A Bayesian Parametric Approach to Handle Nonignorable Missingness in Economic Evaluations

Andrea Gabrio

Research Team: Gianluca Baio (UCL) and Alexina Mason (LSHTM)

University College London
Department of Statistical Science

Collaborations: Michael Daniels (UF)
Rachael Hunter (UCL – PCPH)

ucakgab@.ucl.ac.uk

http://www.ucl.ac.uk/statistics/research/statistics-health-economics/

PRIMENT Statistics, Health Economics and Methodology Seminar

26 June 2018

Outline

Health Economic Evaluation

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources, increasingly often under a Bayesian framework

Health Economic Evaluation

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources, increasingly often under a Bayesian framework

Statistical model

- Estimates relevant **population** parameters $\boldsymbol{\theta}$

Health Economic Evaluation

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources, increasingly often under a Bayesian framework

$$
\begin{aligned}
\Delta_e &= f_e(\boldsymbol{\theta}) \\
\Delta_c &= f_c(\boldsymbol{\theta}) \\
&\dots
\end{aligned}
$$

$$\text{Statistical model} \longrightarrow \text{Economic model}$$

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Combines the parameters to obtain a population average measure for costs and clinical benefits

Health Economic Evaluation

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources, increasingly often under a Bayesian framework

$$\Delta_e = f_e(\boldsymbol{\theta})$$
$$\Delta_c = f_c(\boldsymbol{\theta})$$
$$\ldots$$

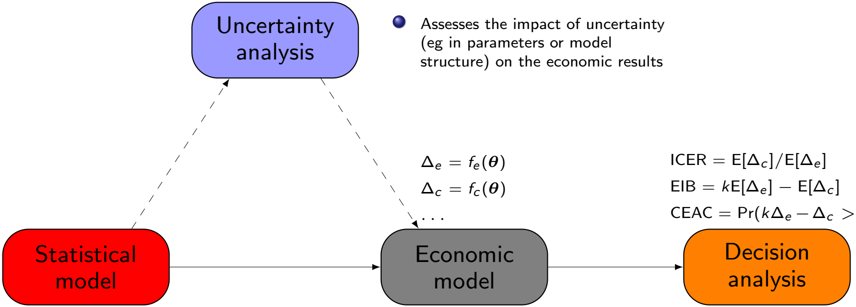
$$\text{ICER} = \text{E}[\Delta_c]/\text{E}[\Delta_e]$$
$$\text{EIB} = k\text{E}[\Delta_e] - \text{E}[\Delta_c]$$
$$\text{CEAC} = \Pr(k\Delta_e - \Delta_c > 0)$$

Statistical model → Economic model → Decision analysis

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Summarises the economic model by computing suitable measures of "cost-effectiveness"

Health Economic Evaluation

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources, increasingly often under a Bayesian framework



- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results

$$\Delta_e = f_e(\boldsymbol\theta)$$
$$\Delta_c = f_c(\boldsymbol\theta)$$
$$\ldots$$

$$\text{ICER} = \text{E}[\Delta_c]/\text{E}[\Delta_e]$$
$$\text{EIB} = k\text{E}[\Delta_e] - \text{E}[\Delta_c]$$
$$\text{CEAC} = \Pr(k\Delta_e - \Delta_c > 0)$$

- Estimates relevant **population** parameters $\boldsymbol\theta$
- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Summarises the economic model by computing suitable measures of "cost-effectiveness"

"Standard" approach — individual level data

		Demographics			HRQL data				Resource use data			
ID	**Trt**	Sex	Age	...	u_0	u_1	...	u_J	c_0	c_1	...	c_J
1	1	M	23	...	0.32	0.66	...	0.44	103	241	...	80
2	1	M	21	...	0.12	0.16	...	0.38	1 204	1 808	...	877
3	2	F	19	...	0.49	0.55	...	0.88	16	12	...	22
...	...	...	...	...	...	...	...	...	...	...	...	...

"Standard" approach — individual level data

		Demographics			HRQL data				Resource use data			
ID	**Trt**	Sex	Age	...	u_0	u_1	...	u_J	c_0	c_1	...	c_J
1	1	M	23	...	0.32	0.66	...	0.44	103	241	...	80
2	1	M	21	...	0.12	0.16	...	0.38	1 204	1 808	...	877
3	2	F	19	...	0.49	0.55	...	0.88	16	12	...	22
...	...	...	...	...	...	...	...	...	...	...	...	...

- The **typical** analysis is based on the following steps:

  1. Compute individual QALYs and total costs as

$$e_i = \sum_{j=1}^{J} (u_{ij} + u_{i\,j-1}) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=1}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$

"Standard" approach — individual level data

		Demographics			HRQL data				Resource use data			
ID	**Trt**	Sex	Age	...	u_0	u_1	...	u_J	c_0	c_1	...	c_J
1	1	M	23	...	0.32	0.66	...	0.44	103	241	...	80
2	1	M	21	...	0.12	0.16	...	0.38	1 204	1 808	...	877
3	2	F	19	...	0.49	0.55	...	0.88	16	12	...	22
...	...	...	...	...	...	...	...	...	...	...	...	...

- The **typical** analysis is based on the following steps:

2. Assume normality and linearity and model **independently** individual QALYs and total costs by controlling for baseline values

$$
\begin{aligned}
e_i &= \alpha_{e0} + \alpha_{e1} u_{0i} + \alpha_{e2}\text{Trt}_i + \varepsilon_{ie}\,[+\ldots], &\quad \varepsilon_{ie} &\sim \text{Normal}(0,\sigma_e)\\
c_i &= \alpha_{c0} + \alpha_{c1} c_{0i} + \alpha_{c2}\text{Trt}_i + \varepsilon_{ic}\,[+\ldots], &\quad \varepsilon_{ic} &\sim \text{Normal}(0,\sigma_c)
\end{aligned}
$$

3. Estimate population average cost and effectiveness differentials and use bootstrap to quantify uncertainty

What's wrong with this?

- Potential **correlation** between costs & utilities
  - Strong positive correlation — effective treatments are innovative and are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.

What's wrong with this?

- Potential **correlation** between costs & utilities
  - Strong positive correlation — effective treatments are innovative and are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
- **Asymmetric** empirical distributions
  - Both outcome variables can be highly skewed
  - Costs are defined on $[0, +\infty)$ and utilities are typically bounded in $[0; 1]$
  - Spikes at one for utilities and at zero for costs may occur

What's wrong with this?

- Potential **correlation** between costs & utilities
  - Strong positive correlation — effective treatments are innovative and are associated with higher unit costs
  - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
- **Asymmetric** empirical distributions
  - Both outcome variables can be highly skewed
  - Costs are defined on $[0, +\infty)$ and utilities are typically bounded in $[0; 1]$
  - Spikes at one for utilities and at zero for costs may occur
- ... and of course **missing data**
  - Missingness may occur in either or both utilities/costs
  - Important to explore the impact on the results of a range of plausible missingness assumptions in sensitivity analysis

A general Bayesian framework

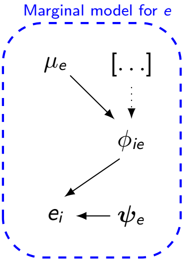
- In general, can account for **correlation** through a joint distribution

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

A general Bayesian framework

- In general, can account for **correlation** through a joint distribution

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



$$e_i \sim p(e \mid \phi_{ie}, \psi_e)$$

$$g(\phi_{ie}) = \mu_e \, [+ \ldots]$$

ϕ_{ie} = location
ψ_e = ancillary

A general Bayesian framework

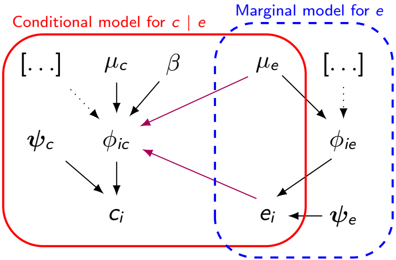
- In general, can account for **correlation** through a joint distribution

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

$c_i \sim p(c \mid e, \phi_{ic}, \psi_c)$

$g(\phi_{ic}) = \mu_c + \beta(e_i - \mu_e)\ [+\ldots]$

ϕ_{ic} = location
ψ_c = ancillary



$e_i \sim p(e \mid \phi_{ie}, \psi_e)$

$g(\phi_{ie}) = \mu_e\ [+\ldots]$

ϕ_{ie} = location
ψ_e = ancillary

A general Bayesian framework

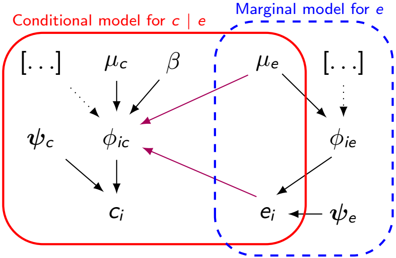
- In general, can account for **correlation** through a joint distribution

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

$c_i \sim p(c \mid e, \phi_{ic}, \boldsymbol{\psi}_c)$

$g(\phi_{ic}) = \mu_c + \beta(e_i - \mu_e)\,[+\ldots]$

ϕ_{ic} = location
$\boldsymbol{\psi}_c$ = ancillary



$e_i \sim p(e \mid \phi_{ie}, \boldsymbol{\psi}_e)$

$g(\phi_{ie}) = \mu_e\,[+\ldots]$

ϕ_{ie} = location
$\boldsymbol{\psi}_e$ = ancillary

- For example:

$$e_i \sim \text{Normal}(\phi_{ie}, \psi_e), \qquad \phi_{ie} = \mu_e\,[+\ldots]$$

$$c_i \mid e_i \sim \text{Normal}(\phi_{ic}, \psi_c), \qquad \phi_{ic} = \mu_c + \beta(e_i - \mu_e)\,[+\ldots]$$

A general Bayesian framework

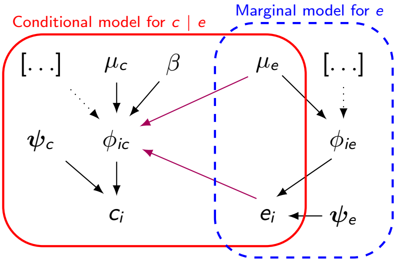
- Flexible enough to use alternative distributions to capture **skewness**

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

$c_i \sim p(c \mid e, \phi_{ic}, \psi_c)$

$g(\phi_{ic}) = \mu_c + \beta(e_i - \mu_e)\,[+\ldots]$

ϕ_{ic} = location
ψ_c = ancillary



$e_i \sim p(e \mid \phi_{ie}, \psi_e)$

$g(\phi_{ie}) = \mu_e\,[+\ldots]$

ϕ_{ie} = location
ψ_e = ancillary

A general Bayesian framework

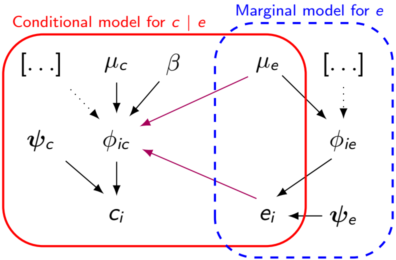
- Flexible enough to use alternative distributions to capture **skewness**

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

$c_i \sim p(c \mid e, \phi_{ic}, \boldsymbol{\psi}_c)$

$g(\phi_{ic}) = \mu_c + \beta(e_i - \mu_e)\,[+\ldots]$

ϕ_{ic} = location
$\boldsymbol{\psi}_c$ = ancillary



$e_i \sim p(e \mid \phi_{ie}, \boldsymbol{\psi}_e)$

$g(\phi_{ie}) = \mu_e\,[+\ldots]$

ϕ_{ie} = location
$\boldsymbol{\psi}_e$ = ancillary

- For example:

$$e_i \sim \text{Beta}(\phi_{ie}\psi_e, (1-\phi_{ie})\psi_e), \qquad \text{logit}(\phi_{ie}) = \mu_e\,[+\ldots]$$

$$c_i \mid e_i \sim \text{Gamma}(\psi_c\phi_{ic}, \psi_c), \qquad \log(\phi_{ic}) = \mu_c + \beta(e_i - \mu_e)\,[+\ldots]$$

A general Bayesian framework

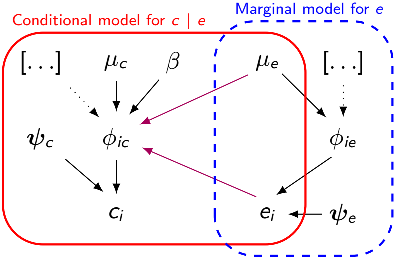
- Can incorporate external information as priors for **missing data**

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

$c_i \sim p(c \mid e, \phi_{ic}, \boldsymbol{\psi}_c)$

$g(\phi_{ic}) = \mu_c + \beta(e_i - \mu_e)\,[+\ldots]$

ϕ_{ic} = location
$\boldsymbol{\psi}_c$ = ancillary



$e_i \sim p(e \mid \phi_{ie}, \boldsymbol{\psi}_e)$

$g(\phi_{ie}) = \mu_e\,[+\ldots]$

ϕ_{ie} = location
$\boldsymbol{\psi}_e$ = ancillary

- For example:

$$e_i \sim \text{Beta}(\phi_{ie}\psi_e, (1-\phi_{ie})\psi_e), \qquad \text{logit}(\phi_{ie}) = \mu_e\,[+\ldots]$$

$$c_i \mid e_i \sim \text{Gamma}(\psi_c\phi_{ic}, \psi_c), \qquad \log(\phi_{ic}) = \mu_c + \beta(e_i - \mu_e)\,[+\ldots]$$

- Combining "modules" and fully characterising uncertainty about deterministic functions of random quantities with MCMC methods

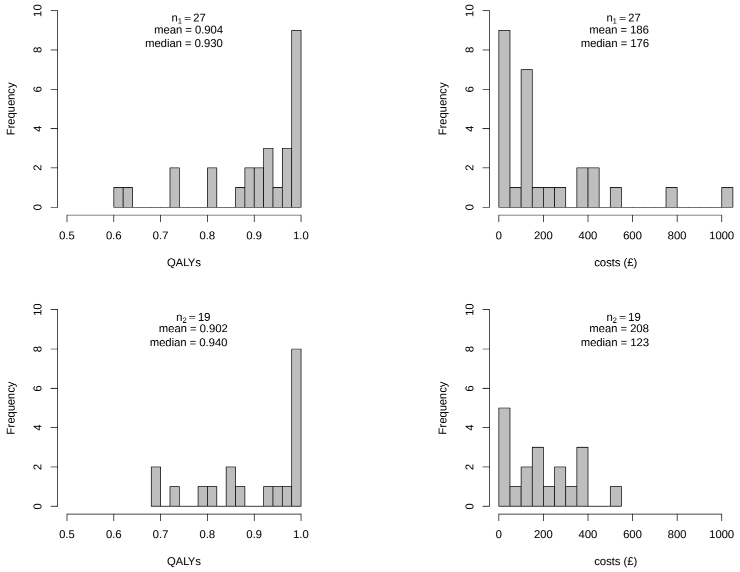
The MenSS Trial

Bailey et al., *Health Tech Ass* 2016; 20(91)

- Pilot RCT that evaluates the cost-effectiveness of a new digital intervention to reduce the incidence of STI in young men with respect to the SOC
  - QALYs calculated from utilities (EQ-5D)
  - Total costs calculated from different components (no baseline)

Time	**Type of outcome**	**observed (%)**	**observed (%)**
		control (n_1=75)	**intervention** (n_2=84)
Baseline	utilities	72 (96%)	72 (86%)
3 months	utilities and costs	34 (45%)	23 (27%)
6 months	utilities and costs	35 (47%)	23 (27%)
12 months	utilities and costs	43 (57%)	36 (43%)
Complete cases	utilities and costs	27 (44%)	19 (23%)

The MenSS Trial: Complete Cases



Modelling

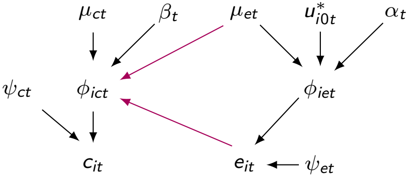
Gabrio et al. (2018). https://arxiv.org/abs/1801.09541

1. **Bivariate Normal**
   - Account for correlation between QALYs and costs

Conditional model for $c \mid e$

$c_{it} \mid e_{it} \sim \text{Normal}(\phi_{ict}, \psi_{ct})$

$\phi_{ict} = \mu_{ct} + \beta_t(e_{it} - \mu_{et})$



Marginal model for e

$e_{it} \sim \text{Normal}(\phi_{iet}, \psi_{et})$

$\phi_{iet} = \mu_{et} + \alpha_t u^*_{i0t}$

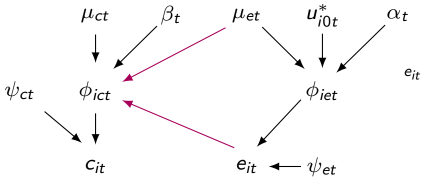
Modelling

Gabrio et al. (2018). https://arxiv.org/abs/1801.09541

1. **Bivariate Normal**
    - Account for correlation between QALYs and costs
2. **Beta-Gamma**
    - Model the relevant ranges: QALYs $\in (0,1)$ and costs $\in (0,\infty)$
    - **But**: needs to rescale observed data $e_{it} = (e_{it} - \epsilon)$ to avoid spikes at 1

Conditional model for $c \mid e$

$$c_{it} \mid e_{it} \sim \text{Gamma}(\psi_{ct}\phi_{ict}, \psi_{ct})$$
$$\log(\phi_{ict}) = \mu_{ct} + \beta_t(e_{it} - \mu_{et})$$



Marginal model for e

$$e_{it} \sim \text{Beta}\left(\phi_{iet}\psi_{et}, (1-\phi_{iet})\psi_{et}\right)$$
$$\phi_{iet} = \mu_{et} + \alpha_t u^*_{i0t}$$

Modelling

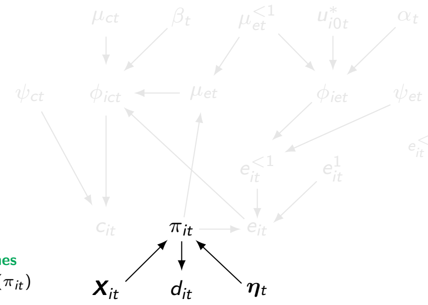
Gabrio et al. (2018). https://arxiv.org/abs/1801.09541

1. **Bivariate Normal**
   - Account for correlation between QALYs and costs
2. **Beta-Gamma**
   - Model the relevant ranges: QALYs $\in (0,1)$ and costs $\in (0,\infty)$
   - **But**: needs to rescale observed data $e_{it} = (e_{it} - \epsilon)$ to avoid spikes at 1
3. **Hurdle model**
   - Model e_{it} as a **mixture** to account for correlation between outcomes, model the relevant ranges and account for structural values

Conditional model for $c \mid e$

$c_{it} \mid e_{it} \sim \text{Gamma}(\psi_{ct}\phi_{ict}, \psi_{ct})$

$\log(\phi_{ict}) = \mu_{ct} + \beta_t(e_{it} - \mu_{et})$



Mixture model for e

$e^1_{it} := 1$

$e^{<1}_{it} \sim \text{Beta}\left(\phi_{iet}\psi_{et}, (1-\phi_{iet})\psi_{et}\right)$

$\text{logit}(\phi_{iet}) = \mu^{<1}_{et} + \alpha_t u^*_{i0t}$

$e_{it} = \pi_{it} e^1_{it} + (1-\pi_{it}) e^{<1}_{it}$

$\mu_{et} = (1-\bar{\pi}_t)\mu^{<1}_{et} + \bar{\pi}_t$

Model for the structural ones

$d_{it} := \mathbb{I}(e_{it} = 1) \sim \text{Bernoulli}(\pi_{it})$

$\text{logit}(\pi_{it}) = \boldsymbol{X}_{it}\boldsymbol{\eta}_t$

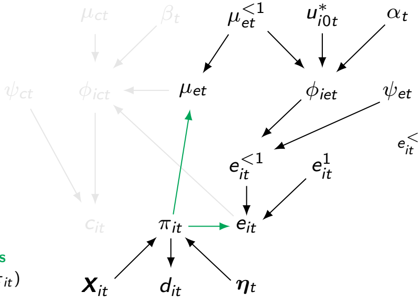
Modelling

Gabrio et al. (2018). https://arxiv.org/abs/1801.09541

1. **Bivariate Normal**
   - Account for correlation between QALYs and costs
2. **Beta-Gamma**
   - Model the relevant ranges: QALYs $\in (0,1)$ and costs $\in (0,\infty)$
   - **But**: needs to rescale observed data $e_{it} = (e_{it} - \epsilon)$ to avoid spikes at 1
3. **Hurdle model**
   - Model e_{it} as a **mixture** to account for correlation between outcomes, model the relevant ranges and account for structural values

Conditional model for $c \mid e$

$$c_{it} \mid e_{it} \sim \text{Gamma}(\psi_{ct}\phi_{ict}, \psi_{ct})$$
$$\log(\phi_{ict}) = \mu_{ct} + \beta_t(e_{it} - \mu_{et})$$



Mixture model for e

$$e_{it}^{1} := 1$$
$$e_{it}^{<1} \sim \text{Beta}\left(\phi_{iet}\psi_{et}, (1-\phi_{iet})\psi_{et}\right)$$
$$\text{logit}(\phi_{iet}) = \mu_{et}^{<1} + \alpha_t u_{i0t}^*$$

$$e_{it} = \pi_{it} e_{it}^{1} + (1-\pi_{it}) e_{it}^{<1}$$

$$\mu_{et} = (1-\bar{\pi}_t)\mu_{et}^{<1} + \bar{\pi}_t$$

Model for the structural ones

$$d_{it} := \mathbb{I}(e_{it} = 1) \sim \text{Bernoulli}(\pi_{it})$$
$$\text{logit}(\pi_{it}) = \boldsymbol{X}_{it}\boldsymbol{\eta}_t$$

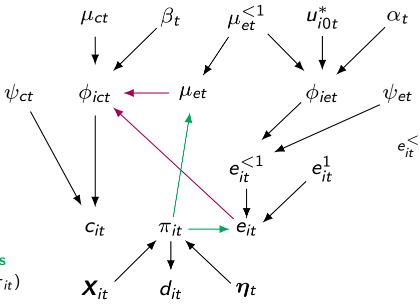
Modelling

Gabrio et al. (2018). https://arxiv.org/abs/1801.09541

1. **Bivariate Normal**
   - Account for correlation between QALYs and costs
2. **Beta-Gamma**
   - Model the relevant ranges: QALYs $\in (0,1)$ and costs $\in (0,\infty)$
   - **But**: needs to rescale observed data $e_{it} = (e_{it} - \epsilon)$ to avoid spikes at 1
3. **Hurdle model**
   - Model e_{it} as a **mixture** to account for correlation between outcomes, model the relevant ranges and account for structural values

Conditional model for $c \mid e$

$$c_{it} \mid e_{it} \sim \text{Gamma}(\psi_{ct}\phi_{ict}, \psi_{ct})$$
$$\log(\phi_{ict}) = \mu_{ct} + \beta_t(e_{it} - \mu_{et})$$



Mixture model for e

$$e^1_{it} := 1$$
$$e^{<1}_{it} \sim \text{Beta}\left(\phi_{iet}\psi_{et}, (1-\phi_{iet})\psi_{et}\right)$$
$$\text{logit}(\phi_{iet}) = \mu^{<1}_{et} + \alpha_t u^*_{i0t}$$

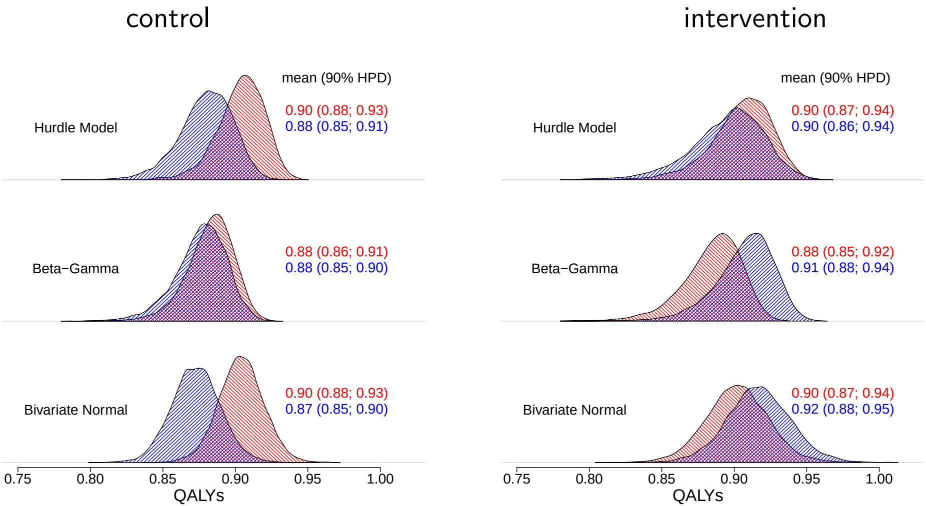
$$e_{it} = \pi_{it} e^1_{it} + (1-\pi_{it}) e^{<1}_{it}$$

$$\mu_{et} = (1-\bar{\pi}_t)\mu^{<1}_{et} + \bar{\pi}_t$$

Model for the structural ones

$$d_{it} := \mathbb{I}(e_{it} = 1) \sim \text{Bernoulli}(\pi_{it})$$
$$\text{logit}(\pi_{it}) = \boldsymbol{X}_{it}\boldsymbol{\eta}_t$$

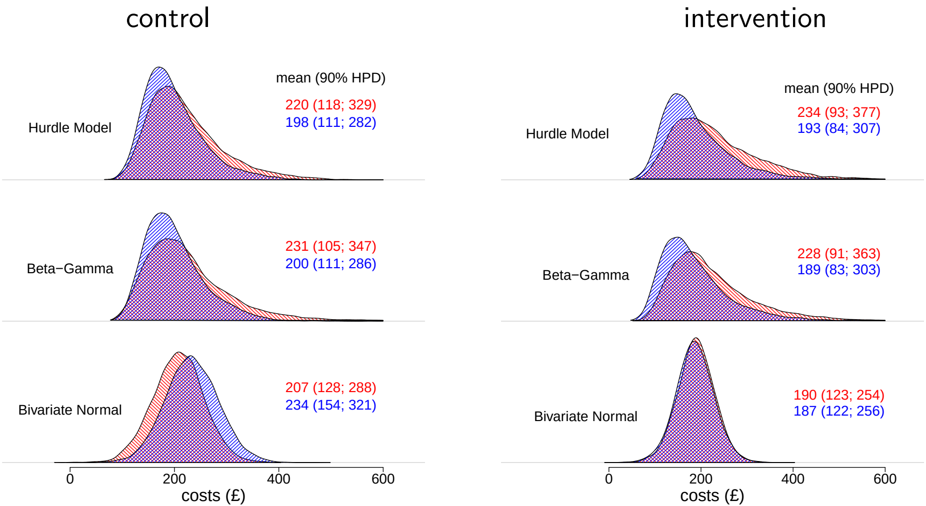
Results: QALYs



Complete Cases

All cases (Missing At Random)

Results: Costs

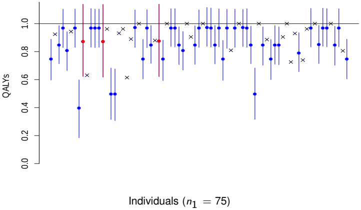


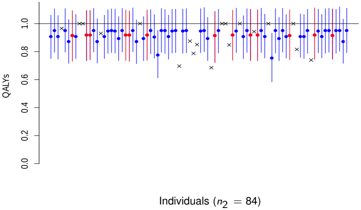
Complete Cases

All cases (Missing At Random)

Imputations (under MAR)

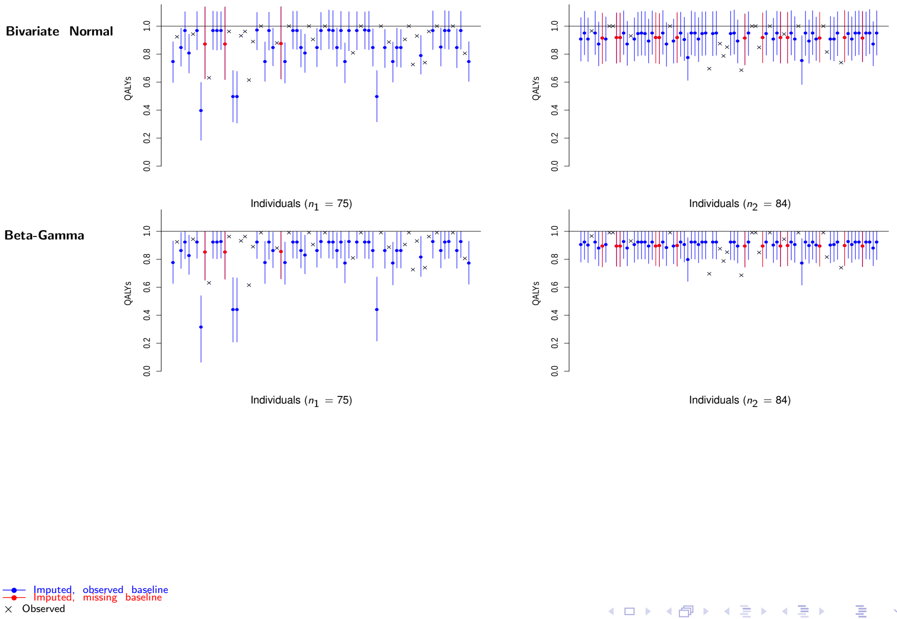
Bivariate Normal



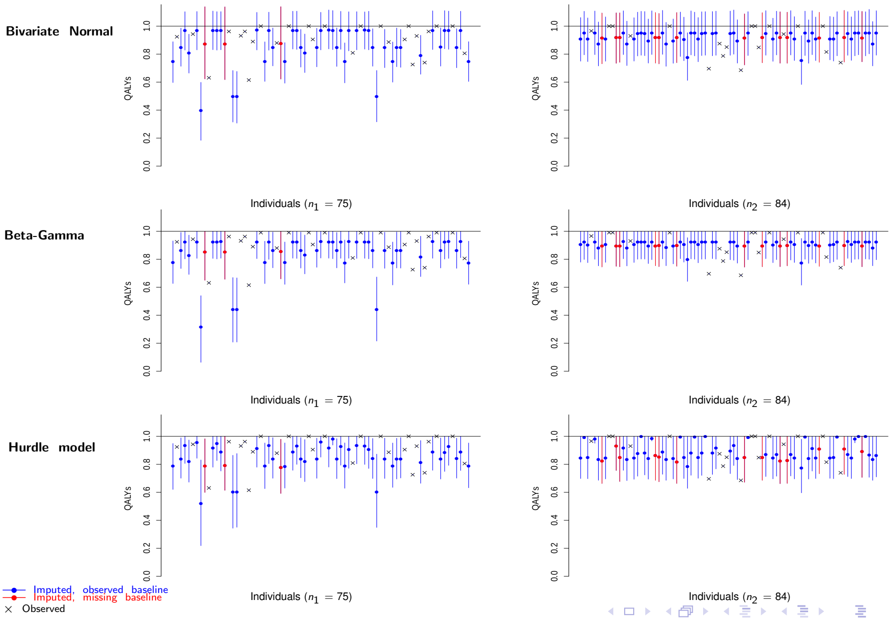


Imputed, observed baseline
Imputed, missing baseline
× Observed

Imputations (under MAR)



Imputations (under MAR)



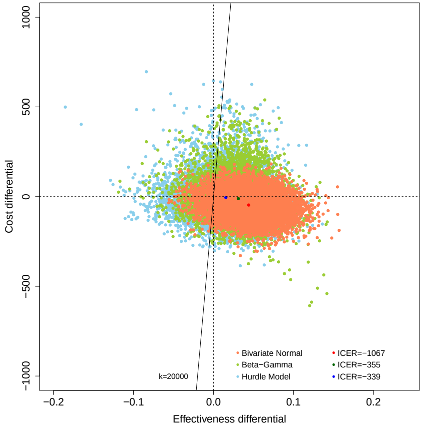
"extreme" MNAR scenarios

- We observe $n^*_{01} = 13$ and $n^*_{02} = 22$ individuals with $u_{0it} = 1$ and $u_{jit} = \text{NA}$, for $j = 1, 2, 3$
- For those individuals, we cannot compute directly the structural one indicator d_{it} and so need to make assumptions/model this
  - Sensitivity analysis to alternative departures from MAR

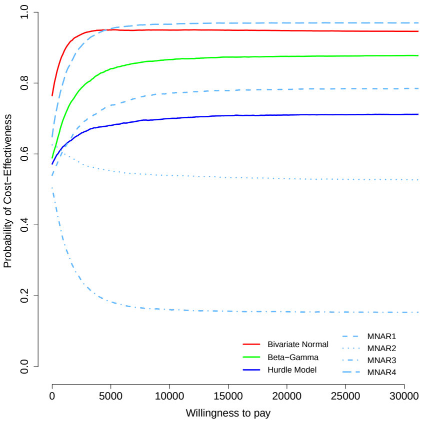
Scenario	Control $(n^*_1 = 13)$	Intervention $(n^*_2 = 22)$
MNAR1	$d_{i1} = 1$	$d_{i2} = 1$
MNAR2	$d_{i1} = 0$	$d_{i2} = 0$
MNAR3	$d_{i1} = 1$	$d_{i2} = 0$
MNAR4	$d_{i1} = 0$	$d_{i2} = 1$

Cost-effectiveness analysis

Cost-Effectiveness Plane



Cost-Effectiveness Acceptability Curve



Discussion

- ILD are subject to some complexities that are typically ignored by the "standard" approach, which could lead to biased results

Discussion

- ILD are subject to some complexities that are typically ignored by the "standard" approach, which could lead to biased results
- A Bayesian approach allows to increase model complexity to jointly account for these with relatively little expansion to the basic model

Discussion

- ILD are subject to some complexities that are typically ignored by the "standard" approach, which could lead to biased results
- A Bayesian approach allows to increase model complexity to jointly account for these with relatively little expansion to the basic model
- MAR can be used as reference assumption but plausible MNAR departures should be explored in sensitivity analysis

Discussion

- ILD are subject to some complexities that are typically ignored by the “standard” approach, which could lead to biased results
- A Bayesian approach allows to increase model complexity to jointly account for these with relatively little expansion to the basic model
- MAR can be used as reference assumption but plausible MNAR departures should be explored in sensitivity analysis
- Possible to expand the framework to a longitudinal setting to handle missingness more efficiently

A longitudinal missingness model

- Advantages
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$

A longitudinal missingness model

- Advantages
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$

A longitudinal missingness model

- Advantages
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$
- Fit model to the joint $p(\boldsymbol{y}, \boldsymbol{r})$
  - Factor $p(\boldsymbol{y}, \boldsymbol{r})$ into $p(\mathbf{y^r}_{obs}, \boldsymbol{r})$ and $p(\mathbf{y^r}_{mis} \mid \mathbf{y^r}_{obs}, \boldsymbol{r})$

A longitudinal missingness model

- Advantages
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$
- Fit model to the joint $p(\boldsymbol{y}, \boldsymbol{r})$
  - Factor $p(\boldsymbol{y}, \boldsymbol{r})$ into $p(\mathbf{y^r}_{obs}, \boldsymbol{r})$ and $p(\mathbf{y^r}_{mis} \mid \mathbf{y^r}_{obs}, \boldsymbol{r})$
  - Integrate out $\mathbf{y^r}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ and estimate the means of $\mathbf{y^r}_{obs}$

A longitudinal missingness model

- Advantages
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$
- Fit model to the joint $p(\boldsymbol{y}, \boldsymbol{r})$
  - Factor $p(\boldsymbol{y}, \boldsymbol{r})$ into $p(\mathbf{y^r}_{obs}, \boldsymbol{r})$ and $p(\mathbf{y^r}_{mis} \mid \mathbf{y^r}_{obs}, \boldsymbol{r})$
  - Integrate out $\mathbf{y^r}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ and estimate the means of $\mathbf{y^r}_{obs}$
  - Identify the means of $\mathbf{y^r}_{mis}$ using:

A longitudinal missingness model

- Advantages
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$
- Fit model to the joint $p(\boldsymbol{y}, \boldsymbol{r})$
  - Factor $p(\boldsymbol{y}, \boldsymbol{r})$ into $p(\mathbf{y^r}_{obs}, \boldsymbol{r})$ and $p(\mathbf{y^r}_{mis} \mid \mathbf{y^r}_{obs}, \boldsymbol{r})$
  - Integrate out $\mathbf{y^r}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ and estimate the means of $\mathbf{y^r}_{obs}$
  - Identify the means of $\mathbf{y^r}_{mis}$ using:
    - The mean estimates of $\mathbf{y^r}_{obs}$
    - Sensitivity parameters $\boldsymbol{\Delta} = (\Delta^u, \Delta^c)$

A longitudinal missingness model

- Advantages
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$
- Fit model to the joint $p(\boldsymbol{y}, \boldsymbol{r})$
  - Factor $p(\boldsymbol{y}, \boldsymbol{r})$ into $p(\mathbf{y^r}_{obs}, \boldsymbol{r})$ and $p(\mathbf{y^r}_{mis} \mid \mathbf{y^r}_{obs}, \boldsymbol{r})$
  - Integrate out $\mathbf{y^r}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ and estimate the means of $\mathbf{y^r}_{obs}$
  - Identify the means of $\mathbf{y^r}_{mis}$ using:
    - The mean estimates of $\mathbf{y^r}_{obs}$
    - Sensitivity parameters $\boldsymbol{\Delta} = (\Delta^u, \Delta^c)$
  - Assess the robustness of the results to plausible MNAR scenarios using different informative priors on $\boldsymbol{\Delta}$

The PBS study

Hassiotis et al., *Br J Psychiatry* 2018; 212(3)

- Multi-centre RCT that evaluates the cost-effectiveness of a new multicomponent intervention (PBS) relative to TAU for individuals suffering from intellectual disability and challenging behaviour
- Both utilities (EQ-5D) and costs (clinic records) are partially-observed

Time	**TAU** (n_1=136)		**PBS** (n_2=108)	
	observed (%)		observed (%)	
	utilities	costs	utilities	costs
Baseline	127 (93%)	136 (100%)	103 (95%)	108 (100%)
6 months	119 (86%)	128 (94%)	102 (94%)	103 (95%)
12 months	125 (92%)	130 (96%)	103 (95%)	104 (96%)
complete cases	108 (79%)		96 (89%)	

Missingness patterns

	TAU ($t = 1$)							PBS ($t = 2$)						
	u_0	c_0	u_1	c_1	u_2	c_2	n_{r1}	u_0	c_0	u_1	c_1	u_2	c_2	n_{r2}
$r = \mathbf{1}$	1	1	1	1	1	1	**108**	1	1	1	1	1	1	**96**
mean	0.678	1546	0.684	1527	0.680	1520		0.726	2818	0.771	2833	0.759	2878	
r	0	1	1	1	1	1	**7**	0	1	1	1	1	1	**5**
mean	–	1310	0.704	1440	0.644	1858		–	2573	0.780	2939	0.849	2113	
r	1	1	0	1	1	1	**4**	1	1	0	1	1	1	**1**
mean	0.709	1620	–	1087	0.737	851		0.467	9649	–	4828	0.259	4930	
r	1	1	1	1	0	1	**2**	1	1	1	1	0	1	**1**
mean	0.564	640	0.648	512	–	286		0.817	3788	0.884	0	–	0	
r	1	1	0	0	1	1	**4**	1	1	0	0	1	1	**1**
mean	0.716	2834	–	–	0.634	679		0.501	3608	–	–	0.872	4781	
r	1	1	0	0	0	0	**4**	1	1	0	0	0	0	**4**
mean	0.434	1528	–	–	–	–		0.760	3086	–	–	–	–	
r	0	1	0	1	1	1	**2**	0	1	0	1	1	1	**0**
mean	–	595	–	397	0.483	69		–	–	–	–	–	–	
r	1	1	1	1	0	0	**2**	1	1	1	1	0	0	**0**
mean	0.743	1434	0.705	1606	–	–		–	–	–	–	–	–	
r	1	1	0	1	0	1	**3**	1	1	0	1	0	1	**0**
mean	0.726	1510	–	432	–	976		–	–	–	–	–	–	

Modelling

- Fit model to completers $\boldsymbol{r} = \mathbf{1}$ and joint set of all other patterns $\boldsymbol{r} \neq \mathbf{1}$ separately for $t = 1, 2$

Modelling

- Fit model to completers $\boldsymbol{r} = \mathbf{1}$ and joint set of all other patterns $\boldsymbol{r} \neq \mathbf{1}$ separately for $t = 1, 2$
- Capture outcome and time dependence through a series of conditional distributions $p(c_{ij} \mid c_{ij-1}, u_{ij-1})$ and $p(u_{ij} \mid c_{ij}, u_{ij-1})$

Modelling

- Fit model to completers $\boldsymbol{r} = \mathbf{1}$ and joint set of all other patterns $\boldsymbol{r} \neq \mathbf{1}$ separately for $t = 1, 2$
- Capture outcome and time dependence through a series of conditional distributions $p(c_{ij} \mid c_{ij-1}, u_{ij-1})$ and $p(u_{ij} \mid c_{ij}, u_{ij-1})$
- Account for skewness using Beta distributions for $u^{\star}_{ij}$ and LogNormal distributions for c_{ij}, with $u^{\star}_{ij} = \frac{u_{ij} - \min(\boldsymbol{u}_j)}{\max(\boldsymbol{u}_j) - \min(\boldsymbol{u}_j)}$

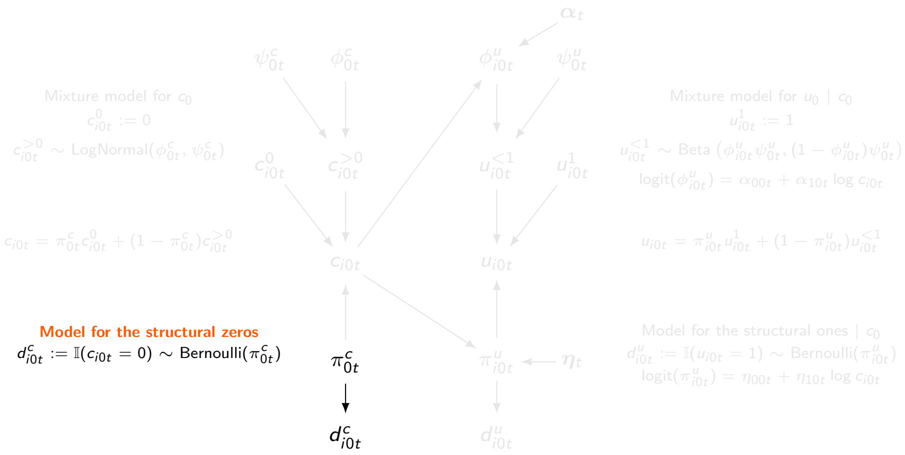
Modelling

- Fit model to completers $\boldsymbol{r} = \mathbf{1}$ and joint set of all other patterns $\boldsymbol{r} \neq \mathbf{1}$ separately for $t = 1, 2$
- Capture outcome and time dependence through a series of conditional distributions $p(c_{ij} \mid c_{ij-1}, u_{ij-1})$ and $p(u_{ij} \mid c_{ij}, u_{ij-1})$
- Account for skewness using Beta distributions for $u^{\star}_{ij}$ and LogNormal distributions for c_{ij}, with $u^{\star}_{ij} = \frac{u_{ij} - \min(\boldsymbol{u}_j)}{\max(\boldsymbol{u}_j) - \min(\boldsymbol{u}_j)}$
- Allow for structural ones in u_{ij} and zeros in c_{ij} using a hurdle form, i.e. $d^{u}_{ij} := \mathbb{I}(u_{ij} = 1)$ and $d^{c}_{ij} := \mathbb{I}(c_{ij} = 0)$

Modelling

Gabrio et al. (2018). https://arxiv.org/abs/1805.07147

- For example, consider the pair $\boldsymbol{y}_{i0t} = (u_{i0t}, c_{i0t})$



Modelling

Gabrio et al. (2018). https://arxiv.org/abs/1805.07147

- For example, consider the pair $\boldsymbol{y}_{i0t} = (u_{i0t}, c_{i0t})$

Mixture model for c_0

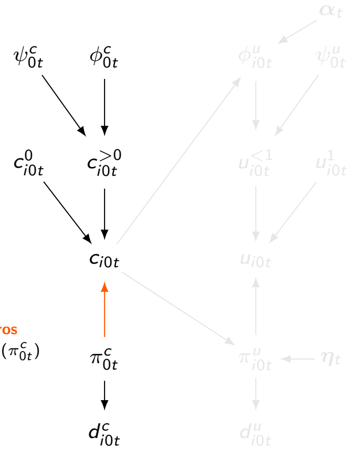
$$c^0_{i0t} := 0$$

$$c^{>0}_{i0t} \sim \text{LogNormal}(\phi^c_{0t}, \psi^c_{0t})$$

$$c_{i0t} = \pi^c_{0t} c^0_{i0t} + (1 - \pi^c_{0t}) c^{>0}_{i0t}$$

Model for the structural zeros

$$d^c_{i0t} := \mathbb{I}(c_{i0t} = 0) \sim \text{Bernoulli}(\pi^c_{0t})$$



Mixture model for $u_0 \mid c_0$

$$u^1_{i0t} := 1$$

$$u^{<1}_{i0t} \sim \text{Beta}\left(\phi^u_{i0t}\psi^u_{0t}, (1 - \phi^u_{i0t})\psi^u_{0t}\right)$$

$$\text{logit}(\phi^u_{i0t}) = \alpha_{00t} + \alpha_{10t} \log c_{i0t}$$

$$u_{i0t} = \pi^u_{i0t} u^1_{i0t} + (1 - \pi^u_{i0t}) u^{<1}_{i0t}$$

Model for the structural ones $\mid c_0$

$$d^u_{i0t} := \mathbb{I}(u_{i0t} = 1) \sim \text{Bernoulli}(\pi^u_{i0t})$$

$$\text{logit}(\pi^u_{i0t}) = \eta_{00t} + \eta_{10t} \log c_{i0t}$$

Modelling

Gabrio et al. (2018). https://arxiv.org/abs/1805.07147

- For example, consider the pair $\boldsymbol{y}_{i0t} = (u_{i0t}, c_{i0t})$

Mixture model for c_0

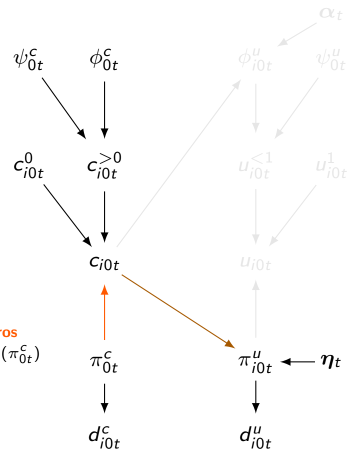
$c^0_{i0t} := 0$

$c^{>0}_{i0t} \sim \text{LogNormal}(\phi^c_{0t}, \psi^c_{0t})$

$c_{i0t} = \pi^c_{0t} c^0_{i0t} + (1 - \pi^c_{0t}) c^{>0}_{i0t}$

Model for the structural zeros

$d^c_{i0t} := \mathbb{I}(c_{i0t} = 0) \sim \text{Bernoulli}(\pi^c_{0t})$



Mixture model for $u_0 \mid c_0$

$u^1_{i0t} := 1$

$u^{<1}_{i0t} \sim \text{Beta}\left(\phi^u_{i0t}\psi^u_{0t}, (1 - \phi^u_{i0t})\psi^u_{0t}\right)$

$\text{logit}(\phi^u_{i0t}) = \alpha_{00t} + \alpha_{10t} \log c_{i0t}$

$u_{i0t} = \pi^u_{i0t} u^1_{i0t} + (1 - \pi^u_{i0t}) u^{<1}_{i0t}$

Model for the structural ones | c_0

$d^u_{i0t} := \mathbb{I}(u_{i0t} = 1) \sim \text{Bernoulli}(\pi^u_{i0t})$

$\text{logit}(\pi^u_{i0t}) = \eta_{00t} + \eta_{10t} \log c_{i0t}$

Modelling

Gabrio et al. (2018). https://arxiv.org/abs/1805.07147

- For example, consider the pair $\boldsymbol{y}_{i0t} = (u_{i0t}, c_{i0t})$

Mixture model for c_0

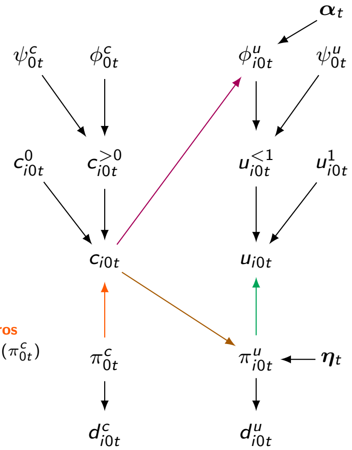
$c^{0}_{i0t} := 0$

$c^{>0}_{i0t} \sim \text{LogNormal}(\phi^{c}_{0t}, \psi^{c}_{0t})$

$c_{i0t} = \pi^{c}_{0t} c^{0}_{i0t} + (1 - \pi^{c}_{0t}) c^{>0}_{i0t}$

Model for the structural zeros

$d^{c}_{i0t} := \mathbb{I}(c_{i0t} = 0) \sim \text{Bernoulli}(\pi^{c}_{0t})$



Mixture model for $u_0 \mid c_0$

$u^{1}_{i0t} := 1$

$u^{<1}_{i0t} \sim \text{Beta}\left(\phi^{u}_{i0t}\psi^{u}_{0t}, (1 - \phi^{u}_{i0t})\psi^{u}_{0t}\right)$

$\text{logit}(\phi^{u}_{i0t}) = \alpha_{00t} + \alpha_{10t} \log c_{i0t}$

$u_{i0t} = \pi^{u}_{i0t} u^{1}_{i0t} + (1 - \pi^{u}_{i0t}) u^{<1}_{i0t}$

Model for the structural ones $\mid c_0$

$d^{u}_{i0t} := \mathbb{I}(u_{i0t} = 1) \sim \text{Bernoulli}(\pi^{u}_{i0t})$

$\text{logit}(\pi^{u}_{i0t}) = \eta_{00t} + \eta_{10t} \log c_{i0t}$

Sensitivity analysis

- Use Monte Carlo integration to derive the mean estimates $\text{E}[\mathbf{y^r}_{obs}]$

Sensitivity analysis

- Use Monte Carlo integration to derive the mean estimates $\mathrm{E}[\mathbf{y^r}_{obs}]$
- Identify the mean estimates $\mathrm{E}[\mathbf{y^r}_{mis}] = \mathrm{E}[\mathbf{y^r}_{obs} + \boldsymbol{\Delta}_j]$

Sensitivity analysis

- Use Monte Carlo integration to derive the mean estimates $E[\mathbf{y^r}_{obs}]$
- Identify the mean estimates $E[\mathbf{y^r}_{mis}] = E[\mathbf{y^r}_{obs} + \mathbf{\Delta}_j]$
- Compute weighted average across $\boldsymbol{r}$ to derive $\boldsymbol{\mu}_{jt} = (\mu^u_{jt}, \mu^c_{jt})$

Sensitivity analysis

- Use Monte Carlo integration to derive the mean estimates $\text{E}[\mathbf{y^r}_{obs}]$
- Identify the mean estimates $\text{E}[\mathbf{y^r}_{mis}] = \text{E}[\mathbf{y^r}_{obs} + \boldsymbol{\Delta}_j]$
- Compute weighted average across $\boldsymbol{r}$ to derive $\boldsymbol{\mu}_{jt} = (\mu^u_{jt}, \mu^c_{jt})$
- Set $\boldsymbol{\Delta}_j = \mathbf{0}$ as benchmark assumption

Sensitivity analysis

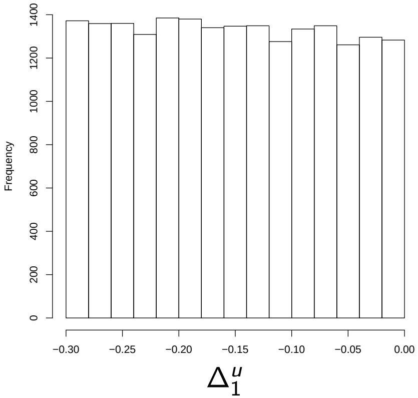
- Use Monte Carlo integration to derive the mean estimates $\text{E}[\mathbf{y^r}_{obs}]$
- Identify the mean estimates $\text{E}[\mathbf{y^r}_{mis}] = \text{E}[\mathbf{y^r}_{obs} + \boldsymbol{\Delta}_j]$
- Compute weighted average across $\boldsymbol{r}$ to derive $\boldsymbol{\mu}_{jt} = (\mu^u_{jt}, \mu^c_{jt})$
- Set $\boldsymbol{\Delta}_j = \mathbf{0}$ as benchmark assumption
- Specify three alternative priors on $\boldsymbol{\Delta}_j = (\Delta^u_j, \Delta^c_j)$, calibrated based on the variability in the observed data at each time j

Priors on sensitivity parameters

- Assumption: $\boldsymbol{u}_{mis} < \boldsymbol{u}_{obs}$ and $\boldsymbol{c}_{mis} > \boldsymbol{c}_{obs}$

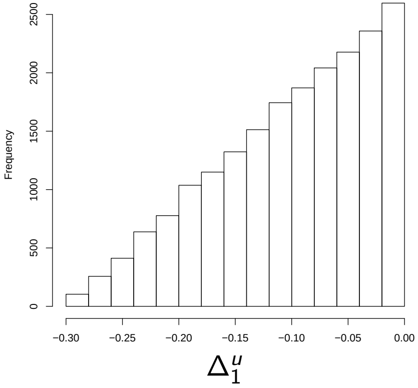
Priors on sensitivity parameters

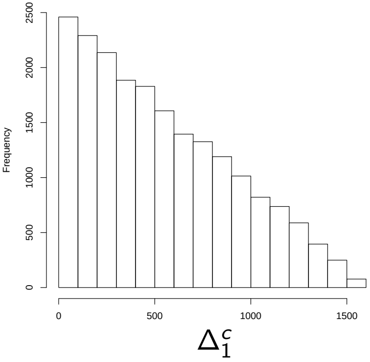
- Assumption: $\boldsymbol{u}_{mis} < \boldsymbol{u}_{obs}$ and $\boldsymbol{c}_{mis} > \boldsymbol{c}_{obs}$
- $\boldsymbol{\Delta}^{\text{flat}}$: Flat between 0 and twice the observed standard deviation



Priors on sensitivity parameters

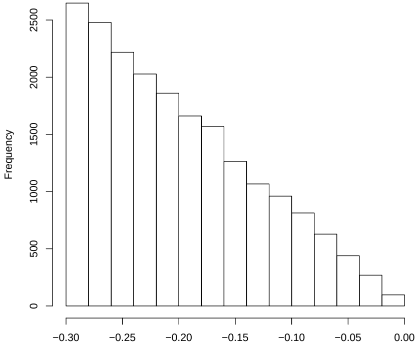
- Assumption: $\boldsymbol{u}_{mis} < \boldsymbol{u}_{obs}$ and $\boldsymbol{c}_{mis} > \boldsymbol{c}_{obs}$
- $\boldsymbol{\Delta}^{\text{skew0}}$: Skewed towards values closer to 0 on the same range as $\boldsymbol{\Delta}^{\text{flat}}$



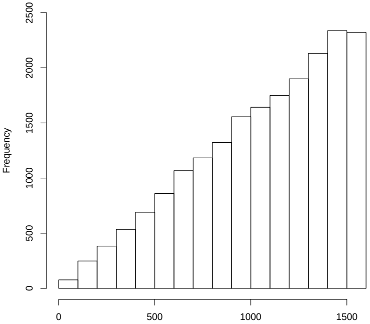


Priors on sensitivity parameters

- Assumption: $\boldsymbol{u}_{mis} < \boldsymbol{u}_{obs}$ and $\boldsymbol{c}_{mis} > \boldsymbol{c}_{obs}$
- $\boldsymbol{\Delta}^{\text{skew1}}$: Skewed towards values far from 0 on the same range as $\boldsymbol{\Delta}^{\text{flat}}$

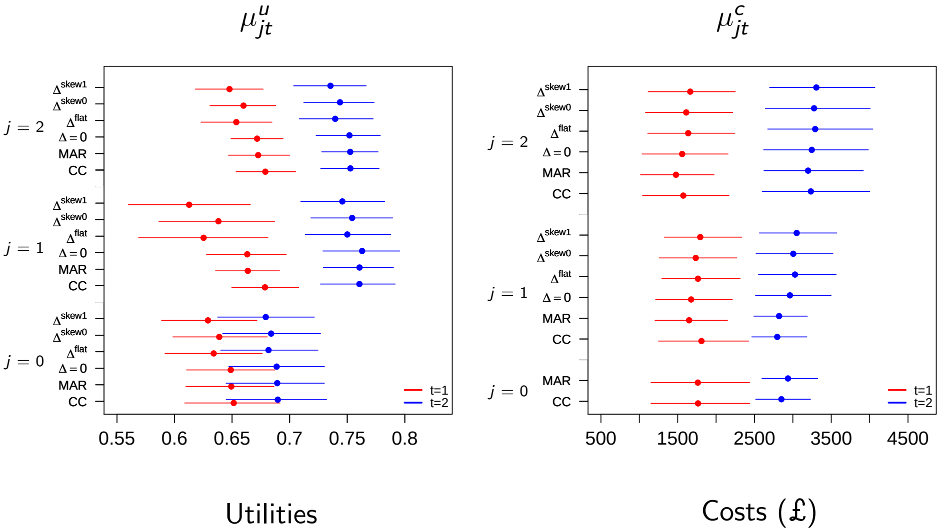


Δ_1^u

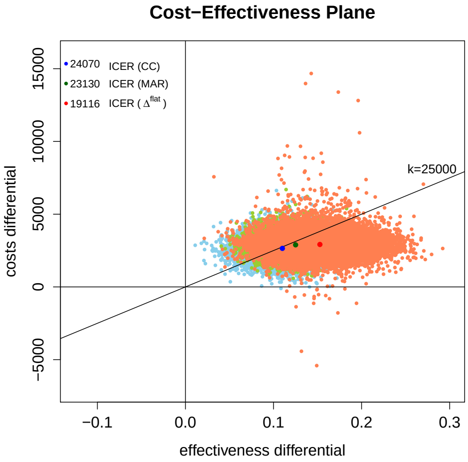


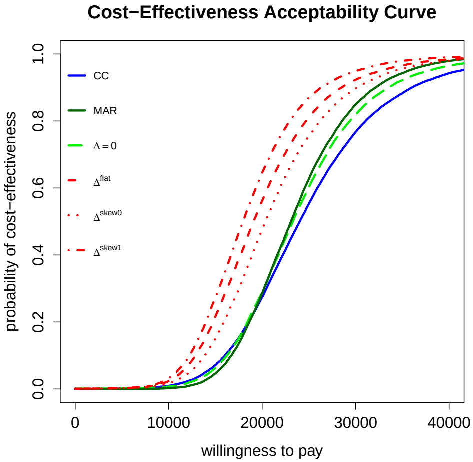
Δ_1^c

Results: means utilities and costs

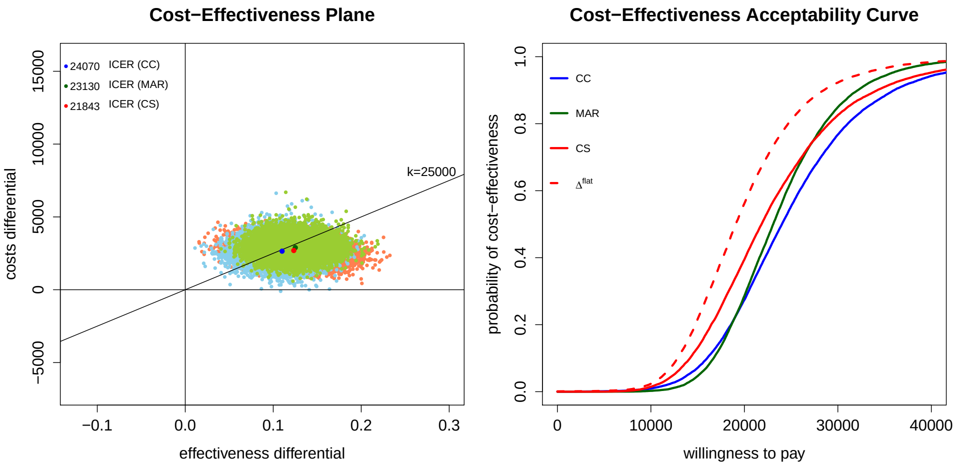


Results: economic evaluation (1)





Results: economic evaluation (2)



Conclusions

- A full Bayesian approach has some advantages for handling missing data in economic evaluation

Conclusions

- A full Bayesian approach has some advantages for handling missing data in economic evaluation

1. Flexibility of the modelling framework
    - Naturally allows the propagation of uncertainty to the economic model
    - Uses modular structure to account for complexities that may bias inferences and mislead the assessment

Conclusions

- A full Bayesian approach has some advantages for handling missing data in economic evaluation

1. Flexibility of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses modular structure to account for complexities that may bias inferences and mislead the assessment
2. Extension of standard "imputation methods"
   - Performs the estimation and imputation tasks simultaneously
   - Uses probabilistic models that can be implemented in standard software (e.g. `OpenBUGS` or `JAGS`)

Conclusions

- A full Bayesian approach has some advantages for handling missing data in economic evaluation

1. Flexibility of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses modular structure to account for complexities that may bias inferences and mislead the assessment
2. Extension of standard "imputation methods"
   - Performs the estimation and imputation tasks simultaneously
   - Uses probabilistic models that can be implemented in standard software (e.g. `OpenBUGS` or `JAGS`)
3. Principled incorporation of external evidence through priors
   - Crucial for conducting sensitivity analysis to MNAR
   - Useful in small/pilot trials where there is limited evidence