Original Research Article

MDM Medical Decision Making

Medical Decision Making
2021, Vol. 41(8) 1033–1048

DOI: 10.1177/0272989X211012348
journals.sagepub.com/home/mdm
SAGE

A Bayesian Framework for Patient-Level Partitioned Survival Cost-Utility Analysis

Andrea Gabrio

Patient-level health economic data collected alongside clinical trials are an important component of the process of technology appraisal. For end-of-life treatments, the modeling of cost-effectiveness data may involve some form of partitioned survival analysis, in which measures of quality of life and survival for pre- and postprogression periods are combined to generate aggregate measures of clinical benefits (e.g., quality-adjusted survival). In addition, resource use data are often collected and costs are calculated for each type of health service (e.g., treatment, hospital, or adverse events costs). A critical problem in these analyses is that effectiveness and cost data present some complexities, such as nonnormality, spikes, and missingness, which should be addressed using appropriate methods to avoid biased results. This article proposes a general Bayesian framework that takes into account the complexities of trial-based partitioned survival cost-utility data to provide more adequate evidence for policy makers. Our approach is motivated by, and applied to, a working example based on data from a trial assessing the cost-effectiveness of a new treatment for patients with advanced non–small-cell lung cancer.

Highlights

- This is the first article proposing a Bayesian modeling framework for patient-level partitioned survival cost-utility analysis that can jointly take into account the typical complexities of the data, such as correlation, skewness, and spikes at zeros.
- The framework is defined using a modular structure that enables a flexible model specification in terms of a sequence of conditional parametric distributions that can be chosen based on the specific characteristics of each modeled variable (e.g., type of effectiveness or cost component).
- The proposed methods overcome the limitations of standard approaches that ignore at least some of the complexities of the data and, although easier to implement and well established among practitioners, may mislead cost-effectiveness decisions of policy makers.

Keywords
Bayesian statistics, economic evaluations, hurdle models, missing data, partitioned survival cost-utility analysis, STAN

Date received: May 26, 2020; accepted: March 26, 2021

The statistical analysis of health economic data is an increasingly important component of clinical trials, which provide one of the earliest opportunities to generate economic data that can be used for decision making.[1] The standard analysis of individual-level data involves the comparison of 2 interventions for which suitable measures of effectiveness and costs are collected on each patient enrolled in the trial, often at different time points during the follow-up. Different types of resource use data (e.g., hospital visits, consultations, scans, number of

Corresponding Author
Andrea Gabrio, Department of Statistical Science, University College London, Gower St, London, WC1E 6BT, UK (ucakgab@ucl.ac.uk).

doses, etc.) are collected for each patient and time point using electronic health records, self-reported questionnaires, or a combination of these. Service use information is combined with unit prices to calculate patient-level costs for different health services and then summed up over the follow-up. The effectiveness is often measured in terms of preference-based health-related quality-of-life instruments (e.g., the EQ-5D questionnaires[2]) and combined with national tariff systems to express the patients' health states in terms of utility scores. For the United Kingdom, utilities are usually measured on a scale from −0.594 (worst imaginable health) to 1 (perfect health).[3] A utility of zero is associated with death, whereas negative utilities denote health states that are valued "worse than death" by the patients. A single metric, called quality-adjusted life-years (QALYs), is then calculated by aggregating the utility scores over the follow-up and represents the health outcome of choice in the economic analysis. A common approach for calculating a QALY is the area under the curve (AUC)[4]:

$$\mathrm{QALY}_{it} = \sum_{j=1}^{J} \left(\frac{u_{ijt} + u_{ij-1t}}{2} \right) \delta_j, \qquad (1)$$

where u_{ijt} is the utility score for the i-th patient in treatment t at the j-th time in the trial, while $\delta_j = (\mathrm{Time}_j - \mathrm{Time}_{j-1})/(\mathrm{Unit\ of\ time})$ is the fraction of the time unit (typically 1 y) between time $j-1$ and j. We note that for patients who die, assumptions have to be made about their utility values at all time points after the time of death so that their QALYs can be computed using Equation 1. Often, a utility of 0 is associated with a state of death at a given time point and is carried over until the last follow-up.

When the primary endpoint of the trial is survival, such as in cancer trials, in which patients may be either dead or still alive at the end of the study, it is possible to combine the information from both utility and survival for each patient into a single measure. More specifically, the patient-level survival at time j is multiplied by his or her corresponding utility values at the same time to formulate a QALY endpoint on an AUC scale, also known as quality-adjusted survival (QAS):

$$\mathrm{QAS}_{it} = \sum_{j=1}^{J} \left(\frac{u_{ijt} + u_{ij-1t}}{2} \right) \left(\frac{s_{ijt} + s_{ij-1t}}{2} \right) \delta_j, \qquad (2)$$

where u_{ijt} and s_{ijt} are the utility and the survival time for the i-th patient in treatment t at the j-th time. The calculation in Equation 2 can be thought as a time-to-event analysis using the QALY as the analysis endpoint.[5]

Partitioned Survival Cost-Utility Analysis

When survival time changes rapidly after the progression of the disease, inferences about mean utilities should take into account the differences between pre- and postprogression responses as well as their dependence relationships. This is the rationale behind partitioned survival analysis, which involves the partitioning of survival data for the time-to-event endpoint, typically overall survival (OS), into 2 components: progression-free survival (PFS) and postprogression survival (PPS), with OS = PFS + PPS. In this context, QAS data can be computed separately for PFS and PPS by multiplying each survival component by the corresponding utilities collected during the pre- and postprogression periods. The partitioning of health-related quality-of-life data based on different components of survival time forms the basis for what is known as partitioned survival cost-utility analysis, in which patient-level QAS based on OS data can be expressed as

$$\mathrm{QAS}_{it}^{\mathrm{OS}} = \mathrm{QAS}_{it}^{\mathrm{PFS}} + \mathrm{QAS}_{it}^{\mathrm{PPS}}, \qquad (3)$$

where $\mathrm{QAS}_{it}^{\mathrm{PFS}}$ and $\mathrm{QAS}_{it}^{\mathrm{PPS}}$ are the QAS computed as in Equation 2 using patient-level utilities and survival times for the pre- and postprogression periods, respectively. In many cases, the different survival components in Equation 3 are analyzed separately using parametric regression models.[6,7] However, direct modeling of $\mathrm{QAS}_{it}^{\mathrm{PPS}}$ is not possible when the utility data are collected only up to progression; in this case, the utilities after disease progressions are usually extrapolated based on some modeling assumptions and OS or PFS data.[8]

We note that the calculation of QAS in Equation 2 and Equation 3 assumes the absence of censoring. In practice, however, some of the patients may be still alive at the end of the trial (censored). When this occurs, unfortunately, the calculation of QAS data based on utility scores may alter survival times and result in informative censoring, which can distort the inferences.[5] For the rest of the article, we will assume that no informative censoring occurs (in our case study, >99% of patients had died during the follow-up) so that standard

Department of Statistical Science, University College London, London, UK (AG). The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author received no financial support for the research, authorship, and/or publication of this article.

partitioned survival cost-utility analysis methods can be assumed to be valid. In the "Discussion" section, we will discuss the potential implications and possible methods to perform the economic assessment in the presence of censored data.

Statistical modeling for trial-based cost-utility data has received much attention in both the health economics and the statistical literature in recent years,[9,10] increasingly often under a Bayesian approach.[11–13] From the statistical point of view, this is a challenging problem because of the generally complex relationships linking the measure of effectiveness (e.g., QALYs) and the associated costs. First, the presence of a bivariate outcome requires the use of appropriate methods to account for the correlation between the variables.[14–16] Second, both utility and cost data are characterized by empirical distributions that are highly skewed, and simplifying assumptions, such as (bivariate) normality of the underlying distributions, are usually not granted. The adoption of parametric distributions that can account for skewness (e.g., beta for the utilities and gamma or log-normal for the costs) has been suggested to improve the fit of the models.[17–19] Third, data may exhibit spikes at 1 or both of the boundaries of the range for the underlying distributions, for example, zero costs and perfect health (i.e., utility of 1), which are difficult to capture with standard parametric models.[19,20] The use of more flexible formulations, known as hurdle models, has been recommended to explicitly account for these "structural" values.[21–23] Hurdle models consist of a mixture of a point mass distribution (the spike) and a parametric model fitted to the natural range of the relevant variable without the boundary values. Finally, individual-level data from clinical trials are almost invariably affected by the problem of missing data. Analyses that are limited to individuals with fully observed data (complete case analysis) are inefficient and yield biased results when the completers are not a random sample of all individuals in the trial. Alternative and more efficient approaches, such as multiple imputation and likelihood-based methods, rely on the less restrictive assumption that missingness can be fully explained based on the observed data, an assumption known as missing at random (MAR).[24,25] However, MAR can never be tested from the data at hand, and when missingness depends on some unobserved data, an assumption known as missing not at random (MNAR), analyses based on the observed data alone will yield biased results. Content-specific knowledge and tailored modeling approaches can be used to make inferences under MNAR, and within a Bayesian approach, informative prior distributions represent a powerful tool for conducting sensitivity analysis to different missingness assumptions.[26]

Outline

In this article, we extend the current methods for modeling trial-based partitioned survival cost-utility data, taking advantage of the flexibility of the Bayesian approach, and specify a joint probabilistic model for the health economic outcomes. We propose a general framework that is able to account for the multiple types of complexities affecting individual-level data (correlation, missingness, skewness, and structural values) while also explicitly modeling the dependence relationships between different types of quality of life and cost components. The article is structured as follows: first, in the next section, we present our modeling framework. In the "Example: The TOPICAL Trial" section, we present our motivating example, and in the "Application to the TOPICAL Study" section, we specify the model to handle the characteristics of the data from the case study. In the "Results" section, we summarize the statistical and health economic results of the analysis. Finally, in the "Discussion" section, we conclude with a discussion.

Methods

Consider a clinical trial in which patient-level information on a set of suitably defined effectiveness and cost variables is collected at J time points on N individuals, who have been allocated to T intervention groups. Assume that the primary endpoint of the trial is OS, whereas the secondary endpoints include PFS, a self-reported health-related quality-of-life questionnaire (e.g., EQ-5D), and health records on different types of services (e.g., drug frequency and dosage, hospital visits, etc.). Following standard health economic notation, we denote with $\boldsymbol{e}_{it}$ and $\boldsymbol{c}_{it}$ the 2 sets of health economic outcomes (effectiveness and costs) collected for the i-th individual in treatment t of the trial. For simplicity, we define $\boldsymbol{e}_{it}$ and $\boldsymbol{c}_{it}$ based on the variables collected from our motivating example that is presented in detail in the "Example: The TOPICAL Trial" section, although the framework can be modified to accommodate different types of outcomes.

The effectiveness outcomes are represented by preprogression ($e_{it}^{PFS} = \text{QAS}^{\text{PFS}}$) and postprogression ($e_{it}^{PPS} = \text{QAS}^{\text{PPS}}$) QAS data calculated using survival and utility data collected up to and beyond progression. We denote the full set of effectiveness variables as $\boldsymbol{e}_{it} = (e_{it}^{\text{PFS}}, e_{it}^{\text{PPS}})$, formed by the pre- and postprogression components.

The cost outcomes are represented by a set of K variables ($c_{it} = c_{it}^k$, for $k = 1, \ldots, K$) calculated based on K different types of health services and associated unit prices. We denote the full set of cost variables as $\boldsymbol{c}_{it} = (c_{it}^1, \ldots, c_{it}^K)$, formed by the K different cost components. Finally, it is also common to have some patient-level information on a set of additional variables $\boldsymbol{x}_{it}$ (for example, on age, sex, or potential comorbidities) that may be included in the economic analysis. Without loss of generality, we assume in the following that only 2 interventions are compared: $t = 1$ is some standard (e.g., currently recommended or applied by the health care provider), and $t = 2$ is a new intervention being suggested to potentially replace the standard.

The objective of the economic evaluation is to perform a patient-level partitioned survival cost-utility analysis by specifying a joint model $p(\boldsymbol{e}_{it}, \boldsymbol{c}_{it}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the full set of model parameters. Among these parameters, interest is in the marginal mean effectiveness and costs $\boldsymbol{\mu} = (\mu_{et}, \mu_{ct})$ that are used to inform the decision-making process. Different approaches can be used to specify $p(\boldsymbol{e}_{it}, \boldsymbol{c}_{it}|\boldsymbol{\theta})$. Here, we express the joint distribution as

$$p(\boldsymbol{e}_{it}, \boldsymbol{c}_{it}|\boldsymbol{\theta}) = p(\boldsymbol{e}_{it}|\boldsymbol{\theta}_e)p(\boldsymbol{c}_{it}|\boldsymbol{e}_{it}, \boldsymbol{\theta}_c), \tag{4}$$

where $p(\boldsymbol{e}_{it}|\boldsymbol{\theta}_e)$ is the marginal distribution of the effectiveness and $p(\boldsymbol{c}_{it}|\boldsymbol{e}_{it}\boldsymbol{\theta}_c)$ is the conditional distribution of the costs given the effectiveness, respectively, indexed by $\boldsymbol{\theta}_e$ and $\boldsymbol{\theta}_c$, with $\boldsymbol{\theta} = (\boldsymbol{\theta}_e, \boldsymbol{\theta}_c)$. Based on previous works,[21,27] we specify the model in terms of a marginal distribution for the effectiveness and a conditional distribution for the costs. A key advantage of using a conditional factorisation, compared with a multivariate marginal approach, is that univariate models for each variable can be flexibly specified to tackle the idiosyncrasies of the data (e.g., nonnormality and spikes) while also capturing the potential correlation between the variables. We now describe how the 2 factors on the right-hand side of Equation 4 can be specified.

Marginal Model for the Effectiveness

For each individual and treatment, we specify a marginal distribution of the effectiveness variables $\boldsymbol{e}_{it} = (e_{it}^{\text{PFS}}, e_{it}^{\text{PPS}})$ using the conditional factorization:

$$p(\boldsymbol{e}_{it}|\boldsymbol{\theta}_{et}) = p(e_{it}^{\text{PFS}}|\boldsymbol{\theta}_{et}^{\text{PFS}})p(e_{it}^{\text{PPS}}|e_{it}^{\text{PFS}}, \boldsymbol{\theta}_{et}^{\text{PPS}}), \tag{5}$$

where $\boldsymbol{\theta}_{et} = (\boldsymbol{\theta}_{et}^{\text{PFS}}, \boldsymbol{\theta}_{et}^{\text{PPS}})$ are the treatment-specific effectiveness parameters formed by the 2 distinct sets that index the marginal distribution of e_{it}^{PFS} and the conditional distribution of $e_{it}^{\text{PPS}}|e_{it}^{\text{PFS}}$. The parameters $\boldsymbol{\theta}_{et}$ can also be expressed in terms of location $\boldsymbol{\phi}_{iet} = (\phi_{iet}^{\text{PFS}}, \phi_{iet}^{\text{PPS}})$ and ancillary $\boldsymbol{\psi}_{et} = (\boldsymbol{\psi}_{et}^{\text{PFS}}, \boldsymbol{\psi}_{et}^{\text{PPS}})$ parameters, the latter comprising some standard deviations $\boldsymbol{\sigma}_{et} = (\sigma_{et}^{\text{PFS}}, \sigma_{et}^{\text{PPS}})$. Modeling of the location parameters as a function of other variables is typically achieved through a generalized linear structure and some link function that relates the expected value of the response to the linear predictors in the model. For example, consider

$$e_{it}^{\text{PFS}} \sim f^{\text{PFS}}(\phi_{iet}^{\text{PFS}}, \boldsymbol{\psi}_{et}^{\text{PFS}}) \quad \text{and} \quad e_{it}^{\text{PPS}}|e_{it}^{\text{PFS}} \sim f^{\text{PPS}}(\phi_{iet}^{\text{PPS}}, \boldsymbol{\psi}_{et}^{\text{PPS}}), \tag{6}$$

where $f^{\text{PFS}}(\cdot)$ and $f^{\text{PPS}}(\cdot)$ are generic parametric distributions that model e_{it}^{PFS} and $e_{it}^{\text{PFS}}|e_{it}^{\text{PFS}}$, respectively. The location parameters are then modeled as

$$\begin{aligned} g(\phi_{iet}^{\text{PFS}}) &= \alpha_{0t}^{\text{PFS}} + [\ldots], \\ g(\phi_{iet}^{\text{PPS}}) &= \alpha_{0t}^{\text{PPS}} + \alpha_{1t}^{\text{PPS}}(e_{it}^{\text{PFS}} - \mu_{et}^{\text{PFS}}) + [\ldots], \end{aligned} \tag{7}$$

where $g(\cdot)$ is the link function, $\boldsymbol{\alpha}^{\text{PFS}} = (\alpha_{0t}^{\text{PFS}}, \ldots)$ and $\boldsymbol{\alpha}^{\text{PPS}} = (\alpha_{0t}^{\text{PPS}}, \alpha_{1t}^{\text{PPS}}, \ldots)$ are the sets of regression parameters indexing the 2 models, and the notation $+ [\ldots]$ indicates that other terms (e.g., quantifying the effect of relevant covariates $\boldsymbol{x}_{it}$) may be included in each model. In the absence of covariates, the quantities $\mu_{et}^{\text{PFS}} = g^{-1}(\alpha_0^{\text{PFS}})$ and $\mu_{et}^{\text{PPS}} = g^{-1}(\alpha_0^{\text{PPS}})$ can be interpreted as the population mean effectiveness for e^{PFS} and e^{PPS}, respectively.

Conditional Model for the Costs Given the Effectiveness

We factor the distribution of $\boldsymbol{c}_{it}|\boldsymbol{e}_{it}$ as the product of a sequence of K conditional cost distributions:

$$p(\boldsymbol{c}_{it}|\boldsymbol{e}_{it}, \boldsymbol{\theta}_{ct}) = p(c_{it}^1|\boldsymbol{e}_{it}, \boldsymbol{\theta}_{ct}^1) \cdots p(c_{it}^K|\boldsymbol{e}_{it}, c_{it}^1, \ldots, c_{it}^{K-1}, \boldsymbol{\theta}_{ct}^K), \tag{8}$$

where $\boldsymbol{\theta}_{ct} = (\boldsymbol{\theta}_{ct}^1, \ldots, \boldsymbol{\theta}_{ct}^K)$ are the treatment-specific parameters that index the K conditional cost distributions. These parameters can be expressed in terms of K location $\boldsymbol{\phi}_{ict} = (\phi_{ict}^1, \ldots, \phi_{ict}^K)$ and ancillary $\boldsymbol{\psi}_{ct} = (\psi_{ct}^1, \ldots, \psi_{ct}^K)$ parameters, the latter including some standard deviations $\boldsymbol{\sigma}_{ct} = (\sigma_{ct}^1, \ldots, \sigma_{ct}^K)$. We can model each conditional cost distribution as

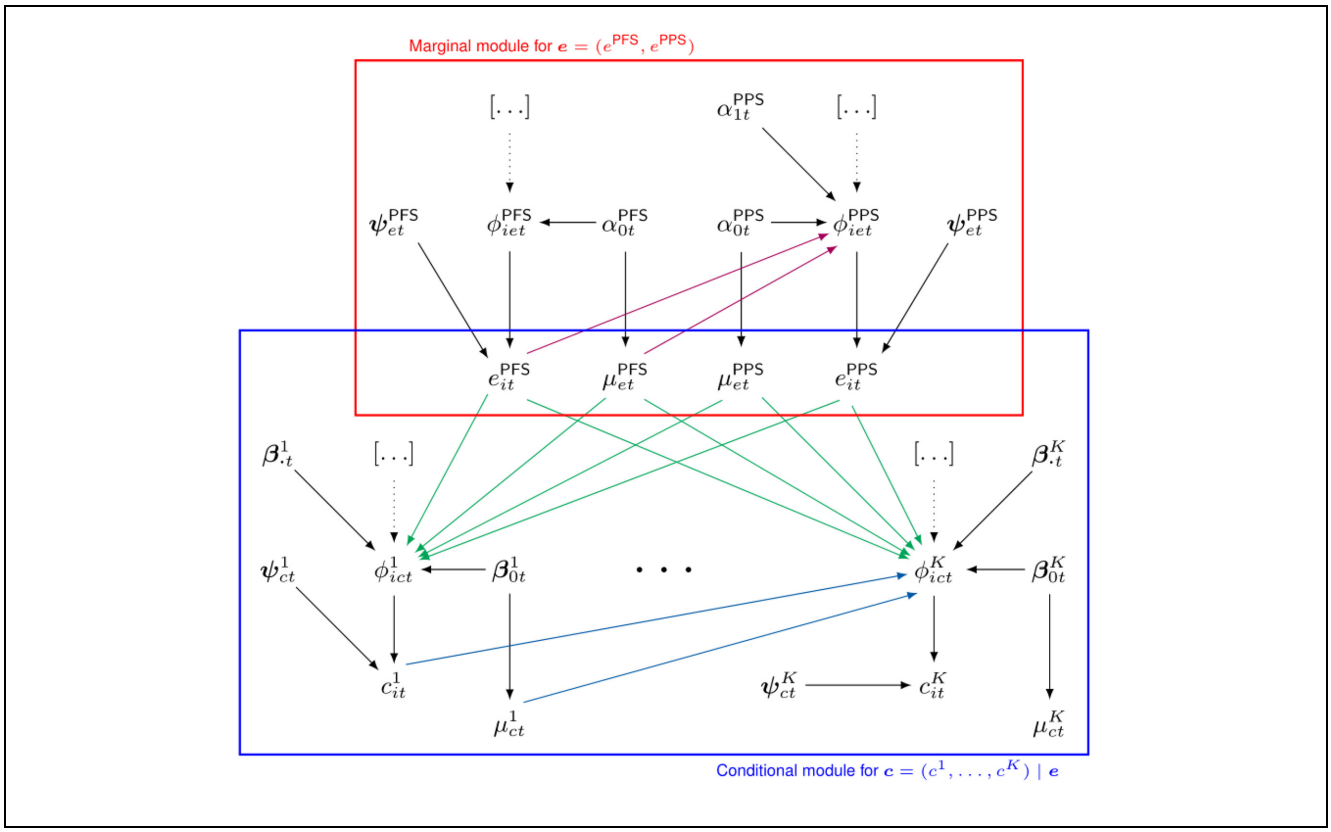

Figure 1 Joint distribution $p(\boldsymbol{e}, \boldsymbol{c})$, expressed in terms of a marginal distribution for the effectiveness variables $\boldsymbol{e} = (e^{\text{PFS}}, e^{\text{PPS}})$ and a conditional distribution for the cost variables $\boldsymbol{c} = (c^1, \dots, c^K)$ given $\boldsymbol{e}$, respectively, indicated with a solid red and blue box. The parameters indexing the corresponding distributions or modules are denoted with different Greek letters, whereas i and t denote the individual and treatment indices. The notation $\boldsymbol{\beta}_{\cdot t}^{1}$ and $\boldsymbol{\beta}_{\cdot t}^{K}$ indicates the set of the conditional mean cost regression parameters for c^1 and c^K, excluding the intercepts. The solid black and colored arrows show the dependence relationships between the parameters within and between different modules, respectively. The 3 large dots indicate the inclusion in the framework of the conditional distributions for the cost variables $c^k|\boldsymbol{e}, c^k, \dots, c^{k-1}$, for $2 < k < K$, omitted for clarity from the figure, whereas the small dots enclosed in the square brackets indicate the potential inclusion of other covariates at the mean level in each module.

$$
\begin{aligned}
c_{it}^{1}|\boldsymbol{e}_{it} &\sim f^{1}(\phi_{ict}^{1}, \boldsymbol{\psi}_{ct}^{1}), \quad \cdots, \\
c_{it}^{K}|\boldsymbol{e}_{it}, c_{it}^{1}, \dots, c_{it}^{K-1} &\sim f^{K}(\phi_{ict}^{K}, \boldsymbol{\psi}_{ct}^{K}),
\end{aligned} \tag{9}
$$

where $f^1(\cdot), \dots, f^K(\cdot)$ denote the distributions associated with the K cost components. The location parameters are modeled as a function of other variables using the generalized linear forms:

$$
\begin{aligned}
g(\phi_{ict}^{1}) &= \beta_{0t}^{1} + \beta_{1t}^{1}(e_{it}^{\text{PFS}} - \mu_{et}^{\text{PFS}}) + \beta_{2t}^{1}(e_{it}^{\text{PPS}} - \mu_{et}^{\text{PPS}}) + [\dots], \\
&\vdots \\
g(\phi_{ict}^{K}) &= \beta_{0t}^{K} + \beta_{1t}^{K}(e_{it}^{\text{PFS}} - \mu_{et}^{\text{PFS}}) + \beta_{2t}^{K}(e_{it}^{\text{PPS}} - \mu_{et}^{\text{PPS}}) + \\
&\quad \beta_{3t}^{K}(c_{it}^{1} - \mu_{ct}^{1}) + \dots + \beta_{K+1,t}^{K}(c_{it}^{K-1} - \mu_{ct}^{K-1}) + [\dots],
\end{aligned} \tag{10}
$$

where $\boldsymbol{\beta}^1 = (\beta_{0t}^1, \beta_{1t}^1, \beta_{2t}^1, \dots), \dots, \boldsymbol{\beta}^K = (\beta_{0t}^K, \beta_{1t}^K, \beta_{2t}^K, \beta_{3t}^K, \dots, \beta_{K+1,t}^K \dots)$ are the sets of regression parameters indexing the K models. Assuming other covariates are absent, the quantities $\mu_{ct}^1 = g^{-1}(\beta_{0t}^1), \dots, \mu_{ct}^K = g^{-1}(\beta_{0t}^K)$ can be interpreted as the K population mean cost components.

Figure 1 provides a visual representation of the proposed modeling framework. The effectiveness and cost distributions are represented in terms of combined "modules" (red and blue boxes), in which the random quantities are linked through logical relationships. Notably, this is general enough to be extended to any suitable distributional assumption as well as to handle covariates in each module. In the following section, we present our motivating example and describe the modeling specification in our analysis.

Example: The TOPICAL Trial

The TOPICAL study was a double-blind, randomized, placebo-controlled, phase III trial conducted in the United Kingdom. Participants were elderly patients with non–small-cell lung cancer receiving best supportive care and considered unfit for chemotherapy because of poor performance status and/or multiple medical comorbidities.[28] Subjects were randomly assigned to receive a control (oral placebo, $t = 1$) or erlotinib (150 mg per day, $t = 2$) until disease progression or unacceptable toxicity. The original trial investigated 350 patients in the active treatment and 320 in the placebo group, with the time horizon of the cost-effectiveness analysis being 1 y. For our analysis, we had access to patient-level QAS and cost data related to a subsample of 300 patients from the original study (150 in the placebo and 150 in the erlotinib group, respectively).

The primary endpoint of the trial was OS; secondary endpoints were PFS (defined as the time between randomization and progression or death) and health-related quality of life measured by the EQ-5D-3L questionnaire, which was collected at monthly intervals up to and beyond progression for each patient. Because all patients progressed/died during the follow-up of the study, no extrapolation of OS and PFS was required. For each individual, PFS and PPS QAS data were obtained by combining his or her available EQ-5D utilities with the corresponding survival times during the follow-up using the formula shown in Equation 2. This implies that QAS may take both negative and positive values according to the utilities and survival observed during the pre- (e_{it}^{PFS}) and postprogression (e_{it}^{PPS}) periods. Given the small proportions of missing EQ-5D data in both treatment groups (reported in the online supplementary material), individual-level QAS was available for more than 80% of the patients.

The costs are calculated on 3 different components: 1) drug (erlotinib), radiotherapy, and additional anticancer treatments, denoted with c_{it}^{drug}; 2) patient management (e.g., hospital visits), denoted with c_{it}^{hos}; 3) and management of treatment-related adverse events (e.g., rash), denoted with c_{it}^{ae}. Resource use was collected monthly on case report forms and combined with unit prices from published sources to derive the costs for each component. Figures 2 and 3 show the histograms of the distributions of the different components of the observed QAS and cost data in both treatment groups, respectively. The number of observations and the empirical mean and standard deviations for each variable are reported in the graphs. The observed distributions of e_{it}^{PFS} and e_{it}^{PPS} show a considerable degree of skewness in both treatment groups, especially for postprogression QAS data. Although most of the values for e_{it}^{PFS} lie in $[0.002; 0.2]$ with no actual observed zero values, the distributions of e_{it}^{PPS} show a lower bound at zero, with about 50% of the individuals in each group being associated with this boundary (structural value). All 31 patients (12 in the control and 19 in the intervention) associated with negative e^{PFS} values (mean $= -0.05$, SD $= 0.12$) have either died or improved to positive e^{PPS} values in the postprogression period. The observed distributions of c_{it}^{drug}, c_{it}^{hos}, and c_{it}^{ae} show a high degree of skewness, especially in the intervention group. All costs are defined on a positive range, but each component has a different variability, with c^{drug} in the intervention being the component associated with the largest standard deviation. The proportions of individuals who are associated with a structural zero cost are 60% (only in the control group) for c_{it}^{drug}, 25% (in each group) for c_{it}^{hos}, and 18% (in each group) for c_{it}^{ae}.

The total number of individuals with fully observed data for all variables (completers) was 249 (83%), whereas among those with partially observed data (51; 27%), most were associated with unobserved values for either e_{it}^{PFS}, e_{it}^{PPS}, or c_{it}^{drug} or a combination of these (29/51 = 57%). A detailed presentation of the missingness patterns is reported in the online supplementary material. We note that missingness in $\boldsymbol{e}$ is only due to incomplete EQ-5D questionnaires (and thus utility scores) and not censoring of survival time, as all patients progressed/died by the time of the analysis. Missingness in $\boldsymbol{c}$ is due to incomplete information from the case report forms on resource use.

Application to the TOPICAL Study

Model Specification

Throughout, we refer to our motivating example to demonstrate the flexibility of the proposed approach for dealing with the complexities of cost-utility data. The selection of the parametric distributions to model each variable is based on relative and absolute measures of fit, including predictive information criteria and predictive checks, which are explained and reported in detail in the "Model Assessment" section. We start by modeling e_{it}^{PFS} using a Gumbel distribution with an identify link function for the mean:

$$\begin{aligned} e_{it}^{\text{PFS}} &\sim \text{Gumbel}(\phi_{et}^{\text{PFS}}, \sigma_{et}^{\text{PFS}}), \\ \phi_{et}^{\text{PFS}} &= \alpha_{0t}^{\text{PFS}}, \end{aligned} \tag{11}$$

where ϕ_{et}^{PFS} and σ_{et}^{PFS} are the mean and standard deviation of e_{it}^{PFS}. The Gumbel distribution has already been recommended for modeling utility data, as it is defined

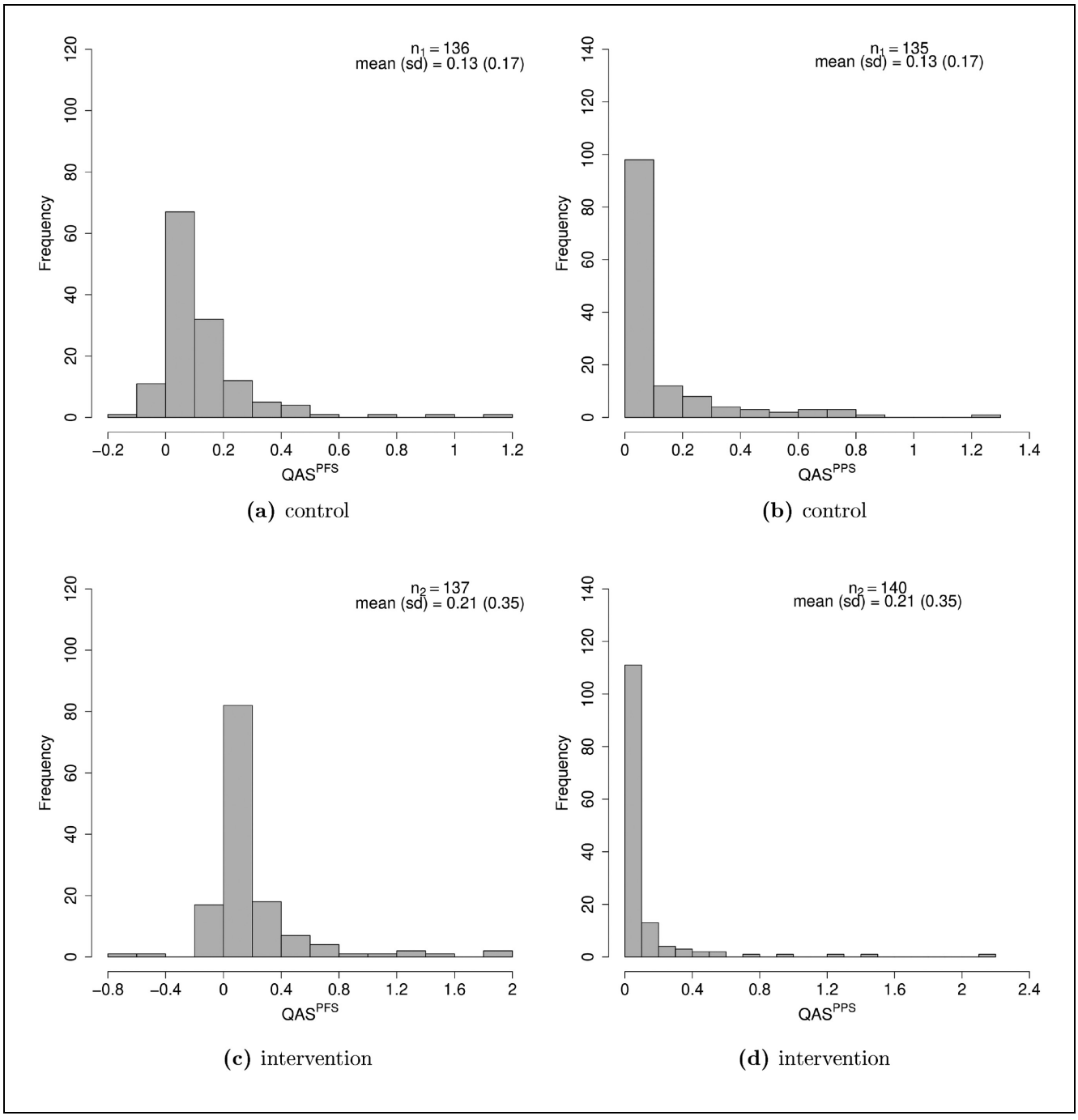

Figure 2 Histograms of the distributions of the pre- and postprogression quality-adjusted survival (QAS) data, in the control (a, b) and intervention (c, d) group. About 50% of the individuals in both groups are associated with zero postprogression survival QAS (highest bars in panels b and d), while no actual zero is observed for progression-free survival QAS, which mainly lies between $[0.002, 0.2]$ (highest bars in panels a and c).

on the real line while also being able to capture skewness.[29] We parameterize the Gumbel distribution in terms of mean and standard deviation to facilitate the specification of the priors on the parameters, compared with using the canonical location a (real) and scale $b>0$ parameters. More specifically, the mean and standard

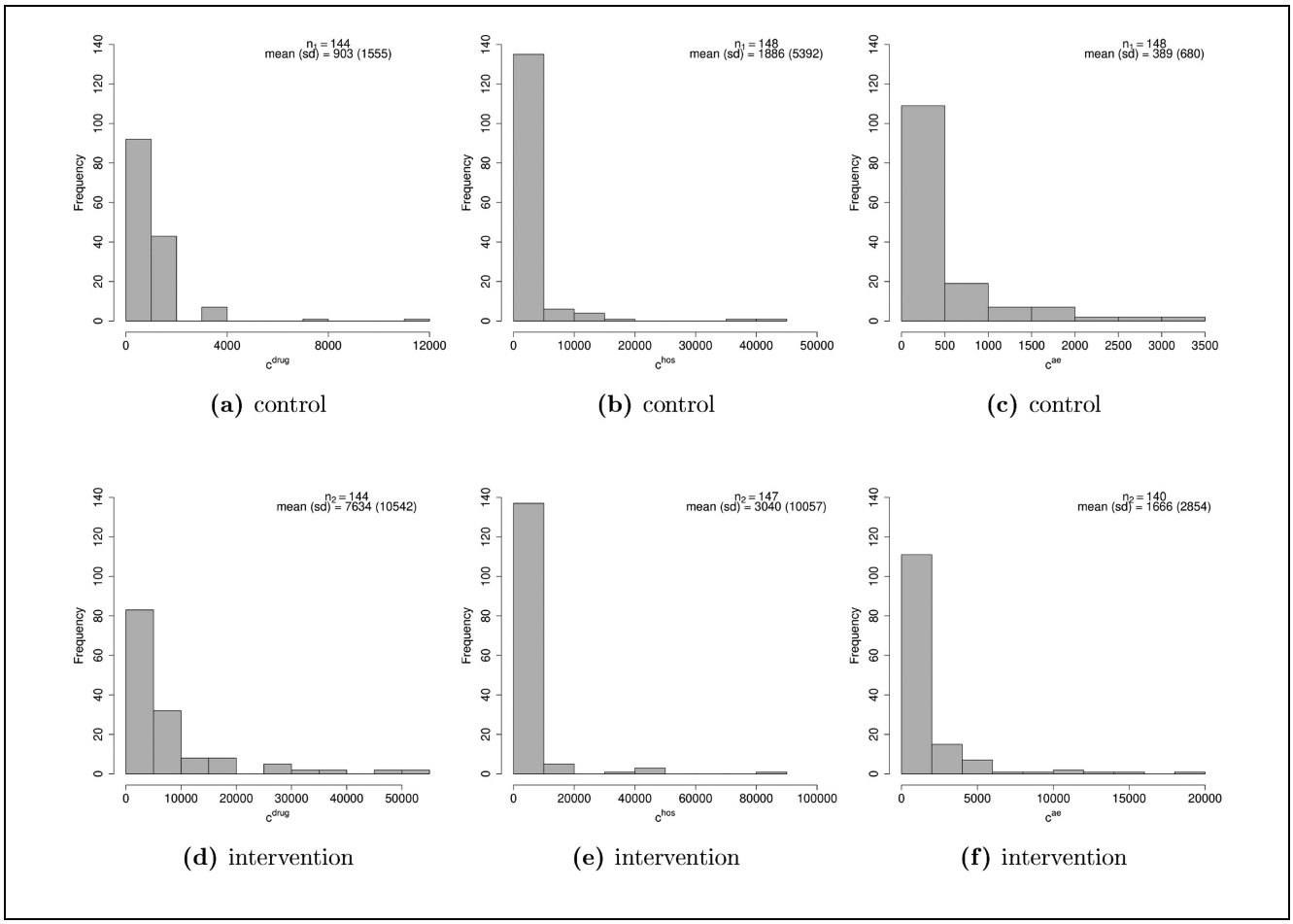

Figure 3 Histograms of the distributions of the 3 cost components (drug, hospital, and adverse events) in the control (a–c) and intervention (d–f) groups (all costs are expressed in pounds).

deviation of the Gumbel distribution are linked to the canonical parameters through the relationships $a = \phi - b\kappa$ and $b = (\sigma\sqrt{6})/\pi$, where κ is the Euler's constant. We refer to the online supplementary material for a detailed presentation of the Gumbel distribution.

When choosing the model for e_{it}^{PPS}, it is important to take into account the considerable proportion of people associated with a zero value in both treatment groups (Figure 2). Here, we specify a hurdle approach that expresses the distribution of e_{it}^{PPS} as a mixture of a point mass distribution at zero and a parametric model for the natural range of the variable excluding the zeros. Specifically, for each subject, we define an indicator variable d_{it}^{PPS} taking value 1 if the i-th individual is associated with $e_{it}^{\text{PPS}} = 0$ and 0 otherwise (i.e., $e_{it}^{\text{PFS}} > 0$). We then model the conditional distribution of $d_{it}^{\text{PPS}} | e_{it}^{\text{PFS}}$ with a Bernoulli distribution using a logit link function for the probability of being associated with a zero:

$$\begin{aligned} d_{it}^{\text{PPS}} | e_{it}^{\text{PFS}} &\sim \text{Bernoulli}(\pi_{iet}^{\text{PPS}}), \\ \text{logit}(\pi_{iet}^{\text{PPS}}) &= \gamma_{0t}^{\text{PPS}} + \gamma_{1t}^{\text{PPS}} e_{it}^{\text{PFS}}, \end{aligned} \tag{12}$$

where π_{iet}^{PPS} is the probability associated with $e_{it}^{\text{PPS}} = 0$, which is expressed as a linear function of e_{it}^{PFS} on the logit scale via the intercept and slope parameters γ_{0t}^{PPS} and γ_{1t}^{PPS}, respectively. Other covariates, which are thought to be strongly associated with the chance of having a zero, can also be included in the logistic regression to improve the estimation of the probabilities. However, in our analysis, the inclusion of any of the baseline variables available in the trial did not lead to substantial changes in the inferences, while also not improving the fit of the model to the observed data compared with Equation 12. Thus, we decided to remove these variables and keep the current specification for the model of d_{it}^{PPS}. We model $e_{it}^{\text{PPS}} | d_{it}^{\text{PPS}} = 0, e_{it}^{\text{PFS}}$

with an exponential distribution using a log-link function for the conditional mean:

$$\begin{aligned} e_{it}^{\text{PPS}} | d_{it}^{\text{PPS}} = 0, e_{it}^{\text{PFS}} &\sim \text{Exponential}(\phi_{iet}^{\text{PPS}}), \\ \log(\phi_{iet}^{\text{PPS}}) &= \alpha_{0t}^{\text{PPS}} + \alpha_{1t}^{\text{PPS}} e_{it}^{\text{PFS}}, \end{aligned} \tag{13}$$

where α_{0t}^{PFS} and α_{1t}^{PPS} are the intercept and slope regression parameters for $e_{it}^{\text{PPS}}>0$, defined on the log scale. Again, the choice of the exponential distribution was made according to the fit to the observed e_{it}^{PPS} after comparing alternative model specifications. We note that the canonical rate parameter r of the exponential distribution can be retrieved from the mean parameter through the relationship $r = \frac{1}{\phi}$.

Next, we specify the conditional distributions of the cost variables $\boldsymbol{c}_{it} = (c_{it}^{\text{drug}}, c_{it}^{\text{hos}}, c_{it}^{\text{ae}})$ using a hurdle approach to handle the zero costs and fit log-normal distributions to the positive cost values (chosen in light of the better fit to the observed data compared with Gamma distributions). For each modeled cost variable, we checked whether the inclusion of any of the available baseline covariates from the trial could lead to some model improvement in terms of fit to the observed data or parameter estimates. However, results from the different model specifications suggest that there is no substantial gain from including these variables, which were therefore removed. We model the conditional distribution of the zero drug cost indicators and drug cost variables given $\boldsymbol{e}_{it}$ as

$$\begin{aligned} d_{it}^{\text{drug}} | \boldsymbol{e}_{it} &\sim \text{Bernoulli}(\pi_{ict}^{\text{drug}}), \\ \text{logit}(\pi_{ict}^{\text{drug}}) &= \delta_{0t}^{\text{drug}} + \delta_{1t}^{\text{drug}} e_{it}^{\text{PFS}} + \delta_{2t}^{\text{drug}} e_{it}^{\text{PPS}}, \\ c_{it}^{\text{drug}} | d_{it}^{\text{drug}} = 0, \boldsymbol{e}_{it} &\sim \text{Lognormal}(\phi_{ict}^{\text{drug}}, \sigma_{ct}^{\text{drug}}), \\ \phi_{ict}^{\text{drug}} &= \beta_{0t}^{\text{drug}} + \beta_{1t}^{\text{drug}} e_{it}^{\text{PFS}} + \beta_{2t}^{\text{drug}} e_{it}^{\text{PPS}}, \end{aligned} \tag{14}$$

where π_{ict}^{drug} is the probability of having $c_{it}^{\text{drug}} = 0$, while ϕ_{ict}^{drug} and $\sigma_{ct}^{\text{drug}}$ are the mean and standard deviation parameters for $c_{it}^{\text{drug}}>0$ on the log scale. The regression parameters $\boldsymbol{\delta}^{\text{drug}} = (\delta_{0t}^{\text{drug}}, \delta_{1t}^{\text{drug}}, \delta_{2t}^{\text{drug}})$ and $\boldsymbol{\beta}^{\text{drug}} = (\beta_{0t}^{\text{drug}}, \beta_{1t}^{\text{drug}}, \beta_{2t}^{\text{drug}})$ capture the dependence between drug costs and the effectiveness variables for the zero and nonzero components, respectively. The conditional distribution of the zero hospital cost indicators and hospital cost variables given $\boldsymbol{e}_{it}$ and c_{it}^{drug} is specified as

$$\begin{aligned} d_{it}^{\text{hos}} | \boldsymbol{e}_{it}, c_{it}^{\text{drug}} &\sim \text{Bernoulli}(\pi_{ict}^{\text{hos}}), \\ \text{logit}(\pi_{ict}^{\text{hos}}) &= \delta_{0t}^{\text{hos}} + \delta_{1t}^{\text{hos}} e_{it}^{\text{PFS}} + \delta_{2t}^{\text{hos}} e_{it}^{\text{PPS}} + \delta_{3t}^{\text{hos}} \log(c_{it}^{\text{drug}}), \\ c_{it}^{\text{hos}} | d_{it}^{\text{hos}} = 0, \boldsymbol{e}_{it}, c_{it}^{\text{drug}} &\sim \text{Lognormal}(\phi_{ict}^{\text{hos}}, \sigma_{ct}^{\text{hos}}), \\ \phi_{ict}^{\text{hos}} &= \beta_{0t}^{\text{hos}} + \beta_{1t}^{\text{hos}} e_{it}^{\text{PFS}} + \beta_{2t}^{\text{hos}} e_{it}^{\text{PPS}} + \beta_{3t}^{\text{hos}} \log(c_{it}^{\text{drug}}), \end{aligned} \tag{15}$$

where π_{ict}^{hos} is the probability of having $c_{it}^{\text{hos}} = 0$, while ϕ_{ict}^{hos} and σ_{ct}^{hos} are the mean and standard deviation parameters for $c_{it}^{\text{hos}}>0$ on the log scale. The regression parameters $\boldsymbol{\delta}^{\text{hos}}$ and $\boldsymbol{\beta}^{\text{hos}}$ capture the dependence between hospital costs, the effectiveness, and the drug cost variables for the zero and nonzero components, respectively. Finally, we specify the conditional distribution of the zero adverse event cost indicators and adverse events cost variables given $\boldsymbol{e}_{it}$, c_{it}^{drug} and c_{it}^{hos} as

$$\begin{aligned} d_{it}^{\text{ae}} | \boldsymbol{e}_{it}, c_{it}^{\text{drug}}, c_{it}^{\text{hos}} &\sim \text{Bernoulli}(\pi_{ict}^{\text{ae}}), \\ \text{logit}(\pi_{ict}^{\text{ae}}) &= \delta_{0t}^{\text{ae}} + \delta_{1t}^{\text{ae}} e_{it}^{\text{PFS}} + \delta_{2t}^{\text{ae}} e_{it}^{\text{PPS}} \\ &\quad + \delta_{3t}^{\text{ae}} \log(c_{it}^{\text{drug}}) + \delta_{4t}^{\text{ae}} \log(c_{it}^{\text{hos}}), \\ c_{it}^{\text{ae}} | d_{it}^{\text{ae}} = 0, \boldsymbol{e}_{it}, c_{it}^{\text{drug}}, c_{it}^{\text{hos}} &\sim \text{Lognormal}(\phi_{ict}^{\text{ae}}, \sigma_{ct}^{\text{ae}}), \\ \phi_{ict}^{\text{ae}} &= \beta_{0t}^{\text{ae}} + \beta_{1t}^{\text{ae}} e_{it}^{\text{PFS}} + \beta_{2t}^{\text{ae}} e_{it}^{\text{PPS}} \\ &\quad + \beta_{3t}^{\text{ae}} \log(c_{it}^{\text{drug}}) + \beta_{4t}^{\text{ae}} \log(c_{it}^{\text{hos}}), \end{aligned} \tag{16}$$

where π_{ict}^{ae} is the probability of having $c_{it}^{\text{ae}} = 0$, while ϕ_{ict}^{ae} and σ_{ct}^{ae} are the mean and standard deviation parameters for $c_{it}^{\text{ae}}>0$ on the log scale. The regression parameters $\boldsymbol{\delta}^{\text{ae}}$ and $\boldsymbol{\beta}^{\text{ae}}$ capture the dependence between adverse events costs, hospital costs, the effectiveness, and the drug cost variables for the zero and nonzero components, respectively.

We note that, in all hurdle model specifications, predictor-specific parameters of logistic and log-linear regressions can be interpreted, respectively, as the change in the log odds for having a zero and in the log-mean for nonzero values, due to a unit variation in the corresponding predictor. In addition, when all predictors are centred, the intercept parameters can be interpreted as the log odds of having a zero value and the marginal mean of nonzero values on the log scale. For all parameters in the model, we specify vague prior distributions: a normal distribution with a large variance on the appropriate scale for the regression parameters (e.g., Normal$(0, 10000)$) and a uniform distribution over a large positive range for the standard deviations (e.g., Uniform$(0, 10000)$). Although the proposed model requires the specification of a relatively large number of parameters, it does not ultimately affect the interpretation of the final analysis, which exclusively focuses on the marginal mean of $\boldsymbol{e}_{it}$ and $\boldsymbol{c}_{it}$.

Estimation of the Marginal Means

When standard parametric distributions are used, the marginal mean effectiveness and cost parameters for each type of modeled variable could be retrieved by simply centering each variable in the effectiveness and cost

modules. However, the use of nonnormal distributions and hurdle models makes it difficult to identify the marginal means in terms of the model parameters. To overcome this problem, we used an alternative approach based on numerical algorithms, known as Markov Chain Monte Carlo (MCMC) methods,[30] to approximate the posterior distributions of the marginal mean parameters. MCMC methods allow sampling from the desired posterior distributions of some parameters of interest via iterative and simulation-based algorithms. Specifically, we fitted the model using a particular type of MCMC algorithm known as Hamiltonian Monte Carlo, and we refer to the online supplementary material for a description of the method and its implementation in our analysis. Once the model is fitted, we save the posterior distributions of all model parameters $p(\boldsymbol{\theta}|\boldsymbol{e}_{it}, \boldsymbol{c}_{it})$ and retrieve the marginal mean effectiveness and cost through the following steps. First, at each iteration of the MCMC output, we use the posterior estimates of the model parameters to draw $l = 1, \ldots, L$ new samples for each type of effectiveness ($\tilde{\boldsymbol{e}}_{tl}$) and cost ($\tilde{\boldsymbol{c}}_{tl}$) variable. Second, at each iteration, we take the average across the newly sampled values for each variable to approximate the posterior distributions of the marginal mean effectiveness and cost parameters. For example, the posterior distribution of the marginal mean preprogression QAS and drug costs are obtain as

$$\mu_{et}^{PFS} = \frac{\sum_{l=1}^{L} \tilde{e}_{tl}^{PFS}}{L} \quad \text{and} \quad \mu_{ct}^{drug} = \frac{\sum_{l=1}^{L} \tilde{c}_{tl}^{drug}}{L}.$$

This approach is known as Monte Carlo integration and allows the approximation of the posterior distributions of the marginal means of the modeled variables by taking the average over a large number of randomly drawn samples from their target distribution. Finally, we derive the overall marginal means $\boldsymbol{\mu} = (\mu_{et}, \mu_{ct})$ by summing up the marginal mean estimates for the different components of the effectiveness and costs, that is,

$$\mu_{et} = \mu_{et}^{\text{PFS}} + \mu_{et}^{\text{PPS}} \quad \text{and} \quad \mu_{ct} = \mu_{ct}^{\text{drug}} + \mu_{ct}^{\text{hos}} + \mu_{ct}^{\text{ae}}, \quad (17)$$

where μ_{et}^{PFS} and μ_{et}^{PPS} are the pre- and postprogression mean QAS, whereas μ_{ct}^{drug}, μ_{ct}^{hos}, and μ_{ct}^{ae} are the means of the three different cost components (drug, hospital, and adverse events) in TOPICAL.

Computation

We fitted the model in STAN,[31] which is a software specifically designed for the analysis of Bayesian models using Hamiltonian Monte Carlo algorithms and which is interfaced with R through the package rstan.[32] Samples from the posterior distribution of the parameters of interest generated by STAN and saved to the R workspace are then used to produce summary statistics and plots. We ran 2 chains with 15,000 iterations per chain, using a burn-in of 3000, for a total sample of 24,000 iterations for posterior inference. For each unknown quantity in the model, we assessed convergence and autocorrelation of the MCMC simulations using diagnostic measures such as density and trace plots, the potential scale reduction factor, and the effective sample size.[33] A summary of the results from these convergence checks for the parameters of the model and the STAN code used to fit the model are provided in the supplementary material.

Model Assessment

We compute 2 relative measures of predictive accuracy to assess the fit of the proposed model specification (denoted as "original") with respect to a second parametric specification (denoted as "alternative"), in which we replace the Gumbel distribution for e_{it}^{PFS} with a logistic distribution, the exponential distribution for $e_{it}^{\text{PPS}}>0$ with a Weibull distribution, and the log-normal distributions for $\boldsymbol{c}_{it}>0$ with Gamma distributions. We specifically rely on the widely applicable information criterion (WAIC)[34] and the leave-one-out information criterion (LOOIC),[35] which provide estimates for the pointwise out-of-sample prediction accuracy from a fitted Bayesian model using the log-likelihood evaluated at the posterior simulations of the parameter values. Both measures can be viewed as an improvement on the popular deviance information criterion[36] in that they use the entire posterior distribution, are invariant to parametrization, and are asymptotically equal to Bayesian cross-validation.[37] These information criteria are obtained based on the model deviance and a penalty for model complexity known as effective number of parameters (p_D) and, when comparing a set of models based on the same data, the one associated with the lowest WAIC or LOOIC is the best-performing, among those assessed.

Results between the 2 alternative specifications are reported in Table 1. For both criteria, the values associated with the "original" specification of the model are systematically lower compared with those from the "alternative" parameterization and result in an overall better fit to the data for the first model. We have also explored alternative model specifications based on different distributions for the effectiveness and cost variables. For e^{PFS}, the distributions compared were normal, Gumbel, and logistic; for all other variablesm the distributions assessed were exponential, Weibull, log-logistic, Gamma, and log-normal. Model selection was performed based on both

Table 1 WAIC, LOOIC, and Effective Number of Parameter (p_D) Estimates for Each Variable in the Model[a]

	Original			Alternative		
Variable	**Distribution**	**WAIC (p_D)**	**LOOIC (p_D)**	**Distribution**	**WAIC (p_D)**	**LOOIC (p_D)**
e^{PFS}	Gumbel	–109 (11)	–107 (12)	Logistic	–68 (8)	–68 (8)
$e^{\text{PPS}}\|e^{\text{PFS}}$	Exponential	34 (10)	35 (10)	Weibull	36 (8)	38 (9)
$c^{\text{drug}}\|\boldsymbol{e}$	Lognormal	3283 (16)	3286 (17)	Gamma	3361 (26)	3365 (28)
$c^{\text{hos}}\|\boldsymbol{e}, c^{\text{drug}}$	Lognormal	3437 (15)	3438 (15)	Gamma	3659 (15)	3660 (16)
$c^{\text{ae}}\|\boldsymbol{e}, c^{\text{drug}}, c^{\text{hos}}$	Lognormal	3208 (22)	3211 (23)	Gamma	3437 (38)	3433 (36)
Total		9853 (74)	9863 (77)		10,425 (95)	10,428 (97)

[a]The "original" and "alternative" model specifications are assessed using different distributions for the pre-/postprogression quality-adjusted survival and the cost data. Total widely applicable information criterion (WAIC), leave-one-out information criterion (LOOIC), and p_D values are reported at the bottom of the table.

predictive information criteria (lowest WAIC and LOOIC) and posterior predictive checks (best visual fit). These comparisons suggested that the original specification was the one associated with the best performance.

We additionally assess the absolute fit of the model using the observed and replicated data, the latter being generated from the posterior predictive distribution using the posterior samples of the parameters in each effectiveness and cost module. We use the posterior estimates of the parameters to sample 10,000 replications of the data, which are then used for model assessment. We computed different types of graphical posterior predictive checks, either in terms of the entire distributions via density and cumulative density plots or in terms of the marginal mean estimates between the real and replicated data (provided in the supplementary material). Overall, these checks suggest a relatively good fit of the model for each modelled variable.

Results

This section presents the results of the analysis from a 2-fold perspective. First, the posterior distribution of the marginal means of each component of the effectiveness ($\mu_{et}^{\text{PFS}}, \mu_{et}^{\text{PPS}}$) and costs ($\mu_{ct}^{\text{drug}}, \mu_{ct}^{\text{hos}}, \mu_{ct}^{\text{ae}}$) as well as the marginal aggregated means (μ_{et}, μ_{ct}) is summarized. Second, the economic results are discussed by computing the probability that the new intervention is cost-effective with respect to the control.

Posterior Estimates

Figure 4 compares the posterior means (squares) and the 50% (thick lines) and 95% (thin lines) highest posterior density (HPD) credible intervals for the marginal means of each effectiveness and cost components, obtained after fitting the model to all cases under an MAR assumption. Results associated with the control ($t = 1$) and intervention ($t = 2$) group are indicated with red and blue colors, respectively. The posterior mean QAS is on average higher for the PFS as compared with the PPS component in both treatment groups. However, both 50% and 95% HPD intervals suggest that the estimates associated with the intervention group have a much higher degree of variability compared with those from the control, especially for the PPS component. The posterior mean costs for each component show that the intervention group is associated with systematically higher values with respect to the control, especially in terms of drug costs, which cover most of the total costs in the intervention. HPD intervals for mean costs show a relatively high degree of skewness, with posterior mean estimates being closer to the upper bounds of the 50% intervals compared with the lower bounds.

We derived the aggregated mean QAS and costs for each treatment group (μ_{et}, μ_{ct}) by summing up the posterior mean estimates of the different components for each type of variable. We then computed the incremental mean estimates between the 2 groups, denoted with $\Delta_e = \mu_{e2} - \mu_{e1}$ and $\Delta_c = \mu_{c2} - \mu_{c1}$, together with the incremental cost-effectiveness ratio (ICER), which represents the cost per QAS gained between the 2 groups. Table 2 shows selected posterior summaries, including means, medians, standard deviations, and 95% HPD intervals, for the marginal and incremental mean estimates. Overall, the posterior results indicate that the new intervention has systematically higher QAS and costs compared with the control, with a positive mean QAS increment of 0.14, a positive mean cost increment of £11,460, and with 95% intervals that exclude zero for both quantities. We note that posterior estimates for the marginal means in the control group show a considerably lower degree of variability (standard deviations of 0.02 and £424) as compared with those from the intervention group (standard deviations of 0.05 and £2628).

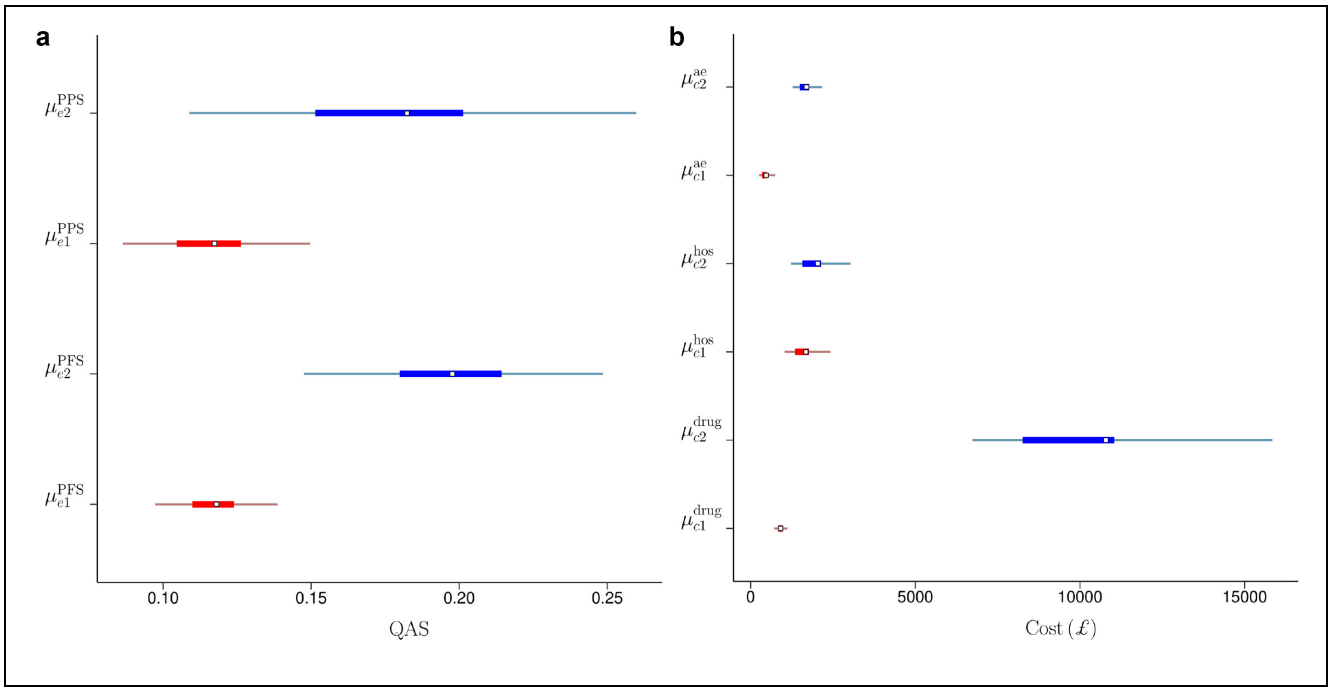

Figure 4 Posterior means (squares), 50% (thick lines) and 95% (thin lines) highest posterior density credible intervals for the marginal means of pre- and postprogression quality-adjusted survival (a) and for the marginal means of the drug, hospital, and adverse events cost (b) in the control (red) and the intervention (blue) group in the TOPICAL trial.

Table 2 Posterior Means, Medians, Standard Deviations, and 95% Highest Posterior Density Credible Intervals for the Marginal (μ_{et}, μ_{ct}) and Incremental (Δ_e, Δ_c) Mean Total Quality-Adjusted Survival and Cost Estimates Associated with the Control ($t = 1$) and Intervention ($t = 2$) Group in the TOPICAL trial[a]

Parameter	Mean	Median	SD	95% CI	
Control ($t = 1$)					
μ_{e1}	0.24	0.23	0.02	0.20	0.27
μ_{c1}	3059	3001	424	2329	3898
Intervention ($t = 2$)					
μ_{e2}	0.38	0.38	0.05	0.29	0.47
μ_{c2}	14,519	14,055	2628	10,235	19,681
Incremental					
Δ_e	0.14	0.14	0.05	0.05	0.24
Δ_c	11,460	11,013	2666	7282	16,983
Incremental cost-effectiveness ratio	79,233				

[a]For clarity, values are rounded up to 2 and 0 decimal places for e and c quantities, respectively. Costs are expressed in £.

Finally, the additional cost per unit of QAS gained is estimated to be roughly £79, 000 for $t = 2$ compared with $t = 1$.

Economic Evaluation

We complete the analysis by assessing the probability of cost-effectiveness for the new intervention with respect to the control. An advantage of using a Bayesian approach is that the economic analysis can be easily performed without the need to use ad hoc methods to represent uncertainty around point estimates (e.g., bootstrapping). Indeed, once the statistical model is fitted to the data, the samples from the posterior distributions of the parameters of interest can be used to compute different types of summary measures of cost-effectiveness.

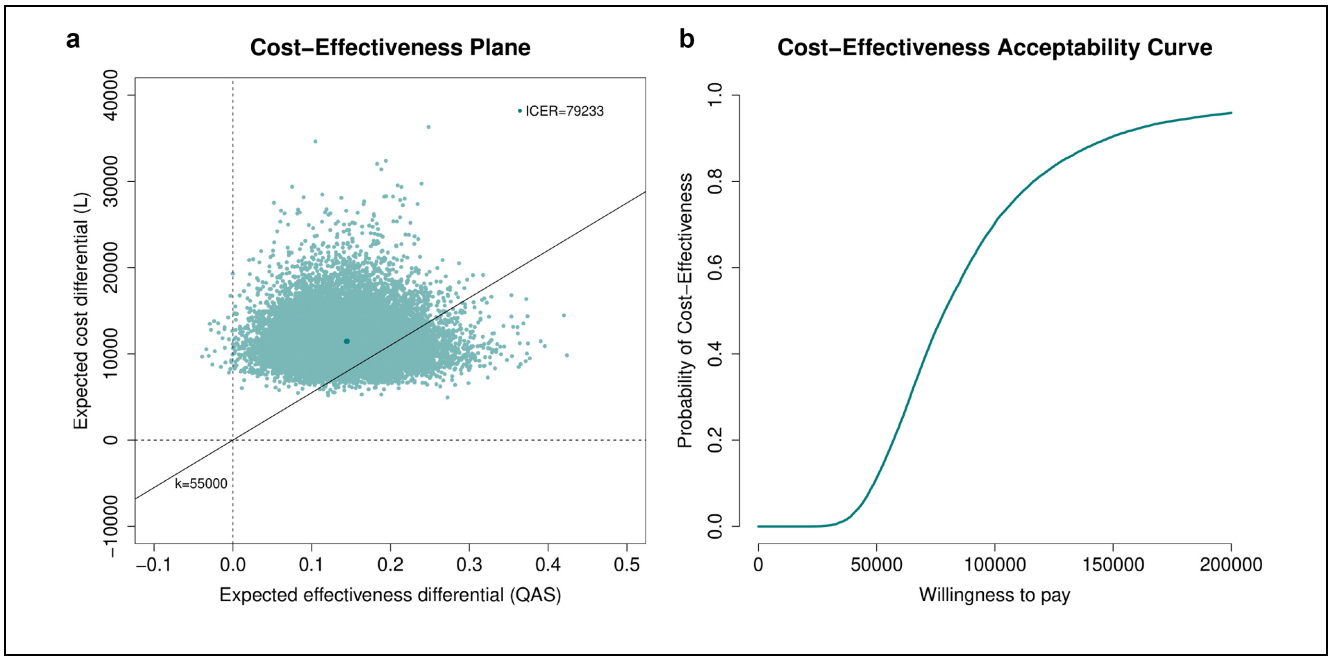

Figure 5 (a) Cost-effectiveness plane and (b) cost-effectiveness acceptability curve (CEAC) graphs associated with the 2 interventions in the TOPICAL trial. In the CEP, the value of the incremental cost-effectiveness ratio is reported (darker green dot), while the portion of the plane on the right-hand side of the straight line passing through the origin (evaluated at k = £55,000) denotes the sustainability area; in the CEAC, the probability of cost-effectiveness is shown for willingness-to-pay threshold values up to £200,000.

We specifically rely on the examination of the cost-effectiveness plane (CEP)[38] and the cost-effectiveness acceptability curve (CEAC)[39] to summarize the economic analysis. Results in terms of the expected incremental benefit are also provided in the online supplementary material. Figure 5a shows the CEP, which is a graphical representation of the joint distribution of the mean effectiveness and cost increments between the 2 groups. The slope of the straight line crossing the plane is the willingness-to-pay threshold (often indicated with k). 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. Current recommendations for generic interventions suggest a value of k between £20,000 and £30,000. However, for end-of-life treatments, such as cancer treatments, the recommended threshold values are typically higher and lie in a range between £50,000 and to 60,000 or greater.[40] Points lying below this straight line fall in the so-called sustainability area[13] and suggest that the new intervention is more cost-effective than the control. In our analysis, almost all samples fall in the north-east quadrant of the plane. This suggests that the intervention is likely to be more effective and more expensive compared with the control. At k = £55,000, the ICER (and the majority of the samples) falls outside the sustainability area, therefore indicating that the new intervention is unlikely to be considered cost-effective at the chosen value of k. Figure 5b shows the CEAC, which is obtained by computing the proportion of points lying in the sustainability area on varying the willingness-to-pay threshold k. The CEAC estimates the probability of cost-effectiveness, thus providing a simple summary of the uncertainty that is associated with the "optimal" decision suggested by the ICER. The graph shows that, as the value of the willingness-to-pay threshold is increased, the chance that the new intervention becomes cost-effective rises up to near full certainty for k = £150,000.

Discussion

In this article, we proposed a general framework for partitioned survival cost-utility analysis using patient-level data (e.g., from a trial), which takes into account the correlation between costs and effectiveness, skewness in the distribution of the observed data, the presence of structural zeros, and missing data. Although alternative

approaches have been proposed in the literature to handle the statistical issues affecting cost-effectiveness data, they had either considered some of these issues separately[19,21,27] or did not specifically focus on partitioned survival analyses.[22,23] The approach developed in the "Methods" section uses a flexible structure that allows for handling the typical idiosyncrasies affecting effectiveness and costs within a joint probabilistic framework. This is a key advantage of the Bayesian approach compared with other approaches, especially in health economic evaluations in which the main objective is not statistical inference per se but rather assessing the uncertainty in decision making induced by the uncertainty in the model inputs.[41,42]

The economic results from our case study should be interpreted with caution, and some potential limitations in terms of the generalizability of the proposed framework should be highlighted. First, our analysis of TOPICAL is based on a subset of the individuals in the original trial (made available to us), and therefore, it is difficult to draw any cost-effectiveness conclusions about the trial from this analysis. Second, although the results are obtained under a MAR assumption, which is typically considered more plausible than just focusing on the complete cases, missingness assumptions can never be checked from the data at hand. It is possible that the assumption of MAR is not tenable, which may therefore introduce some bias. It is recommended that departures from MAR are explored in sensitivity analysis to assess the robustness of the conclusions to some plausible MNAR scenarios.[26] However, given the limitations of our analysis in terms of the interpretation of the trial results and the lack of any external information to guide the choice of the MNAR departures, we decided not to pursue these analyses here. We note that different approaches are available to conduct sensitivity analysis to MNAR, some of which can be implemented within a Bayesian framework, for example, through the elicitation of expert opinions using prior distributions.[26,43]

Finally, although in our analysis no censoring of survival time was observed, in many studies, a considerable proportion of patients may be censored when they do not progress/die during the follow-up. When this occurs, the calculation of patient-level QAS data is typically invalid as it may introduce informative censoring, which distorts the inferences.[5] A possible strategy to deal with censored survival data is to specify 2 different models to separately estimate the marginal mean utilities and the proportion of patients still alive at each follow-up point and then combine these estimates to obtain results on a QAS scale. For example, linear mixed models can be used for estimating the mean utilities, while Kaplan-Meier or other parametric survival functions can be used to estimate the survival probabilities at each time point.[6] In future work, we hope to extend the proposed framework to handle censored survival data and assess the robustness of the results to alternative assumptions, including informative censoring (e.g., using expert opinion).

In conclusion, although our approach may not be applicable to all cases, the data analyzed are very much representative of the typical data used in partitioned survival cost-utility analysis alongside clinical trials. Thus, it is highly likely that the same features apply to other real cases. This is a very important if somewhat overlooked problem, as methods that do not take into account the complexities affecting patient-level data, while being easier to implement and well established among practitioners, may ultimately mislead cost-effectiveness conclusions and bias the decision-making process.

Acknowledgments

We wish to thank the UCL CRUK Cancer Trials Centre for providing a subset of data from the TOPICAL trial. Dr. Andrea Gabrio was affiliated with the Department of Statistical Science, University College London (UK), at the time of the trial and is currently affiliated with the Department of Methodology and Statistics, Faculty of Health Medicine and Social Sciences, Maastricht (NL).

ORCID iD

Andrea Gabrio https://orcid.org/0000-0002-7650-4534

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at http://journals.sagepub.com/home/mdm.

References

1. Glick HA, Doshi JA, Sonnad SS, Polsky D. *Economic Evaluation in Clinical Trials*. Oxford (UK): OUP; 2014.
2. Van Reenen M, Oppe M. EQ-5D-3L user guide basic information on how to use the EQ-5D-3L instrument. 2015. Available from: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-3L_UserGuide_2015.pdf
3. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;34(11):1095–108.
4. Drummond MF, Schulpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford (UK): Oxford University Press; 2005.
5. Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Stat Med*. 1990;9(11):1259–76.

6. Khan I. *Design & Analysis of Clinical Trials for Economic Evaluation & Reimbursement: An Applied Approach Using SAS & STATA*. London: Chapman and Hall/CRC; 2015.
7. Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing treatments using quality-adjusted survival: the q-twist method. *Am Stat*. 1995;49(2):161–9.
8. Williams C, Lewsey JD, Mackay DF, Briggs AH. Estimation of survival probabilities for use in cost-effectiveness analyses: a comparison of a multi-state modeling survival analysis approach with partitioned survival and Markov decision-analytic modeling. *Med Decis Making*. 2017;37(4): 427–39.
9. Willan AR, Briggs AH. *Statistical Analysis of Cost-Effectiveness Data*. Vol. 37. New York: John Wiley & Sons; 2006.
10. Ramsey SD, Willke RJ, Glick H, et al. Cost-effectiveness analysis alongside clinical trials ii—an ISPOR good research practices task force report. *Value Health*. 2015; 18(2):161–72.
11. O'Hagan A, Stevens JW, Montmartin J. Bayesian cost-effectiveness analysis from clinical trial data. *Stat Med*. 2001;20(5):733–53.
12. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Vol. 13. New York: John Wiley & Sons; 2004.
13. Baio G. *Bayesian Methods in Health Economics*. London: Chapman and Hall/CRC; 2012.
14. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Econ*. 2005;14:1217–29.
15. Grieve R, Nixon R, Simon G, Thompson SG. Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized trials. *Med Decis Making*. 2010;30:163–75.
16. Gomes R, Ng ES-W, Grieve R, Nixon R, Carpenter J, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med Decis Making*. 2012;32:350–361.
17. O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Econ*. 2003;12:33–49.
18. Thompson SG, Nixon RM. How sensitive are cost-effectiveness analyses to choice of parametric distributions? *Med Decis Making*. 2005;4:416–23.
19. Basu A, Heckman JJ, Navarro-Lozano S, Urzua S. *Use of Instrumental Variables in the Presence of Heterogeneity and Self-Selection: An Application in Breast Cancer Patients*. HEDG Working Paper 07/07. York (UK): Health Econometrics and Data Group; 2007.
20. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ*. 2011;20(8):897–916.
21. Baio G. Bayesian models for cost-effectiveness analysis in the presence of structural zero costs. *Stat Med*. 2014;33(11): 1900–13.
22. Gabrio A, Mason AJ, Baio G. A full Bayesian model to handle structural ones and missingness in economic evaluations from individual-level data. *Stat Med*. 2019;38(8): 1399–420.
23. Gabrio A, Daniels MJ, Baio G. A Bayesian parametric approach to handle missing longitudinal outcome data in trial-based health economic evaluations. *J R Stat Soc Ser A Stat Soc*. 2019;183:607–629.
24. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons; 1987.
25. Molenberghs G, Fitzmaurice G, Kenward Mg, Tsiatis A, Verbeke G. *Handbook of Missing Data Methodology*. London: Chapman and Hall/CRC; 2014.
26. Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. New York: Chapman and Hall; 2008.
27. O'Hagan A, Stevens JW. A framework for cost-effectiveness analysis from clinical trial data. *Health Econ*. 2001;10:303–15.
28. Lee SM, Khan I, Upadhyay S, et al. First-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy (topical): a double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2012;13(11): 1161–70.
29. Gomes M, Radice R, Brenes JC, Marra G. Copula selection models for non-gaussian outcomes that are missing not at random. *Stat Med*. 2019;38(3):480–96.
30. Brooks S, Gelman A, Jones G, Meng X-L. *Handbook of Markov Chain Monte Carlo*. Boca Raton (FL): CRC Press; 2011.
31. Carpenter B, Gelman A, Hoffman MD, et al. Stan: a probabilistic programming language. *J Stat Softw*. 2017;76(1).
32. Stan Development Team. *Rstan: The r Interface to Stan*. R package version. 2016;2(1).
33. Gelman A, Carlin J, Stern H, Rubin D. *Bayesian Data Analysis*. 2nd ed. New York: Chapman and Hall; 2004.
34. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Machine Learn Res*. 2010;11:3571–94.
35. Vehtari A, Gelman A, Gabry J. Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017;27(5):1413–32.
36. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol*. 2002;64(4):583–639.
37. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput*. 2014;24(6):997–1016.
38. Black WC. A graphic representation of cost-effectiveness. *Med Decis Making*. 1990;10:212–4.

39. Van Hout BA, Al MJ, Gordon GS, Rutten FFH, Kuntz KM. Costs, effects and c/e-ratios alongside a clinical trial. *Health Econ*. 1994;3:309–19.
40. NICE. *Guide to the Methods of Technological Appraisal*. London: NICE; 2013.
41. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ*. 1999;18(3):341–64.
42. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for nice technology assessment: not an optional extra. *Heath Econ*. 2005;27:339–47.
43. Mason AJ, Gomes M, Grieve R, Carpenter JR. A Bayesian framework for health economic evaluation in studies with missing data. *Health Econ*. 2018;27(11):1670–83.