Bayesian parametric models to handle missing longitudinal outcome data in trial-based health economic evaluations

Gianluca Baio

University College London
Department of Statistical Science

g.baio@ucl.ac.uk

Joint work with **Andrea Gabrio (UCL)** and **Michael J Daniels (University of Florida)**

http://www.ucl.ac.uk/statistics/research/statistics-health-economics/
http://www.statistica.it/gianluca
https://github.com/giabaio

12th International Conference of the ERCIM WG on Computational and Methodological Statistics

Invited session: "Sensitivity analysis for uncheckable assumptions"

Monday 16 December 2019

Outline

1. **Health economic evaluation**
   - What is it?
   - How does it work?
2. Statistical modelling
   - Standard approach
   - The importance of being a Bayesian
3. Bayesian modelling for missing data in HTA
   - Modelling & advantages
   - Bayesian nature of dealing with missing data
4. Motivating example
   - Data (and their weird features...) & Bayesian modelling
   - Results
5. Conclusions

Outline

1. Health economic evaluation
   - What is it?
   - How does it work?

2. **Statistical modelling**
   - Standard approach
   - The importance of being a Bayesian

3. Bayesian modelling for missing data in HTA
   - Modelling & advantages
   - Bayesian nature of dealing with missing data

4. Motivating example
   - Data (and their weird features...) & Bayesian modelling
   - Results

5. Conclusions

Outline

1. Health economic evaluation
   - What is it?
   - How does it work?
2. Statistical modelling
   - Standard approach
   - The importance of being a Bayesian
3. **Bayesian modelling for missing data in HTA**
   - Modelling & advantages
   - Bayesian nature of dealing with missing data
4. Motivating example
   - Data (and their weird features...) & Bayesian modelling
   - Results
5. Conclusions

Outline

1. Health economic evaluation
   - What is it?
   - How does it work?
2. Statistical modelling
   - Standard approach
   - The importance of being a Bayesian
3. Bayesian modelling for missing data in HTA
   - Modelling & advantages
   - Bayesian nature of dealing with missing data
4. **Motivating example**
   - Data (and their weird features...) & Bayesian modelling
   - Results
5. Conclusions

Outline

1. Health economic evaluation
   - What is it?
   - How does it work?
2. Statistical modelling
   - Standard approach
   - The importance of being a Bayesian
3. Bayesian modelling for missing data in HTA
   - Modelling & advantages
   - Bayesian nature of dealing with missing data
4. Motivating example
   - Data (and their weird features...) & Bayesian modelling
   - Results
5. **Conclusions**

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

Statistical model

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

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

- Estimates relevant **population** parameters $\boldsymbol\theta$
- Varies with the type of available data (& statistical approach!)

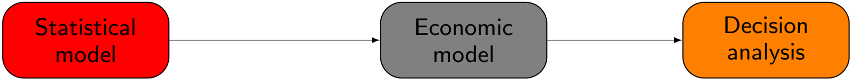
- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

Health technology assessment (HTA)

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

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

ICER $= g(\Delta_e, \Delta_c)$
EIB $= h(\Delta_e, \Delta_c; k)$
$\ldots$

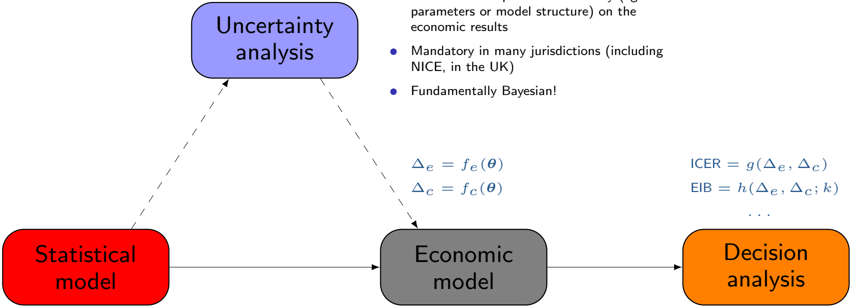


- Estimates relevant **population** parameters $\boldsymbol\theta$
- Varies with the type of available data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

Objective: Combine costs & benefits of a given intervention into a rational scheme for allocating resources



- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!

$\Delta_e = f_e(\boldsymbol\theta)$

$\Delta_c = f_c(\boldsymbol\theta)$

ICER $= g(\Delta_e, \Delta_c)$

EIB $= h(\Delta_e, \Delta_c; k)$

...

- Estimates relevant **population** parameters $\boldsymbol\theta$
- Varies with the type of available data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

		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}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

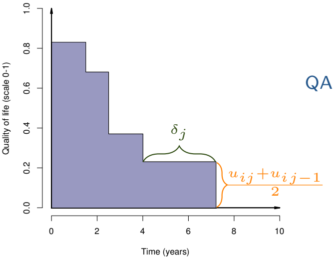
$y_{ij} =$ Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
$u_{ij} =$ Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
$c_{ij} =$ Use of resources (drugs, hospital, GP appointments, ...)

		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}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

y_{ij} = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
u_{ij} = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
c_{ij} = Use of resources (drugs, hospital, GP appointments, ...)

1. Compute individual QALYs and total costs as

$$e_i = \sum_{j=1}^{J} (u_{ij} + u_{ij-1}) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=0}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$



QALY_i = "Area under the curve"

		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}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

y_{ij} = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
u_{ij} = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
c_{ij} = Use of resources (drugs, hospital, GP appointments, ...)

1. Compute individual QALYs and total costs as

$$e_i = \sum_{j=1}^{J} (u_{ij} + u_{ij-1}) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=0}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$

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

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

		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}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

y_{ij} = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
u_{ij} = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
c_{ij} = Use of resources (drugs, hospital, GP appointments, ...)

1. Compute individual QALYs and total costs as

$$e_i = \sum_{j=1}^{J} (u_{ij} + u_{ij-1}) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=0}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$

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 population average cost and effectiveness differentials and use bootstrap to quantify uncertainty

What's wrong with this?...

- Potential correlation between costs & clinical benefits [Individual Level + Aggregated Level Data]
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research ⇒ 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.
  - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**

What's wrong with this?...

- Potential correlation between costs & clinical benefits **[Individual Level + Aggregated Level Data]**
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research $\Rightarrow$ 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.
  - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**

- Joint/marginal normality not realistic **[Mainly ILD]**
  - Costs usually skewed and benefits may be bounded in $[0; 1]$
  - Can use transformation (e.g. logs) — but care is needed when back transforming to the natural scale
  - Should use more suitable models (e.g. Beta, Gamma or log-Normal) — **generally easier under a Bayesian framework**
  - "Structural" values (0 costs, unit utilities)

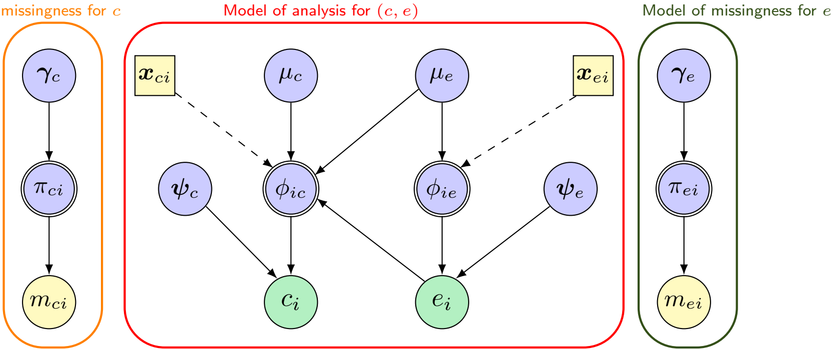
What's wrong with this?...

- Potential correlation between costs & clinical benefits [Individual Level + Aggregated Level Data]
  - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research ⇒ 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.
  - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**

- Joint/marginal normality not realistic [Mainly ILD]
  - Costs usually skewed and benefits may be bounded in $[0; 1]$
  - Can use transformation (e.g. logs) — but care is needed when back transforming to the natural scale
  - Should use more suitable models (e.g. Beta, Gamma or log-Normal) — **generally easier under a Bayesian framework**
  - "Structural" values (0 costs, unit utilities)

- ... and of course **Partially Observed** data
  - Can have item and/or unit non-response
  - Missingness may occur in either or both benefits/costs
  - The missingness mechanisms may also be correlated
  - Focus in decision-making, not inference — **Bayesian approach particularly suited for this!**

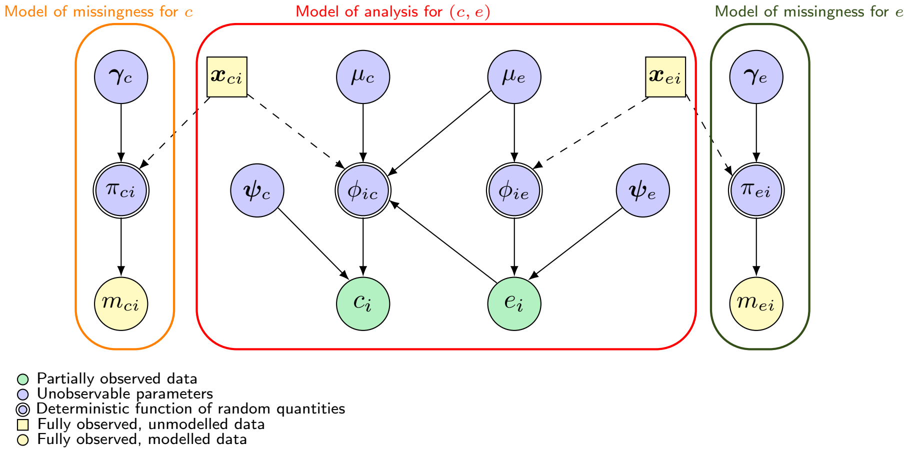
MCAR (e, c)



- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

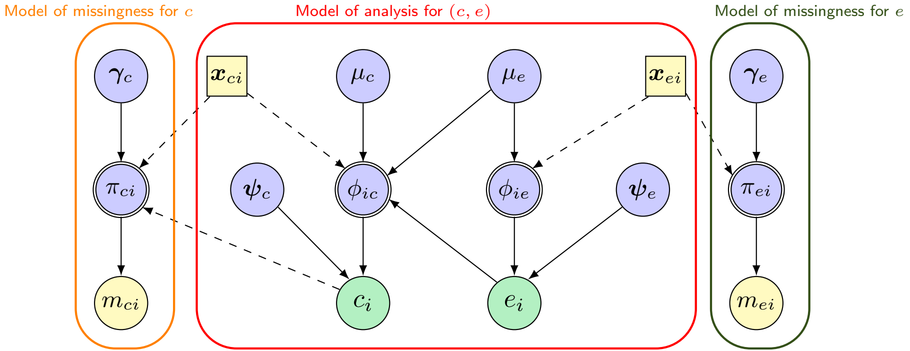
- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$; $\quad \text{logit}(\pi_{ei}) = \gamma_{e0}$
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$; $\quad \text{logit}(\pi_{ci}) = \gamma_{c0}$

MAR (e, c)



- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$; $\quad \text{logit}(\pi_{ei}) = \gamma_{e0} + \sum_{k=1}^{K} \gamma_{ek} x_{eik}$
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$; $\quad \text{logit}(\pi_{ci}) = \gamma_{c0} + \sum_{h=1}^{H} \gamma_{ch} x_{cih}$

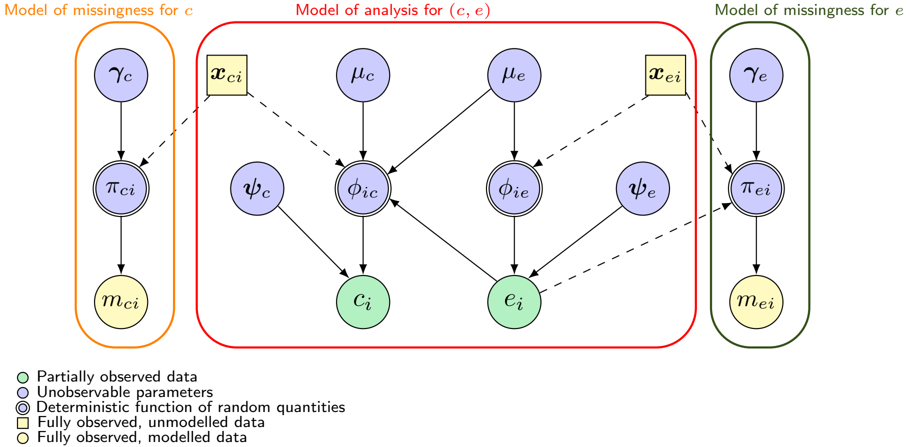
MAR e; MNAR c



- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

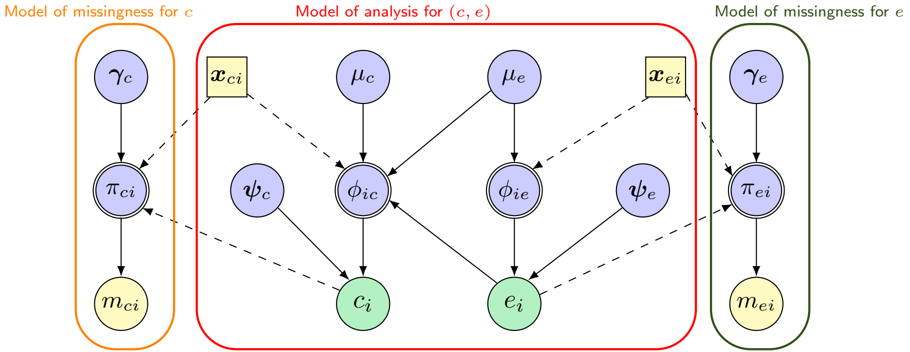
- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$; $\text{logit}(\pi_{ei}) = \gamma_{e0} + \sum_{k=1}^{K} \gamma_{ek} x_{eik}$
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$; $\text{logit}(\pi_{ci}) = \gamma_{c0} + \sum_{h=1}^{H} \gamma_{ch} x_{cih} + \gamma_{cH+1} c_i$

MNAR e; MAR c



- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$; $\text{logit}(\pi_{ei}) = \gamma_{e0} + \sum_{k=1}^{K} \gamma_{ek} x_{eik} + \gamma_{eK+1} e_i$
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$; $\text{logit}(\pi_{ci}) = \gamma_{c0} + \sum_{h=1}^{H} \gamma_{ch} x_{cih}$

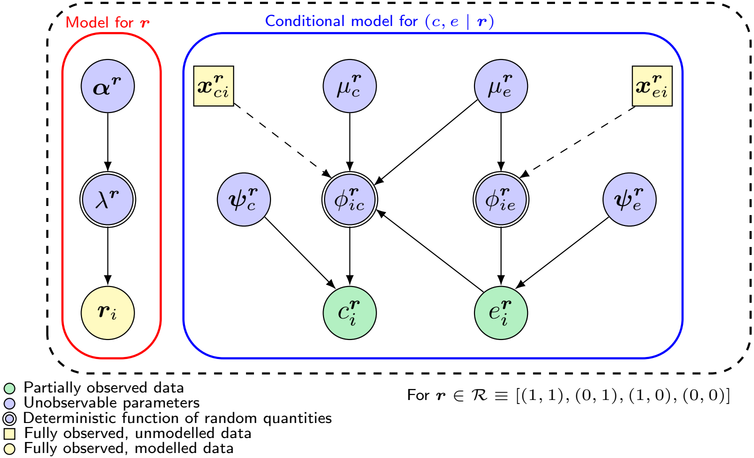
MNAR (e, c)



- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$; $\quad \text{logit}(\pi_{ei}) = \gamma_{e0} + \sum_{k=1}^{K} \gamma_{ek} x_{eik} + \gamma_{eK+1} e_i$
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$; $\quad \text{logit}(\pi_{ci}) = \gamma_{c0} + \sum_{h=1}^{H} \gamma_{ch} x_{cih} + \gamma_{cH+1} c_i$

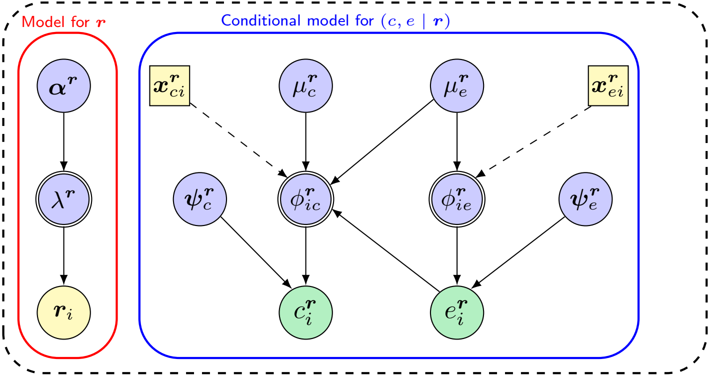
MAR (e, c)



For $\boldsymbol{r} \in \mathcal{R} \equiv [(1,1),(0,1),(1,0),(0,0)]$

- $\boldsymbol{r}_i = (m_{ei}, m_{ci}) \sim \text{Multinomial}(\boldsymbol{\lambda}^{\boldsymbol{r}})$; $\mu_e = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu_e^{\boldsymbol{r}}\lambda^{\boldsymbol{r}}$; $\mu_c = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu_c^{\boldsymbol{r}}\lambda^{\boldsymbol{r}}$
- CC restriction: $\mu_e^{(0,1)} = \mu_e^{(0,0)} = \mu_e^{(1,1)}$; $\mu_c^{(1,0)} = \mu_c^{(0,0)} = \mu_c^{(1,1)}$

MAR e; MNAR c

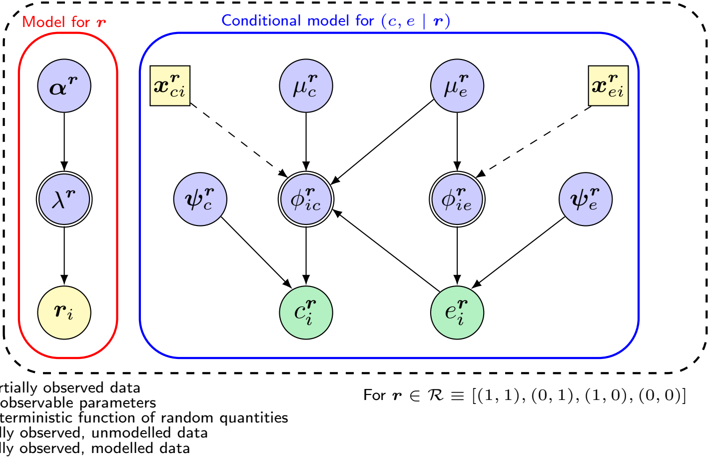


- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

For $\boldsymbol{r} \in \mathcal{R} \equiv [(1,1),(0,1),(1,0),(0,0)]$

- $\boldsymbol{r}_i = (m_{ei}, m_{ci}) \sim \text{Multinomial}(\boldsymbol{\lambda}^{\boldsymbol{r}})$; $\quad \mu_e = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu_e^{\boldsymbol{r}}\lambda^{\boldsymbol{r}}$; $\quad \mu_c = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu_c^{\boldsymbol{r}}\lambda^{\boldsymbol{r}}$
- CC restriction + SP: $\mu_e^{(0,1)} = \mu_e^{(0,0)} = \mu_e^{(1,1)}$; $\quad \mu_c^{(1,0)} = \mu_c^{(0,0)} = \mu_c^{(1,1)} + \Delta_c$

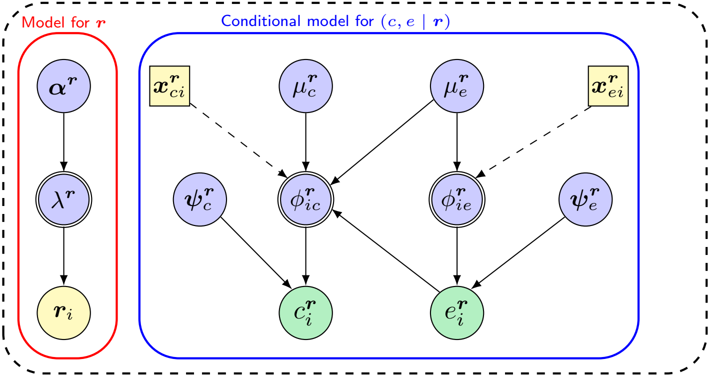
MNAR e; MAR c



For $\boldsymbol{r} \in \mathcal{R} \equiv [(1,1),(0,1),(1,0),(0,0)]$

- $\boldsymbol{r}_i = (m_{ei}, m_{ci}) \sim \text{Multinomial}(\boldsymbol{\lambda}^{\boldsymbol{r}})$; $\quad \mu_e = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu_e^{\boldsymbol{r}}\lambda^{\boldsymbol{r}}$; $\quad \mu_c = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu_c^{\boldsymbol{r}}\lambda^{\boldsymbol{r}}$
- CC restriction + SP: $\mu_e^{(0,1)} = \mu_e^{(0,0)} = \mu_e^{(1,1)} + \Delta_e$; $\quad \mu_c^{(1,0)} = \mu_c^{(0,0)} = \mu_c^{(1,1)}$

MNAR (e, c)



- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

For $\boldsymbol{r} \in \mathcal{R} \equiv [(1,1),(0,1),(1,0),(0,0)]$

- $\boldsymbol{r}_i = (m_{ei}, m_{ci}) \sim$ Multinomial$(\boldsymbol{\lambda}^{\boldsymbol{r}})$; $\quad \mu_e = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu^{\boldsymbol{r}}_e \lambda^{\boldsymbol{r}}$; $\quad \mu_c = \sum_{\boldsymbol{r}\in\mathcal{R}} \mu^{\boldsymbol{r}}_c \lambda^{\boldsymbol{r}}$
- CC restriction + SP: $\mu_e^{(0,1)} = \mu_e^{(0,0)} = \mu_e^{(1,1)} + \Delta_e$; $\quad \mu_c^{(1,0)} = \mu_c^{(0,0)} = \mu_c^{(1,1)} + \Delta_c$

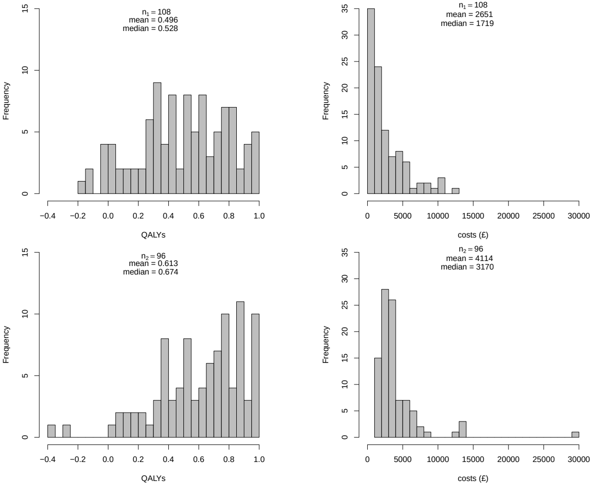
Example: PBS trial

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

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

Time	**Control** (n_1=136)		**Intervention** (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%)	

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



- In reality, the data have a longitudinal nature and particularly in the presence of missing data we have several advantages in fully exploiting it
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$

- In reality, the data have a longitudinal nature and particularly in the presence of missing data we have several advantages in fully exploiting it
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$

- Can use **pattern mixture model**

1. Factorise $p(\boldsymbol{y}, \boldsymbol{r})$ into $p(\boldsymbol{y}^{\boldsymbol{r}}_{obs}, \boldsymbol{r})$ and $p(\boldsymbol{y}^{\boldsymbol{r}}_{mis} \mid \boldsymbol{y}^{\boldsymbol{r}}_{obs}, \boldsymbol{r})$
2. Integrate out $\boldsymbol{y}^{\boldsymbol{r}}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ and estimate the means of $\boldsymbol{y}^{\boldsymbol{r}}_{obs}$
3. Identify the means of $\boldsymbol{y}^{\boldsymbol{r}}_{mis}$ using:
   - The mean estimates of $\boldsymbol{y}^{\boldsymbol{r}}_{obs}$
   - Sensitivity parameters $\boldsymbol{\Delta} = (\Delta_u, \Delta_c)$

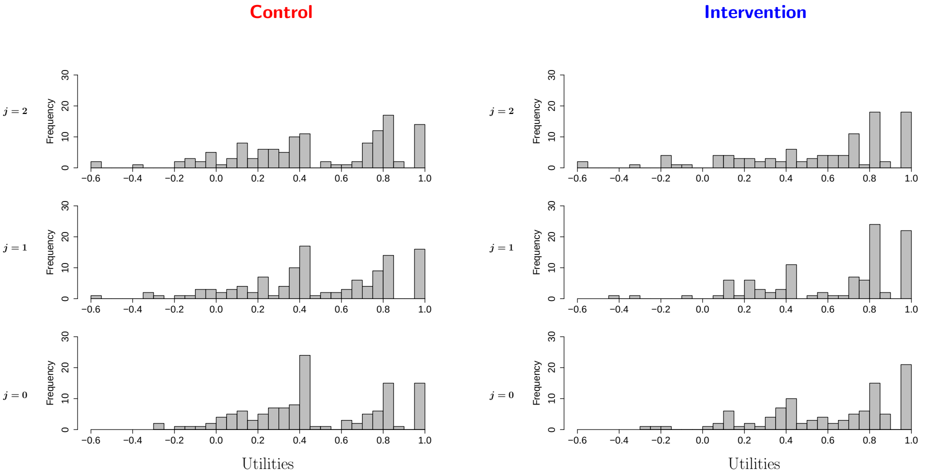
- In reality, the data have a longitudinal nature and particularly in the presence of missing data we have several advantages in fully exploiting it
  - Account for time dependence between outcomes $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$
  - Use all available utility/cost data in each pattern $\boldsymbol{r}_{ij} = (r^u_{ij}, r^c_{ij})$

- Can use **pattern mixture model**

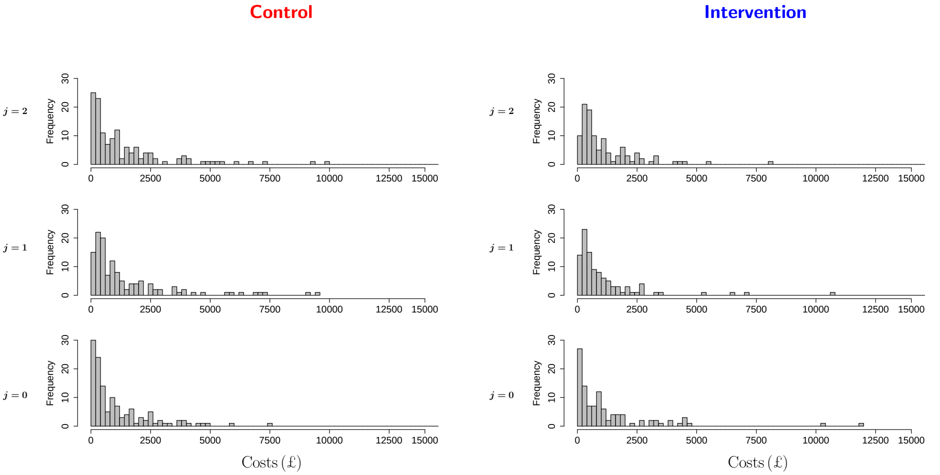
  1. Factorise $p(\boldsymbol{y}, \boldsymbol{r})$ into $p(\boldsymbol{y}^{\boldsymbol{r}}_{obs}, \boldsymbol{r})$ and $p(\boldsymbol{y}^{\boldsymbol{r}}_{mis} \mid \boldsymbol{y}^{\boldsymbol{r}}_{obs}, \boldsymbol{r})$
  2. Integrate out $\boldsymbol{y}^{\boldsymbol{r}}_{mis}$ from $p(\boldsymbol{y}, \boldsymbol{r})$ and estimate the means of $\boldsymbol{y}^{\boldsymbol{r}}_{obs}$
  3. Identify the means of $\boldsymbol{y}^{\boldsymbol{r}}_{mis}$ using:
     - The mean estimates of $\boldsymbol{y}^{\boldsymbol{r}}_{obs}$
     - Sensitivity parameters $\boldsymbol{\Delta} = (\Delta_u, \Delta_c)$

- Assess the robustness of the results to plausible MNAR scenarios using different informative priors on $\boldsymbol{\Delta}$

Example: PBS trial

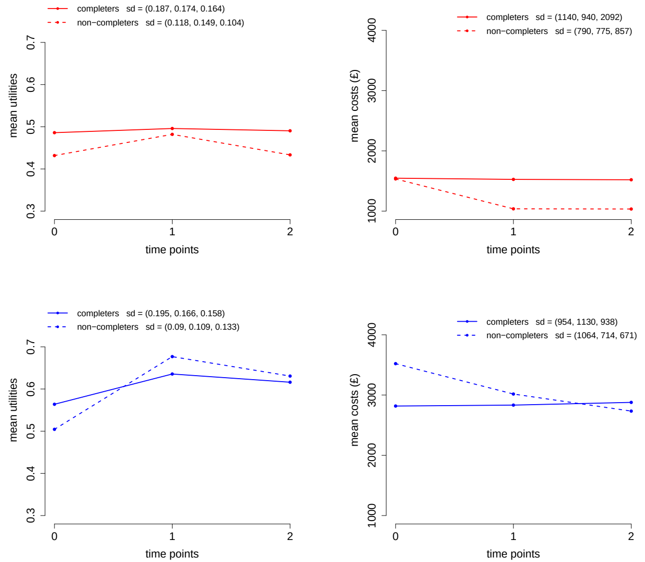


Example: PBS trial



	Control ($t = 1$)							Intervention ($t = 2$)							
	u_0	c_0	u_1	c_1	u_2	c_2	n_1^r	u_0	c_0	u_1	c_1	u_2	c_2	n_2^r	
$\boldsymbol{r}$	1	1	1	1	1	1	**108**	1	1	1	1	1	1	**96**	$\rightarrow \boldsymbol{r} = 1$
mean	0.678	1546	0.684	1527	0.680	1520		0.726	2818	0.771	2833	0.759	2878		
$\boldsymbol{r}$	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		
$\boldsymbol{r}$	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		
$\boldsymbol{r}$	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		
$\boldsymbol{r}$	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		$\boldsymbol{r} \neq 1$
$\boldsymbol{r}$	1	1	0	0	0	0	**4**	1	1	0	0	0	0	**4**	
mean	0.434	1528	–	–	–	–		0.760	3086	–	–	–	–		
$\boldsymbol{r}$	0	1	0	1	1	1	**2**	0	1	0	1	1	1	**0**	
mean	–	595	–	397	0.483	69		–	–	–	–	–	–		
$\boldsymbol{r}$	1	1	1	1	0	0	**2**	1	1	1	1	0	0	**0**	
mean	0.743	1434	0.705	1606	–	–		–	–	–	–	–	–		
$\boldsymbol{r}$	1	1	0	1	0	1	**3**	1	1	0	1	0	1	**0**	
mean	0.726	1510	–	432	–	976		–	–	–	–	–	–		

Example: PBS trial



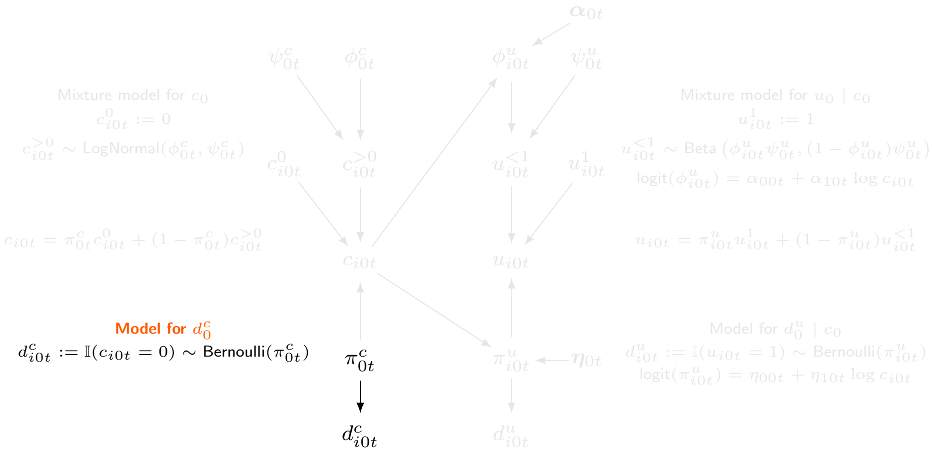
- Fit model to completers $\boldsymbol{r} = \mathbf{1}$ and joint set of all other patterns $\boldsymbol{r} \neq \mathbf{1}$ separately for $t = 1, 2$

- 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 $p(u_{ij} \mid c_{ij}, u_{ij-1})$

- 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 $u^{\star}_{ij} = \dfrac{u_{ij} - \min(\boldsymbol{u}_j)}{\max(\boldsymbol{u}_j) - \min(\boldsymbol{u}_j)}$
  - **LogNormal** distributions for c_{ij}

- 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 $u^\star_{ij} = \dfrac{u_{ij} - \min(\boldsymbol{u}_j)}{\max(\boldsymbol{u}_j) - \min(\boldsymbol{u}_j)}$
  - **LogNormal** distributions for c_{ij}
- Allow for structural **ones** in u_{ij} and **zeros** in c_{ij}
  - Define $d^u_{ij} := \mathbb{I}(u_{ij}=1)$ and $d^c_{ij} := \mathbb{I}(c_{ij}=0)$
  - Use a hurdle model to account for mixture of patients within the groups

- At $j = 0$


α_{0t}
ψ^c_{0t}
ϕ^c_{0t}
ϕ^u_{i0t}
ψ^u_{0t}
Mixture model for c_0
$c^0_{i0t} := 0$
$c^{>0}_{i0t} \sim$ LogNormal$(\phi^c_{0t}, \psi^c_{0t})$
c^0_{i0t}
$c^{>0}_{i0t}$
$u^{<1}_{i0t}$
u^1_{i0t}
Mixture model for $u_0 \mid c_0$
$u^1_{i0t} := 1$
$u^{<1}_{i0t} \sim$ Beta $(\phi^u_{i0t}\psi^u_{0t}, (1-\phi^u_{i0t})\psi^u_{0t})$
logit$(\phi^u_{i0t}) = \alpha_{00t} + \alpha_{10t} \log c_{i0t}$
$c_{i0t} = \pi^c_{0t} c^0_{i0t} + (1-\pi^c_{0t}) c^{>0}_{i0t}$
$u_{i0t} = \pi^u_{i0t} u^1_{i0t} + (1-\pi^u_{i0t}) u^{<1}_{i0t}$
c_{i0t}
u_{i0t}
Model for d^c_0
$d^c_{i0t} := \mathbb{I}(c_{i0t} = 0) \sim$ Bernoulli(π^c_{0t})
π^c_{0t}
π^u_{i0t}
η_{0t}
Model for $d^u_0 \mid c_0$
$d^u_{i0t} := \mathbb{I}(u_{i0t} = 1) \sim$ Bernoulli(π^u_{i0t})
logit$(\pi^u_{i0t}) = \eta_{00t} + \eta_{10t} \log c_{i0t}$
d^c_{i0t}
d^u_{i0t}

- At $j = 0$

Mixture model for c_0

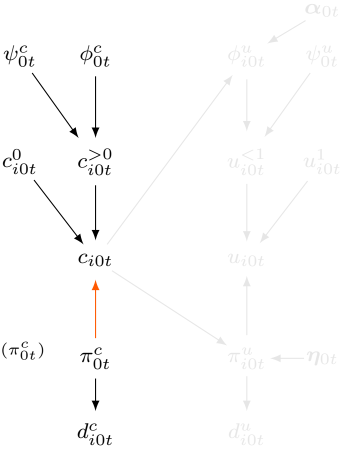
$c^0_{i0t} := 0$

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

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

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 $d^u_0 \mid c_0$

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

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

Modelling

Gabrio et al. JRSS/A (2019). doi:10.1111/rssa.12522

- At $j = 0$

Mixture model for c_0

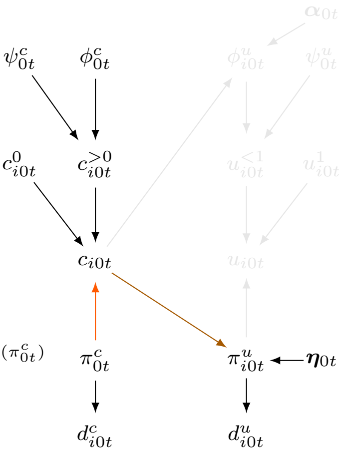
$c^0_{i0t} := 0$

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

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

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 $d^u_0 \mid c_0$

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

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

- At $j = 0$

Mixture model for c_0

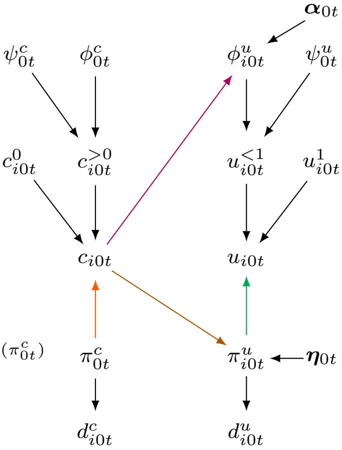
$c^0_{i0t} := 0$

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

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

Model for d^c_0

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



Mixture model for $u_0 \mid c_0$

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

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

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

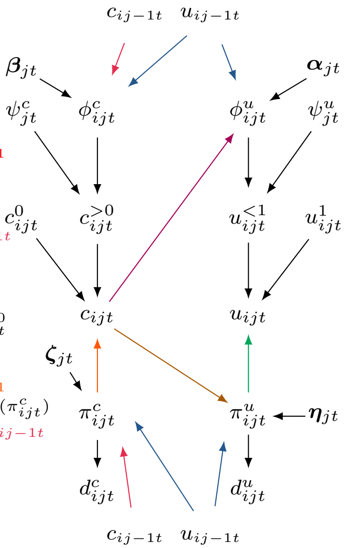
$u_{i0t} = \pi^u_{i0t} u^1_{i0t} + (1 - \pi^u_{i0t}) u^{<1}_{i0t}$

Model for $d^u_0 \mid c_0$

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

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

- At $j = 1, 2$



Mixture model for $c_j \mid c_{j-1}, u_{j-1}$

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

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

$$\log(\phi^c_{ijt}) = \beta_{0jt} + \beta_{1jt} \log c_{ij-1t} + \beta_{2jt} u_{ij-1t}$$

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

Model for $d^c_j \mid c_{j-1}, u_{j-1}$

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

$$\text{logit}(\pi^c_{ijt}) = \zeta_{0jt} + \zeta_{1jt} \log c_{ij-1t} + \zeta_{2jt} u_{ij-1t}$$

Mixture model for $u_j \mid c_j, u_{j-1}$

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

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

$$\text{logit}(\phi^u_{ijt}) = \alpha_{0jt} + \alpha_{1jt} \log c_{ijt} + \alpha_{2jt} u_{ij-1t}$$

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

Model for $d^u_j \mid c_j, u_{j-1}$

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

$$\text{logit}(\pi^u_{ijt}) = \eta_{0jt} + \eta_{1jt} \log c_{ijt} + \eta_{2jt} u_{ij-1t}$$

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

Identifying restrictions and sensitivity parameters

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

Identifying restrictions and sensitivity parameters

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

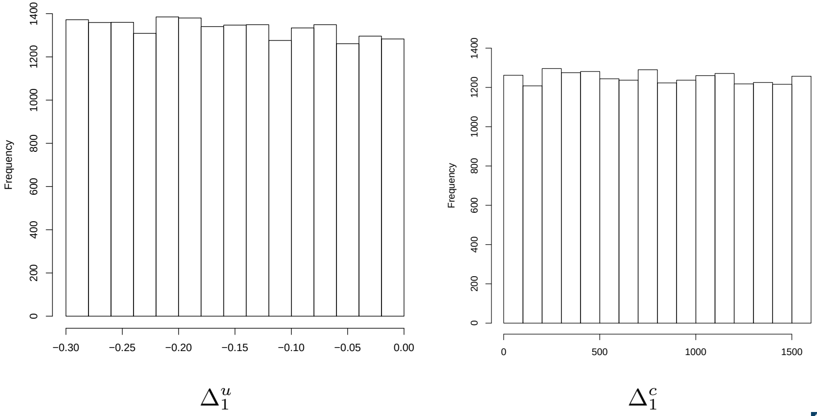
- Use Monte Carlo integration to derive the mean estimates $\mathsf{E}[\boldsymbol{y}^{\boldsymbol{r}}_{obs}]$
- Identify the mean estimates $\mathsf{E}[\boldsymbol{y}^{\boldsymbol{r}}_{mis}] = \mathsf{E}[\boldsymbol{y}^{\boldsymbol{r}}_{obs} + \boldsymbol{\Delta}_j]$
- Compute weighted average across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

Identifying restrictions and sensitivity parameters

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

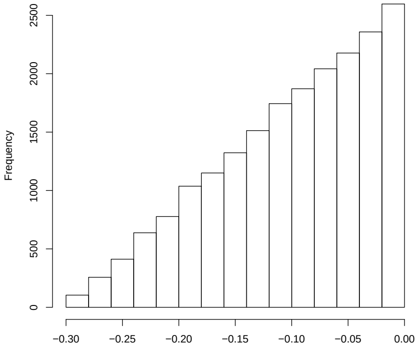
Priors on sensitivity parameters

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

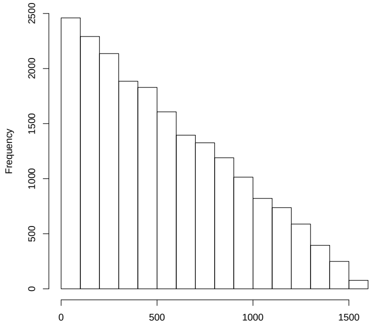


Priors on sensitivity parameters

- 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 $\boldsymbol{\Delta}^{\text{flat}}$



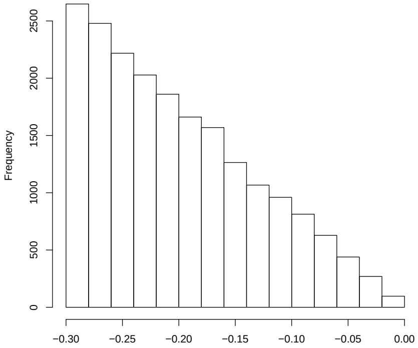
Δ_1^u



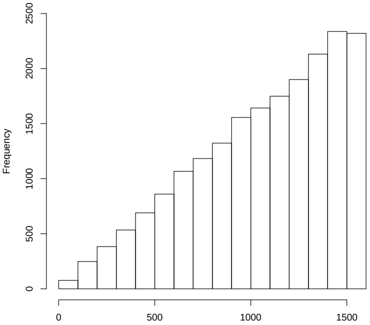
Δ_1^c

Priors on sensitivity parameters

- 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}

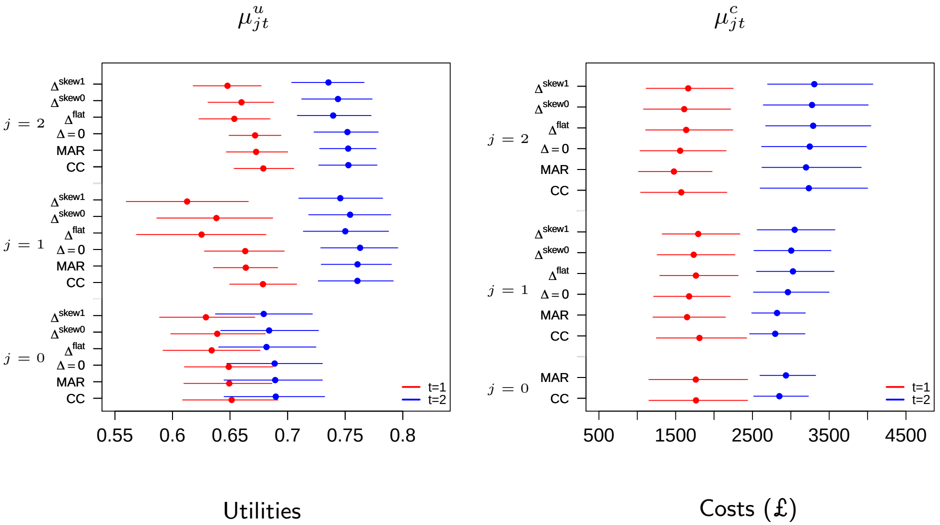


Δ_1^u

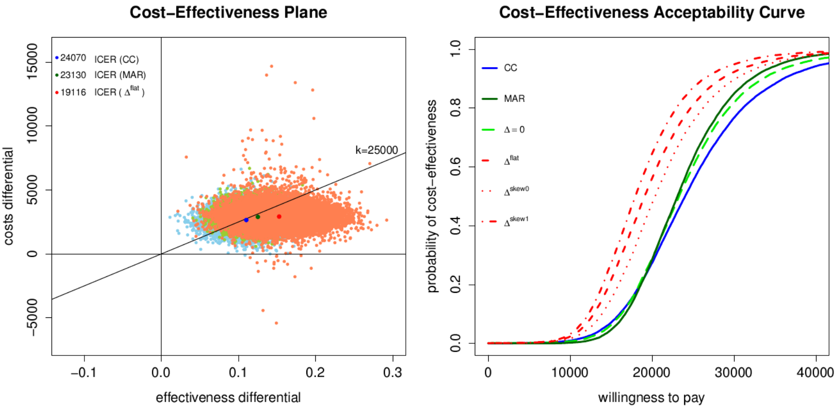


Δ_1^c

Results: means utilities and costs



Cost-effectiveness analysis



1. **Flexibility** of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses a modular structure to account for complexities that may bias inferences and mislead the economic assessment
   - Can extend complex models to a longitudinal framework in a relatively easy way

1. **Flexibility** of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses a modular structure to account for complexities that may bias inferences and mislead the economic assessment
   - Can extend complex models to a longitudinal framework in a relatively easy way
2. **Extension** of standard "imputation methods"
   - Performs the estimation and imputation tasks simultaneously
   - Fits joint models for missing data in a relatively easy way
   - Uses probabilistic appraoches that can be implemented in standard software (e.g. `OpenBUGS` or `JAGS`)

1. **Flexibility** of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses a modular structure to account for complexities that may bias inferences and mislead the economic assessment
   - Can extend complex models to a longitudinal framework in a relatively easy way
2. **Extension** of standard "imputation methods"
   - Performs the estimation and imputation tasks simultaneously
   - Fits joint models for missing data in a relatively easy way
   - Uses probabilistic appraoches 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

Thank you!