multiple imputation by chained equations to explore non-ignorable missing data assumptions for both outcomes. Despite being simple to implement, this method does not account for the complexities of the data or the time dependence between the responses and does not correspond to a valid probabilistic model.

Using a recent randomized trial as our motivating example, we present a Bayesian parametric model for conducting inference on a bivariate health economic longitudinal response. We specify our model to account for the different types of complexities affecting the data while accommodating a sensitivity analysis to explore the effect of alternative missingness assumptions on the inferences and on the decision-making process for health technology assessment.

1.1. Positive behaviour support trial

The positive behaviour support (PBS) study (Hassiotis *et al.*, 2018) is a multicentre RCT that, among its objectives, aimed to evaluate the cost-effectiveness of a new multicomponent intervention (PBS, 108 subjects) relative to treatment as usual (136 subjects) for individuals suffering from mild to severe intellectual disability and challenging behaviour. The primary instruments that were used to assess the clinical benefits and costs were the EQ-5D-3L questionnaires and family or paid carer clinic records respectively. Utilities are derived from the health questionnaires by using the time trade-off algorithm (National Institute for Health and Care Excellence, 2013) and are defined on the interval $[-0.594, 1]$, where 1 represents the perfect health state whereas negative values indicate states that are considered 'worse than death'. Costs, expressed in pounds, are obtained from the clinic records.

Throughout the study, subjects who are associated with a utility of 1, a zero cost or both structural values are observed. Histograms of the empirical distributions of the utilities and costs at each time point in the trial are provided in the web appendix. Measurements were scheduled to be collected at baseline and at 6 and 12 months after baseline. By design, the collection of the baseline data in the PBS trial occurred after randomization and led to relatively large imbalances in the baseline utilities and costs between the two groups. Since this imbalance may be due, at least partially, to the treatment intervention, baseline variables should also be explicitly modelled to avoid misleading inferences.

Let $\mathbf{u}_i = (u_{i0}, \ldots, u_{iJ})$ and $\mathbf{c}_i = (c_{i0}, \ldots, c_{iJ})$ denote the vectors of utilities and costs that were supposed to be observed for subject i at time j in the study, with $j \in \{0, 1, J = 2\}$. We denote by $\mathbf{y}_{ij} = (u_{ij}, c_{ij})$ the bivariate outcome for subject i formed by the utility and cost pair at time j. Both outcomes were partially observed and missingness was non-monotone in the sense that if $\mathbf{y}_{ij}$ was unobserved then $\mathbf{y}_{ij+1}$ could be either observed or unobserved. We group the individuals according to the missingness patterns and denote by $\mathbf{r}_{ij} = (r^u_{ij}, r^c_{ij})$ a pair of indicator variables that take value 1 if the corresponding outcome for subject i at time j is observed and 0 otherwise. We denote by $\mathbf{r}_i = (\mathbf{r}_{i0}; \mathbf{r}_{i1}; \mathbf{r}_{i2})$ the missingness pattern to which subject i belongs, where each pattern is associated with different values for $\mathbf{r}_{ij}$. For example, the pattern $\mathbf{r} = \mathbf{1}$ is associated with the set $\mathbf{r} = (1, 1; 1, 1; 1, 1)$ and corresponds to the completers pattern. We denote by R the total number of observed patterns either in the control (nine patterns) or intervention (six patterns) group. Table 1 reports the missingness patterns in each treatment group as well as the number of individuals and the observed mean responses within each pattern.

Table 1. Missingness patterns for the outcome $\mathbf{y}_j = (u_j, c_j)$ in the PBS study†

	Results for control ($v=1$)						$n_{\mathbf{r}1}$	*Results for intervention ($v=2$)*						$n_{\mathbf{r}2}$
	u_0	c_0	u_1	c_1	u_2	c_2		u_0	c_0	u_1	c_1	u_2	c_2	
$\mathbf{r}=\mathbf{1}$	1	1	1	1	1	1	108	1	1	1	1	1	1	96
Mean	0.486	1546	0.496	1527	0.490	1520		0.564	2818	0.636	2833	0.616	2878	
r	0	1	1	1	1	1	7	0	1	1	1	1	1	5
Mean	—	1310	0.528	1440	0.432	1858		—	2573	0.650	2939	0.760	2113	
r	1	1	0	1	1	1	4	1	1	0	1	1	1	1
Mean	0.536	1620	—	1087	0.581	851		0.151	9649	—	4828	−0.181	4930	
r	1	1	1	1	0	1	2	1	1	1	1	0	1	1
Mean	0.305	640	0.439	512	—	286		0.708	3788	0.815	0	—	0	
r	1	1	0	0	1	1	4	1	1	0	0	1	1	1
Mean	0.547	2834	—	—	0.417	679		0.205	3608	—	—	0.796	4781	
r	1	1	0	0	0	0	4	1	1	0	0	0	0	4
Mean	0.10	1528	—	—	—	—		0.617	3086	—	—	—	—	
r	0	1	0	1	1	1	2	0	1	0	1	1	1	0
Mean	—	595	—	397	0.176	69		—	—	—	—	—	—	
r	1	1	1	1	0	0	2	1	1	1	1	0	0	0
Mean	0.591	1434	0.530	1606	—	—		—	—	—	—	—	—	
r	1	1	0	1	0	1	3	1	1	0	1	0	1	0
Mean	0.564	1510	—	432	—	976		—	—	—	—	—	—	

†For each pattern and treatment group, the number of subjects, $n_{\mathbf{r}t}$, and the observed mean responses at each time $j=0,1,2$ are reported. We denote the absence of response values or individuals within each pattern with a dash.

The number of observed patterns is relatively small and with the exception of the completers ($\mathbf{r}=\mathbf{1}$) the patterns are quite sparse. Baseline costs in both treatment groups are the only fully observed variables, whereas the average proportions of missing utilities and costs are respectively 21% and 10% for the control ($v=1$) and 10% and 8% for the intervention ($v=2$).

1.2. Standard approach to economic evaluation

To perform the economic evaluation, aggregated measures for both utilities and costs are typically derived from the longitudinal responses that are recorded in the study. QALYs, e_{iv}, and total costs, c_{iv}, measures are computed as

$$\begin{aligned} e_{iv} &= \sum_{j=1}^{J} (u_{ijv} + u_{ij-1v}) \frac{\delta_j}{2}, \\ c_{iv} &= \sum_{j=1}^{J} c_{ijv}, \end{aligned} \tag{1}$$

where v denotes the treatment group, and $\delta_j = (\text{Time}_j - \text{Time}_{j-1})/(\text{unit of time})$ is the percentage of the time unit (typically 1 year) which is covered between time $j-1$ and j in the trial. In the PBS trial, since measurements are collected at 6-month intervals, then $\delta_j = (6$ months$)/(12$ months$) = 0.5$ for $j=1,2$. Although no individual died in the trial, we note that a utility of 0 is typically associated with a state of death at a given time point and is carried over until the last follow-up to calculate e_{iv} by using equation (1).

The economic evaluation is then performed by applying some parametric model $p(e_{iv}, c_{iv} \mid \boldsymbol{\theta})$, indexed by a set of parameters $\boldsymbol{\theta}$, to these cross-sectional quantities, typically by using linear regression methods to account for the imbalance in some baseline variables between treatments

(Manca *et al.*, 2005; Van Asselt *et al.*, 2009; European Medicines Agency, 2013). We note that the term cross-sectional here refers to analyses that are based on variables derived from the combination of repeated measurements collected at different times over the trial duration and not on data collected at a single point in time. Finally, QALYs and total costs population mean values are derived from the model:

$$\begin{aligned} \mu_{ev} &= E(e_{iv}|\boldsymbol{\theta}), \\ \mu_{cv} &= E(c_{iv}|\boldsymbol{\theta}). \end{aligned} \tag{2}$$

The differences in μ_{ev} and μ_{cv} between the treatment groups represent the quantities of interest in the economic evaluation and are used in assessing the relative cost-effectiveness of the interventions.

In the original economic evaluation of the PBS study, the quantities in equation (1) were derived on the basis of the longitudinal responses for only the completers, while discarding all other partially observed data. Next, the quantities in equation (2) were obtained under a frequentist approach. The two cross-sectional outcome variables (e_{iv}, c_{iv}) were then modelled independently by using normal linear regression methods to control for differences in baseline values, and including the treatment variable to estimate the mean QALYs and cost differentials between groups.

The modelling approach that was used in the original analysis has the limitation that μ_{ev} and μ_{cv} are derived on the basis of only the completers in the study and does not assess the robustness of the results to a range of plausible missingness assumptions. The model also fails to account for the different complexities that affect the utility and cost data in the trial: from the correlation between variables to the skewness and the presence of structural values (0 for the costs and 1 for the utilities) in both outcomes.

1.3. A longitudinal model to deal with missingness

We propose an alternative approach to deal with a missing bivariate outcome in economic evaluations, while simultaneously allowing for the different complexities that typically affect utility and cost data. Our approach includes a longitudinal model that improves the current practice by taking into account the information from all observed data as well as the time dependence between the responses. The targeted quantities can then be obtained by applying the same formulae in equation (1) to the marginal means at each time for $\mathbf{y}_{ij} = (u_{ij}, c_{ij})$, which can be easily derived from the model. This can be accomplished through the specification of a joint distribution $p(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega})$ for the response and missingness pattern, where $\boldsymbol{\omega}$ is some relevant parameter vector.

We define the data as $\mathbf{y} = (\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}})$ to indicate the subsets that are observed and missing. Next, define $p(\mathbf{y} \mid \boldsymbol{\theta})$ as the response model, parameterized by $\boldsymbol{\theta}$, and $p(\mathbf{r} \mid \mathbf{y}, \boldsymbol{\psi})$ as the missingness model, with parameters $\boldsymbol{\psi}$. Missingness is said to be *ignorable* if the following three conditions hold (Little and Rubin, 2002):

(a) $p(\mathbf{r} \mid \mathbf{y}, \boldsymbol{\psi}) = p(\mathbf{r} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\psi})$, i.e. missingness depends only on the observed responses, a condition known as missingness at random (MAR);
(b) the parameter $\boldsymbol{\omega}$ of the joint model $p(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega})$ can be decomposed as $(\boldsymbol{\theta}, \boldsymbol{\psi})$, with $p(\mathbf{y} \mid \boldsymbol{\theta})$ and $p(\mathbf{r} \mid \mathbf{y}, \boldsymbol{\psi})$;
(c) the parameters of the response and missingness model are *a priori* independent, i.e. $p(\boldsymbol{\omega}) = p(\boldsymbol{\theta})\, p(\boldsymbol{\psi})$.

When any of these conditions is not satisfied, missingness is said to be *non-ignorable*. Often, this is due to the failure of the first condition, which implies that $p(\mathbf{r} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \boldsymbol{\psi}) \neq p(\mathbf{r} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}'_{\text{mis}} \boldsymbol{\psi})$

for $\mathbf{y}_{\text{mis}} \neq \mathbf{y}'_{\text{mis}}$, known as missingness not at random (MNAR). In this case, the joint model $p(\mathbf{y}, \mathbf{r})$ will require untestable assumptions about the missing data in order to be identified. We specify our non-ignorable modelling strategy by using the extrapolation factorization and a pattern mixture approach with identifying restrictions (Little, 1994; Linero and Daniels, 2018). The first factors the joint distribution of the response and missingness $p(\mathbf{y}, \mathbf{r})$ in a way that enables us to separate the distribution of $\mathbf{y}_{\text{obs}}$, for which a flexible parametric model is specified, from the distribution of $\mathbf{y}_{\text{mis}}$, which is left unidentified. The second specifies the distribution of the responses conditionally on the missingness patterns and facilitates the incorporation of external evidence to identify the distribution of $\mathbf{y}_{\text{mis}}$ and to assess the robustness of the results to alternative assumptions.

In this work we present a parametric model with a fully Bayesian framework that can account for both skewness and structural values within a partially observed outcomes setting. A major advantage of adopting a Bayesian approach is the ability to allow for the formal incorporation of external evidence in the analysis through the use of informative prior distributions. This is a crucial element for conducting sensitivity analysis to assess the robustness of the results to a range of plausible missing data assumptions.

1.4. Outline

In Section 2 we describe the general strategy that was used to define the model and the factorization chosen to specify the joint distribution of the cost and utility data and the missingness patterns. In Section 3 we introduce the parametric model that was implemented for the distribution of the observed data and we present alternative specifications to identify the joint model under non-ignorability. In Section 4 we introduce the identifying restrictions that were used and the approach followed to conduct sensitivity analysis. In Section 5 we implement our model to draw inferences on the PBS study under alternative missingness assumptions and summarize the cost-effectiveness results under each scenario from a decision maker's perspective. We close in Section 6 with a discussion.

2. Modelling framework

We define our modelling strategy following Linero and Daniels (2015) and factor the joint distribution for the response and missingness as

$$p(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega}) = p(\mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r} \mid \boldsymbol{\omega})\, p(\mathbf{y}^{\mathbf{r}}_{\text{mis}} \mid \mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r}, \boldsymbol{\omega})$$

where $\mathbf{y}^{\mathbf{r}}_{\text{obs}}$ and $\mathbf{y}^{\mathbf{r}}_{\text{mis}}$ indicate the observed and missing responses within pattern $\mathbf{r}$ respectively. This is the extrapolation *factorization* and factors the joint distribution into two components, of which the extrapolation *distribution* $p(\mathbf{y}^{\mathbf{r}}_{\text{mis}} \mid \mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r}, \boldsymbol{\omega})$ remains unidentified by the data in the absence of unverifiable assumptions about the full data (Daniels and Hogan, 2008).

To specify the observed data distribution $p(\mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r} \mid \boldsymbol{\omega})$ we use a working model p^* for the joint distribution of the response and missingness (Linero and Daniels, 2015). Essentially, the idea is to use the working model $p^*(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega})$ to draw inferences about the distribution of the observed data $p(\mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r} \mid \boldsymbol{\omega})$ by integrating out the missing responses:

$$p(\mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r} \mid \boldsymbol{\omega}) = \int p^*(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega}) \mathrm{d}\mathbf{y}^{\mathbf{r}}_{\text{mis}}.$$

This approach avoids direct specification of the joint distribution of the observed and missing data $p(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega})$, which has the undesirable consequence of identifying the extrapolation

distribution with assumptions that are difficult to check. Indeed, since we use $p^*(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega})$ only to obtain a model for $p(\mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r} \mid \boldsymbol{\omega})$ and not as a basis for inference, the extrapolation distribution is left unidentified. Any inference depending on the observed data distribution may be obtained by using the working model as the true model, with the advantage that it is often easier to specify a model for the full data $p(\mathbf{y}, \mathbf{r})$ compared with a model for the observed data $p(\mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r})$.

We specify p^* by using a pattern mixture approach, factoring the joint $p(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega})$ as the product between the marginal distribution of the missingness patterns $p(\mathbf{r} \mid \boldsymbol{\psi})$ and the distribution of the response conditional on the patterns $p(\mathbf{y} \mid \mathbf{r}, \boldsymbol{\theta})$, respectively indexed by the distinct parameter vectors $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$. If missingness is monotone it is possible to summarize the patterns by dropout time and to model the dropout process directly (Daniels and Hogan, 2008; Gaskins *et al.*, 2016). Unfortunately, as often occurs in trial-based health economic data, missingness in the PBS study is mostly non-monotone and the sparsity of the data in most patterns makes it infeasible to fit the response model within each pattern, with the exception of the completers ($\mathbf{r} = \mathbf{1}$). Thus, we decided to collapse together all the non-completers' patterns ($\mathbf{r} \neq \mathbf{1}$) and to fit the model separately to this aggregated pattern and to the completers.

This strategy assumes that reasons for missingness do not largely differ across the non-completers, which may not be realistic in some cases. However, given the paucity of data in the trial, which makes the identification of the model within each pattern not feasible, we believe that the framework offers a reasonable way to identify the distribution of the responses for those who have some unobserved data without relying on the observations from the completers. Alternative non-ignorable missingness strategies can be considered to handle sparse data, e.g. parametric selection models, but these are not typically well suited for conducting sensitivity analysis (Daniels and Hogan, 2008).

The model can be represented as

$$
\begin{aligned}
p(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\omega}) = p(\mathbf{r} \mid \boldsymbol{\psi})\, & p(\mathbf{y} \mid \mathbf{r} = \mathbf{1}, \boldsymbol{\lambda})^{\mathbb{I}(\mathbf{r}=\mathbf{1})} \\
& \left\{ \prod_{\mathbf{r} \geqslant 2} p(\mathbf{y}^{\mathbf{r}}_{\text{obs}} \mid \mathbf{r}, \boldsymbol{\eta}) \right\}^{\mathbb{I}(\mathbf{r}\neq\mathbf{1})} \Bigg\} \quad \text{observed data distribution} \\
& \left\{ \prod_{\mathbf{r} \geqslant 2} p(\mathbf{y}^{\mathbf{r}}_{\text{mis}} \mid \mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r}, \boldsymbol{\xi}) \right\}^{\mathbb{I}(\mathbf{r}\neq\mathbf{1})} \quad \text{extrapolation distribution}
\end{aligned}
$$

where $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\psi})$, $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are the distinct subsets of $\boldsymbol{\theta}$ that index the response model in the completers and non-completers patterns, and $\boldsymbol{\xi}$ is the subset of $\boldsymbol{\eta}$ that indexes the extrapolation distribution. The joint distribution has three components. The first is given by the model for the patterns and the model for the completers ($\mathbf{r} = \mathbf{1}$), where no missingness occurs. The second component is a model for the observed data in the collapsed patterns $\mathbf{r} \neq \mathbf{1}$ that, together with the first component, form the observed data distribution. The last component is the extrapolation distribution.

Because the targeted quantities of interest (equation (2)) can be derived on the basis of the marginal utility and cost means that, at each time j, in our analysis we do not require the full identification of $p(\mathbf{y}^{\mathbf{r}}_{\text{mis}} | \mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r}, \boldsymbol{\xi})$. Instead, we only partially identify the extrapolation distribution by using *partial identifying restrictions* (Linero and Daniels, 2018). Specifically, we require only the identification of the marginal means for the missing responses in each pattern.

Let $\mathcal{I}^{\mathbf{r}}$ be the indices of the missing observations in pattern $\mathbf{r}$ and let $\mathcal{J}^{\mathbf{r}'}_{\mathbf{r}} \subseteq \mathcal{I}^{\mathbf{r}}$ be the subset of the indices in $\mathcal{I}^{\mathbf{r}}$ for which there are observed responses in $\mathbf{r}'$. We denote by $\mathbf{y}^{\mathbf{r}}_{\text{mis}} = \mathbf{y}^{\mathbf{r}}(\mathcal{I}^{\mathbf{r}})$ the missing

responses in pattern $\mathbf{r}$. Next, we denote by $\mathbf{y}^{\mathbf{r}'}_{\text{obs}}(\mathcal{J}^{\mathbf{r}'}_{\mathbf{r}}) \subseteq \mathbf{y}^{\mathbf{r}'}_{\text{obs}}$ the subset of the observed responses in $\mathbf{r}'$ that corresponds to $\mathbf{y}^{\mathbf{r}}_{\text{mis}}$. We identify the marginal mean of $\mathbf{y}^{\mathbf{r}}_{\text{mis}}$ by using the observed values $\mathbf{y}^{\mathbf{r}'}_{\text{obs}}(\mathcal{J}^{\mathbf{r}'}_{\mathbf{r}})$, averaged across $\mathbf{r}' \neq \mathbf{1}$, and some *sensitivity parameters* $\boldsymbol{\Delta} = (\Delta_u, \Delta_c)$. Therefore, we compute the marginal means by averaging only across the observed components in pattern $\mathbf{r}'$ and ignore the components that are missing:

$$E[\mathbf{y}^{\mathbf{r}}_{\text{mis}}|\mathbf{r}] = E[\underset{\mathbf{r}'\neq\mathbf{1},\mathcal{J}^{\mathbf{r}'}_{\mathbf{r}}}{E}[\mathbf{y}^{\mathbf{r}'}_{\text{obs}}(\mathcal{J}^{\mathbf{r}'}_{\mathbf{r}}) + \boldsymbol{\Delta}|\mathbf{r}']]. \tag{3}$$

We start by setting a benchmark assumption with $\boldsymbol{\Delta} = \mathbf{0}$ and then explore the sensitivity of the results to alternative scenarios by using different prior distributions on $\boldsymbol{\Delta}$, calibrated on the observed data. We note that $\boldsymbol{\Delta} = \mathbf{0}$ does not correspond to an MAR assumption, as with non-monotone patterns the MAR condition within a non-ignorable framework is not transparent and is difficult to identify (Molenberghs *et al.*, 2015; Linero and Daniels, 2018). Thus, $\boldsymbol{\Delta} = \mathbf{0}$ effectively corresponds to a non-ignorable missingness analysis with point priors at 0 on all sensitivity parameters. This provides a convenient benchmark scenario from which departures can be explored by using alternative informative priors on $\boldsymbol{\Delta}$. Once the working model has been fitted to the observed data and the extrapolation distribution has been identified, the overall marginal mean for the response model can be computed by marginalizing over $\mathbf{r}$, i.e. $E[\mathbf{Y}] = \Sigma_{\mathbf{r}} p(\mathbf{r}) E[\mathbf{Y}|\mathbf{r}]$.

3. Model for the missingness patterns and observed response

Following current practice for modelling missingness in a pattern mixture framework, we specified the distribution of the number of patterns by using a multinomial distribution on $\{1, \ldots, R\}$, with the total number of observed patterns R and the probabilities $\psi^{\mathbf{r}}_v$, conditionally on the treatment assignment v (Daniels and Hogan, 2008; Verbeke and Molenberghs, 2009; Molenberghs *et al.*, 2015; Gaskins *et al.*, 2016). We specify a prior for $\psi^{\mathbf{r}}_v$ that gives more weight on the completers' pattern and equal weights to the other patterns. Specifically, we choose a Dirichlet$(1-x, x/R^*, \ldots, x/R^*)$ prior, where x is the expected total dropout rate and $R^* = 64$ is the total number of potential patterns in the study. This is consistent with the design of the study, where the experimenter expects at least $(1-x)\%$ of the individuals to provide complete data, i.e. to fall in $\mathbf{r} = \mathbf{1}$. In practice, this prior is not likely to affect the results as the amount of observed data is enough to learn the posterior of $\psi^{\mathbf{r}}_v$. For comparison, we also consider another specification based on a non-informative Dirichlet$(1, \ldots, 1)$ prior for $\psi^{\mathbf{r}}_v$. Posterior results are robust to the alternative prior choices.

The distribution of the observed responses $\mathbf{y}_{ijv} = (u_{ijv}, c_{ijv})$ is specified in terms of a model for the utility and cost variables at time $j = 0, 1, 2$, which are jointly modelled without using a multilevel approach and separately by treatment group. In particular, the joint distribution for $\mathbf{y}_{ijv}$ is specified as a series of conditional distributions that capture the dependence between utilities and costs as well as the time dependence. We now drop the treatment indicator v for clarity.

Following the recommendations from the published literature, we account for the skewness by using beta and log-normal distributions for the utilities and costs respectively (Basu and Manca, 2012; Ng *et al.*, 2016). Since the beta distribution does not allow for negative values, we scaled the utilities on $[0, 1]$ through the transformation

$$u^*_{ij} = \frac{u_{ij} - \min(\mathbf{u}_j)}{\max(\mathbf{u}_j) - \min(\mathbf{u}_j)}$$

and fitted the model to these transformed variables. To ease the notation we refer to these quantities simply as u_{ij}.

To account for the structural values $u_{ij}=1$ and $c_{ij}=0$ we use a hurdle approach by including in the model the indicator variables $d^u_{ij}:=\mathbb{I}(u_{ij}=1)$ and $d^c_{ij}:=\mathbb{I}(c_{ij}=0)$, which take value 1 if subject i is associated with a structural value at time j and 0 otherwise. The probabilities of observing these values, as well as the mean of each variable, are then modelled conditionally on other variables via linear regressions defined on the logit or log-scale. Specifically, at time $j=1,2$, the probability of observing a 0 and the mean costs are modelled conditionally on the utilities and costs at the previous times, whereas the probability of observing a 1 and the mean utilities are modelled conditionally on the current costs (also at $j=0$) and the utilities at the previous times (only at $j=1,2$). The model can be summarized as follows (for simplicity we omit the subject index i).

At time $j=0$, we model the non-zero costs $c_0\neq 0$ and the indicator $d^c_0:=\mathbb{I}(c_0=0)$ as

$$\begin{gathered} c_0|d^c_0=0\sim\text{log-normal}(\nu^c_0,\tau^c_0),\\ d^c_0\sim\text{Bernoulli}(\pi^c_0)\end{gathered}$$

where ν^c_0 and τ^c_0 are the mean and standard deviation for c_0 given $c_0\neq 0$ on the log-scale, whereas π^c_0 is the probability of a zero cost value. We next model the utilities and the indicator $d^u_0:=\mathbb{I}(u_0=1)$ conditionally on the costs at the same time:

$$\begin{gathered} u_0|d^u_0=0,c_0\sim\text{beta}(\nu^u_0,\sigma^u_0),\\ \text{logit}(\nu^u_0)=\alpha_{00}+\alpha_{10}\log(c_0),\\ d^u_0|c_0\sim\text{Bernoulli}(\pi^u_0),\\ \text{logit}(\pi^u_0)=\gamma_{00}+\gamma_{10}\log(c_0)\end{gathered}$$

where ν^u_0 and σ^u_0 are the mean and standard deviation for u_0 given $u_0\neq\mathbf{1}$ and c_0, whereas π^u_0 is the probability of having a utility value of 1 given c_0. We parameterize the beta distributions in terms of mean and standard deviation to facilitate the specification of the priors on the parameters, compared with using the canonical shape parameters (a,b). Specifically, the mean and standard deviation of the beta distribution are linked to the canonical parameters through the relationships $a=\nu\tau$ and $b=(1-\nu)\tau$, where $\tau=\nu(1-\nu)/\sigma^2-1$.

We use logistic transformations to define a linear dependence for $p(u_0\mid c_0,u_0\neq 1)$ and include the costs on the log-scale (after adding a small constant 0.01) to improve the fit of the model. The sensitivity of the results to different values for this constant (from 0.00001 to 0.1) has been assessed and suggests a general robustness of the inferences.

At time $j=1,2$, we extend the approach illustrated for $j=0$ and make a first-order Markov assumption. For the costs we have

$$\begin{gathered} c_j|d^c_j=0,c_{j-1},u_{j-1}\sim\text{log-normal}(\nu^c_j,\tau^c_j),\\ \nu^c_j=\beta_{0j}+\beta_{1j}\log(c_{j-1})+\beta_{2j}u_{j-1},\\ d^c_j|c_{j-1},u_{j-1}\sim\text{Bernoulli}(\pi^c_j),\\ \text{logit}(\pi^c_j)=\zeta_{0j}+\zeta_{1j}\log(c_{j-1})+\zeta_{2j}u_{j-1}.\end{gathered}$$

Similarly to time $j=0$, the mean, standard deviation and probability parameters for the costs at time j are indicated by ν^c_j, τ^c_j and π^c_j. The regression parameters $\boldsymbol{\beta}_j=(\beta_{0j},\beta_{1j},\beta_{2j})$ and $\boldsymbol{\zeta}_j=(\zeta_{0j},\zeta_{1j},\zeta_{2j})$ capture the dependence between costs at j and the costs and utilities at $j-1$, for the non-zero and zero components respectively. The model for the utilities is

$$
\begin{aligned}
u_j | d_j^u = 0, c_j, u_{j-1} &\sim \text{beta}(\nu_j^u, \sigma_j^u), \\
\text{logit}(\nu_j^u) &= \alpha_{0j} + \alpha_{1j} \log(c_j) + \alpha_{2j} u_{j-1}, \\
d_j^u | c_j, u_{j-1} &\sim \text{Bernoulli}(\pi_j^u), \\
\text{logit}(\pi_j^u) &= \gamma_{0j} + \gamma_{1j} \log(c_j) + \gamma_{2j} u_{j-1}.
\end{aligned}
$$

We denote by ν_j^u, σ_j^u and π_j^u the mean, standard deviation and probability parameters for the utilities at time j, and by $\boldsymbol{\alpha}_j = (\alpha_{0j}, \alpha_{1j}, \alpha_{2j})$ and $\boldsymbol{\gamma}_j = (\gamma_{0j}, \gamma_{1j}, \gamma_{2j})$ the regression parameters that capture the dependence between utilities at j and costs at j and utilities at $j-1$. Although we believe that the proposed Markov assumption provides a reasonable representation of the dependence structure of the data, we acknowledge that alternative specifications could have been used; for example costs at time j could depend on the change in utilities $u_j - u_{j-1}$ rather than on the utilities at j.

For all parameters in the model we specify vague prior distributions: specifically a normal distribution with a large variance on the appropriate scale for the regression parameters and a uniform distribution over a large positive range for the standard deviations. We implement the model and derive the marginal cost and utility means at each time j through Monte Carlo integration. First, we fit the model separately to the completers ($\mathbf{r} = \mathbf{1}$) and the joint set of all other patterns ($\mathbf{r} \neq \mathbf{1}$) for $v = 1, 2$. Second, at each iteration of the posterior distribution, we generate a large number of samples for $\mathbf{y}_{ij} = (c_{ij}, u_{ij})$ based on the posterior values for the parameters of the utility and cost models in the Markov chain Monte Carlo (MCMC) output. Third, we approximate the posterior distribution of the marginal means for each $\mathbf{r}$ by taking the expectation over these sampled values at each iteration. Finally, we derive the overall marginal means $\boldsymbol{\mu}_{jv} = (\mu_{jv}^c, \mu_{jv}^u)$ as weighted averages across the marginal means in each pattern, using the posterior $\psi_v^{\mathbf{r}}$ as weights. We note that, although the model of the observed data distribution requires the specification of a relatively large number of parameters, which may make the model difficult to interpret, this, however, does not ultimately affect the final analysis, which exclusively focuses on the marginal means that are retrieved via Monte Carlo integration.

4. Identifying restrictions and sensitivity parameters

As mentioned in Section 2, we use partial identifying restrictions to link the observed data distribution $p(\mathbf{y}_{\text{obs}}, \mathbf{r})$ to the extrapolation distribution $p(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}, \mathbf{r})$ and consider interpretable deviations from a benchmark scenario to assess how inferences are driven by our assumptions (Linero and Daniels, 2018). Specifically, we identify the marginal mean of the missing responses in each pattern $\mathbf{y}_{\text{mis}}^{\mathbf{r}}$ by averaging across the corresponding components that are observed and add the sensitivity parameters $\boldsymbol{\Delta}_j$ (equation (3)).

We define $\boldsymbol{\Delta}_j = (\Delta_{c_j}, \Delta_{u_j})$ to be time-specific location shifts at the marginal mean in each pattern and set $\boldsymbol{\Delta}_j = \mathbf{0}$ as the benchmark scenario (Daniels and Hogan, 2000). We then explore departures from this benchmark by using alternative priors on $\boldsymbol{\Delta}_j$, which are calibrated by using the observed standard deviations for costs and utilities at each time j to define the amplitude of the departures from $\boldsymbol{\Delta}_j = \mathbf{0}$ (Section 5.3).

5. Application to the positive behaviour support study

5.1. Computation

We fitted the model by using `JAGS` (Plummer, 2010), which is software specifically designed for the analysis of Bayesian models by using MCMC simulation (Brooks *et al.*, 2011), which can be interfaced with R through the package `R2jags` (Su and Yajima, 2015). Samples from

the posterior distribution of the parameters of interest generated by JAGS and saved to the R work space are then used to produce summary statistics and plots. We ran two chains with 20000 iterations per chain, using a burn-in of 5000, for a total sample of 30000 iterations for posterior inference. For each unknown quantity in the model, we assessed convergence and autocorrelation of the MCMC simulations by using diagnostic measures including the *potential scale reduction factor* and the *effective sample size* (Gelman *et al.*, 2004).

In the non-completers pattern ($\mathbf{r} \neq \mathbf{1}$), we set to 0 the regression parameters (ζ_{11}, ζ_{21}) and $(\gamma_{10}, \gamma_{11}, \gamma_{21})$ for the model fitted to the control and intervention group respectively. This simplification was required because, among the non-completers, there is only one observed $c_j = 0$ at time $j = 1$ in the control group and one observed $u_j = 1$ at time $j = \{0, 1\}$ in the intervention group. We therefore drop from the model the dependence between the probabilities of having a structural value at these times and the variables at the previous or same times to ensure the convergence of the algorithm and to avoid identifiability problems.

5.2. Model assessment

We computed the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) to assess the fit of the model with respect to an alternative parametric specification, where the log-normal distributions are replaced with gamma distributions for the cost variables. The DIC is a measure of comparative predictive ability based on the model deviance and a penalty for model complexity known as the effective number of parameters, p_D. When comparing a set of models based on the same data, the model that is associated with the lowest DIC is the best performing, among those assessed. There are different ways of constructing the DIC in the presence of missing data, which means that its use and interpretation are not straightforward (Celeux *et al.*, 2006; Daniels and Hogan, 2008; Mason *et al.*, 2012). In our analysis, we consider a DIC based on the observed data under MAR as its value does not depend on the values of the sensitivity parameters (Wang and Daniels, 2011). Because the sampling distribution of the observed data was not available in closed form, we computed it by using Monte Carlo integration. Results between the two alternative specifications are reported in Table 2.

The DIC components for the costs are systematically lower when log-normal distributions are used compared with gamma distributions (lower values are shown in italics in Table 2), and result in an overall better fit to the data for the first model.

Table 2. DIC and p_D based on the observed data likelihood for each variable in the model†

Variable	*Results for gamma model*		*Results for log-normal model*	
	DIC	p_D	*DIC*	p_D
c_0	2147.91	2.05	*2133.39*	1.97
$u_0 \mid c_0$	−377.52	2.87	−377.62	2.82
$c_1 \mid c_0, u_0$	1904.53	4.16	*1827.45*	4.13
$u_1 \mid u_0, c_1$	−468.02	5.37	−468.19	5.32
$c_2 \mid c_1, u_1$	1913.69	4.65	*1856.23*	4.36
$u_2 \mid u_1, c_2$	−454.07	5.87	−453.47	5.99
Total	4667	25	*4518*	25

†Two models are assessed either assuming log-normal or gamma distributions for the cost variables (lower DIC values are shown in italics). Total DIC and p_D are also reported at the bottom of the table.

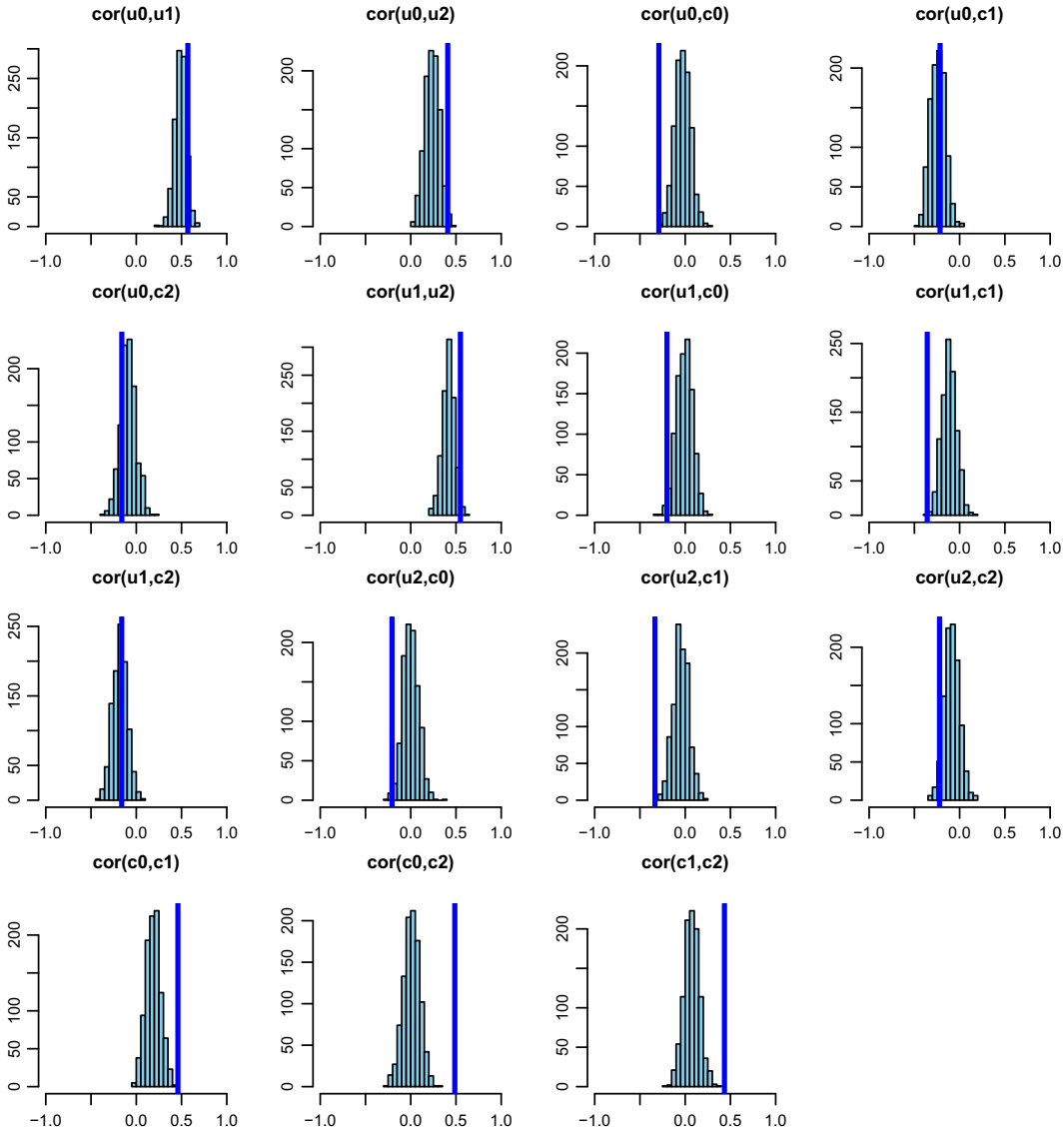


Fig. 1. Posterior predictive distributions for the pairwise correlation between utilities and costs variables in the control arm across 1000 observed replicated data sets (histograms) compared with the estimates based on the observed data in the real data set (|)

We also assess the absolute fit of the model by using posterior predictive checks based on observed data replications (Xu *et al.*, 2016). We sample from the posterior predictive distribution $p(\tilde{\mathbf{y}}, \tilde{\mathbf{r}}|\mathbf{y}^{\mathbf{r}}_{\text{obs}}, \mathbf{r}, \boldsymbol{\omega})$. Conditionally on the replicated patterns $\tilde{\mathbf{r}}$, we define the replicated observed data in each pattern as $\tilde{\mathbf{y}}^{\tilde{\mathbf{r}}}_{\text{obs}} = \{\tilde{\mathbf{y}}_j : \tilde{\mathbf{r}}_j = \mathbf{1}\}$, i.e. the components of $\tilde{\mathbf{y}}$ for which the corresponding missing data indicators at time j in the replicated patterns $\tilde{\mathbf{r}}$ are equal to 1.

We compute the rank correlations between each pair of variables for each replicated data set and compare them with the corresponding values from the real data set. The results, which are shown in Figs 1 and 2 suggest that the parametric model proposed captures most of the correlations well both in the control (Fig. 1) and in the intervention (Fig. 2) group.

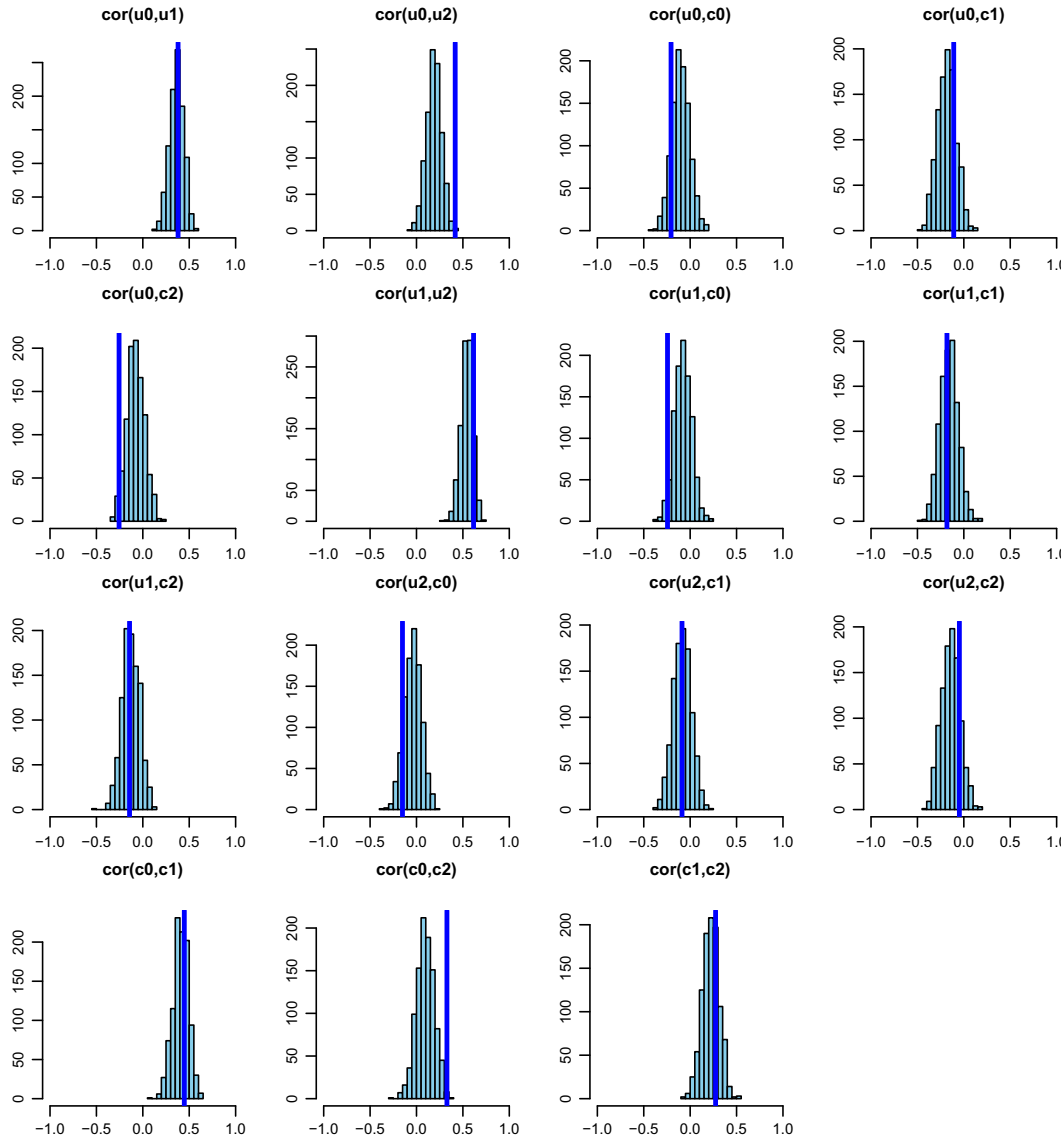


Fig. 2. Posterior predictive distributions for the pairwise correlation between utilities and costs variables in the intervention arm across 1000 observed replicated data sets (histograms) compared with the estimates based on the observed data in the real data set (|)

5.3. Priors on sensitivity parameters

We consider three alternative sets of priors on $\boldsymbol{\Delta}_j = (\Delta_{u_j}, \Delta_{c_j})$, calibrated on the basis of the variability in the observed data at each time j. The three types of prior that were used are as follows:

(a) L-$\boldsymbol{\Delta}^{\text{flat}}$, flat between 0 and twice the observed standard deviation,

$$\Delta_{c_j} \sim \text{uniform}\{0, 2\ \text{sd}(c_j)\},$$
$$\Delta_{u_j} \sim \text{uniform}\{-2\ \text{sd}(u_j), 0\};$$

(b) L-$\boldsymbol{\Delta}^{\text{skew0}}$, skewed towards values closer to 0, over the same range as L-$\boldsymbol{\Delta}^{\text{flat}}$,

$$\Delta_{c_j} = 2\,\text{sd}(c_j)(1-\sqrt{U}),$$
$$\Delta_{u_j} = -2\,\text{sd}(u_j)(1-\sqrt{U});$$

(c) L-$\boldsymbol{\Delta}^{\text{skew1}}$, skewed towards values far from 0, over the same range as L-$\boldsymbol{\Delta}^{\text{flat}}$,

$$\Delta_{c_j} = 2\,\text{sd}(c_j)\sqrt{U},$$
$$\Delta_{u_j} = -2\,\text{sd}(u_j)\sqrt{U},$$

where $U \sim \text{uniform}(0, 1)$ and $\text{sd}(u_j)$ and $\text{sd}(c_j)$ are the standard deviations computed on the observed utilities and costs at time j for the non-completers ($\mathbf{r} \neq \mathbf{1}$). We choose these priors because we believe that departures from L-$\boldsymbol{\Delta} = 0$ for both outcomes are not likely to be larger than twice the observed standard deviations at each time j. A graphical representation of the densities that are associated with the three alternative priors that were used is provided in the web appendix.

5.4. Results

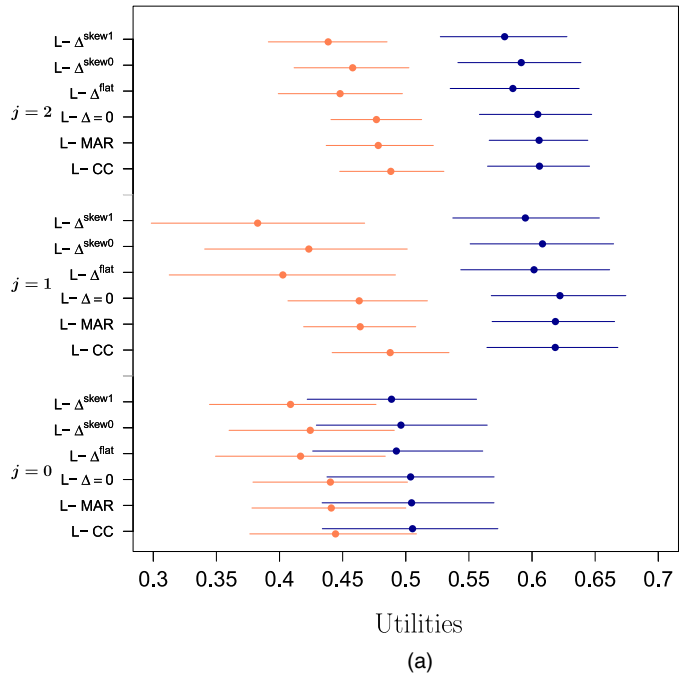
This section summarizes and discusses the results from the posterior distribution of the main quantities of interest in the model, namely the marginal mean utility and cost parameters at each time point $\boldsymbol{\mu}_{jv}$ and the marginal mean QALYs and total costs (μ_{ev}, μ_{cv}) evaluated over the trial duration.

Fig. 3 compares the posterior means and 95% highest posterior density (HPD) credible intervals for $\boldsymbol{\mu}_{jv} = (\mu^u_{jv}, \mu^c_{jv})$ obtained from fitting the model under six alternative scenarios: completers, L-CC, all cases assuming ignorability, L-MAR, and using the extrapolation factorization under the benchmark, L-$\boldsymbol{\Delta}_j = \mathbf{0}$, and three departure scenarios, L-$\boldsymbol{\Delta}^{\text{flat}}$, L-$\boldsymbol{\Delta}^{\text{skew0}}$ and L-$\boldsymbol{\Delta}^{\text{skew1}}$. Since baseline costs are fully observed, only the estimates under L-CC and L-MAR are shown for $\boldsymbol{\mu}^c_0$. Results that are associated with the control and intervention group are indicated in light and dark colours respectively.

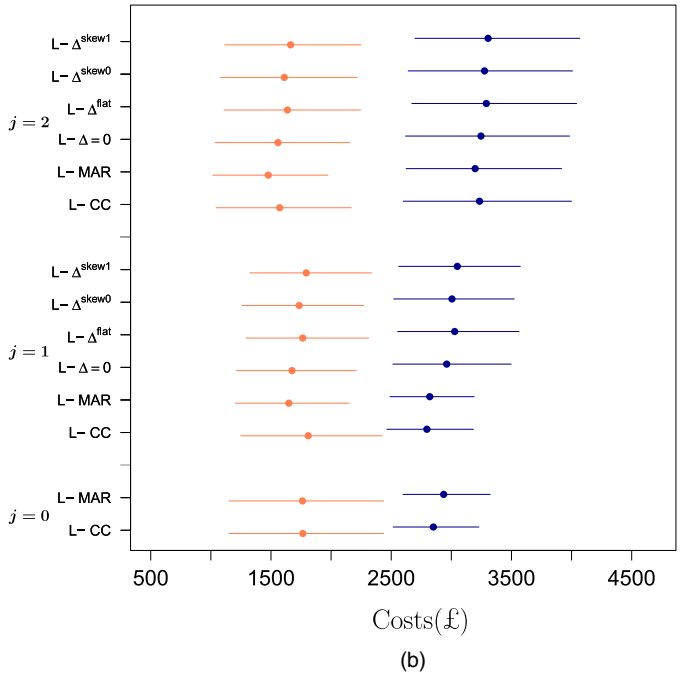
The distributions of both $\boldsymbol{\mu}^u_j$ and $\boldsymbol{\mu}^c_j$ show values that are higher in the intervention compared with the control at each time j and show relatively similar mean estimates across L-CC, L-MAR and L-$\boldsymbol{\Delta} = \mathbf{0}$. However, under the other non-ignorable missingness scenarios, mean utilities or costs are on average 3% (L-$\boldsymbol{\Delta}^{\text{flat}}$), 4% (L-$\boldsymbol{\Delta}^{\text{skew0}}$) and 5% (L-$\boldsymbol{\Delta}^{\text{skew1}}$) lower or higher compared with L-$\boldsymbol{\Delta} = \mathbf{0}$ in the control group. In the intervention group, mean utilities or costs are on average 1% (L-$\boldsymbol{\Delta}^{\text{flat}}$), 1.5% (L-$\boldsymbol{\Delta}^{\text{skew0}}$) and 2.5% (L-$\boldsymbol{\Delta}^{\text{skew1}}$) lower or higher compared with $\boldsymbol{\Delta} = \mathbf{0}$. In spite of the variations between the mean estimates, overall, the HPD intervals for each marginal mean utility and cost parameter do not show evidence of large discrepancies across all scenarios.

We then derived the QALYs and total costs means μ_{ev} and μ_{cv} by applying equation (1) to the cost and utility marginal means $\boldsymbol{\mu}_{jv}$ obtained from the model. We also compare the estimates that are derived from our approach with those obtained from two cross-sectional models fitted on e_i and c_i, computed only on the basis of the completers in the study. The first, denoted by CS-IND, was specified following the approach that was used in the original analysis of the PBS study (assuming independent normal distributions for the two outcomes and including baseline adjustments), but implemented within a Bayesian framework. The second, denoted by CS-HUR, is specified by using a beta–log-normal distribution with a hurdle approach to handle both unit QALYs and zero total costs.

Table 3 shows the posterior means and 95% HPD credible intervals that were associated with the targeted quantities under all scenarios for both treatment groups.



(a)



(b)

Fig. 3. Posterior means and 95% HPD intervals for the marginal utility and cost means in the control (—●—, $t = 1$) and intervention (—●—, $t = 2$) group at each time j in the study across alternative assumptions (six scenarios are compared; completers, L-CC, ignorability, L-MAR, and non-ignorability using the extrapolation factorization under the benchmark assumption (L-$\boldsymbol{\Delta} = \mathbf{0}$) and under the three scenarios described in Section 5.3 (L-$\boldsymbol{\Delta}^{\text{flat}}$, L-$\boldsymbol{\Delta}^{\text{skew0}}$ and L-$\boldsymbol{\Delta}^{\text{skew1}}$); since the baseline costs are fully observed in both groups, only the results under L-CC and L-MAR are displayed for μ_0^c): (a) μ_{jv}^u; (b) μ_{jv}^c

Table 3. Posterior means and 95% HPD credible intervals for μ_{et} and μ_{ct} in the control ($v = 1$) and intervention ($v = 2$) group under alternative scenarios: cross-sectional (CS-IND and CS-HUR), L-CC, L-MAR, L-$\boldsymbol{\Delta} = \mathbf{0}$, L-$\boldsymbol{\Delta}^{\text{flat}}$, L-$\boldsymbol{\Delta}^{\text{skew0}}$ and L-$\boldsymbol{\Delta}^{\text{skew1}}$

Scenario	μ_{e1}		μ_{e2}		μ_{c1}		μ_{c2}	
	Mean	*95% credible interval*	*Mean*	*95% credible interval*	*Mean*	*95% credible interval*	*Mean*	*95% credible interval*
CS-IND	0.49	(0.45; 0.52)	0.61	(0.57; 0.65)	3073	(2188; 3915)	5768	(5115; 6413)
CS-HUR	0.49	(0.46; 0.52)	0.60	(0.56; 0.63)	3283	(2459; 4261)	5919	(5385; 6506)
L-CC	0.48	(0.45; 0.51)	0.59	(0.55; 0.62)	3382	(2583; 4246)	6031	(5281; 6889)
L-MAR	0.46	(0.43; 0.49)	0.59	(0.56; 0.62)	3125	(2483; 3846)	6018	(5314; 6806)
L-$\boldsymbol{\Delta} = \mathbf{0}$	0.46	(0.42; 0.49)	0.59	(0.56; 0.62)	3233	(2489; 4041)	6208	(5364; 7142)
L-$\boldsymbol{\Delta}^{\text{flat}}$	0.42	(0.36; 0.47)	0.57	(0.53; 0.61)	3400	(2616; 4196)	6318	(5462; 7271)
L-$\boldsymbol{\Delta}^{\text{skew0}}$	0.43	(0.39; 0.48)	0.58	(0.54; 0.61)	3345	(2605; 4173)	6281	(5409; 7200)
L-$\boldsymbol{\Delta}^{\text{skew1}}$	0.40	(0.35; 0.45)	0.56	(0.53; 0.60)	3457	(2678; 4250)	6355	(5522; 7332)

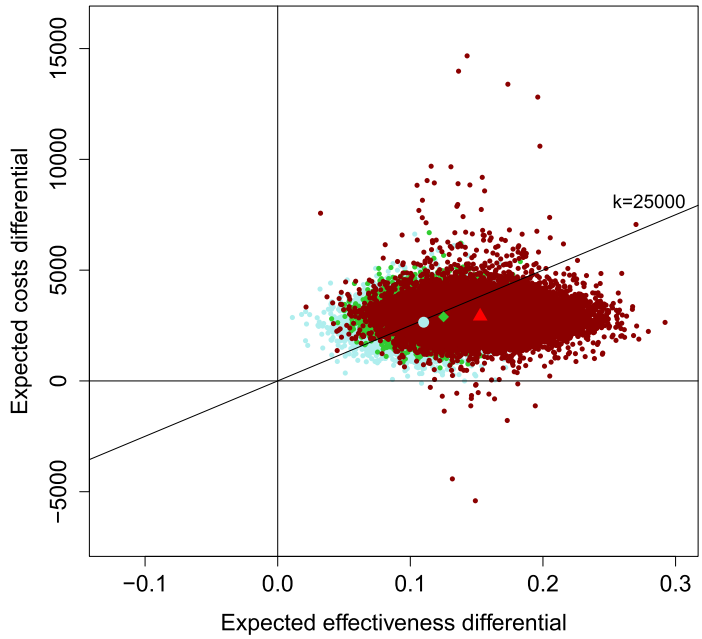
The estimates under the two cross-sectional models (CS-IND and CS-HUR) are relatively close to those derived from the longitudinal model fitted only to the completers (L-CC). These estimates are generally higher with respect to those from the model that additionally incorporates the evidence from the non-completers (L-MAR), both for the QALYs and for the total costs. The estimates under the non-ignorable missingness scenarios are systematically lower for mean QALYs and are systematically higher for mean total costs compared with the other scenarios in both treatment groups, although the HPD intervals for μ_{ev} and μ_{cv} show a considerable overlap across all scenarios. These variations are larger in the control, which has higher proportions of non-completers with respect to the intervention.

5.5. Economic evaluation

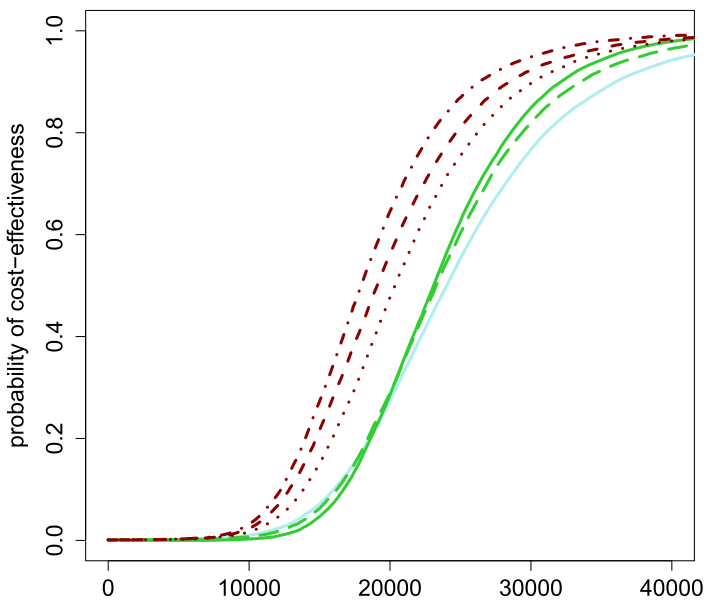
We complete the analysis by assessing the cost-effectiveness of the new intervention with respect to the control, comparing the results under the cross-sectional, CS-IND, complete-case, L-CC, ignorable, L-MAR, benchmark non-ignorable, L-$\boldsymbol{\Delta} = \mathbf{0}$, and the three alternative non-ignorable departure scenarios. We specifically rely on the examination of the cost-effectiveness plane (CEP) (Black, 1990) and the cost-effectiveness acceptability curve (CEAC) (Van Hout *et al.*, 1994) to summarize the economic analysis.

The CEP (Fig. 4(a)) is a graphical representation of the joint distribution for the population-average effectiveness and costs increments between the two arms, indicated respectively as $\mu_{e2} - \mu_{e1}$ and $\mu_{c2} - \mu_{c1}$. Each dot in the plane corresponds to the value of the parameters drawn from the posterior distributions of the expected mean QALYs and total costs, evaluated at each iteration of the MCMC output.

We show the results under only three scenarios (lighter to darker colours for L-CC, L-MAR and L-$\boldsymbol{\Delta}^{\text{flat}}$) for clarity and visualization. The results for the other non-ignorable missingness scenarios are available in the web appendix. The slope of the straight line crossing the plane is the 'willingness-to-pay' threshold (which is often indicated as k). Following current recommendations for trial-based economic analyses, we choose a willingness-to-pay threshold of £25000 per QALY gained (National Institute for Health and Care Excellence, 2013). This can be considered as the amount of budget that the decision maker is willing to spend to increase the health outcome of 1 unit and effectively is used to trade clinical benefits for money. Points lying below



(a)



(b)

Fig. 4. (a) CEPs and (b) CEACs associated with alternative missingness scenarios: in the CEPs, the incremental cost-effectiveness ratios based on the results from L-CC (●, 24070), L-MAR (◆, 23130) and L-$\boldsymbol{\Delta}^{\text{flat}}$ (▲, 19116) are indicated whereas the portion of the plane on the right-hand side of the straight line passing through the plot (evaluated at $k = £25000$) denotes the sustainability area; for the CEACs, in addition to the results under L-CC (——) and L-MAR (——), the probability values for the alternative scenarios are represented also (— —, L-$\boldsymbol{\Delta} = \mathbf{0}$; – – –, L-$\boldsymbol{\Delta}^{\text{flat}}$; · · · · ·, L-$\boldsymbol{\Delta}^{\text{skew0}}$; –·–·–, L-$\boldsymbol{\Delta}^{\text{skew1}}$)

this straight line fall in the so-called *sustainability area* (Baio, 2012) and suggest that the active intervention is more cost-effective than the control. In the graph, we also show the incremental cost-effectiveness ratio (ICER) computed under each scenario, as darker coloured dots. This is defined as

$$\text{ICER} = \frac{E[\mu_{c2} - \mu_{c1}]}{E[\mu_{e2} - \mu_{e1}]}$$

and quantifies the cost per incremental unit of effectiveness.

For all three scenarios almost all samples fall in the north-east quadrant and their ICERs fall in the sustainability area. This suggests that under all scenarios the intervention is likely to be cost-effective by producing both QALY gains and cost savings compared with the control. However, the ICERs that are associated with L-CC and L-MAR are only slightly lower than the threshold value of $k = £25000$ and are therefore associated with more uncertain cost-effectiveness decisions compared with L-$\boldsymbol{\Delta}^{\text{flat}}$. This is also indicated by the full distribution of the CEP which is shifted to the right under L-$\boldsymbol{\Delta}^{\text{flat}}$ with respect to the other two scenarios (therefore suggesting a more favourable conclusion for the intervention).

The CEAC (Fig. 4(b)) is obtained by computing the proportion of points lying in the sustainability area on varying the willingness-to-pay threshold k. On the basis of standard practice in routine analyses, we consider a range for k up to £40000 per QALY gained. The CEAC estimates the probability of cost-effectiveness, thus providing a simple summary of the uncertainty that is associated with the 'optimal' decision making that is suggested by the ICER. The results under L-CC and L-MAR are reported by using full curves. In addition, the results that were derived under non-ignorability are reported by using different broken curves.

The CEACs under L-CC, L-MAR and the benchmark scenarios show a similar trend and indicate a probability of cost-effectiveness below 0.65 of the new intervention for values of k up to £40000. However, under the other scenarios, the curve is shifted upwards by an average probability of 0.2 (L-$\boldsymbol{\Delta}^{\text{flat}}$), 0.15 (L-$\boldsymbol{\Delta}^{\text{skew0}}$) and 0.25 (L-$\boldsymbol{\Delta}^{\text{skew1}}$) and suggests a more favourable cost-effectiveness assessment.

We finally compare the economic results under our longitudinal approach with respect to those derived from the two cross-sectional models CS-IND and CS-HUR. Fig. 5 shows the CEPs (Fig. 5(a)) that are associated with the CS-IND, CS-HUR and L-CC scenarios, respectively indicated with darker to lighter coloured dots. In the CEACs (Fig. 5(b)), in addition to the probability values that are associated with these scenarios, the results from L-MAR and L-$\boldsymbol{\Delta}^{\text{flat}}$ are indicated by using broken and full curves for the cross-sectional and longitudinal models respectively.

The distribution of the posterior samples in the CEP (Fig. 5(a)) show relatively small differences between the scenarios, with the ICER of CS-IND being the lowest among those compared. In the CEAC (Fig. 5(b)), the acceptability curve for CS-IND is generally higher than those for CS-HUR, L-CC and L-MAR for most willingness-to-pay values but remains systematically lower with respect to L-$\boldsymbol{\Delta}^{\text{flat}}$, L-$\boldsymbol{\Delta}^{\text{skew0}}$ and L-$\boldsymbol{\Delta}^{\text{skew1}}$ (in Fig. 5 we show only the results for L-$\boldsymbol{\Delta}^{\text{flat}}$ for clarity).

Overall, both inferences and cost-effectiveness conclusions are sensitive to the assumptions about the missing values, which can lead to a considerable change in the output of the decision process and the cost-effectiveness conclusions. More specifically, the results from L-CC are relatively close to those from the cross-sectional analyses (CS-IND and CS-HUR). When the information from all observed cases is incorporated in the model (L-MAR and L-$\boldsymbol{\Delta} = \mathbf{0}$), the results suggest a more cost-effective intervention, although the magnitude of this change is relatively small. Finally, under the three non-ignorable missingness scenarios assessed (L-$\boldsymbol{\Delta}^{\text{flat}}$,

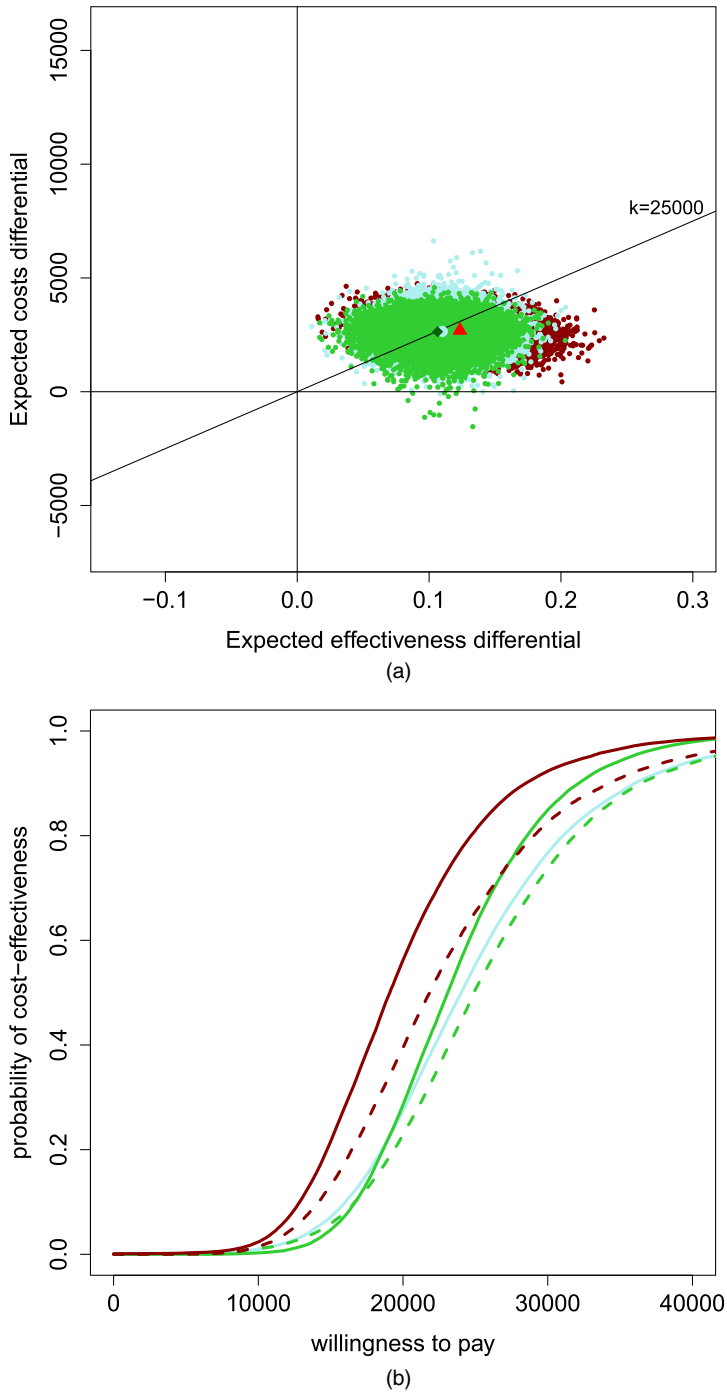


Fig. 5. (a) CEPs and (b) CEACs associated with alternative scenarios; in the CEPs, the ICERs based on the results from CS-IND (▲, 21843), CS-HUR (◆, 24761) and L-CC (●, 24070) are indicated whereas the portion of the plane on the right-hand side of the straight line passing through the plot (evaluated at $k = £25000$) denotes the sustainability area; for the CEACs, in addition to the results under CS-IND (– – –) and CS-HUR (– – –), the probability values for L-CC (———), L-MAR (———) and L-Δ^{flat} (———) are also shown

L-$\boldsymbol{\Delta}^{\text{skew0}}$ and L-$\boldsymbol{\Delta}^{\text{skew1}}$), a considerably more favourable economic assessment for the new intervention is obtained.

We note that the results from these analyses should be interpreted with caution. Indeed, by construction, CEACs are concerned only with currently available information and do not consider explicitly the possibility of gathering additional evidence, therefore providing only a partial evaluation of the overall decision process (Koerkamp *et al.*, 2007). Despite their limitations, however, CEACs represent the standard tools that are used by practitioners in trial-based analyses to report the economic results, particularly for small trials, where the objective is to provide a preliminary assessment of cost-effectiveness. Since all analyses in this paper are based on a trial, we follow current practice and use CEACs as the main tool to summarize the uncertainty about the cost-effectiveness of the intervention (National Institute for Health and Care Excellence, 2013; Ramsey *et al.*, 2015).

6. Discussion

Missingness represents a threat to economic evaluations as, when dealing with partially observed data, any analysis makes assumptions about the missing values that cannot be verified from the data at hand. Trial-based analyses are typically conducted on cross-sectional quantities, e.g. QALYs and total costs, which are derived on the basis of only the observed data from the completers in the study. This is an inefficient approach which may discard a substantial proportion of the sample, especially when there is a relatively large number of time points, where individuals are more likely to have some missing value or to drop out from the study. In addition, when there are systematic differences between the responses of the completers and non-completers, which is typically so when dealing with self-reported outcomes in trial-based analyses, the results based only on the former may be biased and mislead the final assessment. A further concern is that routine analyses typically rely on standard models that ignore or at best fail to account properly for potentially important features in the data such as correlation, skewness and the presence of structural values.

In this paper, we have proposed an alternative approach for conducting parametric Bayesian inference under non-ignorable missingness for a longitudinal bivariate outcome in health economic evaluations, while accounting for typical data features such as skewness and structural values in both utilities and costs. The analysis of the PBS data shows the benefits of using our approach compared with a standard cross-sectional model and a considerable impact of alternative MNAR assumptions on the final decision-making conclusions, suggesting a more cost-effective intervention compared with the results that are obtained under ignorability, L-MAR.

There are two main potential limitations of the framework proposed. First, the choice of aggregating all non-completers into a single pattern to handle the sparsity of the data is relatively simple to implement but rests on the critical assumption that reasons for missingness should not largely differ across these patterns, which, however, may not be realistic in more general situations. Second, the model may become computationally challenging to implement when the number of variables (i.e. time points) increases. Alternative approaches could be used to overcome these limitations. For example, the sparsity of the data in certain missingness patterns could be handled by specifying shared prior distributions across the patterns to identify the parameters in each pattern by sharing the information from the observed data across all or some of the patterns (Gaskins *et al.*, 2016). A full Baysian non-parametric specification could also be used to handle sparse data (Linero and Daniels, 2015).

Gomes *et al.* (2019) have recently proposed a non-ignorable missingness approach for

bivariate outcomes which combines multiple-imputation methods with a copula selection model to allow for non-normal outcomes while simultaneously imputing missing outcome data under MNAR. However, this approach relies on parametric assumptions about the joint distribution of the observed and missing data, which are typically difficult to check and do not allow us to introduce sensitivity parameters (Daniels and Hogan, 2008). By contrast, our approach is based on the extrapolation factorization which allows a flexible specification of the model for the observed data. The extrapolation distribution is then separately identified only up to the marginal mean with partial identifying restrictions using the marginal means that are estimated from the non-completers. As an alternative approach, we could have used the marginal mean estimates from the completers, but we considered those of the incompleters as a more reasonable default MNAR assumption. Next, we used sensitivity parameters to characterize the uncertainty about the missing data within each pattern. Our framework may be particularly useful when external sources of information about missingness are available (e.g. expert opinion) as they can be formally incorporated in the model and their effect on the results can be transparently assessed by using different priors on the sensitivity parameters.

Although the priors for the sensitivity parameters in Section 5.3 were not derived from a formal elicitation of expert opinion, given that it was not available in the PBS study, they offer a convenient framework to assess the robustness of the conclusions to differing MNAR departures based on the magnitude of quantities (e.g. the variance) from the observed data. In our analysis we considered only departures where missing values were assumed to be associated with higher utilities and lower costs compared with the observed data at each time point (exploring different variations). We choose these departures on the basis of a discussion with the people who are involved in the trial and to provide a reasonable number of scenarios to explore. However, in general cases these assumptions may not be plausible and alternative departures could be considered on the basis of the available information about missingness.

Our framework represents a considerable step forward for the handling of missingness in economic evaluations compared with the current practice, which typically relies on methods that assume an ignorable MAR assumption and rarely conducts sensitivity analysis to MNAR departures. Nevertheless, further improvements are certainly possible. For example, a potential area for future work is to increase the flexibility of our approach through a semiparametric or non-parametric specification for the observed data distribution, which would allow a weakening of the model assumptions and probably further improve the fit of the model to the observed data and address sparse patterns in an automated way. As for the extrapolation distribution, alternative identifying restrictions that introduce the sensitivity parameters via the conditional mean (rather than the marginal mean) could be considered, and their effect on the conclusions assessed in a sensitivity analysis.

Acknowledgements

Dr Michael J. Daniels is partially supported by US National Institutes of Health grant CA-183854.

Dr Gianluca Baio is partially supported as the recipient of an unrestricted research grant sponsored by the Mapi Group at University College London.

Dr Andrea Gabrio was partially funded in his doctoral programme at University College London by a research grant sponsored by Foundation Blanceflor Boncompagni Ludovisi, *née* Bildt.

References

Baio, G. (2012) *Bayesian Methods in Health Economics*. Boca Raton: Chapman and Hall–CRC.

Baio, G. (2014) Bayesian models for cost-effectiveness analysis in the presence of structural zero costs. *Statist. Med.*, **33**, 1900–1913.
Basu, A. and Manca, A. (2012) Regression estimators for generic health-related quality of life and quality-adjusted life years. *Med. Decsn Makng*, **1**, 56–69.
Black, W. (1990) A graphic representation of cost-effectiveness. *Med. Decsn Makng*, **10**, 212–214.
Briggs, A. (2000) Handling uncertainty in cost-effectiveness models. *PharmacoEconomics*, **22**, 479–500.
Briggs, A., Schulpher, M. and Claxton, K. (2006) *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press.
Brooks, S., Gelman, A., Jones, G. and Meng, X. (2011) *Handbook of Markov Chain Monte Carlo*. Boca Raton: CRC Press.
Celeux, G., Forbes, S., Robert, C. and Titterington, D. (2006) Deviance information criteria for missing data models. *Baysn Anal.*, **1**, 651–674.
Claxton, K. (1999) The irrelevance of inference: a decision making approach to stochastic evaluation of health care technologies. *J. Hlth Econ.*, **18**, 342–364.
Cooper, N., Sutton, A., Mugford, M. and Abrams, K. (2003) Use of Bayesian Markov Chain Monte Carlo methods to model cost-of-illness based on general recommended guidelines. *Med. Decsn Makng*, **23**, 38–53.
Daniels, M. and Hogan, J. (2000) Reparameterizing the pattern mixture model for sensitivity analysis under informative dropout. *Biometrics*, **56**, 1241–1248.
Daniels, M. and Hogan, J. (2008) *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. New York: Chapman and Hall.
Diaz-Ordaz, K., Kenward, M. G. and Grieve, R. (2014) Handling missing values in cost effectiveness analyses that use data from cluster randomized trials. *J. R. Statist. Soc.* A, **177**, 457–474.
European Medicines Agency (2013) Guideline on adjustment for baseline covariates. Committee for Medicinal Products for Human Use, European Medicines Agency, London. (Available from `http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/06/WC500144946.pdf`.)
Gabrio, A., Mason, A. and Baio, G. (2017) Handling missing data in within-trial cost-effectiveness analysis: a review with future recommendations. *PharmacoEconomicsOpen*, **1**, 79–97.
Gabrio, A., Mason, A. and Baio, G. (2019) A full Bayesian model to handle structural ones and missingness in economic evaluations from individual level data. *Statist. Med.*, **38**, 1399–1420.
Gaskins, J., Daniels, M. and Marcus, B. (2016) Bayesian methods for nonignorable dropout in joint models in smoking cessation studies. *J. Am. Statist. Ass.*, **111**, 1454–1465.
Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2004) *Bayesian Data Analysis*, 2nd edn. New York: Chapman and Hall.
Gomes, R., Ng, E., Grieve, R., Nixon, R., Carpenter, J. and Thompson, S. (2012) Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med. Decsn Makng*, **32**, 350–361.
Gomes, M., Radice, R., Camarena Brenes, J. and Marra, G. (2019) Copula selection models for nongaussian outcomes that are missing not at random. *Statist. Med.*, **38**, 480–496.
Hassiotis, A., Poppe, M., Strydom, A., Vickerstaff, V., Hall, I., Crabtree, J., Omar, R., King, M., Hunter, R., Bosco, A., Biswas, A., Ratti, V., Blickwedel, J., Cooper, V., Howie, W. and Crawford, M. (2018) Positive behaviour support training for staff for treating challenging behaviour in people with intellectual disabilities: a cluster rct. *Hlth Technol. Assessmnt*, **22**, 1–110.
Jackson, C. H., Thompson, S. G. and Sharples, L. D. (2009) Accounting for uncertainty in health economic decision models by using model averaging. *J. R. Statist. Soc.* A, **172**, 383–404.
Koerkamp, B., Hunink, M., Stijnen, T., Hammitt, J., Kuntz, K. and Weinstein, M. (2007) Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. *Med. Decsn Makng*, **27**, 101–111.
Leurent, B., Gomes, M. and Carpenter, J. (2018a) Missing data in trial-based cost-effectiveness analysis: an incomplete journey. *Hlth Econ.*, **6**, 1024–1040.
Leurent, B., Gomes, M., Faria, R., Morris, S., Grieve, R. and Carpenter, J. (2018b) Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *PharmacoEconomics*, **36**, 889–901.
Linero, A. and Daniels, M. (2015) A flexible Bayesian approach to monotone missing data in longitudinal studies with nonignorable missingness with application to an acute schizophrenia clinical trial. *J. Am. Statist. Ass.*, **110**, 45–55.
Linero, A. and Daniels, M. (2018) Bayesian approaches for missing not at random outcome data: the role of identifying restrictions. *Statist. Sci.*, **33**, 198–213.
Little, R. (1994) A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471–483.
Little, R. and Rubin, D. (2002) *Statistical Analysis with Missing Data*, 2nd edn. New York: Wiley.
Manca, A., Hawkins, N. and Sculpher, M. (2005) Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Hlth Econ.*, **14**, 487–496.
Mason, A., Richardson, S. and Best, N. (2012) Two-pronged strategy for using DIC to compare selection models with non-ignorable missing responses. *Baysn Anal.*, **7**, 109–146.
Molenberghs, G., Fitzmaurice, G., Kenward, M., Tsiatis, A. and Verbeke, G. (2015) *Handbook of Missing Data Methodology*. Boca Raton: Chapman and Hall.
Molenberghs, G., Kenward, M. and Lesaffre, E. (1997) The analysis of longitudinal ordinal data with non-random drop-out. *Biometrika*, **84**, 33–44.

National Institute for Health and Care Excellence (2013) *Guide to the Methods of Technological Appraisal*. London: National Institute for Health and Care Excellence.
Ng, E., Diaz-Ordaz, K., Grieve, R., Nixon, R., Thompson, S. and Carpenter, J. (2016) Multilevel models for cost-effectiveness analyses that use cluster randomised trial data: an approach to model choice. *Statist. Meth. Med. Res.*, **25**, 2036–2052.
Nixon, R. and Thompson, S. (2005) Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Hlth Econ.*, **14**, 1217–1229.
Noble, S., Hollingworth, W. and Tilling, K. (2012) Missing data in trial-based cost-effectiveness analysis: the current state of play. *Hlth Econ.*, **21**, 187–200.
O'Hagan, A., McCabe, C., Hakehurst, R., Brennan, A., Briggs, A., Claxton, K., Fenwick, E., Fryback, D., Schulpher, M., Spiegelhalter, D. and Willan, A. (2004) Incorporation of uncertainty in health economic modelling studies. *PharmacoEconomics*, **23**, 529–536.
O'Hagan, A. and Stevens, J. (2001) A framework for cost-effectiveness analysis from clinical trial data. *Hlth Econ.*, **10**, 303–315.
Plummer, M. (2010) JAGS: just another Gibbs sampler. (Available from `http://www-fis.iarc.fr/~martyn/software/jags/`.)
Ramsey, S., Willke, R., Glick, H., Reed, S., Augustovski, F., Johnsson, B., Briggs, A. and Sullivan, S. (2015) Cost-effectiveness analysis alongside clinical trials II—an ISPOR good research practices task force report. *Val. Hlth*, **18**, 161–172.
Rubin, D. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
Scharfstein, D., Rotnitzky, A. and Robins, J. (1999) Adjusting for nonignorable dropout using semiparametric nonresponse models. *J. Am. Statist. Ass.*, **94**, 1135–1146.
Sculpher, M., Claxton, K., Drummond, M. and McCabe, C. (2005) Whither trial-based economic evaluation for health decision making? *Hlth Econ.*, **15**, 677–687.
Spiegelhalter, D., Abrams, K. and Myles, J. (2004) *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Chichester: Wiley.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.
Su, Y. and Yajima, M. (2015) Package 'R2jags'. (Available from `https://cran.r-project.org/web/packages/R2jags/index.html`.)
Thompson, S. and Nixon, R. (2005) How sensitive are cost-effectiveness analyses to choice of parametric distributions? *Med. Decsn Makng*, **4**, 416–423.
Van Asselt, A., van Mastrigt, G., Dirksen, C., Arntz, A., Severens, J. and Kessels, A. (2009) How to deal with cost differences at baseline. *PharmacoEconomics*, **27**, 519–528.
Van Hout, B., Al, M., Gordon, G., Rutten, F. and Kuntz, K. (1994) Costs, effects and C/E-Ratios alongside a clinical trial. *Hlth Econ.*, **3**, 309–319.
Vansteelandt, S., Goetghebeur, E., Kenward, M. and Molenberghs, G. (2006) Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statist. Sin.*, **16**, 953–979.
Verbeke, G. and Molenberghs, G. (2009) *Linear Mixed Models for Longitudinal Data*. New York: Springer Science and Business Media.
Wang, C. and Daniels, M. (2011) A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics*, **67**, 810–818.
Xu, D., Chatterjee, A. and Daniels, M. (2016) A note on posterior predictive checks to assess model fit for incomplete data. *Statist. Med.*, **35**, 5029–5039.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Web appendix to A Bayesian parametric approach to handle missing longitudinal outcome data in trial-based health economic evaluations'.