A Bayesian Parametric Approach to Handle Missing Longitudinal Outcome Data in Trial-Based Health Economic Evaluations

Andrea Gabrio, [1] Michael J. Daniels [2] and Gianluca Baio [1]

[1] Department of Statistical Science, University College London (UCL),
[2] Department of Biostatistics, University of Florida (UF)

ucakgab@.ucl.ac.uk

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

5th International Clinical Trials Methodology Conference, PS1C - O4

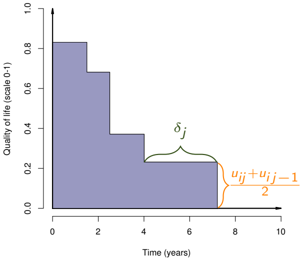
Brighton, 07 October 2019

"Standard" Statistical modelling **Individual level** data

		Demographics			HRQL data				Resource use data				Clinical outcome			
ID	**Trt**	Sex	Age	...	u_0	u_1	...	u_J	c_0	c_1	...	c_J	y_0	y_1	...	y_J
1	1	M	23	...	0.32	0.66	...	0.44	103	241	...	80	y_{10}	y_{11}	...	y_{1J}
2	1	M	21	...	0.12	0.16	...	0.38	1 204	1 808	...	877	y_{20}	y_{21}	...	y_{2J}
3	2	F	19	...	0.49	0.55	...	0.88	16	12	...	22	y_{30}	y_{31}	...	y_{3J}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

1. Compute individual QALYs and total costs as

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



QALY_i = "Area under the curve"

"Standard" Statistical modelling

Individual level data

		Demographics			HRQL data				Resource use data				Clinical outcome			
ID	**Trt**	Sex	Age	...	u_0	u_1	...	u_J	c_0	c_1	...	c_J	y_0	y_1	...	y_J
1	1	M	23	...	0.32	0.66	...	0.44	103	241	...	80	y_{10}	y_{11}	...	y_{1J}
2	1	M	21	...	0.12	0.16	...	0.38	1 204	1 808	...	877	y_{20}	y_{21}	...	y_{2J}
3	2	F	19	...	0.49	0.55	...	0.88	16	12	...	22	y_{30}	y_{31}	...	y_{3J}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

1. Compute individual QALYs and total costs as

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

2. (Often implicitly) assume normality and linearity and model **independently** individual QALYs and total costs by controlling for baseline values

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

3. Estimate mean differentials and use bootstrap to quantify **uncertainty**

What's wrong with this?

Statistical issues

- Potential **correlation** between costs & utilities
  - Ignoring a strong correlation may inflate variability in the estimates
- **Asymmetric** empirical distributions
  - 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
  - Modelling (e_i, c_i) is inefficient as partially-observed (u_{ij}, c_{ij}) are ignored
  - Inference often based on the observed data alone – **Missing At Random** – untestable
  - Plausible **Missing Not At Random** departures should be explored in sensitivity analysis

A Bayesian longitudinal missing data model

- **Key features**:
  - Jointly model $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$ to account for correlation
  - Use parametric distributions to deal with skewness (e.g. Beta and LogNormal) with an hurdle approach to handle spikes at one and zero
  - Use all data in each missingness pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$ to fit the model

A Bayesian longitudinal missing data model

- **Key features**:
  - Jointly model $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$ to account for correlation
  - Use parametric distributions to deal with skewness (e.g. Beta and LogNormal) with an hurdle approach to handle spikes at one and zero
  - Use all data in each missingness pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$ to fit the model

- Conduct sensitivity analysis to MNAR using a **pattern mixture model** for $p(\boldsymbol{y}, \boldsymbol{r})$:
  - Integrate out $\mathbf{y^r}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ to estimate $\text{E}[\mathbf{y^r}_{obs} \mid \boldsymbol{r}]$
  - Use sensitivity parameters to identify $\text{E}[\mathbf{y^r}_{mis} \mid \mathbf{y^r}_{obs}, \boldsymbol{r}] = \text{E}[\mathbf{y^r}_{obs} \mid \boldsymbol{r}] + \boldsymbol{\Delta}$
  - Assess the robustness of the results to plausible MNAR assumptions using different informative priors on $\boldsymbol{\Delta} = (\Delta^u, \Delta^c)$

A Bayesian longitudinal missing data model

- **Key features**:
  - Jointly model $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$ to account for correlation
  - Use parametric distributions to deal with skewness (e.g. Beta and LogNormal) with an hurdle approach to handle spikes at one and zero
  - Use all data in each missingness pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$ to fit the model
- Conduct sensitivity analysis to MNAR using a **pattern mixture model** for $p(\boldsymbol{y}, \boldsymbol{r})$:
  - Integrate out $\mathbf{y^r}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ to estimate $\text{E}[\mathbf{y^r}_{obs} \mid \boldsymbol{r}]$
  - Use sensitivity parameters to identify $\text{E}[\mathbf{y^r}_{mis} \mid \mathbf{y^r}_{obs}, \boldsymbol{r}] = \text{E}[\mathbf{y^r}_{obs} \mid \boldsymbol{r}] + \boldsymbol{\Delta}$
  - Assess the robustness of the results to plausible MNAR assumptions using different informative priors on $\boldsymbol{\Delta} = (\Delta^u, \Delta^c)$
- Focus in **decision-making**, not inference – Bayesian approach particularly suited
  - Uncertainty about unobserved quantities is fully characterised using MCMC
  - Priors can be used to incorporate external evidence

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 the control for individuals suffering from intellectual disability and challenging behaviour
- Both utilities (EQ-5D) and costs (clinic records) are partially-observed

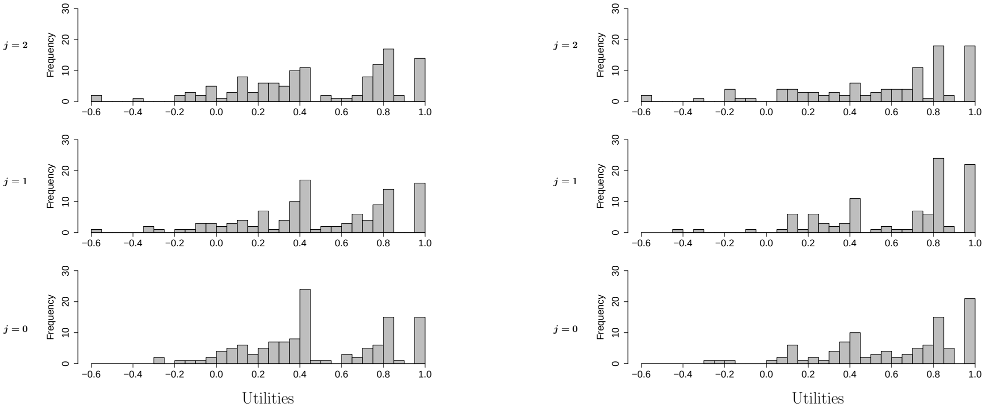
Time	**Control** ($t = 1$)		**Intervention** ($t = 2$)	
	observed (%)		observed (%)	
	utilities	**costs**	**utilities**	**costs**
Baseline ($j = 0$)	127 (93%)	136 (100%)	103 (95%)	108 (100%)
6 months ($j = 1$)	119 (86%)	128 (94%)	102 (94%)	103 (95%)
12 months ($j = 2$)	125 (92%)	130 (96%)	103 (95%)	104 (96%)
complete cases	108 (79%)		96 (89%)	

The PBS study

utility distributions

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

Control **Intervention**

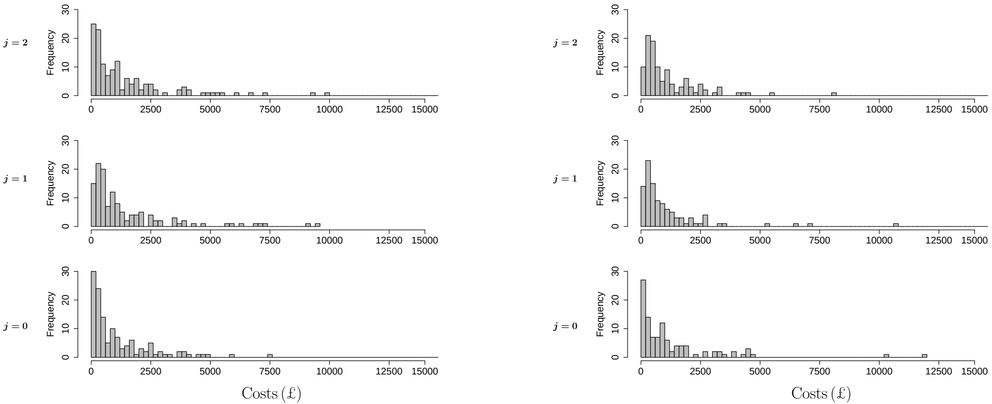


The PBS study

cost distributions

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

Control **Intervention**



Modelling the data from PBS

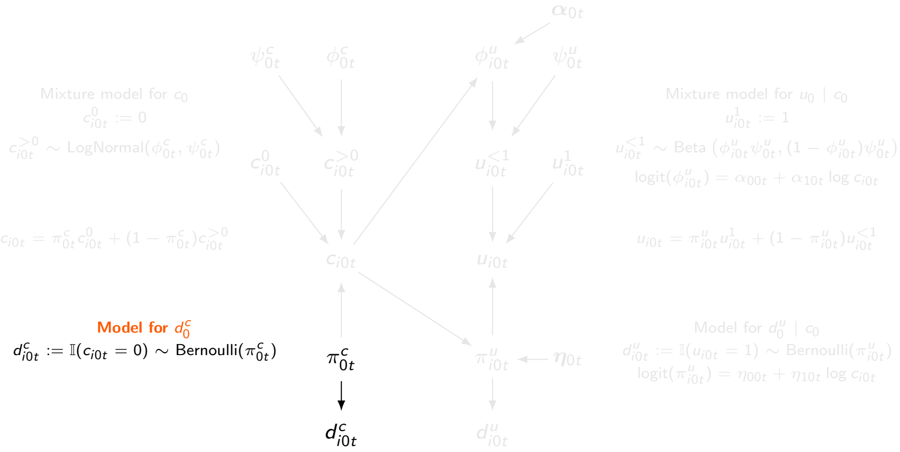
pattern mixture model

- Fit the model to the completers and the set of all other patterns separately for $t = 1, 2$
- Capture correlation between variables through a series of conditional distributions $p(c_{ij} \mid c_{ij-1}, u_{ij-1})$ and $p(u_{ij} \mid c_{ij}, u_{ij-1})$ using regressions on the log and logit scale
- Account for skewness using **Beta** distributions for u_{ij} and **LogNormal** distributions for c_{ij}
- Allow for **ones** in u_{ij} and **zeros** in c_{ij} using a hurdle form, i.e. modelling the indicators $d^u_{ij} := \mathbb{I}(u_{ij} = 1)$ and $d^c_{ij} := \mathbb{I}(c_{ij} = 0)$ using logistic regressions

Model structure

Gabrio et al. (2019). https://arxiv.org/abs/1805.07147

- At $j = 0$



Model structure

Gabrio et al. (2019). https://arxiv.org/abs/1805.07147

- At $j = 0$

Mixture model for c_0

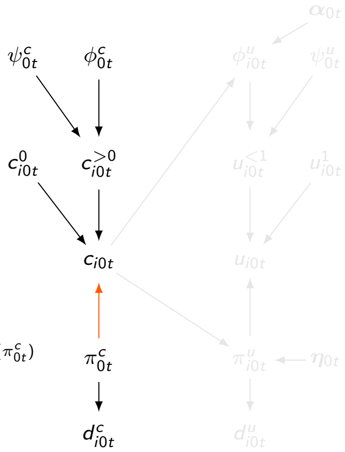
$$c^{0}_{i0t} := 0$$

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

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

Model for d^{c}_{0}

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



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 $d^{u}_{0} \mid c_0$

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

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

Model structure

Gabrio et al. (2019). https://arxiv.org/abs/1805.07147

- At $j = 0$

Mixture model for c_0

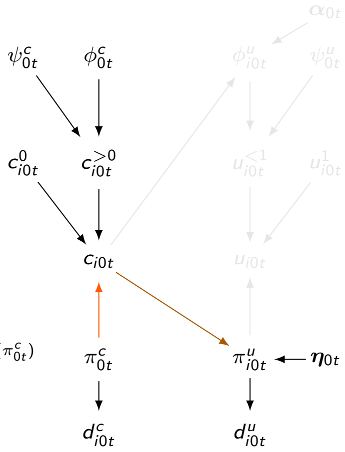
$c^{0}_{i0t} := 0$

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

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

Model for d^{c}_{0}

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



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 $d^{u}_{0} \mid c_0$

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

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

Model structure

Gabrio et al. (2019). https://arxiv.org/abs/1805.07147

- At $j = 0$

Mixture model for c_0

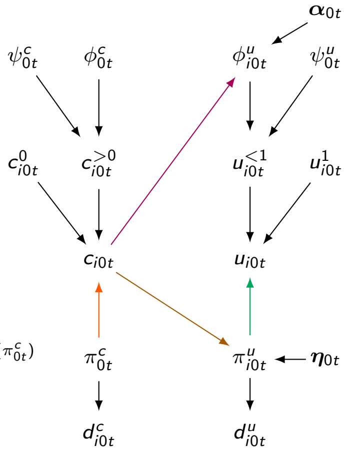
$c^{0}_{i0t} := 0$

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

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

Model for d^{c}_{0}

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



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 $d^{u}_{0} \mid c_0$

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

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

Sensitivity analysis to MNAR

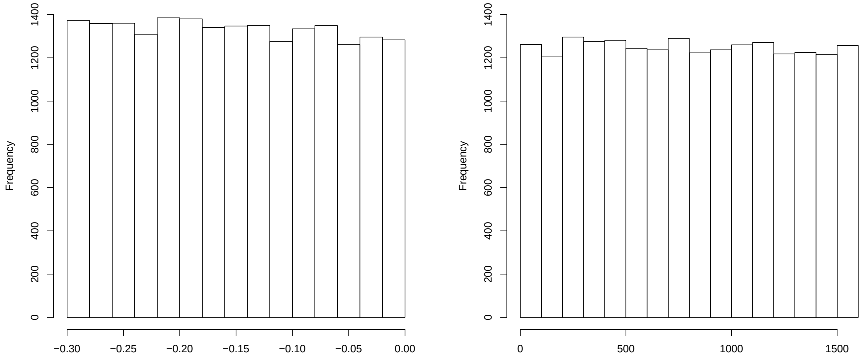
sensitivity parameters

- Use **Monte Carlo integration** to derive $\text{E}[\mathbf{y^r}_{obs} \mid \boldsymbol{r}]$
- Add **sensitivity parameters** to identify $\text{E}[\mathbf{y^r}_{mis} \mid \boldsymbol{r}] = \text{E}[\mathbf{y^r}_{obs} \mid \boldsymbol{r}] + \boldsymbol{\Delta}_j$
- Compute weighted averages over $\boldsymbol{r}$ to obtain the mean estimates $\boldsymbol{\mu}_{jt} = (\mu^u_{jt}, \mu^c_{jt})$
- Set $\boldsymbol{\Delta}_j = \mathbf{0}$ as benchmark assumption ($\approx$ MAR)
- Specify three MNAR priors on $\boldsymbol{\Delta}_j = (\Delta^u_j, \Delta^c_j)$, calibrated on the variability in the observed data at each time j

Priors on sensitivity parameters

MNAR priors

- Assumption: $u_{mis} < u_{obs}$ and $c_{mis} > c_{obs}$
- Δ^{flat}: Flat between 0 and twice the observed standard deviation

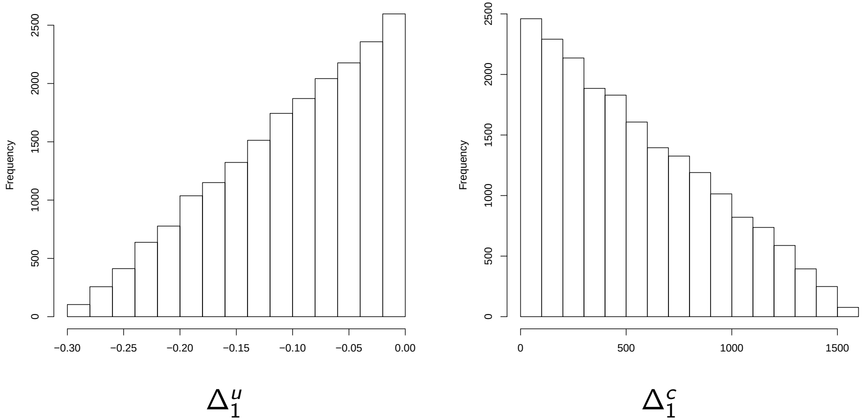


Δ_1^u Δ_1^c

Priors on sensitivity parameters

MNAR priors

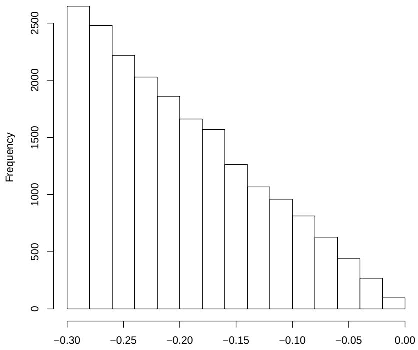
- Assumption: $u_{mis} < u_{obs}$ and $c_{mis} > c_{obs}$
- $\boldsymbol{\Delta}^{\text{skew0}}$: Skewed towards values closer to 0 on the same range as $\boldsymbol{\Delta}^{\text{flat}}$



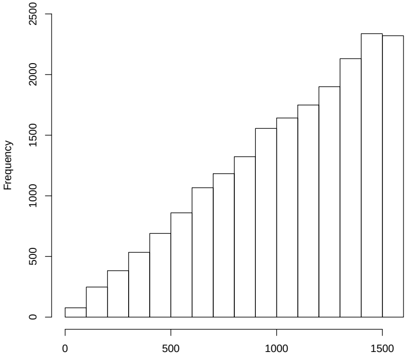
Priors on sensitivity parameters

MNAR priors

- Assumption: $u_{mis} < u_{obs}$ and $c_{mis} > c_{obs}$
- Δ^{skew1}: Skewed towards values far from 0 on the same range as Δ^{flat}



Δ_1^u

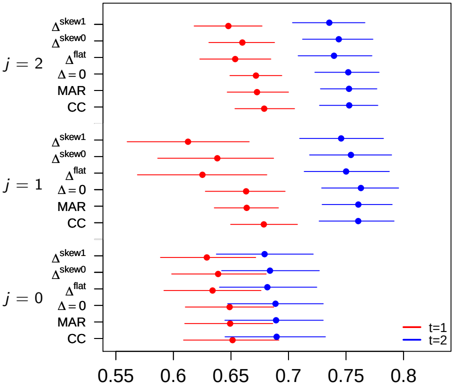


Δ_1^c

Means utilities and costs

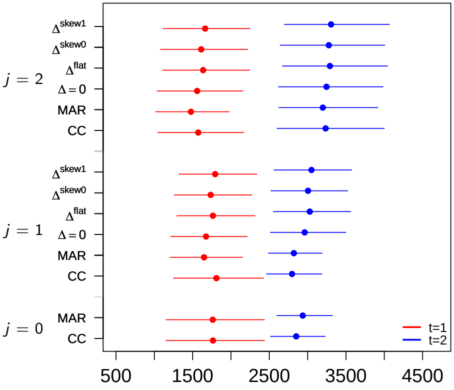
posterior estimates

μ^u_{jt}



Utilities

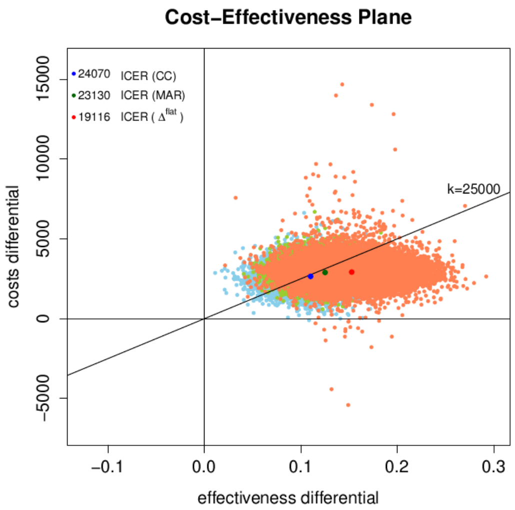
μ^c_{jt}

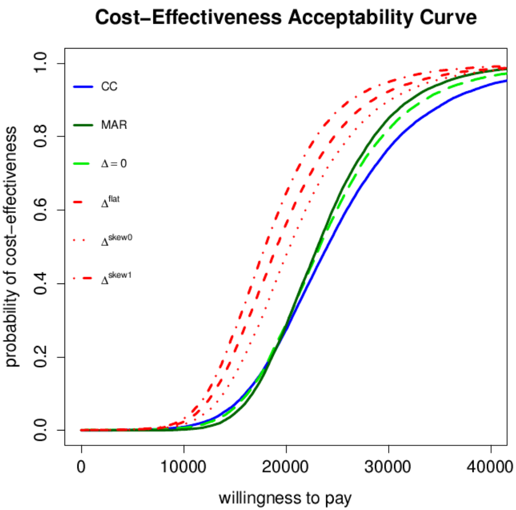


Costs (£)

Cost-effectiveness analysis

CEA results





Discussion

1. **Flexibility** of the modelling framework
    - Naturally allows the propagation of uncertainty to the economic model
    - Uses a modular structure to account for data complexities in a relatively easy way
    - Exploit all available evidence using a longitudinal approach
2. **Extension** of standard "imputation methods"
    - Performs the estimation and imputation tasks simultaneously
    - Fitting joint models for missing data straightforward with MCMC
    - 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