

# Bayesian models for cost-effectiveness analysis in the presence of structural zero costs

Gianluca Baio\*<sup>†</sup>

Bayesian modelling for cost-effectiveness data has received much attention in both the health economics and the statistical literature, in recent years. Cost-effectiveness data are characterised by a relatively complex structure of relationships linking a suitable measure of clinical benefit (e.g. quality-adjusted life years) and the associated costs. Simplifying assumptions, such as (bivariate) normality of the underlying distributions, are usually not granted, particularly for the cost variable, which is characterised by markedly skewed distributions. In addition, individual-level data sets are often characterised by the presence of structural zeros in the cost variable. Hurdle models can be used to account for the presence of excess zeros in a distribution and have been applied in the context of cost data. We extend their application to cost-effectiveness data, defining a full Bayesian specification, which consists of a model for the individual probability of null costs, a marginal model for the costs and a conditional model for the measure of effectiveness (given the observed costs). We presented the model using a working example to describe its main features. © 2013 The Authors. *Statistics in Medicine* published by John Wiley & Sons, Ltd.

**Keywords:** cost-effectiveness models; Bayesian mixture models; zero costs

## 1. Introduction

Modelling for cost-effectiveness data has received much attention in both the health economics and the statistical literature in recent years [1, 2], increasingly often under a Bayesian statistical approach [3–6]. From the statistical point of view, this is an interesting problem because of the generally complex structure of relationships linking a suitable measure of clinical benefit (e.g. quality-adjusted life years (QALYs)) and the associated costs. In addition, simplifying assumptions, such as (bivariate) normality of the underlying distributions, are usually not granted, particularly for the cost variables.

In fact, costs are typically characterised by a markedly skewed distribution, which is generally due to the presence of a small proportion of individuals incurring large costs. To accommodate this feature, several models have been suggested and implemented. Among them, the most popular are probably represented by the log-Normal and Gamma distributions [7, 8], which are well suited to describe right skewed data.

However, in addition, individual-level data sets (such as those collected in clinical trials) are often characterised by the presence of *structural* zeros in the cost variable: this amounts to observing a proportion of subjects for whom the observed cost is equal to zero. This may occur, for instance, in a study where the control intervention is treatment as usual and the disease being investigated is not life threatening; thus, it is possible to observe some patients who do not experience any major event and thus may not require any treatment at all.

Under these circumstances, the use of log-Normal or Gamma models becomes impractical, because these distributions are defined for strictly positive arguments. A simple solution is to add a small constant

Department of Statistical Science, University College London, London, U.K.

\*Correspondence to: Gianluca Baio, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, U.K.

<sup>†</sup>E-mail: gianluca@stats.ucl.ac.uk

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

$\epsilon$  to the entire set of observed values for the cost variable, thus artificially re-scaling it in the open interval  $(0, \infty)$  [9]. Although it is very easy to implement, this strategy is potentially problematic, because the results are likely to be strongly affected by the actual choice of the scaling parameter  $\epsilon$ . In particular, there is no real guidance as to ‘how small’ the value  $\epsilon$  should be in order to minimise its influence on the economic results. In addition, this fails to recognise that the underlying data generating process characterising the individuals with observed zero costs is most likely different than that for those with observed positive values (e.g. the former group of people may be healthier to start with).

Alternatively, it is possible to use specific strategies to model data including structural zero costs that overcome this issue, for example, *hurdle* models [10]; extensive treatment of this topic in the health economics literature is given in [9, 11, 12], while applications include [13, 14]. In a nutshell, the idea is to build a ‘pattern’ model that predicts the probability of a given individual being associated with a null cost; this is typically performed using a logistic regression as a function of a set of relevant covariates. Then, for the individuals incurring a positive cost, a regression model is fitted to estimate the average cost, which effectively is a mixture of the two components.

With the notable exception of [15] (who applied a bivariate Normal model to estimate survival and partially measured costs), hurdle models have been mainly used to either estimate the effect of relevant covariates or to predict future costs, without explicit reference to a measure of clinical benefit. The evaluation of the costs, however, is only one side of a comprehensive cost-effectiveness analysis, which needs to simultaneously account for the expected clinical benefits as well. As mentioned earlier, because costs and benefits are typically correlated, it is necessary to produce a multivariate model that can cater for this situation.

In this paper, we aim at extending the two-part model to produce a general framework able to account for (i) structural zero costs and (ii) correlation between costs and clinical benefits. We take advantage of the flexibility of Bayesian models, which allow to specify several components that can then be linked to induce correlation among the different modules. We consider three components; the first one is a model that predicts the probability that each individual is associated with zero costs. The second module is a marginal model for the costs, which is expressed as a mixture of two components, depending on the observed value for the costs. Finally, the third module is a conditional model for the variable of effectiveness, given the observed value for the costs.

We structured this paper as follows: first, in Section 2, we set out our modelling framework. We then present the data and specific model used to analyse a case study in Section 3, discussing the specific model in Section 3.1 and the results in Section 3.2. Section 4 reviews our main conclusions.

## 2. Modelling framework

Consider a data set  $\mathcal{D}$  including information on a set of  $n$  individuals. This may arise in the case of a randomised clinical trial, or from observational data obtained from registries of clinical practice. We assume that, for each subject,  $\mathcal{D}$  contains at least two variables  $(e, c)$  measuring suitably defined clinical benefits and the associated costs. As we will show in the following, it is helpful to assume that the study also records some additional information at the individual level, for example, age, sex or potential co-morbidities. We also note that, even in the case of RCTs where these variables are not essential in the estimation of the treatment effects (by virtue of randomisation), they are usually measured and included in the final data set.

For each intervention or treatment  $t = 0, \dots, (T - 1)$  under consideration, we can define a subset  $\mathcal{D}_t$  with sample size  $n_t$ , so that  $\mathcal{D} = \bigcup_t \mathcal{D}_t$  and  $n = \sum_t n_t$ . We partition the observed data as  $\mathcal{D}_t = (\mathcal{D}_t^{\text{null}} \cup \mathcal{D}_t^{\text{pos}})$ , where  $\mathcal{D}_t^{\text{pos}}$  includes the  $n_t^{\text{pos}}$  individuals generating a positive cost. Consequently,  $n_t^{\text{null}} = n_t - n_t^{\text{pos}}$  is the number of subjects with structural zero cost. Without loss of generality, we assume in the following that only two interventions are being considered:  $t = 0$  is some standard (e.g. currently recommended or applied by the health care provider), and  $t = 1$  is a new intervention being suggested to potentially replace the standard.

### 2.1. Pattern model for $c = 0$

We estimate the probability that each individual has a null observed cost, as a function of  $J$  relevant covariates. For each subject in  $i = 1, \dots, n_t$ , we define an indicator  $d_{it}$  taking value 1 if that individual is observed to have a null cost, and 0 otherwise. We model this variable as

$$d_{it} \sim \text{Bernoulli}(\pi_{it})$$

$$\text{logit}(\pi_{it}) = \beta_{0t} + \sum_{j=1}^J \beta_{jt} Z_{ij}^t \tag{1}$$

where  $\pi_{it}$  indicates the required probability. Both for computational and practical reasons (which we describe later), it generally helps to centre the covariates, that is, for each treatment group, instead of the originally observed covariate  $X_j^t$ , we include in (1) its centred version  $Z_{ij}^t = X_{ij}^t - E[X_j^t]$ . Of course, this construction implies that  $E[Z_j^t] = 0$ .

Within the Bayesian framework, we give the coefficients  $\beta_t = (\beta_{0t}, \beta_{1t}, \dots, \beta_{Jt})$  suitable probability distributions. One simple choice is to use independent minimally informative Normal distributions (*i.e.* centred on 0 and with large variance), but of course, other choices are possible. This, however, does not usually impact to a (too) large extent on the results, especially in the presence of at least moderately large data sets.

Nevertheless, it is worth noting that, in cases where the number of subjects with structural zeros is very small, *separation* (*i.e.* the fact that a linear combination of the predictors is perfectly predictive of the outcome) is potentially an issue. A possible solution is to model the coefficients using Cauchy priors centred on 0 and with a small scale parameter, which leads to more stable estimates [16].

Under the assumptions specified earlier, the quantity

$$p_t = \frac{\exp(\beta_{0t})}{1 + \exp(\beta_{0t})}$$

represents the estimated overall probability of having a null cost for the ‘average’ individual (*i.e.* one with the values of the covariates set to 0, their mean). Subgroup analyses would be possible by selecting the combination of modalities for the covariates that define the required individual profile. Moreover, the model in (1) can be extended to include individual structured (‘random’) effects, for example, in the case of clustering over time in repeated measurement data.

## 2.2. Marginal model for the costs

In the second module, we model the observed costs by specifying a single distributional form for the two components (subjects with null or positive costs). For each treatment, we index this by two different sets of parameters  $\theta_t = (\theta_t^{\text{null}}, \theta_t^{\text{pos}})$ , which depend on the value taken by  $d_{it}$ . In particular, we define

$$c_{it} | d_{it} \sim \begin{cases} p(c_{it} | d_{it} = 1) = p(c_{it} | \theta_t^{\text{null}}), & \text{for } i \in \mathcal{D}_t^{\text{null}} \\ p(c_{it} | d_{it} = 0) = p(c_{it} | \theta_t^{\text{pos}}), & \text{for } i \in \mathcal{D}_t^{\text{pos}} \end{cases}$$

In a sense, the parameters  $\theta_t^{\text{null}}$  are redundant; because we *know* that the subjects in  $\mathcal{D}_t^{\text{null}}$  have cost identically equal to 0, strictly speaking, there is no need to model them using a probability distribution—in fact,  $p(c_{it} | \theta_t^{\text{null}})$  is a degenerate distribution on 0. However, by using this strategy, we simplify the overall model (*e.g.* the link with the effectiveness module), because we can use a single probabilistic assumption to describe the cost of *all* the subject. For example, when coding the MCMC algorithm to estimate the parameters of interest, there is no need to distinguish between  $\mathcal{D}_t^{\text{null}}$  and  $\mathcal{D}_t^{\text{pos}}$ . Thus, despite including an extra set of parameters that are not necessarily used in the overall estimation, we do so at a relatively small cost and gain in modelling flexibility.

**2.2.1. Gamma model.** At this stage, we can choose any suitable distribution for  $p(c_{it} | d_{it})$ . For example, we can model the costs for both components using a Gamma distribution

$$c_{it} | d_{it} \sim \text{Gamma}(\eta_{t,d_{it}}, \lambda_{t,d_{it}})$$

where the nested index  $d_{it}$  takes values 0,1 for patients with positive and null costs, respectively. Thus,  $\theta_t^{\text{pos}} = (\eta_{t0}, \lambda_{t0})$  and  $\theta_t^{\text{null}} = (\eta_{t1}, \lambda_{t1})$ , where, for  $s, t = 0, 1$ ,  $\eta_{ts}$  is the *shape*,  $\lambda_{ts}$  is the *rate* of the Gamma density.

To complete our model, we need to define suitable priors on the *original-scale* parameters  $\theta_t$ . However, for many distributions, this is not a straightforward task, because they are defined on scales on which it is usually difficult to formalise some prior knowledge: for example, in the current case, we should be able to quantify our uncertainty about the rate and scale of the costs for the two mixture components.

Conversely, it is much easier to think in terms of some *natural-scale* parameters  $\omega_t = (\psi_{ts}, \zeta_{ts})$ , representing the mean and standard deviation of the costs on the natural scale. Because, in general, there is a unique deterministic relationship  $\omega_t = h(\theta_t)$  linking the natural-scale to the original-scale parameters, which define the mathematical form of the distribution, defining a prior on  $\omega_t$  will automatically imply one for  $\theta_t$ .

For example, by the mathematical properties of the Gamma density, the elements of  $\omega_t$  on which we set the priors are defined as

$$\psi_{ts} = \frac{\eta_{ts}}{\lambda_{ts}} \quad \text{and} \quad \zeta_{ts} = \sqrt{\frac{\eta_{ts}}{\lambda_{ts}^2}} \tag{2}$$

In particular, we need to model the costs of the patients in  $\mathcal{D}_t^{\text{null}}$  using a Gamma distribution that is identically 0. But if we choose  $\psi_{t1} = w$  and  $\lambda_{t1} = W$ , with  $w, W \rightarrow 0$  (e.g.  $w = W = 0.000001$ ), then we are effectively ensuring that for the patients with observed null value, the cost is estimated to be identically 0—and as a matter of fact, this prior is so informative that virtually no amount of evidence can modify it in the posterior.

As for the patients with positive costs, we need to assume a non-degenerate prior on  $\omega_t^{\text{pos}} = (\psi_{t0}, \zeta_{t0})$  to obtain a reasonable model. Just as an example, one may assume  $\psi_{t0} \sim \text{Uniform}(0, H_\psi)$  and  $\zeta_{t0} \sim \text{Uniform}(0, H_\zeta)$  for suitably selected values  $H_\psi, H_\zeta$ , which in general may depend on the treatment  $t$ . Of course, genuine information should be used to form this prior, and other formulations may be more appropriate. But in any case, inverting the deterministic relationships in (2), it is easy to derive

$$\eta_{ts} = \psi_{ts} \lambda_{ts} \quad \text{and} \quad \lambda_{ts} = \frac{\psi_{ts}}{\zeta_{ts}^2}$$

and thus, the distributions selected for  $\omega_t$  automatically induce the priors for  $\theta_t$ .

Notice that these will in general not be vague at all, even in case the priors for  $(\psi_{t0}, \zeta_{t0})$  are chosen to be minimally informative, as in the example earlier. In fact, by assuming a flat prior on the natural-scale parameters, we are implying some information on the original-scale parameters of the assumed Gamma distribution. This is not a problem: in this way, we are including substantive information on the *relevant* parameters (i.e. the mean and standard deviation on the natural scale). The resulting posterior distributions will be of course affected by the assumptions we make in the priors; but, by definition, however informative the implied priors for  $(\eta_{t0}, \lambda_{t0})$  turn out to be, this will by necessity be consistent with the substantive knowledge (or lack thereof) that we are assuming for the natural-scale parameters.

**2.2.2. Log-Normal model.** We can use the same rationale to encode different distributional assumptions. For example, we could use a log-Normal model to describe sampling variations in the observed costs. In this case, we have that

$$c_{it} \mid d_{it} \sim \text{log-Normal}(\eta_{t,d_{it}}, \lambda_{t,d_{it}})$$

where  $\eta_{ts}$  and  $\lambda_{ts}$  are now the population average and standard deviation of the cost on the log scale. By the basic properties of the log-Normal distribution, for each subgroup  $s = 0, 1$  the mean and the standard deviation of the cost on the natural scale can be computed as functions of the original-scale parameters as

$$\psi_{ts} = \exp\left(\eta_{ts} + \frac{\lambda_{ts}^2}{2}\right) \quad \text{and} \quad \zeta_{ts} = \sqrt{(\exp(\lambda_{ts}^2) - 1) \exp(2\eta_{ts} + \lambda_{ts}^2)} \tag{3}$$

Again, setting  $\psi_{t1} = w$  and  $\zeta_{t1} = W$ , for  $w = W = 0.000001$ , implies that for the individuals in  $\mathcal{D}_t^{\text{null}}$ , the cost on the natural scale is effectively 0 with no substantial variability. In a similar fashion to what has shown earlier, we can define the non-degenerate prior for the individuals with positive costs for instance by selecting again uniform priors for  $(\psi_{t0}, \zeta_{t0})$ . We can then invert the relationships in (3) to obtain

$$\eta_{ts} = \log(\psi_{ts}) - \frac{1}{2} \log\left[1 + \left(\frac{\zeta_{ts}}{\psi_{ts}}\right)^2\right] \quad \text{and} \quad \lambda_{ts} = \sqrt{\log\left[1 + \left(\frac{\zeta_{ts}}{\psi_{ts}}\right)^2\right]}$$

and thus induce the priors for  $\theta_t$ . Again, this allows us to formulate any available knowledge or assess the impact of using vague specifications on a scale that is easier to manipulate.

2.2.3. *Computation of the average cost.* Of course, none of the distributional assumptions discussed earlier is essential, and it is possible to express the available prior information in different ways and using other parametric models. Nevertheless, the general framework still applies, and one can use a single distribution to represent the observed costs in both the components of the population, simply by cleverly modelling the parameters.

In addition, regardless on the underlying marginal model, once the two components of  $\psi_t = (\psi_{t0}, \psi_{t1})$  have been estimated, it is then possible to derive the overall average cost in the population by computing the weighted average

$$\begin{aligned} \mu_{ct} &= (1 - p_t)\psi_{t0} + p_t\psi_{t1} \\ &= (1 - p_t)\psi_{t0} \end{aligned} \tag{4}$$

because  $\psi_{t1}$  will be invariably estimated as 0. The weights of the mixture are given by the estimated probability associated with each of the two classes, derived from the pattern model. In effect, the population average cost is obtained by down-weighting the estimated average for  $\mathcal{D}_t^{\text{pos}}$ , to account for the presence of the structural zeros.

### 2.3. Conditional model for the measure of clinical benefit

The final module consists in modelling the measure of clinical benefit  $e$  so that correlation between the two dimensions of the health economic evaluation is accounted for. One possible way of doing so is to factorise the joint distribution  $p(e, c)$  in the product of a marginal and a conditional distribution. Intuitively, it is easier to think of this factorisation in terms of  $p(e)p(c | e)$ , that is assuming that the observed costs somehow depend on the value taken by the measure of effectiveness. This construction has been used, for example, in [3].

However, because we are merely modelling a *probabilistic* structure (i.e. we are not claiming any *causal* relationship), it is equally reasonable to factorise the joint distribution in terms of a marginal density for the costs and a conditional density for the benefits given the costs, that is,  $p(e, c) = p(c)p(e | c)$ —in this sense, we refer to the model of Section 2.2 as *marginal* and to the one in the current section as *conditional*.

The distribution  $p(e|c)$  is chosen according to the nature of the effectiveness variable. For example, if  $e$  were expressed in terms of QALYs over a long period, it should be a continuous density defined in  $\mathbb{R}^+$ . But, as discussed in [7], whatever this choice, one can always describe its mean  $\phi_{it}$  (which represents the conditional average effectiveness, given the costs) through a regression model

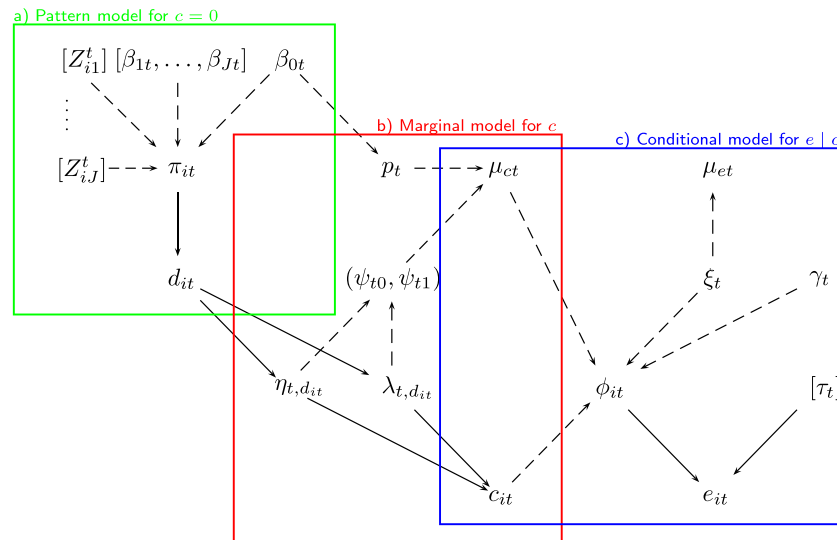
$$g(\phi_{it}) = \xi_t + \gamma_t(c_{it} - \mu_{ct}) \tag{5}$$

defined in terms of a suitable link function  $g(\cdot)$ . The form of the link function obviously depends on the scale in which  $\phi_{it}$  is defined; for example, if  $\phi_{it}$  were modelled on the natural scale of  $e$ , then  $g(\cdot)$  would be the identity function.

In (5), the coefficient  $\mu_{ct}$  is the population average cost obtained in the mixture model of (4), whereas the coefficients  $\xi_t$  and  $\gamma_t$  represent respectively the population (marginal) average effectiveness and the level of correlation between effectiveness and (the centred version of the) costs. Notice that these are quantified on the scale defined by the link function. Thus, in order to estimate the marginal average effectiveness on the natural scale, it is necessary to compute the inverse transformation  $\mu_{et} = g^{-1}(\xi_t)$ . To complete the full Bayesian model, the parameters  $(\xi_t, \gamma_t)$  as well as any other nuisance parameter characterising  $p(e | c)$  are given appropriate prior distributions.

Figure 1 shows a graphical representation of the general model structure, highlighting the links among the three modules. Dashed connections indicate logical relationships among nodes (variables), whereas solid connections represent probabilistic relationships or dependence. For instance, the individual probability of zero costs  $\pi_{it}$  is deterministically related to the coefficients  $\beta_{0t}, \dots, \beta_{Jt}$ , and thus, the arrows that connect them are typeset as dashed. Conversely, the zero cost indicator  $d_{it}$  depends probabilistically on the parameter  $\pi_{it}$ , which explains why the arrow connecting them is typeset as solid. Variables enclosed in square brackets are not strictly necessary, and we could build a model without including them—for example, we may not have access to the covariates  $Z_{1t}^t, \dots, Z_{Jt}^t$ .

The three modules are connected by means of these relationships: for example, the nodes  $d_{it}$  and  $p_t$  feed into module (b) from module (a), inducing correlation between them and propagating uncertainty throughout the model. The complete structure of Figure 1 encodes the assumption that the joint



**Figure 1.** A graphical representation of the full Bayesian model accounting for (a) the pattern model for  $c = 0$ , (b) the marginal model for the costs and (c) the conditional model for the clinical benefit (given the observed costs). Dashed connections indicate logical relationships, whereas solid connections indicate probabilistic relationships. Nodes enclosed in brackets may not be used: for example, the covariates  $Z_1^t, \dots, Z_J^t$  may not be observed, and hence, the coefficients  $\beta_{1t}, \dots, \beta_{Jt}$  are not included in the pattern model; similarly, the parameter  $\tau_t$  may not be needed in the conditional model for  $e$  (e.g. if a Bernoulli distribution is considered, only  $\phi_{it}$  is necessary).

model for the zero cost indicator  $d$ , the observed costs  $c$  and the observed benefits  $e$  is of the form  $p(d, e, c) = p(d)p(c | d)p(e | c)$ .

#### 2.4. Economic evaluation

Once the model is fitted to the observed data  $\mathcal{D}$ , it is possible to directly use the posterior distributions for  $(\mu_{et}, \mu_{ct})$  to perform the health economic evaluation. For example, we can construct suitable health economic summaries, such as the *increment in mean effectiveness*  $\Delta_e := \mu_{e1} - \mu_{e0}$  and the *increment in mean cost*  $\Delta_c := \mu_{c1} - \mu_{c0}$ .

After having obtained the required posteriors, for instance using an MCMC procedure, one can post-process the output (e.g. using the R package BCEA [6, 17]) and perform the economic analysis. This includes constructing the cost-effectiveness plane, which describes the posterior joint distribution of  $(\Delta_e, \Delta_c)$ ; the incremental cost-effectiveness ratio ICER =  $\frac{E[\Delta_c]}{E[\Delta_e]}$ , which estimates the cost per added unit of effectiveness; and the expected incremental benefit EIB =  $kE[\Delta_e] - E[\Delta_c]$ , which is used to perform the decision analysis upon deterministically varying the willingness-to-pay threshold  $k^\ddagger$ .

Moreover, it is helpful to conduct a probabilistic sensitivity analysis, for example, in terms of the cost-effectiveness acceptability curve CEAC =  $\Pr(k\Delta_e - \Delta_c > 0)$  and the analysis of the expected value of information [5, 6], in order to assess the impact of parameters uncertainty on the decision process. Finally, as different distributional assumptions for the costs may also have an impact on the decision-making, it is generally advisable to perform a structural sensitivity analysis [18]. Our framework allows these to be performed in a straightforward way.

#### 2.5. Sensitivity to the parameters specification for $(\psi_{t1}, \zeta_{t1})$

The choice of the values  $w$  and  $W$  in the models for  $(\psi_{t1}, \zeta_{t1})$  described earlier is potentially a delicate issue, as the results may be sensitive to their specification. In fact, the estimation for the main parameters  $(\mu_{et}, \mu_{ct})$  is not really affected by this choice, provided that the encoded relationships between

<sup>‡</sup>The willingness-to-pay  $k$  is used to put the cost and effectiveness differentials on the same scale, and it represents the cost that the decision maker is willing to pay to increment the effectiveness measure by one unit. If  $EIB > 0$ , then, for a given  $k$ ,  $t = 1$  is more cost-effective than  $t = 0$ . More details are presented in [1–6].

them really induce  $\psi_{t1}, \zeta_{t1} \rightarrow 0$ . In particular, the overall cost average will be unbiased because we are directly implying a constraint on the mean for the subjects in  $\mathcal{D}_t^{\text{null}}$ . The constraint on the standard deviation ensures that also  $p(c | \theta_t^{\text{null}})$ , as well as the priors for  $\psi_{t1}$  and  $\zeta_{t1}$ , is a degenerate distribution, which contributes to model fitting.

On the other hand, it is worth noticing that, as is reasonable, in general, different values for  $w$  and  $W$  do have an impact on measures of model fit, such as the deviance inflation criterion (DIC) [19]. This is essentially because the population is really made by two groups, one of which shows costs that are identically null. Thus, the faster the rate of convergence to 0 for the degenerate distributions  $p(c | \theta_t^{\text{null}})$ , the better the fit to the observed data and therefore the smaller the resulting DIC. But the implications in terms of the resulting estimated values (and hence the resulting economic model) are immaterial.

## 2.6. Links with the missing data literature

There are some parallels between our model specification and the problem of missing data. In particular, we can think of our structure as some special case of a ‘pattern-mixture model’ [20–22]. To extend this further, the case in which the probability of structural zero  $\pi_{it}$  is estimated using the intercept only (i.e. when no covariates are included) effectively assumes a mechanism of *zero completely at random* (ZCAR), in which the chance of observing an individual associated with zero costs is completely independent on any other variable. This is similar to the missing completely at random assumption in the missing data literature.

Of course, this is rarely justifiable; in all but trivial cases, observations associated with a zero cost will have some feature that sets them apart from the other subjects. For instance, in a primary care setting, we can think of those individual as being healthier at baseline and thus less likely to consume any health care resource. Consequently, it is generally necessary to assume that the structural zero mechanism is rather *zero at random* (ZAR); this means that the individual chance of having zero cost depends only on a specified set of completely observed variables.

Of course, the possibility that other, unobserved covariates influence the mechanism by which the zero costs arise cannot be ruled out—we can term this *zero not at random* (ZNAR). Under ZNAR, observed data alone are not sufficient to estimate the relevant quantities without bias. Content-specific knowledge as well as methods from the probabilistic causal inference and econometric literature (e.g. instrumental variable estimation or imputation) can be brought to bear to try and balance the subgroups (i.e. the subjects with positive and those with zero costs) with respect to the observed covariates. Moreover, the use of informative prior distributions and thorough sensitivity analysis become essential in this case [23]. These can be formally included in the general framework of our model.

In any case, we acknowledge that the problem of zero costs is slightly simpler than the missing data case; in fact, in the presence of structural zeros, we *know* what the distribution of the outcome variable ( $c$ ) is for the individuals in the two subgroups (positive or null costs). Conversely, in the missing data problem, we observe that some subjects have no recorded data for the outcome, but we do not know what that value should be. Thus, it is harder to model missing data, than it is to deal with structural zeros.

## 3. Example: the TOPICAL trial

The TOPICAL study is a double-blind, randomised, placebo-controlled, phase III trial, conducted at 78 centres in the UK in patients with non-small-cell lung cancer [24]. Subjects were randomly assigned to receive oral placebo (which we indicate with  $t = 0$ ) or erlotinib (150 mg per day,  $t = 1$ ) until disease progression or unacceptable toxicity. The original trial investigated 350 patients in the active treatment and 320 in the placebo group.

Total QALYs gained (which we use as the measure of clinical effectiveness), costs and some individual baseline characteristics are measured and available to us, for a subsample of 228 patients (120 in the placebo and 108 in the erlotinib group, respectively). For each treatment, the covariates are  $X_1^t =$  age,  $X_2^t =$  sex (coded as female = 0 and male = 1),  $X_3^t =$  the baseline stage of the disease (coded as Stage IIIb = 0 and Stage IV = 1, indicating progressively worse conditions) and  $X_4^t =$  a measure of pre-progression quality of life. In the control group, 16% of the patients are associated with null costs, whereas in the active treatment, this proportion is only 4%.

We apply our model first assuming a ZCAR mechanism, that is, without considering the observed covariates in the pattern model and then relaxing this assumption to consider a ZAR mechanism, that is, assuming that the observed set of covariates  $X^t = (X_1^t, \dots, X_4^t)$  is sufficient to explain away the potential unbalance in the baseline characteristics of the subjects with positive or null costs.

### 3.1. Model specification

If we assume a ZCAR mechanism, we model the individual probability of structural zero as  $\text{logit}(\pi_{it}) = \beta_{0t}$ , whereas when assuming a ZAR mechanism, we simply extend this by including the covariates in  $\mathbf{X}^t$ , as in (1). In both cases, we implemented the ‘robust’ Cauchy(0, 2.5) specification for the prior on the regression coefficients. As mentioned earlier, this is advisable to avoid that the posterior estimation be highly unstable.

As for the costs, we use both a Gamma and a log-Normal specification, as described in Sections 2.2.1 and 2.2.2. In both models, we set  $H_\psi = 50\,000$  and  $H_\xi = 15\,000$ ; these values are set to encode vague knowledge on the financial impact of the interventions (notice that these are the *total* costs). In addition, we consider  $w = W = 0.000001$  and perform a sensitivity analysis to this choice to assess its impact on the health economic outcome.

Finally, we model the effectiveness measure using a Beta regression, which, in line with [25], we specify as follows:

$$\begin{aligned} e_{it} | c_{it} &\sim \text{Beta}(\phi_{it}\tau_t, (1 - \phi_{it})\tau_t) \\ \text{logit}(\phi_{it}) &= \xi_t + \gamma_t(c_{it} - \mu_{ct}) \\ \xi_t, \gamma_t, \log(\tau_t) &\stackrel{iid}{\sim} \text{Normal}(0, 10\,000) \end{aligned} \quad (6)$$

In (6), the parameter  $\phi_{it}$  represents the conditional subject-specific average QALYs, whereas the parameter  $\tau_t$  is the conditional precision (inverse variance), which we assume constant across the subjects within each treatment arm. The actual measure of effectiveness (i.e. the marginal population average QALYs under either treatment) can be then retrieved on the correct scale by applying the inverse logit transformation

$$\mu_{et} = g^{-1}(\xi_t) = \frac{\exp(\xi_t)}{1 + \exp(\xi_t)}$$

### 3.2. Results

We fitted the models of Section 3.1 using the R package `BCEs0` [26], which implements the general framework described in Section 2 under a set of possible distributional assumptions. In `BCEs0`, the user needs to (i) specify a data list including the observed values for  $(e, c)$  under the two treatment options, the fixed parameters  $H_\psi$  and  $H_\xi$  and possibly the matrices including the values for the covariates  $\mathbf{X}_1^t, \dots, \mathbf{X}_J^t$ ; (ii) select a distribution for the costs (implemented choices are Gamma, log-Normal and Normal); and (iii) select a distribution for the measure of effectiveness (Beta, Bernoulli, Gamma and Normal are currently implemented).

`BCEs0` will then write the `JAGS` [27] model for the selected specification to a text file, call the library `R2jags` [28] (which connects `JAGS` to R in background) and perform the MCMC analysis. The resulting simulations from the posterior distributions are saved to the R workspace and can be used for the health economic evaluation. Because the model file is saved in the working directory, it is possible to use it as a template and modify the model assumptions, if required.

We ran 10 000 iterations, using a burn-in of 5000 and retaining one iteration every 10, resulting in a sample of 1000 iterations, which we used to produce the posterior analysis. For each variable in the model, we assessed convergence of the MCMC sampler by the analysis of the potential scale reduction [29], as well as the effective sample size.

**3.2.1. Zero completely at random mechanism.** Table I presents summary statistics from the posterior distributions of the main parameters in the ZCAR mechanism model, for both specifications of the cost variable. In both models, treatment  $t = 1$  is associated with both higher costs and higher QALYs, on average. The average costs are substantially larger for this arm of the trial. As is possible to see, for both treatments, there is a significant difference between  $\mu_{ct}$ , the overall average cost and  $\psi_{t0}$ , the average cost for the subjects in  $\mathcal{D}_t^{\text{pos}}$ . In the treatment arm, the log-Normal/Beta model produces estimations of the costs that are slightly lower than those produced by the Gamma/Beta specification. Conversely, in the control arm, the estimated average costs are slightly higher in the log-Normal/Beta model than in the Gamma/Beta model. The estimation of the effectiveness measure is very similar in both specifications and for both arms.



**Table I.** Posterior summaries for selected parameters for the Gamma/Beta and log-Normal/Beta models, assuming zero completely at random.

Parameter	Gamma/Beta model				Log-Normal/Beta model			
	Mean	SD	95% interval		Mean	SD	95% interval	
$p_0$	0.17	0.04	0.11	0.24	0.17	0.03	0.11	0.24
$\psi_{00}$	4069.95	512.85	3190.65	5166.28	4312.52	461.62	3358.93	5176.79
$\mu_{c0}$	3373.55	444.88	2571.21	4315.12	3583.45	411.49	2770.08	4385.66
$\mu_{e0}$	0.21	0.02	0.18	0.25	0.22	0.02	0.18	0.25
$p_1$	0.04	0.02	0.01	0.09	0.04	0.02	0.01	0.08
$\psi_{10}$	10 356.47	1060.49	8463.40	12 653.51	9321.01	717.66	7884.13	10681.00
$\mu_{c1}$	9930.72	1032.05	8082.63	12 155.24	8939.05	707.12	7551.40	10284.65
$\mu_{e1}$	0.23	0.02	0.19	0.27	0.22	0.02	0.19	0.25

**Table II.** Posterior summaries for selected parameters for the Gamma/Beta and log-Normal/Beta models, assuming zero at random.

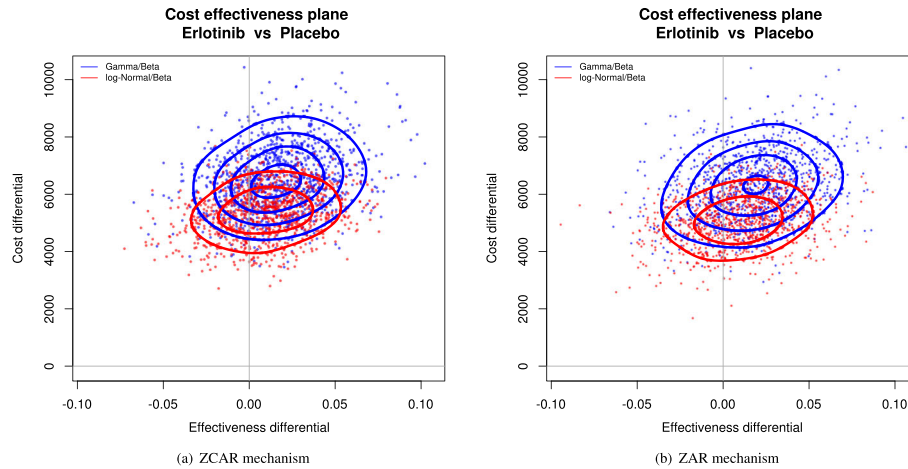
Parameter	Gamma/Beta model				Log-Normal/Beta model			
	Mean	SD	95% interval		Mean	SD	95% interval	
$\beta_{00}$ (intercept)	-2.70	0.53	-3.88	-1.78	-2.68	0.53	-3.86	-1.78
$\beta_{10}$ (age)	-0.03	0.04	-0.10	0.05	-0.03	0.04	-0.10	0.05
$\beta_{20}$ (sex)	0.63	0.57	-0.47	1.8	0.62	0.60	-0.48	1.88
$\beta_{30}$ (stage)	0.09	0.61	-1.15	1.20	0.06	0.59	-1.05	1.26
$\beta_{40}$ (QALY)	-1.61	0.50	-2.70	-0.73	-1.58	0.51	-2.72	-0.72
$p_0$	0.07	0.03	0.02	0.14	0.07	0.03	0.02	0.14
$\psi_{00}$	4104.42	556.05	3159.00	5370.27	4322.24	467.200	3342.10	5193.25
$\mu_{c0}$	3817.95	537.16	2905.75	4989.01	4014.76	467.52	3068.24	4903.59
$\mu_{e0}$	0.21	0.02	0.12	0.25	0.21	0.02	0.18	0.25
$\beta_{01}$ (intercept)	-3.86	0.66	-5.34	-2.73	-3.85	0.67	-5.31	-2.73
$\beta_{11}$ (age)	-0.09	0.09	-0.28	0.12	-0.09	0.10	-0.27	0.10
$\beta_{21}$ (sex)	-0.35	0.99	-2.23	1.63	-0.27	0.94	-2.16	1.73
$\beta_{31}$ (stage)	0.61	1.13	-1.43	3.21	0.63	1.12	-1.24	3.15
$\beta_{41}$ (QALY)	-0.12	0.31	-0.81	0.39	-0.14	0.31	-0.88	0.37
$p_1$	0.02	0.01	0.00	0.06	0.02	0.01	0.00	0.06
$\psi_{10}$	10376.91	1035.29	8550.78	12571.45	9320.26	710.00	7777.58	10659.33
$\mu_{c1}$	10119.80	1022.73	8367.69	12329.24	9086.38	701.03	7594.49	10362.48
$\mu_{e1}$	0.23	0.02	0.19	0.27	0.22	0.02	0.19	0.26

3.2.2. *Zero at random mechanism.* Table II shows a summary of the posterior distributions for the main parameters in the model, under the ZAR assumption. In this case, we also present the results for the regression coefficients that are used in the pattern model to estimate the individual and then the marginal probability of zero costs.

In the placebo group, the baseline quality of life level seems to be an important predictor for the zero mechanism, with lower values associated with a higher propensity to present zero costs; this makes intuitive sense, because those patients are presumably the most frail and thus are more likely to die during the course of the trial before consuming health resources.

In terms of the actual estimation for the population average values of cost and effectiveness, there are some differences with respect to the ZCAR model; most notably, the impact of the covariates selected in the pattern model produce different values for the marginal probability of zero costs in the two treatment groups. This obviously influences the estimation of the two population averages  $\mu_{cs}$ , obtained by mixing the average costs for the patients in  $\mathcal{D}_t^{\text{null}}$  and those in  $\mathcal{D}_t^{\text{pos}}$  (in fact, notice that the estimations for  $\psi_{00}$  and  $\psi_{10}$  do not differ by much in Tables I and II).

3.2.3. *Health economic evaluation.* Despite these important differences, the substance of the economic evaluation, in this particular case, is not modified dramatically. Figure 2 shows the cost-effectiveness



**Figure 2.** Cost-effectiveness analysis of erlotinib versus placebo, via the cost-effectiveness plane. Panel (a) shows the distribution of  $(\Delta_e, \Delta_c)$  under the assumption of zero completely at random and for both the Gamma/Beta (in blue) and the log-Normal/Beta (in red) models. Panel (b) shows the same quantities under the assumption of zero at random. In either case, the contour plots are extremely similar, indicating very limited sensitivity to the zero mechanisms as well as to the distributional assumptions for the cost distribution.

plane deriving from the implementation of the modelling framework under ZCAR (a) and ZAR (b) for the Gamma/Beta and the log-Normal/Beta models, in blue and red, respectively.

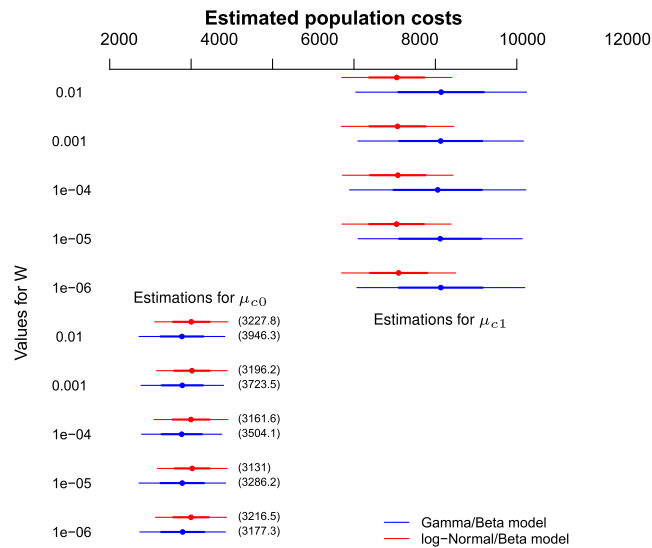
In the graphs, the posterior distributions for  $\Delta_e$  and  $\Delta_c$  are plotted on the  $x$ -axis and on the  $y$ -axis, respectively. These quantify the incremental benefits deriving from using the active treatment  $t = 1$  instead of the placebo  $t = 0$ . In both panels (a) and (b), and for both specifications of the costs, the entire distribution of  $\Delta_c$  is positive, indicating that  $t = 1$  is more expensive than the placebo—in this case, the incremental cost is quite large, as the bulk of the distribution is far from 0 along the  $y$ -axis.

The distribution of the effectiveness differential is positive, on average (as confirmed by the inspection of Tables I and II). However, as is often the case for life-threatening diseases (such as the type of cancer considered in the TOPICAL study), it is difficult to see a very large difference in terms of QALYs. This is typically because patients often die within a short time after the conclusion of the trial or even during the study, because of the seriousness of their condition. Thus, it is impossible to accrue sufficient gains in QALY so as to generate a relatively small incremental cost-effectiveness ratio, especially in case of very expensive interventions.

In this particular case, it is interesting to note that both models would effectively give the same answer in terms of the economic evaluation, deriving the same qualitative result, that is, that the new intervention is likely to produce gains in QALYs, at a cost that is relatively high. It is worth noticing that, specifically in a case such as this, the actual decision about which intervention should be implemented is based on considerations about the societal value of the interventions that often go beyond the precepts of standard health economic evaluation.

We can compare the relative fit of the two models using the DIC: that is, 3180.0 and 3177.3 in the ZCAR and ZAR, respectively, for the Gamma/Beta, and 3225.3 and 3216.5 for the log-Normal model, which under both assumptions regarding the zero mechanism indicate preference for the former specification.

**3.2.4. Sensitivity to the choice of  $(w, W)$ .** We have run the models using different values for the parameters  $(w, W)$ , to assess their impact on the cost estimation. As an example, Figure 3 shows the results of the sensitivity analysis on  $W$ , holding  $w$  fixed to its default value of 0.000001 for all cases and under the ZAR assumption; in particular, in addition to the default setting  $W = 0.000001$ , we have selected values of  $W = (0.00001, 0.0001, 0.001, 0.01)$ , which indicate decreasing precision for the (increasingly less) degenerate distribution  $p(c | \theta_t^{\text{null}})$ . In the graph, we report the posterior mean and both a 50% and 95% posterior credible intervals for the average costs (the dark and light lines, respectively). The results for  $t = 0$  are depicted on the left side, whereas those for  $t = 1$  are on the right side. The numbers in brackets are the estimated DIC for each model specification (of course, these are features of the overall model rather than of the treatment arm, so we only report them once in the graph).



**Figure 3.** Sensitivity analysis for the choice of the parameter  $W$ . The dots represent the posterior means for the estimated costs  $\mu_{c0}$  and  $\mu_{c1}$  (on the left and the right of the panel, respectively). The light dark and lines indicate the 50% and 95% credible intervals, respectively. Dots and lines in blue indicate the cost estimation from the Gamma/Beta model, whereas those in red indicate the log-Normal/Beta model. The number in brackets represent the deviance inflation criterion for each model. Within models, in all cases, the results are substantially identical and do not depend at all on the selected value of  $W$ .

As is possible to see, the point and the interval estimate of the average costs are effectively unchanged in all the cases. As mentioned earlier, under the default specification of the parameters ( $w, W$ ), the Gamma model is preferred, according to the DIC. Conversely, when we consider increasingly lower values for the standard deviation  $W$ , the log-Normal model shows better fit. The gain against the Gamma model does increase as  $W$  becomes smaller. This seems to indicate that the log-Normal model for the costs is more robust to ‘mis-specifications’ of the variability in the degenerate distribution for the costs associated with the subjects in  $\mathcal{D}_t^{\text{null}}$ . It is also worth noticing that, under our framework, it is easy to understand the meaning (and implication) of the chosen assumption for the parameters ( $w, W$ ), because they represent a clear feature of the selected models, that is, the mean and standard deviation. This makes it easier to run sensitivity analyses to this aspect. We also notice for completeness that the DIC is one possible, convenient way of measuring model fit, but other possibilities can be explored.

#### 4. Discussion

In this paper, we have defined and discussed a general framework to handle cost-effectiveness analysis using individual level data (e.g. from an RCT) in the presence of structural zeros in the cost variable. This is a challenging situation because cost-effectiveness models are characterised by a relatively complex structure, which require the formal inclusion of correlation between the outcomes. In addition, because of the asymmetry in the cost distributions, we also need to model them using suitable formulations.

The framework developed in Section 2 uses a flexible structure and allows the cost distributions to be modelled using a single specification for both the subjects with null and positive observed values. The parameters of the cost distributions are defined differently in the two components of the mixture; for the individuals with observed null costs, the specification implies that the final estimation is identically 0, by using a degenerate distribution induced by the extremely informative prior. The final estimation of the overall population mean costs is a weighted average of the two components and the correlation between costs and clinical benefits is ensured by the model structure. While using up a set of parameters that are not relevant for inference, this construction allows increased flexibility and easier implementation, for example, in terms of the coding of the MCMC algorithm.

The choice of the pattern model for  $c = 0$  is of course crucial. The assumption of ZCAR, much as its counterpart in the missing data case (missing completely at random), is hardly ever tenable; the fact that some individuals are associated with zero costs is intuitively due to some particular features (e.g.

in terms of baseline characteristics) that are not necessarily similar in those with positive costs. Thus, in order to avoid the introduction of bias, it is advisable to at least entertain the assumption of ZAR. Of course, it is basically impossible to rule out the possibility of some residual unobserved confounders (i.e. a ZNAR mechanism), but this is rather a general problem, than one specific to our framework. As such, we can apply usual methods to help limit the impact of confounding and the structure of Section 2.1 can be extended to deal with this problem.

The R package BCEs0 can be used, at least as a first approximation, to build a model consistent with the general framework. The choice of possible distributions for  $(e, c)$  is limited to what we consider to be the most likely situations. However, a translation into the JAGS language (which is effectively identical when applied to other software such as OpenBUGS) is automatically generated. Thus, the user can easily modify the ‘template’ model file to cater for their specific needs (e.g. adding a different distribution for  $e | c$ , including structured effects in the pattern model, or modifying the priors). The output of BCEs0 can be easily post-processed to produce standardise economic analysis, for example, by integrating it with packages such as BCEA.

Finally, unlike simpler but less efficient solutions to the problem of structural zeros in cost-effectiveness analysis, the model of Section 2 explicitly accounts for the fact that subjects with observed zero costs are likely to show features that set them apart from those with positive costs. Moreover, it is robust to the choice of the relevant parameters. In particular, it is quite easy to tune them to ensure that the null component of the mixture for the cost distributions is indeed identically 0. By selecting the priors on the parameters  $\omega_{t1}$  defined in Sections 2.2.1 and 2.2.2, it is easy to verify the level by which the distribution  $p(c_{it} | \theta_t^{\text{null}})$  degenerates to 0, which means that it is easy to assess the impact of the model assumptions on the economic results. As suggested in Section 3.2.4, although model fit may vary depending on the parameter configuration, the estimation of the costs is insensitive to this aspect, thus rendering our framework extremely robust.

## Appendix: the R package BCEs0

The package has a main function `bces0`, which takes the following arguments.

- `data`: a named list with arguments
  - `e0, e1`: the individual values of the measure of effectiveness under either treatment;
  - `c0, c1`: the individual values of the costs under either treatment;
  - `H.psi`: a vector including the upper limit for the default uniform prior for the mean of the cost non-degenerate distribution. The first value is used for  $t = 0$ , whereas the second is used for  $t = 1$ ;
  - `H.zeta`: a vector including the upper limit for the default uniform prior for the standard deviation of the cost non-degenerate distribution. The first value is used for  $t = 0$ , whereas the second is used for  $t = 1$ ;
  - `X0`: an optional matrix including some individually measured covariates under treatment  $t = 0$  to be used in the pattern model—if available;
  - `X1`: an optional matrix including some individually measured covariates under treatment  $t = 1$  to be used in the pattern model—if available;
- `dist.c`: a string specifying the assumed distribution for the cost variable. Possible choices are ‘gamma’, ‘logn’ and ‘norm’, implementing the Gamma, log-Normal and Normal models, respectively. The Normal model is not recommended usually but could be useful if the original cost variable has been pre-processed and transformed in some suitable scale to induce at least approximate normality;
- `dist.e`: a string specifying the assumed distribution for the effectiveness variable. Possible choices are ‘beta’, ‘gamma’, ‘bern’ and ‘norm’. These implement the Beta, Gamma, Bernoulli and Normal distributions. The Beta model can be used for effectiveness measures defined in  $[0; 1]$ ; for example, QALYs measured in a 1-year horizon, whereas the Gamma model can describe effectiveness measures defined as positive quantities, for example, QALYs over a long period of time. The Bernoulli distribution can model effectiveness measures defined as binary variables, for example, dead/alive, whereas the Normal distribution is again not recommended, usually, but could be useful

if the original effectiveness variable has been pre-processed and transformed in some suitable scale to induce at least approximate normality);

- `w, W`: the values for the parameters  $w$  and  $W$ , which are used to induce a degenerate distribution for the component with null costs. The default choice is  $w = W = 0.000001$ . These have no real impact on the model convergence and economic results (provided that they induce a suitable degenerate prior with mean and SD close enough to 0) but may have an impact on measures of model fit (e.g. DIC), which may be used for model averaging and structural probabilistic sensitivity analysis;
- `n.iter`: the number of MCMC iterations to be run (default value = 10 000);
- `n.burnin`: the number of MCMC iterations to be discarded in the burn-in period (default value = 5000);
- `n.chains`: the number of Markov chains to be used in the process;
- `robust`: a logical value (default TRUE) to indicate whether a robust (e.g. Cauchy) specification should be used for the regression coefficients in the pattern model. If FALSE, then a minimally informative Normal distribution is applied;
- `model.file`: a string with the name of the `.txt` to which the JAGS code representing the assumptions specified by the user is written. The default choice is `model.txt` in the current working directory.

If no covariates are specified in data, then only the intercepts  $\beta_{0t}$  will be used to estimate the probabilities of zero cost,  $p_t$ . If the covariates are given in the data list, `bces0` will check if they are centred and if not will compute and use  $Z_{ij}^t = X_{ij}^t - E[X_j^t]$  in the pattern model. Then it will use the R library `R2jags` (which is automatically loaded with `BCEs0`) to run the MCMC model in background using JAGS (which of course needs to be installed—Chapter 4 of [6] describes in details how to make a Bayesian analysis using R and JAGS).

We then stored the results of the Bayesian model in an R object, which is made available to the current workspace and can be then used to perform a full economic analysis, for example, using `BCEA`. We then can edit the model file to specify different models/assumptions (e.g. including individual structured effects or different prior distributions).

## Acknowledgements

We wish to thank the UCL CRUK Cancer Trials Centre for providing a subset of data from the TOPICAL trial, and the editor and two anonymous reviewers for providing insightful and thought-provoking comments on an earlier version of the paper.

## References

1. Briggs A, Sculpher M, Claxton K. *Decision Modelling For Health Economic Evaluation*. Oxford University Press: Oxford, UK, 2006.
2. Willan A, Briggs A. *The Statistical Analysis of Cost-Effectiveness Data*. John Wiley and Sons: Chichester, UK, 2006.
3. O'Hagan A, Stevens J. A framework for cost-effectiveness analysis from clinical trial data. *Health Economics* 2001; **10**:303–315.
4. O'Hagan A, Stevens J, Montmartin J. Bayesian cost effectiveness analysis from clinical trial data. *Statistics in Medicine* 2001; **20**:733–753.
5. Spiegelhalter D, Abrams K, Myles J. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley and Sons: Chichester, UK, 2004.
6. Baio G. *Bayesian Methods in Health Economics*. Chapman Hall/CRC Press: Boca Raton, FL, US, 2012.
7. Thompson S, Nixon R. How sensitive are cost-effectiveness analysis to choice of parametric distributions? *Medical Decision Making* 2005; **14**:421–428.
8. Nixon R, Thompson S. Incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Economics* 2005; **14**:1217–1229.
9. Cooper N, Lambert P, Abrams K, Sutton A. Predicting costs over time using Bayesian Markov chain Monte Carlo methods: an application to early inflammatory polyarthritis. *Health Economics* 2007; **16**:37–56.
10. Ntzoufras I. *Bayesian Modelling Using WinBUGS*. John Wiley and Sons: New York, NY, 2009.
11. Cooper N, Sutton A, Mugford M, Abrams K. Use of Bayesian Markov chain Monte Carlo methods to model cost-of-illness data. *Medical Decision Making* 2003; **1**:38–53.
12. Mihaylova B, Briggs A, O'Hagan A, Thompson S. Review of statistical methods for analysing healthcare resources and costs. *Health Economics* 2011; **20**:897–916.
13. Tooze J, Grunwald G, Jones K. Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research* 2002; **11**:341–355.

14. Härkänen T, Maljanen T, Lindfors O, Virtala E, Knekt P. Confounding and missing data in cost-effectiveness analysis: comparing different methods. *Health Economics Review* 2013; **28**:3–8.
15. Lambert P, Billingham L, Cooper N, Sutton A, Abrams K. Estimating the cost-effectiveness of an intervention in a clinical trial when partial cost information is available: a Bayesian approach. *Health Economics* 2008; **17**:67–81.
16. Gelman A, Jakulin A, Pittau M, Su Y. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2008; **2**(4):1360–1383.
17. Baio G. BCEA: a package to run Bayesian cost-effectiveness analysis in R, 2013. (Available from: [www.statistica.it/gianluca/BCEA](http://www.statistica.it/gianluca/BCEA)) [Accessed on 8 December 2013].
18. Jackson C, Boijke L, Thompson S, Claxton K, Sharples L. A framework for addressing structural uncertainty in decision models. *Medical Decision Making* 2011; **31**:662–674.
19. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B* 2002; **64**(4):583–639.
20. Little R. A class of pattern-mixture models for multivariate incomplete data. *Biometrika* 1994; **81**:471–483.
21. Daniels M, Hogan J. *Missing Data in Longitudinal Studies*. Chapman Hall/CRC Press: Boca Raton, FL, US, 2008.
22. Carpenter J, Kenward M. *Multiple Imputation and its Applications*. John Wiley and Sons: New York, NY, 2013.
23. Mason A, Richardson S, Plewis I, Best N. Strategy for modelling non-random missing data mechanisms in observational studies using Bayesian methods. *Journal of Official Statistics* 2012; **28**:279–302.
24. Lee S, Khan I, Upadhyay S, Lewanski C, Falk S, Skailes S, Marshall E, Woll P, Hatton M, Lal R, Jones R, Toy E, Chao D, Middleton G, Bulley S, Ngai Y, Rudd R, Hackshaw A, Boshoff C. First-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy (TOPICAL): a double-blind, placebo-controlled, phase 3 trial. *The Lancet Oncology* 2012; **13**(11):1161–1170.
25. Figueroa J, Arellano-Valle R, Ferrari S. Mixed beta regression: a Bayesian perspective. *Computational Statistics & Data Analysis* 2013; **61**:137–147.
26. Baio G. BCEs0: A R package for Bayesian models for cost-effectiveness analysis in the presence of structural zero costs, 2013. (Available from: [www.statistica.it/gianluca/BCEs0](http://www.statistica.it/gianluca/BCEs0)) [Accessed on 8 December 2013].
27. Plummer M. JAGS: Just Another Gibbs Sampler, 2010. (Available from: [www.mcmc-jags.sourceforge.net/](http://www.mcmc-jags.sourceforge.net/)) [Accessed on 8 December 2013].
28. Su Y, Yajima M. R2 JAGS user manual: a package to call JAGS from R, 2010. (Available from: [www.cran.r-project.org/web/packages/R2jags/](http://www.cran.r-project.org/web/packages/R2jags/)) [Accessed on 8 December 2013].
29. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. *Statistical Sciences* 1992; **7**:457–511.