A Full Bayesian Model to Handle Structural Ones and Missingness in Economic Evaluations from Individual-Level Data

Health Economics Symposium, UCL, February 2018

Andrea Gabrio[1], Alexina J. Mason[2], and Gianluca Baio[1]

[1] *Department of Statistical Science, University College London*

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

1 Introduction

In routine analyses, trial-based CEAs typically assume normality and independence for the underlying cost and effectiveness data $(e; c)$. There are several potential issues with this setting:

- Clinical outcome (e.g. QALYs) and costs tend to be correlated, as more effective treatments have typically higher unit costs, but may also reduce overall implementation cost;
- Joint/marginal normality is not realistic as costs are usually skewed and benefits may be bounded in $[0; 1]$ and, possibly, bimodal;
- Data may exhibit spikes at one or both of the boundaries of the range for the underlying distribution. For example, some patients in a trial may not accrue any cost at all (i.e. $c_i = 0$) or may be associated with perfect health, i.e. unit QALY;
- Individual level outcome data from RCTs are almost invariably affected by a substantial proportion of missingness ($> 25\%$) and simple methods discarding the missing values (e.g. CCA) may lead to biased results.

Objective: Provide a general Bayesian framework that can jointly handle a number of features affecting effectiveness and cost data while assessing the impact of missingness uncertainty on decision-making.

2 Example – The MenSS trial

The MenSS trial is a pilot RCT conducted in the UK to evaluate the cost-effectiveness of a new digital intervention to reduce the incidence of Sexually Transmitted Infections (STIs) in young men. A total of 159 men at risk of STIs were randomised to receive either usual clinical care only ($n_1 = 75$), or a combination of usual care and the new intervention ($n_2 = 84$). Utility (u_{ij}) and resource use (c_{ij}) data are computed from participant responses to questionnaires collected at baseline ($j = 0$ – only for the utilities) and 3, 6 and 12 months ($j = 1, 2, J = 3$). QALYs and Total costs derived from these measures show:

- A high degree of skewness in the distributions of QALYs and costs in both treatment groups;
- High spikes at 1 in the QALYs for both groups due to many individuals incurring a perfect health status;
- High levels of missingness which lead to 27 (control group) and 19 (intervention group) complete cases.

3 Modelling Framework

We can specify the joint distribution of the effectiveness and cost variables $p(e, c)$ as:

$$p(e, c) = p(c)p(e \mid c) = p(e)p(c \mid e),$$

where, for example, $p(e)$ is the *marginal* distribution of the effectiveness and $p(c \mid e)$ is the *conditional* distribution of costs given the effectiveness.

3.1 Marginal Model for the Effectiveness

For each individual we consider a marginal distribution $p(e_i \mid \boldsymbol{\theta}_e)$ indexed by a set of parameters $\boldsymbol{\theta}_e$ comprising a *location* ϕ_{ie} and a set of *ancillary* parameters $\boldsymbol{\psi}_e$ typically including some measure of *marginal* variance, σ^2_e. We can model the location parameter using a generalised linear structure, e.g.

$$g_e(\phi_{ie}) = \alpha_0 \; [+ \ldots],$$

where α_0 is the intercept and the notation $[+ \ldots]$ indicates that other terms may or may not be included in the model. In the absence of covariates or assuming that a centered version $x^*_i = (x_i - \bar{x})$ is used, the parameter $\mu_e = g_e^{-1}(\alpha_0)$ represents the population average effectiveness.

3.2 Conditional Model for the Costs

For the costs, we consider a conditional model $p(c_i \mid e_i, \boldsymbol{\theta}_c)$, which explicitly depends on the effectiveness, as well as on a set of quantities $\boldsymbol{\theta}_c$, again comprising a location and ancillary parameters. The location can be modelled as a function of the effectiveness variable as:

$$g_c(\phi_{ic}) = \beta_0 + \beta_1(e_i - \mu_e) \; [+ \ldots].$$

Here, $(e_i - \mu_e)$ is the centered version of the effectiveness, while β_1 quantifies the correlation between c and e. Assuming other covariates are either also centered or absent, $\mu_c = g_c^{-1}(\beta_0)$ is the population average cost.

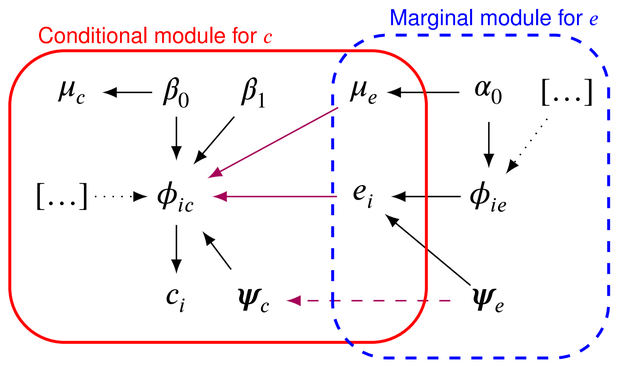


Figure 1: Graphical representation of the framework

4 Analysis Models

Three alternative specifications of the general structure depicted in Figure 1 are compared for modelling the QALYs and costs from the MenSS trial. In all cases, baseline utilities are included in the model of the QALYs as covariates. The models are:

1) Normal marginal for e and Normal conditional for c – **Bivariate Normal**

2) Beta marginal for e and Gamma conditional for c – **Beta-Gamma**

3) Beta marginal for e and Gamma conditional for c with hurdle approach for $e = 1$ – **Hurdle Model**

4.1 Hurdle Model

Hurdle models are mixture models defined by two components: the first one is a mass distribution at the spike, while the second is a parametric model applied to the natural range of the relevant variable. We can expand the model by defining an indicator variable d_{ie} taking value 1 if the $i-$th individual is associated with a unit QALYs ($e_i = 1$) and 0 otherwise ($e_i < 1$). This is then modelled as

$$\begin{aligned} d_{ie} := \mathbb{I}(e_i = 1) &\sim \text{Bernoulli}(\pi_{ie}) \\ \text{logit}(\pi_{ie}) &= \gamma_0 + \gamma_1(u_{i0} - \bar{u}_0) \; [+ \ldots], \end{aligned}$$

where π_{ie} is the individual probability of unit QALYs, which is estimated on the logit scale as a function of a baseline parameter γ_0 and the centred baseline utilities, whose effect is captured by the parameter γ_1. The quantity $\bar{\pi}_e = \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}$ represents the estimated marginal probability of unit QALYs.

4.2 Missingness Scenarios

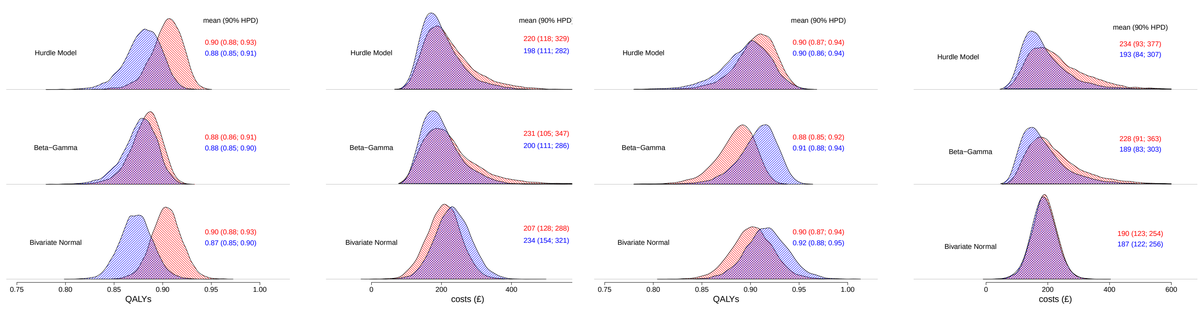
For each model we assume a complete cases scenario and then extend the structure to all cases (complete and missing), considering both MAR (for all models) or alternative MNAR scenarios (for the Hurdle Model only).

- **MAR** – No change to the model structure is required under MAR, but baseline utilities (u_{i0}) need to be explicitly modelled using suitable distributions to impute the missing values (Normal, Beta and Beta-Hurdle for the three specifications, respectively);
- **MNAR** – We observe that $n^*_1 = 13$ (control) and $n^*_2 = 22$ (intervention) individuals have $u_{i0} = 1$ and $u_{ij} = NA$, for $j > 1$. For those individuals, we cannot compute directly the value of d_{ie} and so need to make assumptions/model this → Sensitivity analysis to alternative MNAR departures from MAR. Table 1 shows the four MNAR scenarios considered:

Scenario	**Control** ($n^*_1 = 13$)	**Intervention** ($n^*_2 = 22$)
MNAR1	$d_{ie} = 1$	$d_{ie} = 1$
MNAR2	$d_{ie} = 0$	$d_{ie} = 0$
MNAR3	$d_{ie} = 1$	$d_{ie} = 0$
MNAR4	$d_{ie} = 0$	$d_{ie} = 1$

Table 1: MNAR scenarios considered in the MenSS study for the Hurdle Model.

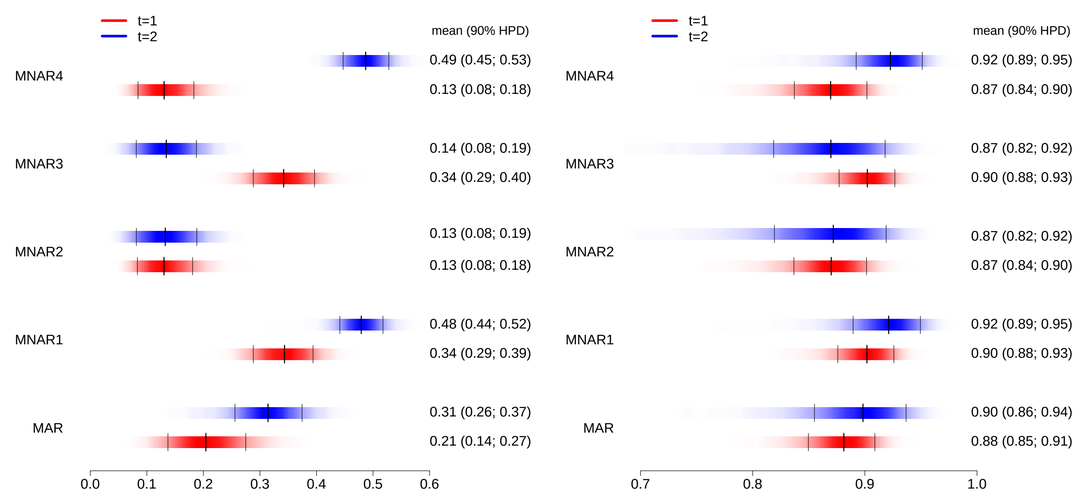
5 Results (Complete and All Cases – MAR)



(a) Control (b) Intervention

Figure 2: Posterior mean QALYs and costs distributions in both arms for the three models

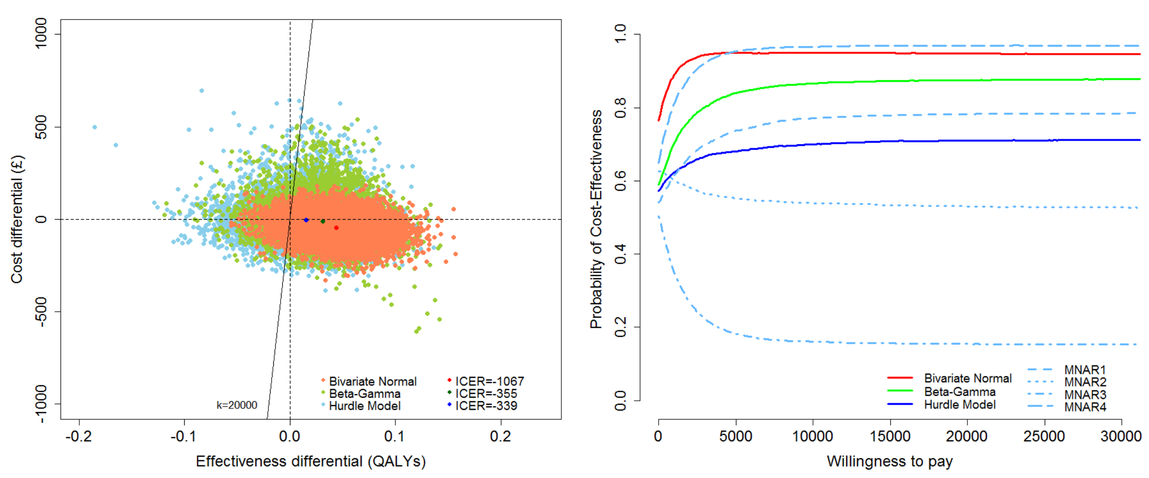
6 Sensitivity Analysis – MNAR



(a) Probability of structural one ($\bar{\pi}_e$) (b) Mean QALYs (μ_e)

Figure 3: Posterior density strips for π_e and μ_e in both arms under MAR and four MNAR scenarios

7 Economic Evaluation



(a) Cost-Effectiveness Plane) (b) Cost-Effectiveness Acceptability Curve)

Figure 4: CEP and CEAC associated with the three models under MAR and four MNAR scenarios

References

Bailey, J., Webster, R., Hunter, R., Griffin, M., N, F., Rait, G., Estcourt, C., Michie, S., Anderson, J., Stephenson, J., Gerressu, M., Sinag Ang, C., and Murray, E. (2016). The mens's safer sex project: intervention development and feasibility randomised controlled trial of an interactive digital intervention to increase condom use in men. *Health Technology Assessment*, 20.

Baio, G. (2012). *Bayesian Methods in Health Economics*. Chapman and Hall/CRC, University College London, London, UK.

Nixon, R. and Thompson, S. (2005). Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Economics*, 14:1217–1229.

Ntzoufras, I. (2009). *Bayesian Modelling Using WinBUGS*. John Wiley and Sons, New York, US.