Methoden voor het omgaan met ontbrekende gegevens bij de beoordeling van gezondheidstechnologie

(Thanks/blame to Google Translate)

Andrea Gabrio

Maastricht University
Faculty of Health Medicine and Life Sciences
Department of Methodology and Statistics

email: a.gabrio@maastrichtuniversity.nl
website: https://agabrioblog.onrender.com
GitHub page: https://github.com/AnGabrio

HTA research seminar series, VU, Amsterdam

Friday 17 June 2022

Outline

Part 1

Introduction to modelling in HTA

Back to Table of content

Disclaimer...

Before I begin:

Disclaimer...

Before I begin:

- **Statistics** is an *all-encompassing discipline*:

  Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields.

Disclaimer...

Before I begin:

- **Statistics** is an *all-encompassing discipline*:

  Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields.

- The thought process of statistics is unique and it is difficult for others to understand its contribution to the fields:

  The **success** *of Statistics relies on developing skills to make it easier for colleagues in other disciplines to appreciate our contribution.*

Disclaimer...

Before I begin:

- **Statistics** is an *all-encompassing discipline*:

  Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields.

- The thought process of statistics is unique and it is difficult for others to understand its contribution to the fields:

  The **success** *of Statistics relies on developing skills to make it easier for colleagues in other disciplines to appreciate our contribution.*

- My **personal** view of the world:

  Statisticians should be in charge of everything.

Disclaimer...

Before I begin:

- **Statistics** is an *all-encompassing discipline*:
  Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields.
- The thought process of statistics is unique and it is difficult for others to understand its contribution to the fields:
  The **success** *of Statistics relies on developing skills to make it easier for colleagues in other disciplines to appreciate our contribution.*
- My **personal** view of the world:
  Statisticians should be in charge of everything.
- So I probably will be very annoying throughout the presentation [1]

[1] But luckily no non-Statistician has been harmed in the making of this slides

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!)

Health technology assessment (HTA)

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

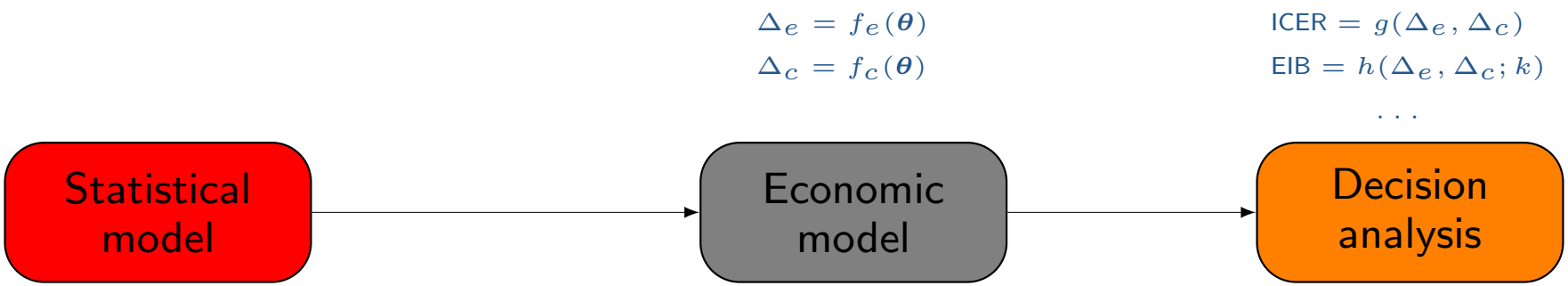
$$\Delta_e = f_e(\boldsymbol{\theta})$$
$$\Delta_c = f_c(\boldsymbol{\theta})$$

- 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



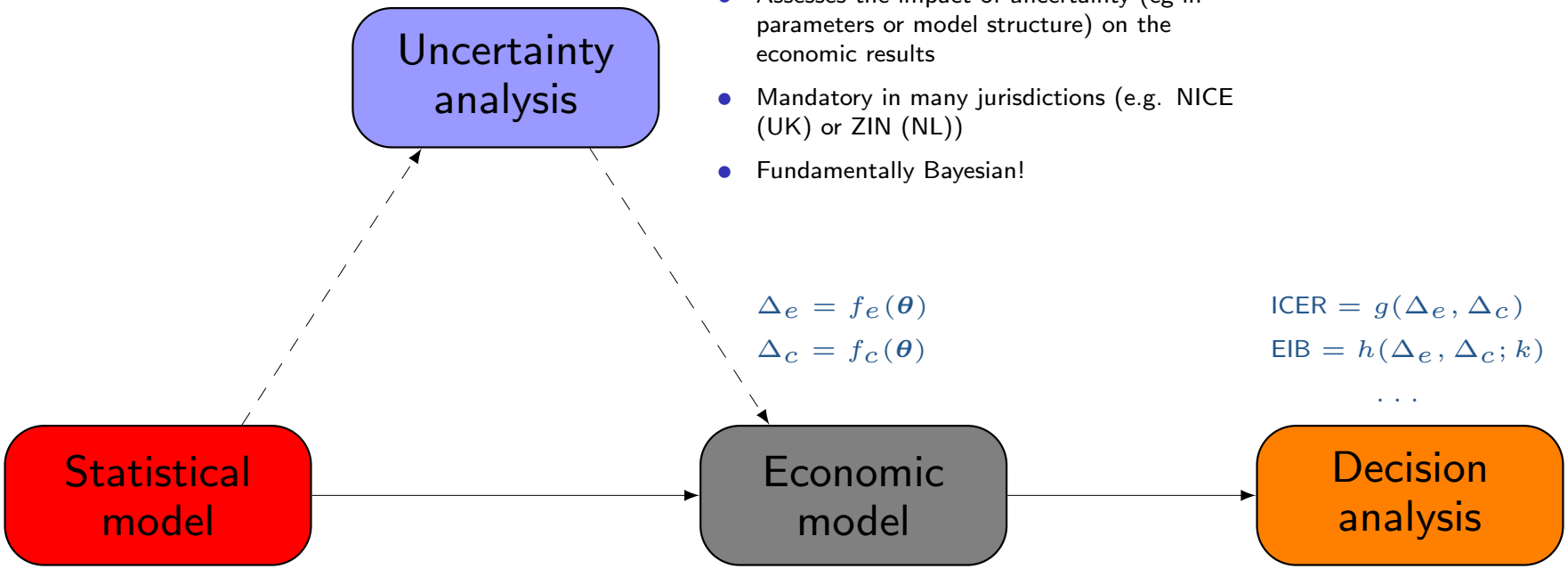
- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

- 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

Health technology assessment (HTA)

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 (e.g. NICE (UK) or ZIN (NL))
- Fundamentally Bayesian!

- 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}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

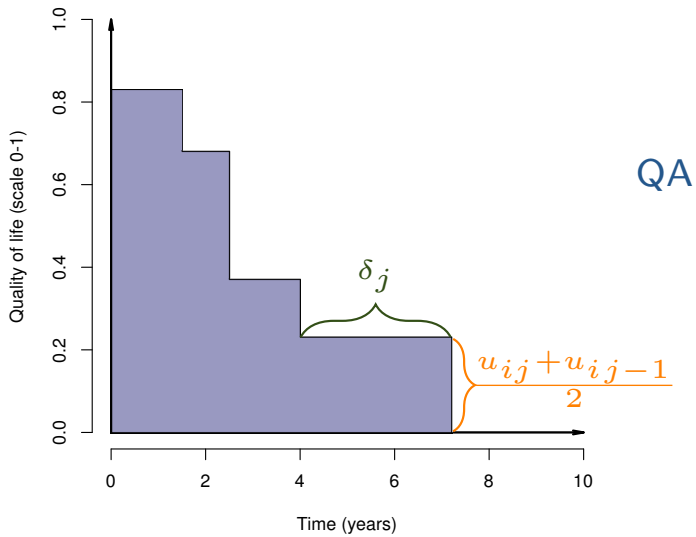
y_{ij} = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
u_{ij} = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
c_{ij} = Use of resources (drugs, hospital, GP appointments, ...)

		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

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



QALY_i = "Area under the curve"

		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

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

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

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

		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

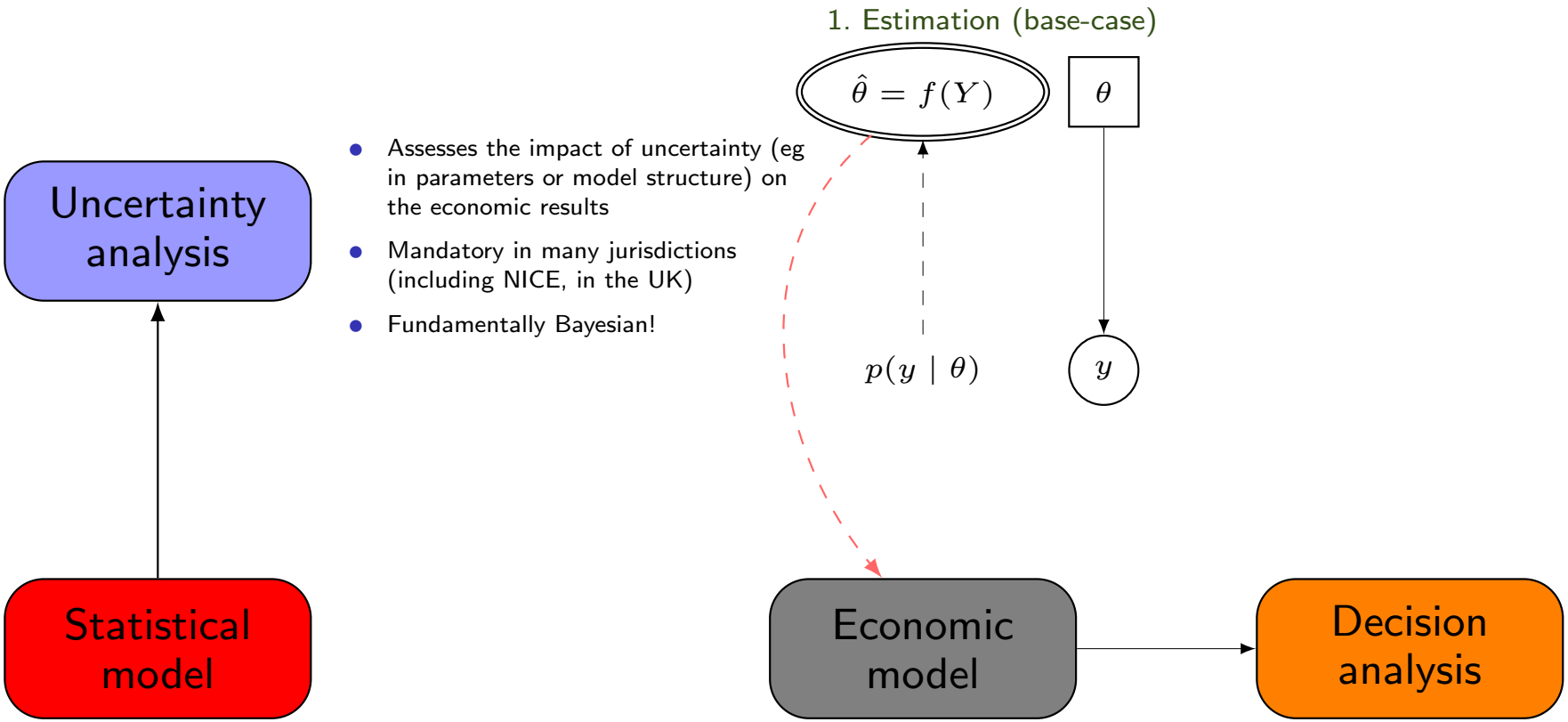
$$e_i = \sum_{j=1}^{J} (u_{ij} + u_{ij-1}) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=1}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$

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

"Standard" approach to HTA — "Two-stage"



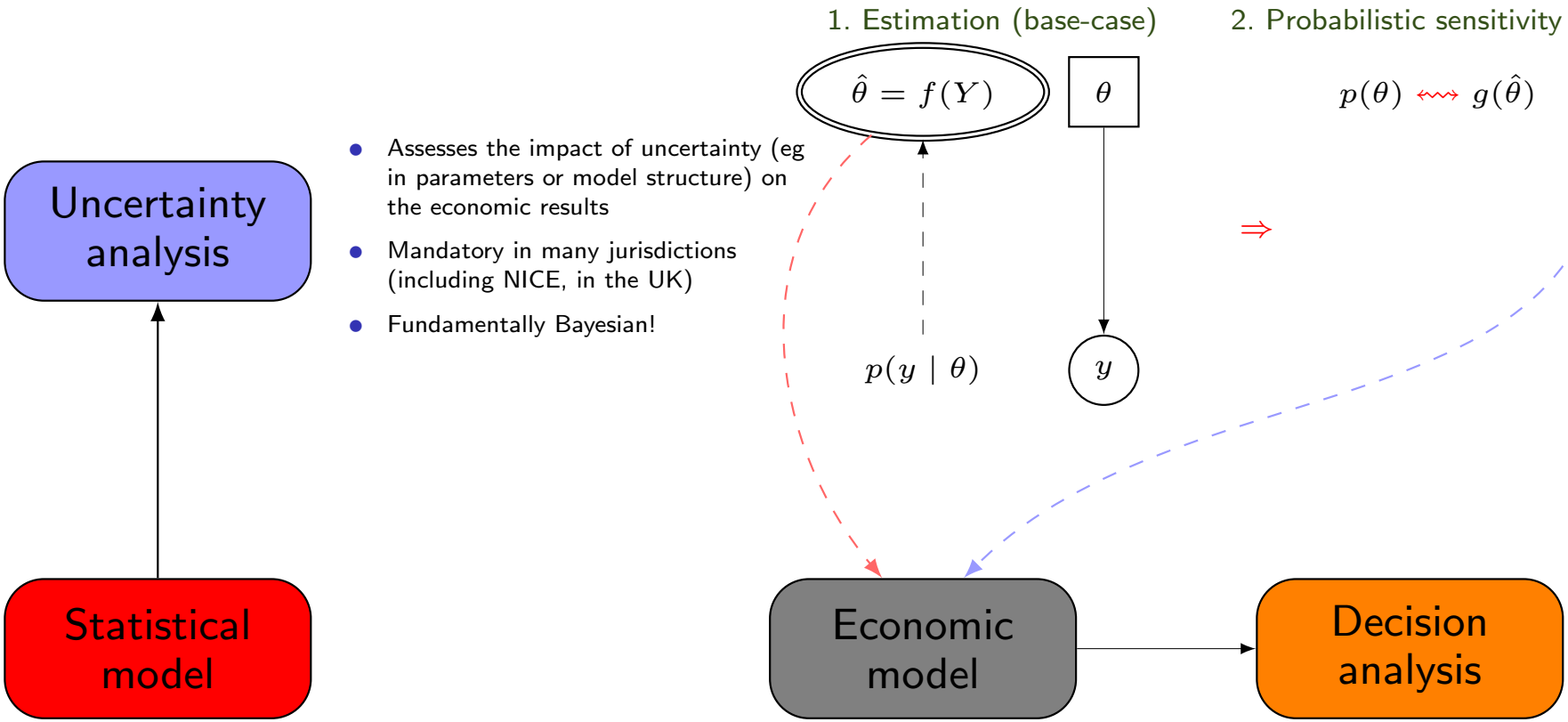
- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!

- 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

"Standard" approach to HTA — "Two-stage"



- 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!

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

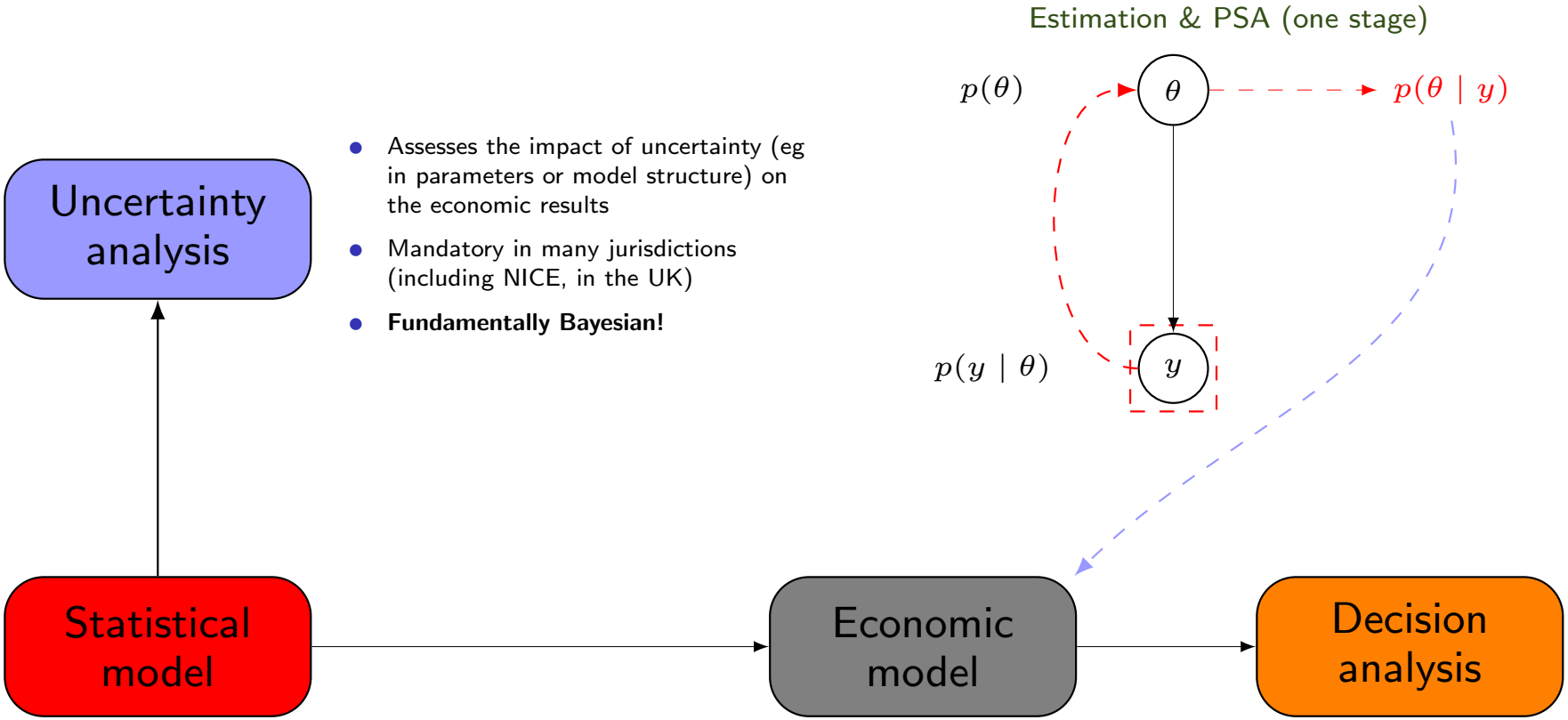
- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

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

"Two-stage approach" (Spiegelhalter, Abrams & Myles, 2004)

The Sherlock conundrum...

Bayesian approach to HTA



- 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!**

- 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

"Integrated approach" Spiegelhalter, Abrams & Myles (2004)
Baio, Berardi & Heath (2017)

Bayesians do it better...

- Potential correlation between costs & clinical benefits
  - 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!**

Bayesians do it better...

- Potential correlation between costs & clinical benefits
  - 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
  - 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**

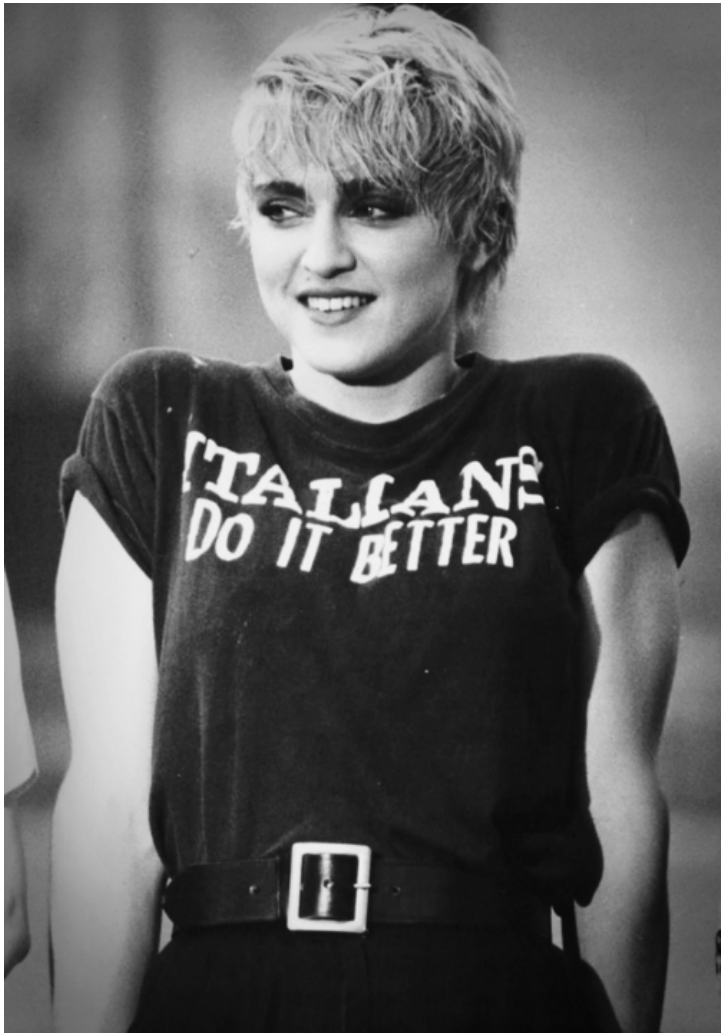
Bayesians do it better...

- Potential correlation between costs & clinical benefits
  - 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
  - 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**

- ... 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!**

Bayesians do it better...



Part 2

Missing data

Back to Table of content

Background

The problems with missing data...

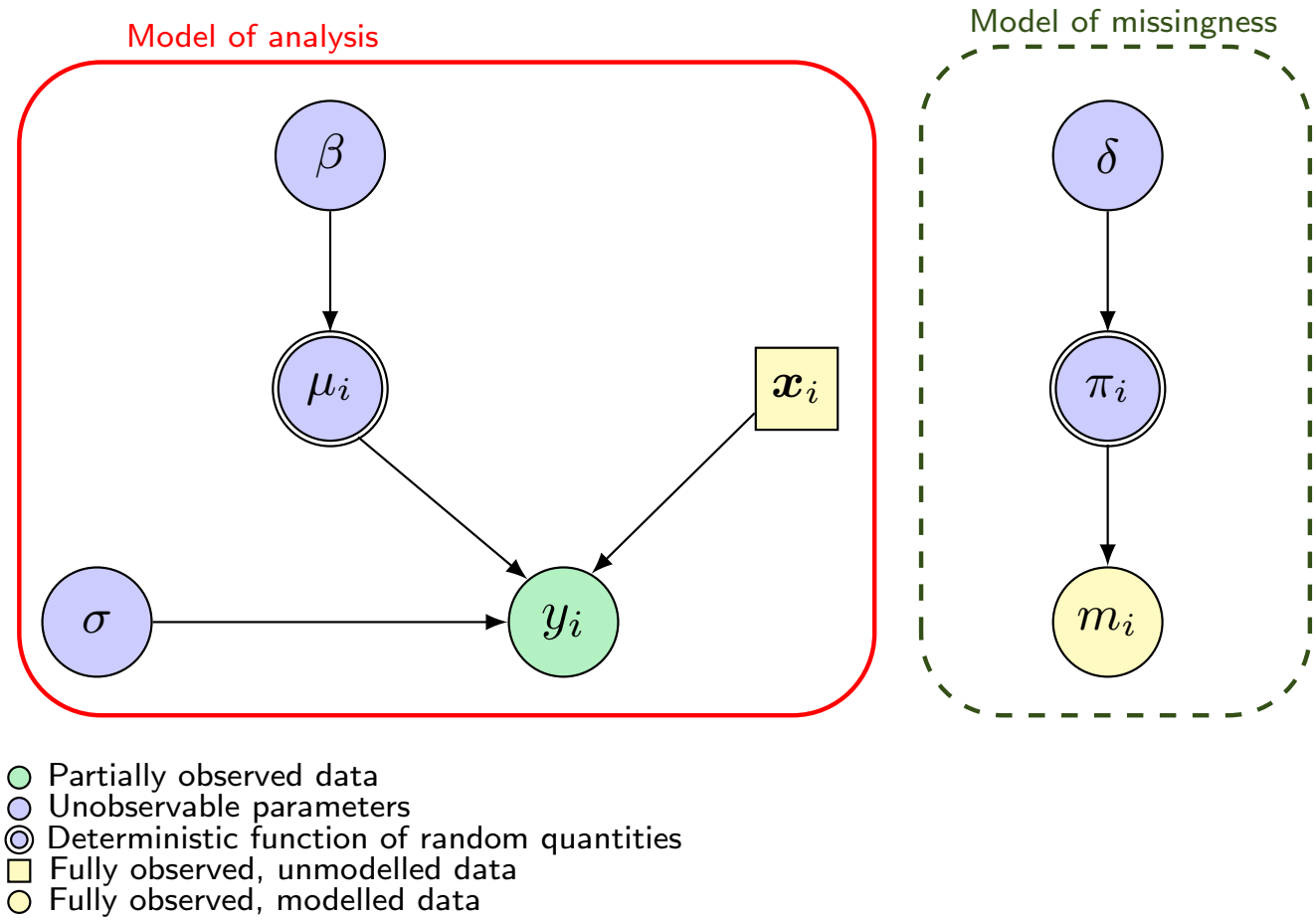
- We plan to observe n_{planned} data points, but end up with a (much) lower number of observations n_{observed}
  - What is the proportion of missing data? Does it matter?...
- We typically don't know **why** the unobserved points are missing and **what** their value might have been
  - Missingness can be differential in treatment/exposure groups

Background

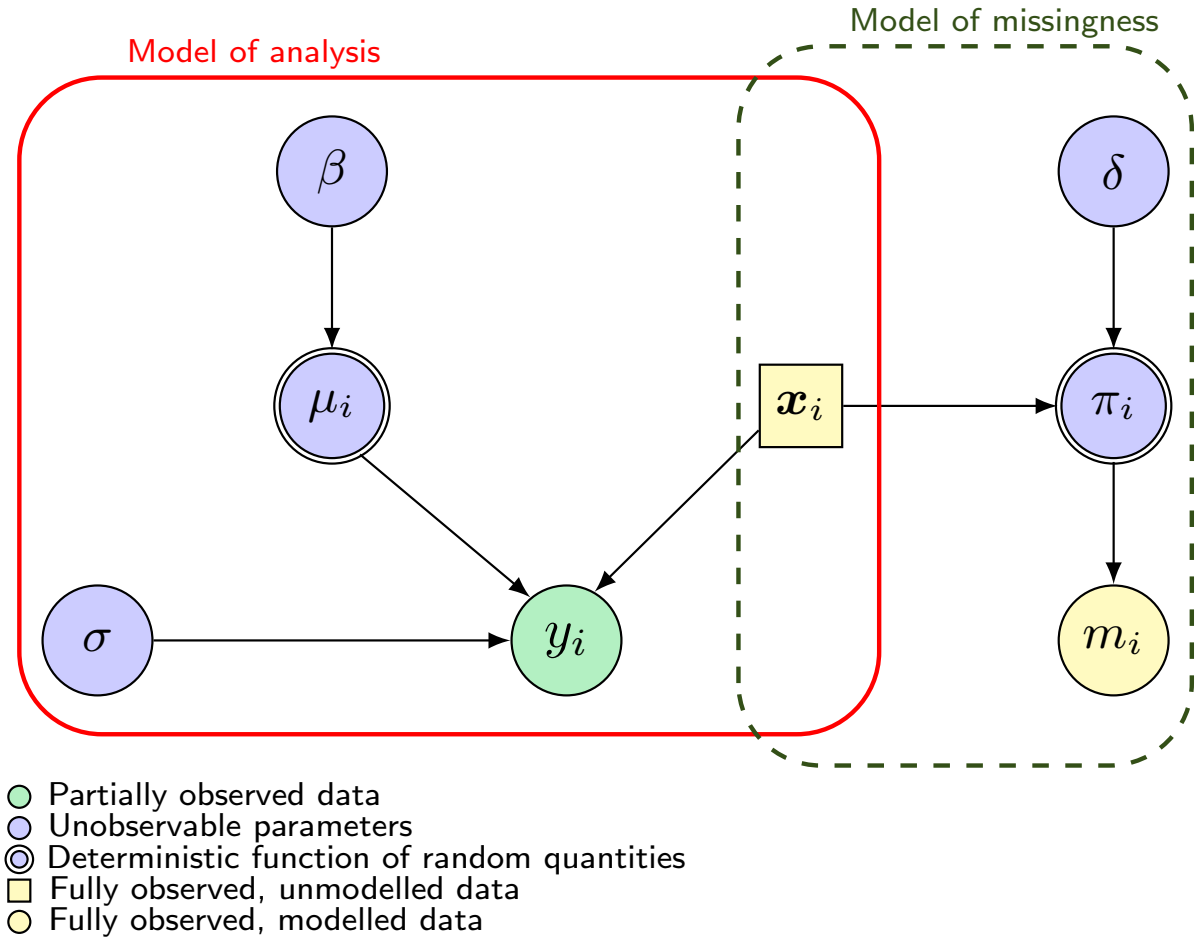
The problems with missing data...

- We plan to observe n_{planned} data points, but end up with a (much) lower number of observations n_{observed}
  - What is the proportion of missing data? Does it matter?...
- We typically don't know **why** the unobserved points are missing and **what** their value might have been
  - Missingness can be differential in treatment/exposure groups
- ... Basically, not very very much we can do about it!
  - Any modelling based on at least some **untestable** assumptions
  - Cannot check model fit to unobserved data
  - Have to accept inherent uncertainty in our analysis!

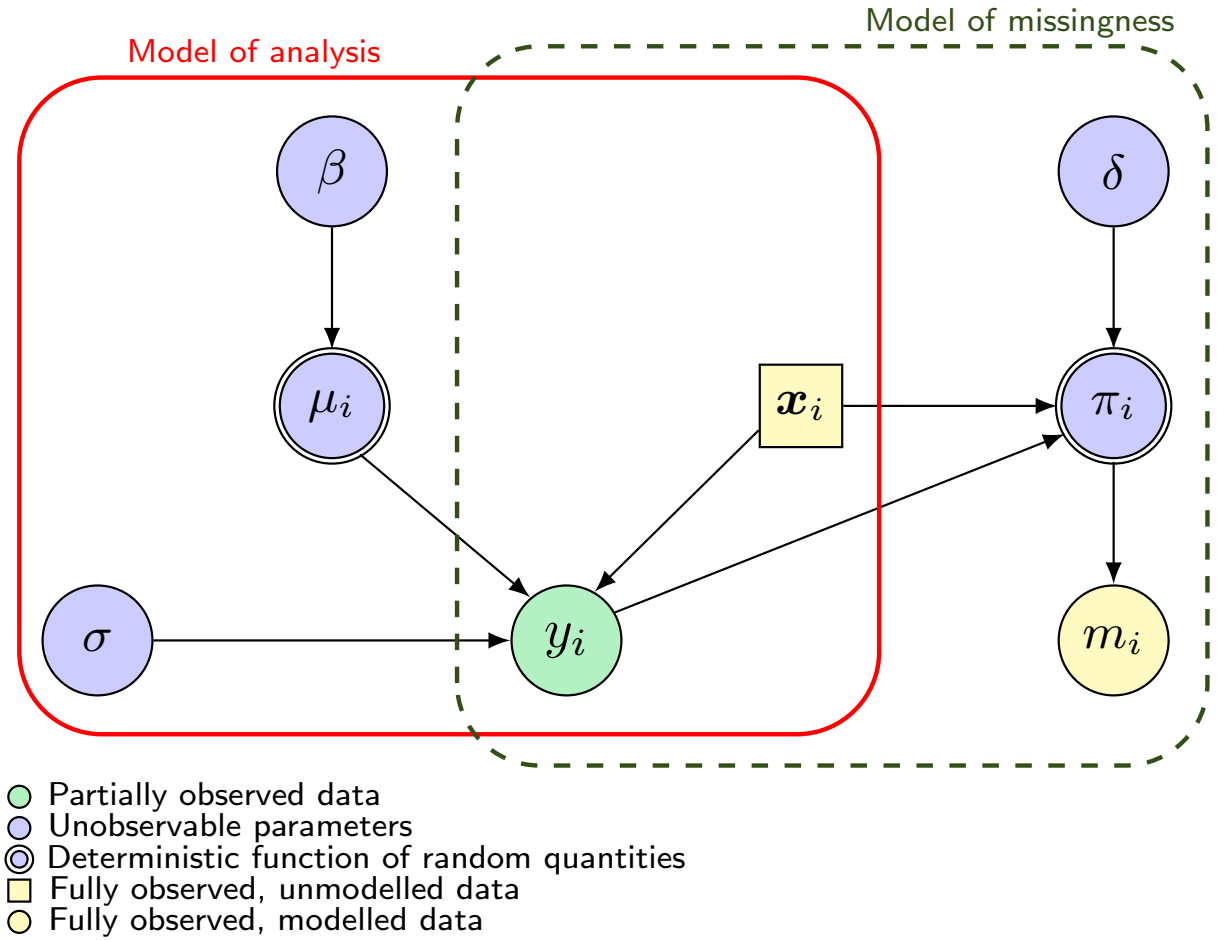
Missing data **mechanism**: Missing Completely At Random (MCAR)



- $y_i =$ Outcome subject to missingness
- $m_i = 1$ if y_i missing or 0 if y_i observed ("missingness indicator")
- $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{\mathrm{MoA}}, \boldsymbol{\theta}^{\mathrm{MoM}}\right) =$ model parameters
  - $\boldsymbol{\theta}^{\mathrm{MoA}} = (\beta, \sigma)$
  - $\boldsymbol{\theta}^{\mathrm{MoM}} = \delta$



- $y_i =$ Outcome subject to missingness
- $m_i = 1$ if y_i missing or 0 if y_i observed ("missingness indicator")
- $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{\mathrm{MoA}}, \boldsymbol{\theta}^{\mathrm{MoM}}\right) =$ model parameters
  - $\boldsymbol{\theta}^{\mathrm{MoA}} = (\beta, \sigma)$
  - $\boldsymbol{\theta}^{\mathrm{MoM}} = \delta$



- $y_i =$ Outcome subject to missingness
- $m_i = 1$ if y_i missing or 0 if y_i observed ("missingness indicator")
- $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{\mathrm{MoA}}, \boldsymbol{\theta}^{\mathrm{MoM}}\right) =$ model parameters
  - $\boldsymbol{\theta}^{\mathrm{MoA}} = (\beta, \sigma)$
  - $\boldsymbol{\theta}^{\mathrm{MoM}} = \delta$

(Bayesian) Modelling for missing data

- Effectively, need to model a bivariate outcome (y, m), depending on the model parameters

$$
\begin{aligned}
p(y, m \mid \boldsymbol{\theta}) &= p\left(y \mid m, \boldsymbol{\theta}^{\mathrm{MoA}}\right) p\left(m \mid \boldsymbol{\theta}^{\mathrm{MoM}}\right) && \text{(Pattern mixture model)} \\
&= p\left(m \mid y, \boldsymbol{\theta}^{\mathrm{MoM}}\right) p\left(y \mid \boldsymbol{\theta}^{\mathrm{MoA}}\right) && \text{(Selection model)}
\end{aligned}
$$

  - Common assumption: the two blocks of model parameters are independent (at least a priori)

(Bayesian) Modelling for missing data

- Effectively, need to model a bivariate outcome (y, m), depending on the model parameters

$$
\begin{aligned}
p(y, m \mid \boldsymbol{\theta}) &= p\left(y \mid m, \boldsymbol{\theta}^{\mathrm{MoA}}\right) p\left(m \mid \boldsymbol{\theta}^{\mathrm{MoM}}\right) && \text{(Pattern mixture model)} \\
&= p\left(m \mid y, \boldsymbol{\theta}^{\mathrm{MoM}}\right) p\left(y \mid \boldsymbol{\theta}^{\mathrm{MoA}}\right) && \text{(Selection model)}
\end{aligned}
$$

  - Common assumption: the two blocks of model parameters are independent (at least a priori)

- **Pattern mixture models**
  - Needs to model the full possible missingness "patterns" m using a marginal distribution
  - Models for data more natural
- **Selection models**
  - Models directly the marginal distribution of the observable data
  - Needs to figure out how the missingness model may be affected by it

- Decisions in HTA are often informed by within-trial CEA

- Decisions in HTA are often informed by within-trial CEA
- Missing data occur frequently in RCTs in both effectiveness and cost outcomes

- Decisions in HTA are often informed by within-trial CEA
- Missing data occur frequently in RCTs in both effectiveness and cost outcomes
- **CCA** has historically represented the standard approach in health economics

- Decisions in HTA are often informed by within-trial CEA
- Missing data occur frequently in RCTs in both effectiveness and cost outcomes
- **CCA** has historically represented the standard approach in health economics
  - Easy to implement but **inefficient** and generally inadequate for handling missingness
  - May yield **biased** inferences and lead to **incorrect** cost-effectiveness conclusions
  - Alternative approaches (e.g. **MI**) have become more popular among practitioners

- Decisions in HTA are often informed by within-trial CEA
- Missing data occur frequently in RCTs in both effectiveness and cost outcomes
- **CCA** has historically represented the standard approach in health economics
  - Easy to implement but **inefficient** and generally inadequate for handling missingness
  - May yield **biased** inferences and lead to **incorrect** cost-effectiveness conclusions
  - Alternative approaches (e.g. **MI**) have become more popular among practitioners
- Guidelines on missing data handling in CEA have started to appear in the literature
  - The analysis should be based on **plausible** assumption for the missing data mechanism
  - The choice of the method should **fit** with the assumed mechanism
  - **Sensitivity analysis** should be conducted to assess the robustness of the conclusions to alternative assumptions

Current solutions to deal **with HTA** data

Alternative strategies have been proposed to deal with the typical complexities of HTA data:

Current solutions to deal **with HTA** data

Alternative strategies have been proposed to deal with the typical complexities of HTA data:

- Seemingly unrelated regression or proper joint modelling allow to capture the **correlation** between the outcomes (Willan et al. (2004). *Health economics*, 13(5), 461-475)
- Alternative parametric distributions (e.g. Beta or Gamma) improve the model fit to **skewed** data (Basu et al. (2012). *Medical Decision Making*, 32(1), 56-69)
- Two-part regression/hurdle models deal with **spikes** in the observed data (Gabrio et al. (2019). *Statistics in medicine*, 38(8), 1399-1420)

Current solutions to deal **with HTA** data

Alternative strategies have been proposed to deal with the typical complexities of HTA data:

- Seemingly unrelated regression or proper joint modelling allow to capture the **correlation** between the outcomes (Willan et al. (2004). *Health economics*, 13(5), 461-475)
- Alternative parametric distributions (e.g. Beta or Gamma) improve the model fit to **skewed** data (Basu et al. (2012). *Medical Decision Making*, 32(1), 56-69)
- Two-part regression/hurdle models deal with **spikes** in the observed data (Gabrio et al. (2019). *Statistics in medicine*, 38(8), 1399-1420)

These methods are almost exclusively implemented at the level of QALYs/total costs to simplify the model specification. However, this strategy:

- Is **inefficient** as information from incomplete cases is lost
- Is potentially **biased**, unless the observed cases are representative of the study population (MAR)

Part 3

A longitudinal missingness model in HTA

Back to Table of content

Motivation

- 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 modelling framework to a **longitudinal** setting to handle missingness more efficiently

General notation

- Let $\boldsymbol{Y}_{ij} = (U_{ij}, C_{ij})$ be the vector of utility and cost variables available for the i-th person at the j-th time in the study
- Let $\boldsymbol{R}_{ij} = (R^u_{ij}, R^c_{ij})$ be the corresponding vector of missingness indicators, e.g. $R^u_{ij} = \mathbb{I}(U_{ij}\text{is obs})$, and let
  - $\boldsymbol{Y}_{\boldsymbol{r}} = (Y_{ij} : R_{ij} = 1)$ denote the observed responses
  - $\boldsymbol{Y}_{\bar{\boldsymbol{r}}} = (Y_{ij} : R_{ij} = 0)$ denote the missing responses

General notation

- Let $\boldsymbol{Y}_{ij} = (U_{ij}, C_{ij})$ be the vector of utility and cost variables available for the i-th person at the j-th time in the study
- Let $\boldsymbol{R}_{ij} = (R^u_{ij}, R^c_{ij})$ be the corresponding vector of missingness indicators, e.g. $R^u_{ij} = \mathbb{I}(U_{ij}\text{is obs})$, and let
  - $\boldsymbol{Y}_{\boldsymbol{r}} = (Y_{ij} : R_{ij} = 1)$ denote the observed responses
  - $\boldsymbol{Y}_{\bar{\boldsymbol{r}}} = (Y_{ij} : R_{ij} = 0)$ denote the missing responses
- Define a joint probability model $p(\boldsymbol{y}, \boldsymbol{r} \mid \boldsymbol{\omega})$ under a benchmark missing data assumption:
  - Often this is **MAR**, i.e. when $p(\boldsymbol{r} \mid \boldsymbol{y}, \boldsymbol{\psi}) = p(\boldsymbol{r} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{\psi})$

General notation

- Let $\boldsymbol{Y}_{ij} = (U_{ij}, C_{ij})$ be the vector of utility and cost variables available for the i-th person at the j-th time in the study
- Let $\boldsymbol{R}_{ij} = (R^u_{ij}, R^c_{ij})$ be the corresponding vector of missingness indicators, e.g. $R^u_{ij} = \mathbb{I}(U_{ij}\text{is obs})$, and let
  - $\boldsymbol{Y}_{\boldsymbol{r}} = (Y_{ij} : R_{ij} = 1)$ denote the observed responses
  - $\boldsymbol{Y}_{\bar{\boldsymbol{r}}} = (Y_{ij} : R_{ij} = 0)$ denote the missing responses
- Define a joint probability model $p(\boldsymbol{y}, \boldsymbol{r} \mid \boldsymbol{\omega})$ under a benchmark missing data assumption:
  - Often this is **MAR**, i.e. when $p(\boldsymbol{r} \mid \boldsymbol{y}, \boldsymbol{\psi}) = p(\boldsymbol{r} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{\psi})$
  - If we use an individual summary function of $\boldsymbol{r}$, e.g. $d(\boldsymbol{r})$, such that $p(\boldsymbol{r} \mid \boldsymbol{y}, d(\boldsymbol{r}), \boldsymbol{\kappa}) = p(\boldsymbol{r} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\kappa})$, then this defines a **Partial MAR** assumption, with $\boldsymbol{\kappa} \in \boldsymbol{\psi}$

General notation

- Let $\boldsymbol{Y}_{ij} = (U_{ij}, C_{ij})$ be the vector of utility and cost variables available for the i-th person at the j-th time in the study
- Let $\boldsymbol{R}_{ij} = (R^u_{ij}, R^c_{ij})$ be the corresponding vector of missingness indicators, e.g. $R^u_{ij} = \mathbb{I}(U_{ij}\text{is obs})$, and let
  - $\boldsymbol{Y}_{\boldsymbol{r}} = (Y_{ij} : R_{ij} = 1)$ denote the observed responses
  - $\boldsymbol{Y}_{\bar{\boldsymbol{r}}} = (Y_{ij} : R_{ij} = 0)$ denote the missing responses
- Define a joint probability model $p(\boldsymbol{y}, \boldsymbol{r} \mid \boldsymbol{\omega})$ under a benchmark missing data assumption:
  - Often this is **MAR**, i.e. when $p(\boldsymbol{r} \mid \boldsymbol{y}, \boldsymbol{\psi}) = p(\boldsymbol{r} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{\psi})$
  - If we use an individual summary function of $\boldsymbol{r}$, e.g. $d(\boldsymbol{r})$, such that $p(\boldsymbol{r} \mid \boldsymbol{y}, d(\boldsymbol{r}), \boldsymbol{\kappa}) = p(\boldsymbol{r} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\kappa})$, then this defines a **Partial MAR** assumption, with $\boldsymbol{\kappa} \in \boldsymbol{\psi}$
  - An intuitive choice for $d(\boldsymbol{r})$ is the **dropout time**, i.e. time of the last observed response for each individual

General notation

- Let $\boldsymbol{Y}_{ij} = (U_{ij}, C_{ij})$ be the vector of utility and cost variables available for the i-th person at the j-th time in the study
- Let $\boldsymbol{R}_{ij} = (R^u_{ij}, R^c_{ij})$ be the corresponding vector of missingness indicators, e.g. $R^u_{ij} = \mathbb{I}(U_{ij}\text{is obs})$, and let
  - $\boldsymbol{Y}_{\boldsymbol{r}} = (Y_{ij} : R_{ij} = 1)$ denote the observed responses
  - $\boldsymbol{Y}_{\bar{\boldsymbol{r}}} = (Y_{ij} : R_{ij} = 0)$ denote the missing responses
- Define a joint probability model $p(\boldsymbol{y}, \boldsymbol{r} \mid \boldsymbol{\omega})$ under a benchmark missing data assumption:
  - Often this is **MAR**, i.e. when $p(\boldsymbol{r} \mid \boldsymbol{y}, \boldsymbol{\psi}) = p(\boldsymbol{r} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{\psi})$
  - If we use an individual summary function of $\boldsymbol{r}$, e.g. $d(\boldsymbol{r})$, such that $p(\boldsymbol{r} \mid \boldsymbol{y}, d(\boldsymbol{r}), \boldsymbol{\kappa}) = p(\boldsymbol{r} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\kappa})$, then this defines a **Partial MAR** assumption, with $\boldsymbol{\kappa} \in \boldsymbol{\psi}$
  - An intuitive choice for $d(\boldsymbol{r})$ is the **dropout time**, i.e. time of the last observed response for each individual
  - If **intermittent** missingness is not informative, then inferences only require a model for $p(\boldsymbol{y}, d(\boldsymbol{r}))$

Modelling framework

- To allow a MNAR analysis we use a PMM specification and rely on a PMAR assumption, conditional on dropout time $d(\boldsymbol{r})$, to express the joint model as

$$p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}) = p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\omega}_E) p(\boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}_O)$$

  - $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\omega}_E)$ is the **extrapolation distribution** (not identified)
  - $p(\boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}_O)$ is the **observed data distribution** (identified)

Modelling framework

- To allow a MNAR analysis we use a PMM specification and rely on a PMAR assumption, conditional on dropout time $d(\boldsymbol{r})$, to express the joint model as

$$p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}) = p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\omega}_E) p(\boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}_O)$$

  - $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\omega}_E)$ is the **extrapolation distribution** (not identified)
  - $p(\boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}_O)$ is the **observed data distribution** (identified)

- **Problem**: Since we have a bivariate outcome $\boldsymbol{y} = (u, c)$, then dropout time may differ between the types of responses, i.e. we have $d(\boldsymbol{r}^u)$ and $d(\boldsymbol{r}^c)$

Modelling framework

- To allow a MNAR analysis we use a PMM specification and rely on a PMAR assumption, conditional on dropout time $d(\boldsymbol{r})$, to express the joint model as

$$p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}) = p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\omega}_E) p(\boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}_O)$$

  - $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}), \boldsymbol{\omega}_E)$ is the **extrapolation distribution** (not identified)
  - $p(\boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}) \mid \boldsymbol{\omega}_O)$ is the **observed data distribution** (identified)

- **Problem**: Since we have a bivariate outcome $\boldsymbol{y} = (u, c)$, then dropout time may differ between the types of responses, i.e. we have $d(\boldsymbol{r}^u)$ and $d(\boldsymbol{r}^c)$

- **Solution**: Define two new dropout indicators for each individual
  - $d^{min} = \min(d(\boldsymbol{r}^u), d(\boldsymbol{r}^c))$ – minimum of the two indicators
  - $d^{max} = \max(d(\boldsymbol{r}^u), d(\boldsymbol{r}^c))$ – maximum of the two indicators

- Use these in the model to specify PMAR conditional on two aggregated patterns: 1) $d^{min} = d^{max}$; 2) $d^{min} \neq d^{max}$

Modelling framework

- The joint model $p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega})$ can be split into five components:

1 The model for the patterns and for the completers:

$$p(d(\boldsymbol{r}) \mid \boldsymbol{\pi}) \left[p(\boldsymbol{y} \mid \boldsymbol{r} = \mathbf{1}, \boldsymbol{\lambda})\right]^{\mathbb{I}(d^{min} = d^{max} = J)}$$

Modelling framework

- The joint model $p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega})$ can be split into five components:

1 The model for the patterns and for the completers:

$$p(d(\boldsymbol{r}) \mid \boldsymbol{\pi}) \left[p(\boldsymbol{y} \mid \boldsymbol{r} = \mathbf{1}, \boldsymbol{\lambda})\right]^{\mathbb{I}(d^{min}=d^{max}=J)}$$

2 The model for $\boldsymbol{y_r}$ when $d^{min} = d^{max}$:

$$\left[\prod_{d^{min}=d^{max}} p(\boldsymbol{y_r} \mid \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\eta})\right]^{\mathbb{I}(d^{min}=d^{max}<J)}$$

Modelling framework

- The joint model $p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega})$ can be split into five components:

1 The model for the patterns and for the completers:

$$p(d(\boldsymbol{r}) \mid \boldsymbol{\pi}) \left[p(\boldsymbol{y} \mid \boldsymbol{r} = \mathbf{1}, \boldsymbol{\lambda})\right]^{\mathbb{I}(d^{min} = d^{max} = J)}$$

2 The model for $\boldsymbol{y_r}$ when $d^{min} = d^{max}$:

$$\left[\prod_{d^{min} = d^{max}} p(\boldsymbol{y_r} \mid \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\eta})\right]^{\mathbb{I}(d^{min} = d^{max} < J)}$$

3 The model for $\boldsymbol{y_r}$ when $d^{min} \neq d^{max}$:

$$\left[\prod_{d^{min} \neq d^{max}} p(\boldsymbol{y_r} \mid \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\upsilon})\right]^{\mathbb{I}(d^{min} \neq d^{max} < J)}$$

Modelling framework

- The joint model $p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega})$ can be split into five components:

4 The model for $\boldsymbol{y}_{\bar{\boldsymbol{r}}}$ when $d^{min} = d^{max}$:

$$\left[\prod_{d^{min}=d^{max}} p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\xi}) \right]^{\mathbb{I}(d^{min}=d^{max}<J)}$$

Modelling framework

- The joint model $p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega})$ can be split into five components:

4 The model for $\boldsymbol{y}_{\bar{\boldsymbol{r}}}$ when $d^{min} = d^{max}$:

$$\left[\prod_{d^{min}=d^{max}} p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\xi}) \right]^{\mathbb{I}(d^{min}=d^{max}<J)}$$

5 The model for $\boldsymbol{y}_{\bar{\boldsymbol{r}}}$ when $d^{min} \neq d^{max}$:

$$\left[\prod_{d^{min}\neq d^{max}} p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\xi}) \right]^{\mathbb{I}(d^{min}\neq d^{max}<J)}$$

Modelling framework

- The joint model $p(\boldsymbol{y}, d(\boldsymbol{r}) \mid \boldsymbol{\omega})$ can be split into five components:

4 The model for $\boldsymbol{y}_{\bar{\boldsymbol{r}}}$ when $d^{min} = d^{max}$:

$$\left[\prod_{d^{min}=d^{max}} p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\xi}) \right]^{\mathbb{I}(d^{min}=d^{max}<J)}$$

5 The model for $\boldsymbol{y}_{\bar{\boldsymbol{r}}}$ when $d^{min} \neq d^{max}$:

$$\left[\prod_{d^{min}\neq d^{max}} p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{r} \neq \mathbf{1}, \boldsymbol{d}, \boldsymbol{\xi}) \right]^{\mathbb{I}(d^{min}\neq d^{max}<J)}$$

- This representation is used to define the identification strategy of $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}))$ through **identifying restrictions** and **sensitivity parameters**

Model for dropout

- Model $\boldsymbol{D}_i = (D_i^{min}, D_i^{max})$ using a multinomial distribution with dropout probabilities $\boldsymbol{\pi}_{t_i}$ defined on $\{1, \ldots, J^2\}$, where $J^2 =$ number of dropout patterns

Model for dropout

- Model $\boldsymbol{D}_i = (D_i^{min}, D_i^{max})$ using a multinomial distribution with dropout probabilities $\boldsymbol{\pi}_{t_i}$ defined on $\{1, \ldots, J^2\}$, where J^2 = number of dropout patterns

- The prior on $\boldsymbol{\pi}_{t_i}$ incorporates prior knowledge about the expected dropout probabilities for CEA studies:

$$\boldsymbol{\pi}_{t_i} \sim \text{Dirichlet}\left(1-x, \frac{x/2}{J-1}, \ldots, \frac{x/2}{J-1}, \frac{x/2}{(J^2-J)}, \ldots, \frac{x/2}{(J^2-J)}\right),$$

Model for dropout

- Model $\boldsymbol{D}_i = (D_i^{min}, D_i^{max})$ using a multinomial distribution with dropout probabilities $\boldsymbol{\pi}_{t_i}$ defined on $\{1, \ldots, J^2\}$, where J^2 = number of dropout patterns

- The prior on $\boldsymbol{\pi}_{t_i}$ incorporates prior knowledge about the expected dropout probabilities for CEA studies:

$$\boldsymbol{\pi}_{t_i} \sim \text{Dirichlet}\left(1-x, \frac{x/2}{J-1}, \ldots, \frac{x/2}{J-1}, \frac{x/2}{(J^2-J)}, \ldots, \frac{x/2}{(J^2-J)}\right),$$

  - $1-x$ is the prior probability for completing the study (given an expected dropout rate of $x\%$)
  - $\frac{x/2}{J-1}$ is the prior dropout probability for each of the $J-1$ patterns where $D_i^{min} = D_i^{max}$
  - $\frac{x/2}{(J^2-J)}$ is the prior dropout probability for each of the J^2-J patterns where $D_i^{min} \neq D_i^{max}$

Model for dropout

- Model $\boldsymbol{D}_i = (D_i^{min}, D_i^{max})$ using a multinomial distribution with dropout probabilities $\boldsymbol{\pi}_{t_i}$ defined on $\{1, \ldots, J^2\}$, where J^2 = number of dropout patterns

- The prior on $\boldsymbol{\pi}_{t_i}$ incorporates prior knowledge about the expected dropout probabilities for CEA studies:

$$\boldsymbol{\pi}_{t_i} \sim \text{Dirichlet}\left(1-x, \frac{x/2}{J-1}, \ldots, \frac{x/2}{J-1}, \frac{x/2}{(J^2-J)}, \ldots, \frac{x/2}{(J^2-J)}\right),$$

  - $1-x$ is the prior probability for completing the study (given an expected dropout rate of $x\%$)
  - $\frac{x/2}{J-1}$ is the prior dropout probability for each of the $J-1$ patterns where $D_i^{min} = D_i^{max}$
  - $\frac{x/2}{(J^2-J)}$ is the prior dropout probability for each of the J^2-J patterns where $D_i^{min} \neq D_i^{max}$

- Alternative: a noninformative Dirichlet$(1, \ldots, 1)$

Model for the observed response

- We choose a nonparametric approach as it allows to:
  - weaken **distributional assumptions**
  - account for more complex **functional forms** (e.g. non-linearity)

Model for the observed response

- We choose a nonparametric approach as it allows to:
  - weaken **distributional assumptions**
  - account for more complex **functional forms** (e.g. non-linearity)
- We model the observed responses $\boldsymbol{y}_{ij} = (u_{ij}, c_{ij})$ through a **Dirichlet process mixture of normals** (Escobar et al., 1995)

$$\begin{aligned} \boldsymbol{y}_i &\sim \text{Normal}\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right), \\ (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) &\sim G, \\ G &\sim DP(\alpha, G_0). \end{aligned}$$

  - G is an uncountable set of Dirichlet processes, indexed by the baseline distribution G_0 and the concentration parameter α

Model for the observed response

- Dependence is introduced using the **stick-breaking representation** of each element in this set (Rasmussen, 2000), which allows to re-write and approximate the model as a finite mixture of $k = 1, \ldots, K$ normals

$$\boldsymbol{y}_i \sim \sum_{k=1}^{K} \nu^{(k)} \text{Normal}\left(\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}\right),$$

Where $\nu^{(k)} = V^{(k)} \prod_{j<k} \left(1 - V^{(j)}\right)$ and $V^{(k)} \sim \text{Beta}(1, \alpha)$

  - When $K = 1$, the model corresponds to a multivariate normal
  - When K large, the approximation to G is improved.

Model for the observed response

- Dependence is introduced using the **stick-breaking representation** of each element in this set (Rasmussen, 2000), which allows to re-write and approximate the model as a finite mixture of $k = 1, \ldots, K$ normals

$$\boldsymbol{y}_i \sim \sum_{k=1}^{K} \nu^{(k)} \text{Normal}\left(\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}\right),$$

  Where $\nu^{(k)} = V^{(k)} \prod_{j<k} \left(1 - V^{(j)}\right)$ and $V^{(k)} \sim \text{Beta}(1, \alpha)$
  - When $K = 1$, the model corresponds to a multivariate normal
  - When K large, the approximation to G is improved.

- Can be reparameterised using the **generalised autoregressive coefficients** (GARP) and **innovation variances** (IV) decomposition of MVN within each mixture component k (Taddy, 2008)

Model for the observed response

- Express the joint normal distribution of $\boldsymbol{y}_i$ as the product of a sequence of normal distributions

Model for the observed response

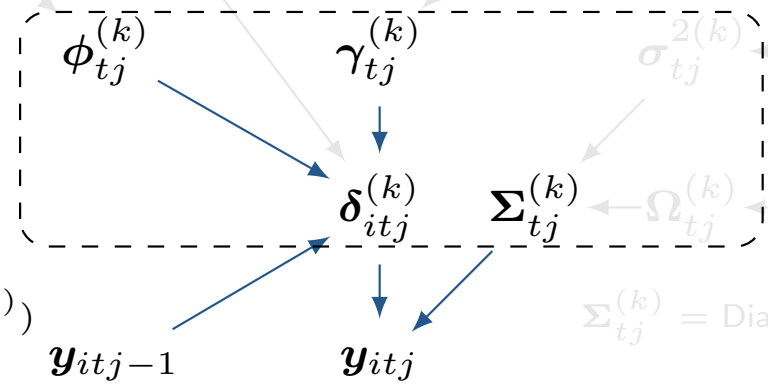
- Express the joint normal distribution of $\boldsymbol{y}_i$ as the product of a sequence of normal distributions
- Assuming a **first-order Markov dependence structure**

$$p(\boldsymbol{y}_i) = f^{(k)}(\boldsymbol{y}_i) = f_1^{(k)}(\boldsymbol{y}_{i1}) \prod_{j=2}^{J} f_j^{(k)}(\boldsymbol{y}_{ij} \mid \boldsymbol{y}_{ij-1}),$$

  - $f_1^{(k)}(\boldsymbol{y}_{i1})$ is the density of the bivariate distribution at $j = 1$ (baseline)
  - $f_j^{(k)}(\boldsymbol{y}_{ij} \mid \boldsymbol{y}_{ij-1})$ is the density of the bivariate response $\boldsymbol{y}_{ij}$ at time j for subjects who have their responses observed at time $j-1$ conditional on $\boldsymbol{y}_{ij-1}$

Modelling framework

- Consider the pair $\boldsymbol{y}_{itj} = (u_{itj}, c_{itj})$ for individual i at time j in group t:



$k = 1, \ldots, K$

Model for the responses

$$f_{tj}^{(k)}(\boldsymbol{y}_{itj} \mid \boldsymbol{y}_{itj-1}) \sim \text{Normal}(\boldsymbol{\delta}_{itj}^{(k)}, \boldsymbol{\Sigma}_{tj}^{(k)})$$

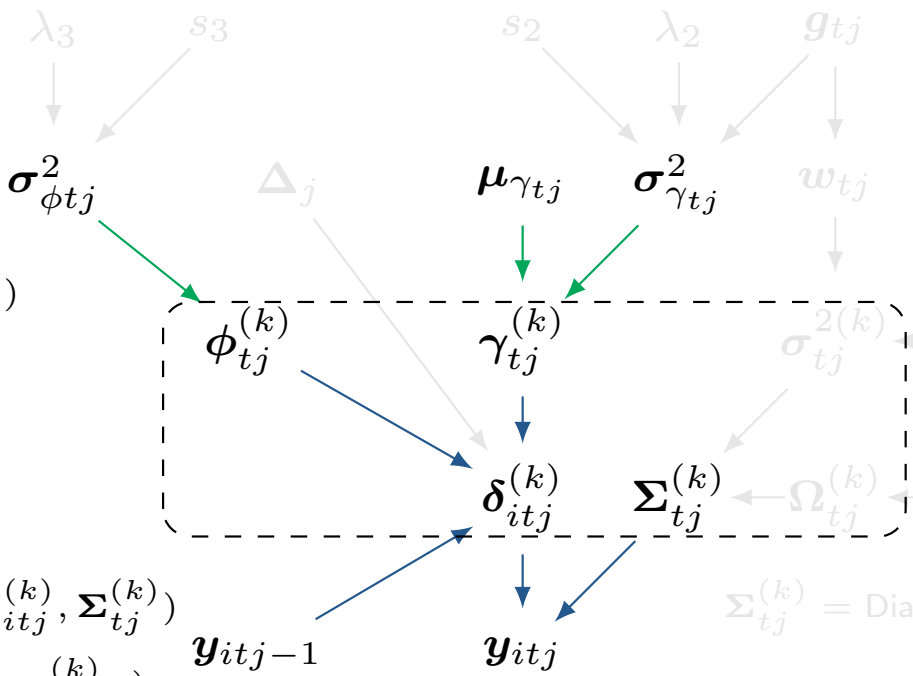
$$\boldsymbol{\delta}_{itj}^{(k)} = \boldsymbol{\gamma}_{tj}^{(k)} + \boldsymbol{\phi}_{tj}^{(k)}(\boldsymbol{y}_{itj-1} - \boldsymbol{\gamma}_{tj-1}^{(k)})$$

Modelling framework

- Consider the pair $\boldsymbol{y}_{itj} = (u_{itj}, c_{itj})$ for individual i at time j in group t:

Priors on the location

$$\boldsymbol{\gamma}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{\mu}_{\gamma_{tj}}, \boldsymbol{\sigma}^2_{\gamma_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\phi}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\sigma}^2_{\phi_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\mu}_{\gamma_{tj}}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{0.5I})$$
$$\boldsymbol{\sigma}^{2(k)}_{\gamma_{tj}} \sim \text{Gamma}^{-1}(s_2, s_2\lambda_2 \boldsymbol{g}_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{\phi_{tj}} \sim \text{Gamma}^{-1}(s_3, s_3\lambda_3)$$



$k = 1, \ldots, K$

Model for the responses

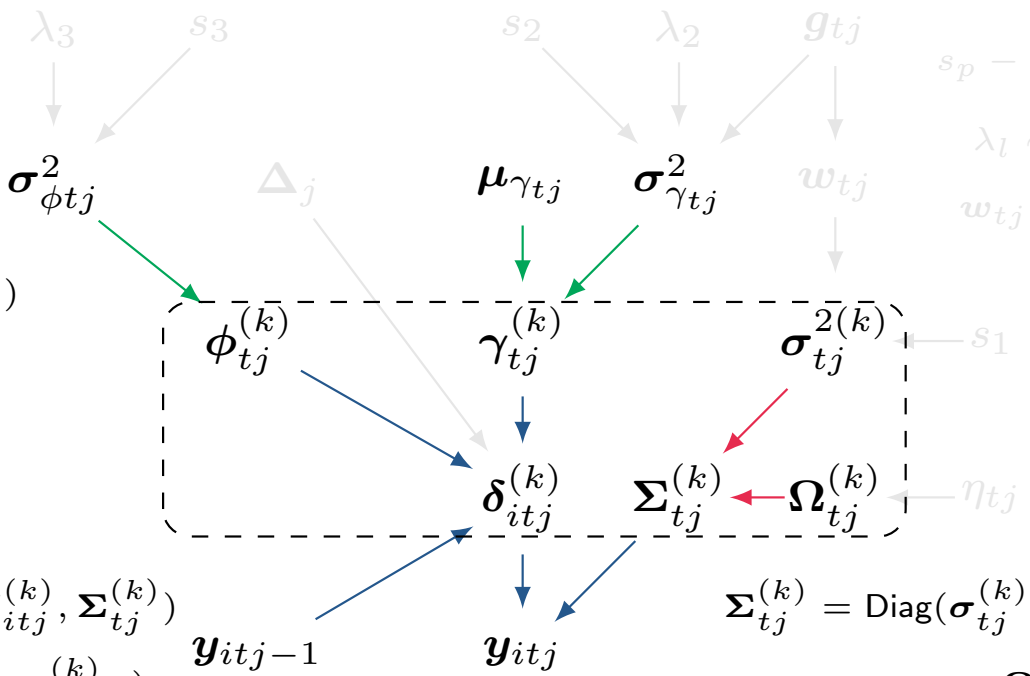
$$f_{tj}^{(k)}(\boldsymbol{y}_{itj} \mid \boldsymbol{y}_{itj-1}) \sim \text{Normal}(\boldsymbol{\delta}_{itj}^{(k)}, \boldsymbol{\Sigma}_{tj}^{(k)})$$
$$\boldsymbol{\delta}_{itj}^{(k)} = \boldsymbol{\gamma}_{tj}^{(k)} + \boldsymbol{\phi}_{tj}^{(k)}(\boldsymbol{y}_{itj-1} - \boldsymbol{\gamma}_{tj-1}^{(k)})$$

Modelling framework

- Consider the pair $\boldsymbol{y}_{itj} = (u_{itj}, c_{itj})$ for individual i at time j in group t:

Priors on the location

$$\boldsymbol{\gamma}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{\mu}_{\gamma_{tj}}, \boldsymbol{\sigma}^2_{\gamma_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\phi}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\sigma}^2_{\phi_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\mu}_{\gamma_{tj}}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{0.5I})$$
$$\boldsymbol{\sigma}^{2(k)}_{\gamma_{tj}} \sim \text{Gamma}^{-1}(s_2, s_2\lambda_2 \boldsymbol{g}_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{\phi_{tj}} \sim \text{Gamma}^{-1}(s_3, s_3\lambda_3)$$



Hyperpriors

$$s_p - 2 \sim \text{Gamma}^{-1}(1, 1)$$
$$\lambda_l \sim \text{Gamma}^{-1}(1, 1)$$
$$\boldsymbol{w}_{tj} \sim \text{Gamma}(1, \tfrac{2}{\boldsymbol{g}_{tj}})$$

$k = 1, \ldots, K$

Model for the responses

$$f_{tj}^{(k)}(\boldsymbol{y}_{itj} \mid \boldsymbol{y}_{itj-1}) \sim \text{Normal}(\boldsymbol{\delta}_{itj}^{(k)}, \boldsymbol{\Sigma}_{tj}^{(k)})$$
$$\boldsymbol{\delta}_{itj}^{(k)} = \boldsymbol{\gamma}_{tj}^{(k)} + \boldsymbol{\phi}_{tj}^{(k)}(\boldsymbol{y}_{itj-1} - \boldsymbol{\gamma}_{tj-1}^{(k)})$$

Priors on the variance

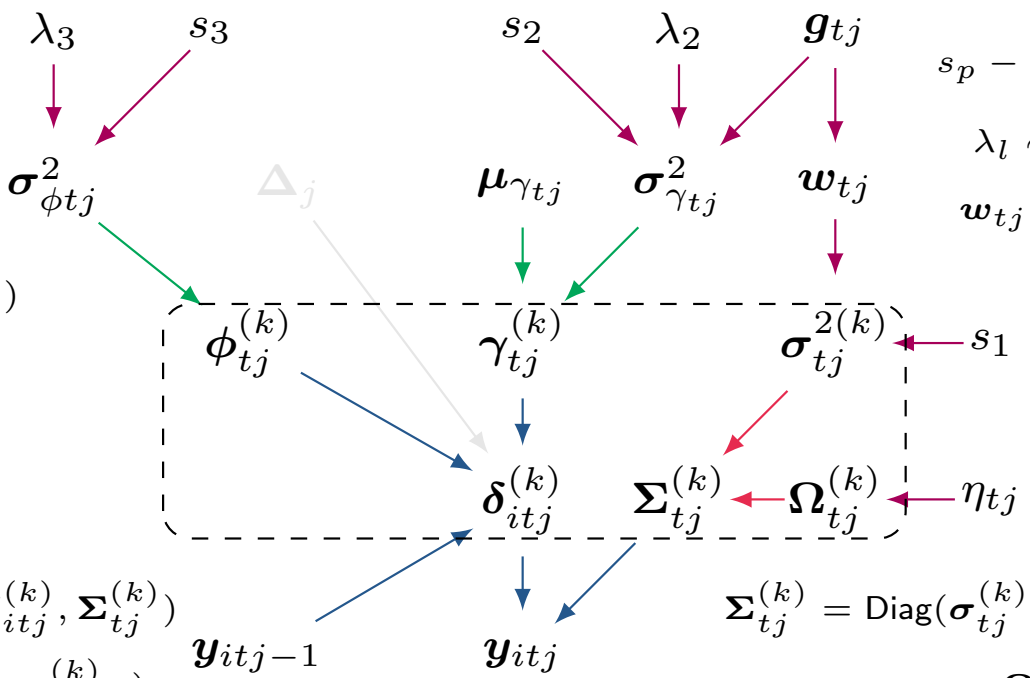
$$\boldsymbol{\Sigma}_{tj}^{(k)} = \text{Diag}(\boldsymbol{\sigma}_{tj}^{(k)}) \times \boldsymbol{\Omega}_{tj}^{(k)} \times \text{Diag}(\boldsymbol{\sigma}_{tj}^{(k)})$$
$$\boldsymbol{\Omega}_{tj}^{(k)} \sim \text{LKJ}(\eta_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{tj} \sim \text{Gamma}^{-1}(s_1, s_1 \boldsymbol{w}_{tj})$$

Modelling framework

- Consider the pair $\boldsymbol{y}_{itj} = (u_{itj}, c_{itj})$ for individual i at time j in group t:

Priors on the location

$$\boldsymbol{\gamma}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{\mu}_{\gamma_{tj}}, \boldsymbol{\sigma}^2_{\gamma_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\phi}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\sigma}^2_{\phi_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\mu}_{\gamma_{tj}}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{0.5I})$$
$$\boldsymbol{\sigma}^{2(k)}_{\gamma_{tj}} \sim \text{Gamma}^{-1}(s_2, s_2\lambda_2 \boldsymbol{g}_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{\phi_{tj}} \sim \text{Gamma}^{-1}(s_3, s_3\lambda_3)$$



Hyperpriors

$$s_p - 2 \sim \text{Gamma}^{-1}(1, 1)$$
$$\lambda_l \sim \text{Gamma}^{-1}(1, 1)$$
$$\boldsymbol{w}_{tj} \sim \text{Gamma}(1, \tfrac{2}{\boldsymbol{g}_{tj}})$$

$$k = 1, \ldots, K$$

Model for the responses

$$f^{(k)}_{tj}(\boldsymbol{y}_{itj} \mid \boldsymbol{y}_{itj-1}) \sim \text{Normal}(\boldsymbol{\delta}^{(k)}_{itj}, \boldsymbol{\Sigma}^{(k)}_{tj})$$
$$\boldsymbol{\delta}^{(k)}_{itj} = \boldsymbol{\gamma}^{(k)}_{tj} + \boldsymbol{\phi}^{(k)}_{tj}(\boldsymbol{y}_{itj-1} - \boldsymbol{\gamma}^{(k)}_{tj-1})$$

Priors on the variance

$$\boldsymbol{\Sigma}^{(k)}_{tj} = \text{Diag}(\boldsymbol{\sigma}^{(k)}_{tj}) \times \boldsymbol{\Omega}^{(k)}_{tj} \times \text{Diag}(\boldsymbol{\sigma}^{(k)}_{tj})$$
$$\boldsymbol{\Omega}^{(k)}_{tj} \sim \text{LKJ}(\eta_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{tj} \sim \text{Gamma}^{-1}(s_1, s_1 \boldsymbol{w}_{tj})$$

Identification strategy - **benchmark scenario**

- We identify $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}))$ under a benchmark assumption using **identifying restrictions**, from which deviations are explored via **sensitivity parameters** (Daniels et al., 2008)

- We impose different restrictions for each component of $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}))$ conditional on the dropout patterns $\boldsymbol{d} = (d^{min}, d^{max})$

Identification strategy - **benchmark scenario**

- We identify $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}))$ under a benchmark assumption using **identifying restrictions**, from which deviations are explored via **sensitivity parameters** (Daniels et al., 2008)

- We impose different restrictions for each component of $p(\boldsymbol{y}_{\bar{\boldsymbol{r}}} \mid \boldsymbol{y}_{\boldsymbol{r}}, d(\boldsymbol{r}))$ conditional on the dropout patterns $\boldsymbol{d} = (d^{min}, d^{max})$

  - For $d^{min} = d^{max}$, we impose the MAR restrictions (Molenberghs et al., 2007)
  - For $d^{min} \neq d^{max}$, the unidentified distributions at $d^{min} < j \leq d^{max}$ are identified using all the observed distributions at j up to d^{max}
  - For $d^{min} \neq d^{max}$, the unidentified distributions at time $j > d^{max}$ are identified using the **MAR** restrictions for $j, \ldots, J$

- The unidentified distributions at $j > d^{max}$ for the patterns $d^{min} = d^{max}$ and $d^{min} \neq d^{max}$, and those at $d^{min} < j < d^{max}$ for the patterns $d^{min} \neq d^{max}$ are identified using a mixture over the distributions at the same time for all the identified patterns $s = j, \ldots, J$ and $s = j, \ldots, d^{max}$, respectively.

Analysis under MNAR

- The benchmark assumption is a reasonable starting point **but** observed data may not fully explain missingness:

Analysis under MNAR

- The benchmark assumption is a reasonable starting point **but** observed data may not fully explain missingness:
  - Patients who leave the study may have worse health states (i.e. lower utilities) than those that continue on, even after conditioning on their past measurements

Analysis under MNAR

- The benchmark assumption is a reasonable starting point **but** observed data may not fully explain missingness:
  - Patients who leave the study may have worse health states (i.e. lower utilities) than those that continue on, even after conditioning on their past measurements

- Explore **MNAR** departures from the benchmark using time-specific location shifts $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ – **sensitivity parameters**

Analysis under MNAR

- The benchmark assumption is a reasonable starting point **but** observed data may not fully explain missingness:
  - Patients who leave the study may have worse health states (i.e. lower utilities) than those that continue on, even after conditioning on their past measurements
- Explore **MNAR** departures from the benchmark using time-specific location shifts $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ – **sensitivity parameters**
- **Objective**: assess the robustness of results to *plausible* deviations from the benchmark scenario under partial ignorability conditional on $\boldsymbol{d} = (d^{min}, d^{max})$

Analysis under MNAR

- The benchmark assumption is a reasonable starting point **but** observed data may not fully explain missingness:
  - Patients who leave the study may have worse health states (i.e. lower utilities) than those that continue on, even after conditioning on their past measurements
- Explore **MNAR** departures from the benchmark using time-specific location shifts $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ – **sensitivity parameters**
- **Objective**: assess the robustness of results to *plausible* deviations from the benchmark scenario under partial ignorability conditional on $\boldsymbol{d} = (d^{min}, d^{max})$
- **Strategy**: incorporate $\boldsymbol{\Delta}_j$ within the conditional means of the extrapolation distributions identified under the bechmark scenario

Analysis under MNAR

- The benchmark assumption is a reasonable starting point **but** observed data may not fully explain missingness:
  - Patients who leave the study may have worse health states (i.e. lower utilities) than those that continue on, even after conditioning on their past measurements
- Explore **MNAR** departures from the benchmark using time-specific location shifts $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ – **sensitivity parameters**
- **Objective**: assess the robustness of results to *plausible* deviations from the benchmark scenario under partial ignorability conditional on $\boldsymbol{d} = (d^{min}, d^{max})$
- **Strategy**: incorporate $\boldsymbol{\Delta}_j$ within the conditional means of the extrapolation distributions identified under the bechmark scenario
- **Elicitation**: use available knowledge/information from available data/experts to elicit informative priors on $\boldsymbol{\Delta}_j$

Analysis under MNAR

- The benchmark assumption is a reasonable starting point **but** observed data may not fully explain missingness:
  - Patients who leave the study may have worse health states (i.e. lower utilities) than those that continue on, even after conditioning on their past measurements
- Explore **MNAR** departures from the benchmark using time-specific location shifts $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ – **sensitivity parameters**
- **Objective**: assess the robustness of results to *plausible* deviations from the benchmark scenario under partial ignorability conditional on $\boldsymbol{d} = (d^{min}, d^{max})$
- **Strategy**: incorporate $\boldsymbol{\Delta}_j$ within the conditional means of the extrapolation distributions identified under the bechmark scenario
- **Elicitation**: use available knowledge/information from available data/experts to elicit informative priors on $\boldsymbol{\Delta}_j$
  - $\boldsymbol{\Delta}_j$ represent expected deviations of mean $\boldsymbol{y} = (\boldsymbol{u}, \boldsymbol{c})$ values at time $j = 2, 3, \ldots$ after dropout compared to estimates obtained from observed data
  - Thanks to identification strategy, it is possible to elicit different values of $\boldsymbol{\Delta}_j$ for each dropout pattern $\boldsymbol{d} = (d^{min}, d^{max})$

Sensitivity analysis

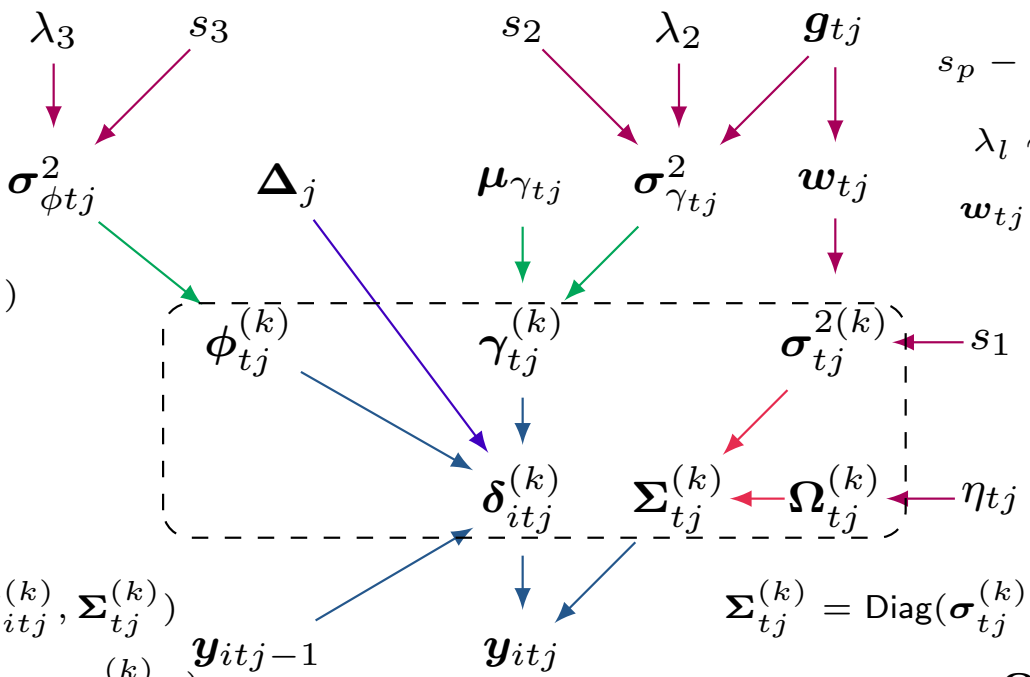
- For $d^{min} = d^{max}$ at $j > d^{max}$:

$$f_{d,j}(\boldsymbol{y}_{ij} \mid \boldsymbol{y}_{ij-1}, d^{min}, d^{max}) \equiv \tilde{f}_{d,j}(\boldsymbol{y}_{ij} + \boldsymbol{\Delta}_j \mid \boldsymbol{y}_{ij-1}, d^{min}, d^{max}),$$

- $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$

Priors on the location

$$\boldsymbol{\gamma}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{\mu}_{\gamma_{tj}}, \boldsymbol{\sigma}^2_{\gamma_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\phi}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\sigma}^2_{\phi_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\mu}_{\gamma_{tj}}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{0.5I})$$
$$\boldsymbol{\sigma}^{2(k)}_{\gamma_{tj}} \sim \text{Gamma}^{-1}(s_2, s_2\lambda_2 \boldsymbol{g}_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{\phi_{tj}} \sim \text{Gamma}^{-1}(s_3, s_3\lambda_3)$$



Hyperpriors

$$s_p - 2 \sim \text{Gamma}^{-1}(1, 1)$$
$$\lambda_l \sim \text{Gamma}^{-1}(1, 1)$$
$$\boldsymbol{w}_{tj} \sim \text{Gamma}(1, \tfrac{2}{\boldsymbol{g}_{tj}})$$

$k = 1, \ldots, K$

Model for the responses

$$f_{tj}^{(k)}(\boldsymbol{y}_{itj} \mid \boldsymbol{y}_{itj-1}) \sim \text{Normal}(\boldsymbol{\delta}_{itj}^{(k)}, \boldsymbol{\Sigma}_{tj}^{(k)})$$
$$\boldsymbol{\delta}_{itj}^{(k)} = \boldsymbol{\gamma}_{tj}^{(k)} + \boldsymbol{\Delta}_j + \boldsymbol{\phi}_{tj}^{(k)}(\boldsymbol{y}_{itj-1} - \boldsymbol{\gamma}_{tj-1}^{(k)})$$

Priors on the variance

$$\boldsymbol{\Sigma}_{tj}^{(k)} = \text{Diag}(\boldsymbol{\sigma}_{tj}^{(k)}) \times \boldsymbol{\Omega}_{tj}^{(k)} \times \text{Diag}(\boldsymbol{\sigma}_{tj}^{(k)})$$
$$\boldsymbol{\Omega}_{tj}^{(k)} \sim \text{LKJ}(\eta_{tj})$$
$$\boldsymbol{\sigma}_{tj}^{2(k)} \sim \text{Gamma}^{-1}(s_1, s_1 \boldsymbol{w}_{tj})$$

Sensitivity analysis

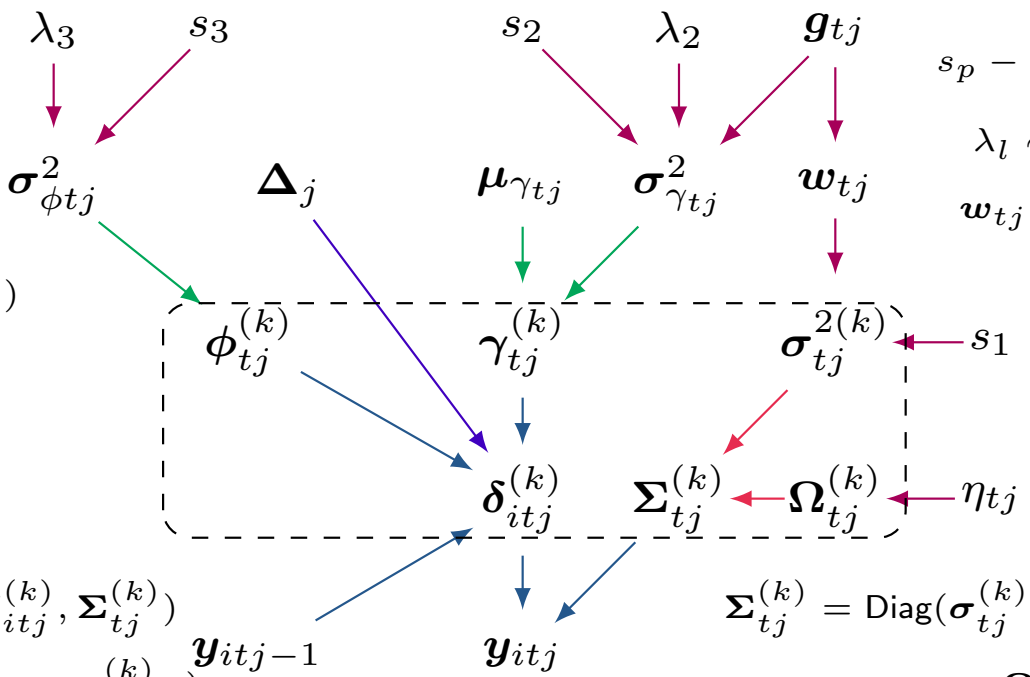
- For $d^{min} \neq d^{max}$ at $d^{min} < j \leq d^{max}$ and $j > d^{max}$:

$$f_{d,j}(\boldsymbol{y}_{ij} \mid \boldsymbol{y}_{ij-1}, d^{min}, d^{max}) \equiv \tilde{f}_{d,j}(\boldsymbol{y}_{ij} + \boldsymbol{\Delta}_j \mid \boldsymbol{y}_{ij-1}, d^{min}, d^{max}),$$

- $\boldsymbol{\Delta}_j = (\Delta_j^u, 0)$ if $(d^{min} = d^u, d^{max} = d^c)$; $\boldsymbol{\Delta}_j = (0, \Delta_j^c)$ if $(d^{min} = d^c, d^{max} = d^u)$

Priors on the location

$$\boldsymbol{\gamma}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{\mu}_{\gamma_{tj}}, \boldsymbol{\sigma}^2_{\gamma_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\phi}_{tj}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\sigma}^2_{\phi_{tj}} \boldsymbol{I})$$
$$\boldsymbol{\mu}_{\gamma_{tj}}^{(k)} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{0.5I})$$
$$\boldsymbol{\sigma}^{2(k)}_{\gamma_{tj}} \sim \text{Gamma}^{-1}(s_2, s_2\lambda_2 \boldsymbol{g}_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{\phi_{tj}} \sim \text{Gamma}^{-1}(s_3, s_3\lambda_3)$$



Hyperpriors

$$s_p - 2 \sim \text{Gamma}^{-1}(1, 1)$$
$$\lambda_l \sim \text{Gamma}^{-1}(1, 1)$$
$$\boldsymbol{w}_{tj} \sim \text{Gamma}(1, \tfrac{2}{\boldsymbol{g}_{tj}})$$

$$k = 1, \ldots, K$$

Model for the responses

$$f_{tj}^{(k)}(\boldsymbol{y}_{itj} \mid \boldsymbol{y}_{itj-1}) \sim \text{Normal}(\boldsymbol{\delta}_{itj}^{(k)}, \boldsymbol{\Sigma}_{tj}^{(k)})$$
$$\boldsymbol{\delta}_{itj}^{(k)} = \boldsymbol{\gamma}_{tj}^{(k)} + \boldsymbol{\Delta}_j + \boldsymbol{\phi}_{tj}^{(k)}(\boldsymbol{y}_{itj-1} - \boldsymbol{\gamma}_{tj-1}^{(k)})$$

Priors on the variance

$$\boldsymbol{\Sigma}_{tj}^{(k)} = \text{Diag}(\boldsymbol{\sigma}_{tj}^{(k)}) \times \boldsymbol{\Omega}_{tj}^{(k)} \times \text{Diag}(\boldsymbol{\sigma}_{tj}^{(k)})$$
$$\boldsymbol{\Omega}_{tj}^{(k)} \sim \text{LKJ}(\eta_{tj})$$
$$\boldsymbol{\sigma}^{2(k)}_{tj} \sim \text{Gamma}^{-1}(s_1, s_1 \boldsymbol{w}_{tj})$$

Posterior computation

- Posterior computation is carried out at each iteration of the posterior distribution
  1. We draw samples $(\boldsymbol{\pi}, \boldsymbol{\theta})$ from the posterior given the observed data $(\boldsymbol{y_r}, d(\boldsymbol{r}))$ and augmenting the missing data $\boldsymbol{y_{\bar{r}}}$
  2. We calculate the distributions of desired functionals of $p(\boldsymbol{y} \mid \boldsymbol{\omega})$

Posterior computation

- Posterior computation is carried out at each iteration of the posterior distribution
  1. We draw samples $(\boldsymbol{\pi}, \boldsymbol{\theta})$ from the posterior given the observed data $(\boldsymbol{y_r}, d(\boldsymbol{r}))$ and augmenting the missing data $\boldsymbol{y_{\bar{r}}}$
  2. We calculate the distributions of desired functionals of $p(\boldsymbol{y} \mid \boldsymbol{\omega})$

- Specifically, we are interested in the computation of:

$$\mathsf{E}\left[\mathcal{T}(\boldsymbol{Y}) \mid \boldsymbol{\omega}\right] = \int \mathcal{T}(\boldsymbol{y}) p(\boldsymbol{y} \mid \boldsymbol{\omega}) d\boldsymbol{y},$$

where $\mathcal{T}(\boldsymbol{Y})$ is some transformation such that $\mathcal{T}_j(\boldsymbol{Y}_j) = \mathcal{T}_j(\boldsymbol{Y}_j \mid \overline{\boldsymbol{Y}}_{j-1}, \boldsymbol{\Delta}_j)$ at the j-th time in the study. For example,

Posterior computation

- Posterior computation is carried out at each iteration of the posterior distribution
  1. We draw samples $(\boldsymbol{\pi}, \boldsymbol{\theta})$ from the posterior given the observed data $(\boldsymbol{y_r}, d(\boldsymbol{r}))$ and augmenting the missing data $\boldsymbol{y_{\bar{r}}}$
  2. We calculate the distributions of desired functionals of $p(\boldsymbol{y} \mid \boldsymbol{\omega})$

- Specifically, we are interested in the computation of:

$$\mathrm{E}\left[\mathcal{T}(\boldsymbol{Y}) \mid \boldsymbol{\omega}\right] = \int \mathcal{T}(\boldsymbol{y}) p(\boldsymbol{y} \mid \boldsymbol{\omega}) d\boldsymbol{y},$$

  where $\mathcal{T}(\boldsymbol{Y})$ is some transformation such that $\mathcal{T}_j(\boldsymbol{Y}_j) = \mathcal{T}_j(\boldsymbol{Y}_j \mid \overline{\boldsymbol{Y}}_{j-1}, \boldsymbol{\Delta}_j)$ at the j-th time in the study. For example,
  - If $\mathcal{T}_j(\boldsymbol{Y}_j \mid \overline{\boldsymbol{Y}}_{j-1}, \boldsymbol{0}) = \boldsymbol{Y}_j$, then deviations of $\boldsymbol{\Delta}_j$ from $\boldsymbol{0}$ represent deviations of the assumed model from MAR
  - Although $\mathrm{E}\left[\mathcal{T}(\boldsymbol{Y}) \mid \boldsymbol{\omega}\right]$ are not directly available, we can derive them by **Monte Carlo integration**

Priors on the sensitivity parameters

- Priors on $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ are calibrated using information on the scale of the data as an intuitive starting point (Linero et al., 2015) – e.g. residual standard deviations for $\boldsymbol{u}$ and $\boldsymbol{c}$ pulled across time

Priors on the sensitivity parameters

- Priors on $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ are calibrated using information on the scale of the data as an intuitive starting point (Linero et al., 2015) – e.g. residual standard deviations for $\boldsymbol{u}$ and $\boldsymbol{c}$ pulled across time
  - **Assumption**: It is thought unlikely that deviations from the benchmark would exceed a standard deviation for both outcomes

Priors on the sensitivity parameters

- Priors on $\boldsymbol{\Delta}_j = (\Delta_j^u, \Delta_j^c)$ are calibrated using information on the scale of the data as an intuitive starting point (Linero et al., 2015) – e.g. residual standard deviations for $\boldsymbol{u}$ and $\boldsymbol{c}$ pulled across time
  - **Assumption**: It is thought unlikely that deviations from the benchmark would exceed a standard deviation for both outcomes
- We marginally specify the prior distributions of each sensitivity parameter Δ_j using normal distributions based on a linear AR(1) or **first order mean autocorrelation structure**:

$$\boldsymbol{\Delta}_j \sim \text{Normal}\left(\boldsymbol{\Delta}_0 + \boldsymbol{\rho}\boldsymbol{\Delta}_{j-1}, \boldsymbol{\sigma}_{\Delta_j}\right),$$

  - $\boldsymbol{\rho} = (\rho^u, \rho^c)$ is the two-vector of autoregressive coefficients capturing the temporal dependence between the shifts after dropout
  - $\boldsymbol{\sigma}_{\Delta_j} = (\sigma^u_{\Delta_j}, \sigma^c_{\Delta_j})$ is set to sd$(\boldsymbol{y})$, obtained from the observed $\boldsymbol{u}$ and $\boldsymbol{c}$
  - Set $\mid \boldsymbol{\rho} \mid < \mathbf{1}$ and start the process at $\boldsymbol{\Delta}_j = \mathbf{0}$ (dropout time)

Priors on the sensitivity parameters

- We specify the multivariate dependence in terms of a correlation matrix parametrised by ρ_{time}, which determines the correlation between $\boldsymbol{\Delta}_j$

Priors on the sensitivity parameters

- We specify the multivariate dependence in terms of a correlation matrix parametrised by ρ_{time}, which determines the correlation between $\boldsymbol{\Delta}_j$
- We elicit a **first-order autoregressive structure**:

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{|j-s|} \\ \rho_1 & 1 & \cdots & \vdots \\ \vdots & \cdots & 1 & \vdots \\ \rho_{|j-s|} & \cdots & \cdots & 1 \end{pmatrix}$$

Priors on the sensitivity parameters

- We specify the multivariate dependence in terms of a correlation matrix parametrised by ρ_{time}, which determines the correlation between $\boldsymbol{\Delta}_j$
- We elicit a **first-order autoregressive structure**:

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \boldsymbol{\rho}_1 & \cdots & \boldsymbol{\rho}_{|j-s|} \\ \boldsymbol{\rho}_1 & 1 & \cdots & \vdots \\ \vdots & \cdots & 1 & \vdots \\ \boldsymbol{\rho}_{|j-s|} & \cdots & \cdots & 1 \end{pmatrix}$$

- We explore alternative assumptions about the autocorrelation structure by giving $\boldsymbol{\rho}$ point-mass priors and varying the priors over the grid $\{0.1, 0.5, 0.9\}$

Priors on the sensitivity parameters

- We specify the multivariate dependence in terms of a correlation matrix parametrised by ρ_{time}, which determines the correlation between $\boldsymbol{\Delta}_j$
- We elicit a **first-order autoregressive structure**:

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \boldsymbol{\rho}_1 & \cdots & \boldsymbol{\rho}_{|j-s|} \\ \boldsymbol{\rho}_1 & 1 & \cdots & \vdots \\ \vdots & \cdots & 1 & \vdots \\ \boldsymbol{\rho}_{|j-s|} & \cdots & \cdots & 1 \end{pmatrix}$$

- We explore alternative assumptions about the autocorrelation structure by giving $\boldsymbol{\rho}$ point-mass priors and varying the priors over the grid $\{0.1, 0.5, 0.9\}$
  - **Assumption**: individuals are associated with a progressive increase in their mean utility decrement/cost increment

Part 4

Application to the PBS study

Back to Table of content

- 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

- 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

- Utilities (EQ-5D) and costs (clinic records) are collected at baseline, 6 and 12 months follow-up

- 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

- Utilities (EQ-5D) and costs (clinic records) are collected at baseline, 6 and 12 months follow-up

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%)	

The PBS study – Implementation

- We implement our nonparametric framework to the PBS data and:
  - we compare the fit of the proposed model to alternative parametric choices from the literature via **information criteria**
  - we assess the fit of the proposed model to the observed data under the benchmark scenario via **posterior predictive checks**
- We compare the results and CEA conclusions obtained across 6 alternative missingness assumptions
- **MAR**: fit model without PMM (ignorable MAR)
- **IR MAR**: fit model with PMM using $d = d^{min}$ (MAR restrictions)
- **IR PMAR**: benchmark scenario (partial MAR restrictions)
- **MNAR**: benchmark + $\boldsymbol{\Delta}_j \neq \mathbf{0}$ under 3 scenarios
  - low temporal correlation: $\boldsymbol{\rho} = 0.1$ (**MNAR L**)
  - medium temporal correlation: $\boldsymbol{\rho} = 0.5$ (**MNAR M**)
  - high temporal correlation: $\boldsymbol{\rho} = 0.9$ (**MNAR H**)

The PBS study – model assessment (PIC)

- We compute three **relative measures** of predictive accuracy to compare the fit of the model to alternative specifications: **WAIC**(Watanabe et al., 2010); **LOOIC**(Vehtari et al., 2017); **LPML**(Geisser et al, 1979)

The PBS study – model assessment (PIC)

- We compute three **relative measures** of predictive accuracy to compare the fit of the model to alternative specifications: **WAIC**(Watanabe et al., 2010); **LOOIC**(Vehtari et al., 2017); **LPML**(Geisser et al, 1979)
- We compare the DPM (under MAR) with three parametric specifications:
  1. A multivariate Normal fitted jointly to all variables (MVN)
  2. A multivariate parametric model proposed by Gabrio et al. (2019), using Betas for the utility and Gammas for the cost variables (BG)
  3. A similar model to 2 but replacing Gammas with LogNormals for the costs (BLN)

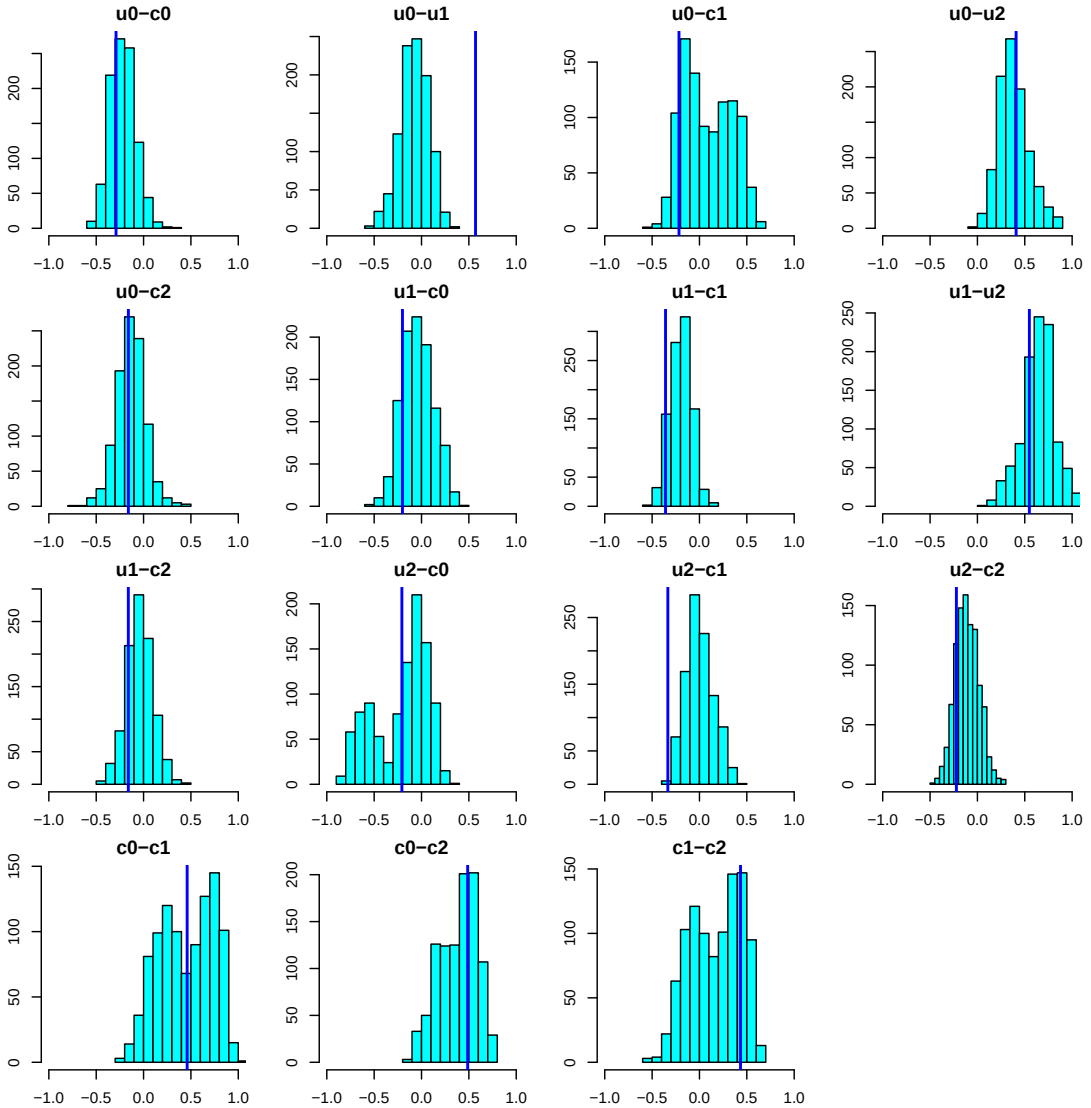
The PBS study – model assessment (PIC)

- We compute three **relative measures** of predictive accuracy to compare the fit of the model to alternative specifications: **WAIC**(Watanabe et al., 2010); **LOOIC**(Vehtari et al., 2017); **LPML**(Geisser et al, 1979)
- We compare the DPM (under MAR) with three parametric specifications:
  1. A multivariate Normal fitted jointly to all variables (MVN)
  2. A multivariate parametric model proposed by Gabrio et al. (2019), using Betas for the utility and Gammas for the cost variables (BG)
  3. A similar model to 2 but replacing Gammas with LogNormals for the costs (BLN)

Model	**WAIC (lpd; p_{waic})**	**LOOIC (lpd; p_{loo})**	**LPML**
MVN	13513 (-6757; 150)	13327 (-6664; 138)	-6663
BG	12696 (-6348, 198)	12518 (-6259; 156)	-6260
BLN	11850 (-5925, 346)	10918 (-5459; 258)	-5460
DPM	8154 (-4077, 361)	8155 (-4077; 361)	-4077

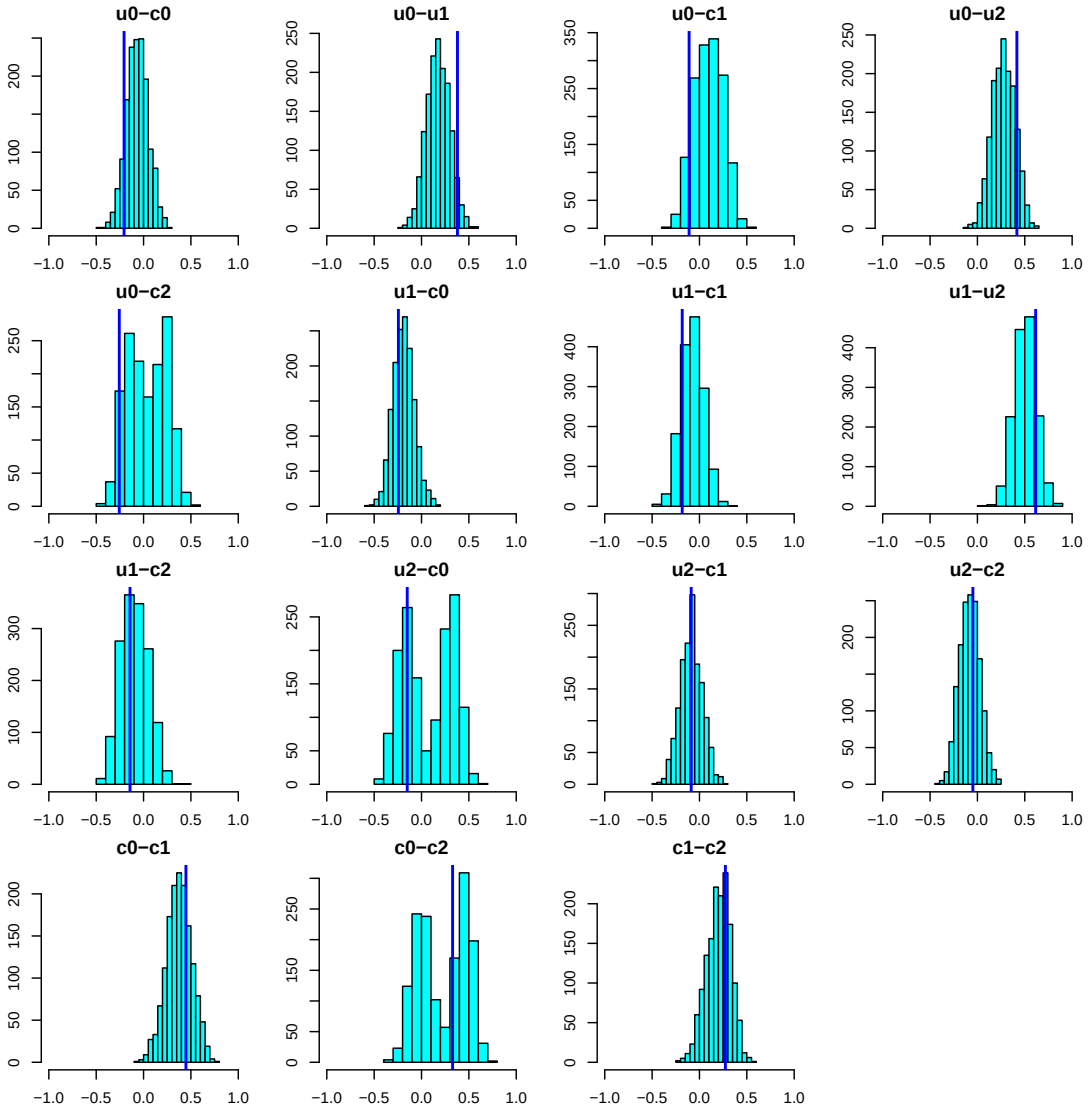
The PBS study – model assessment (PPC – control)

- We also assess the absolute fit of the model using **posterior predictive checks** based on data replications from the posterior predictive $p(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{r}} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{r}, \boldsymbol{\omega})$.

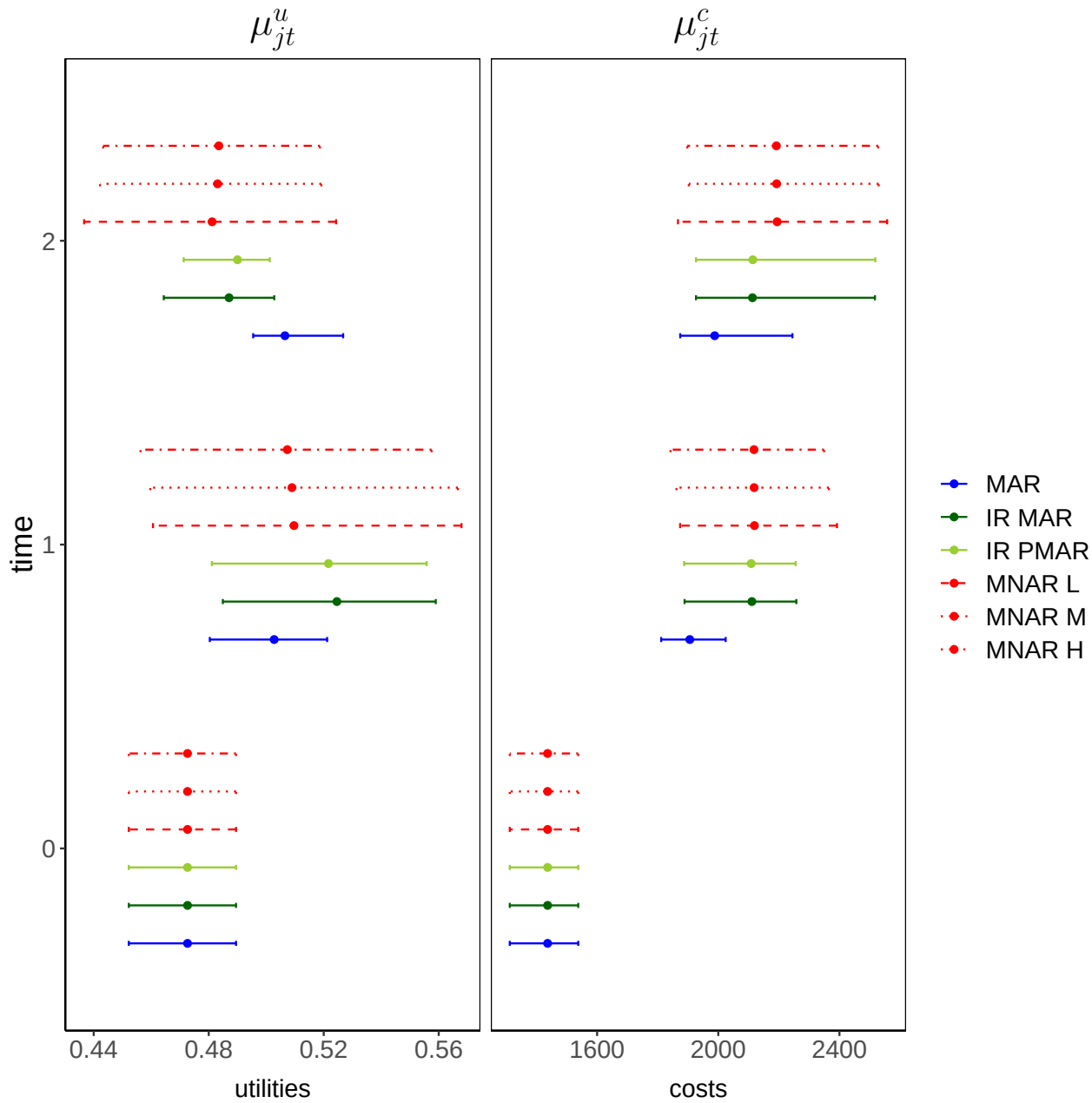


The PBS study – model assessment (PPC - intervention)

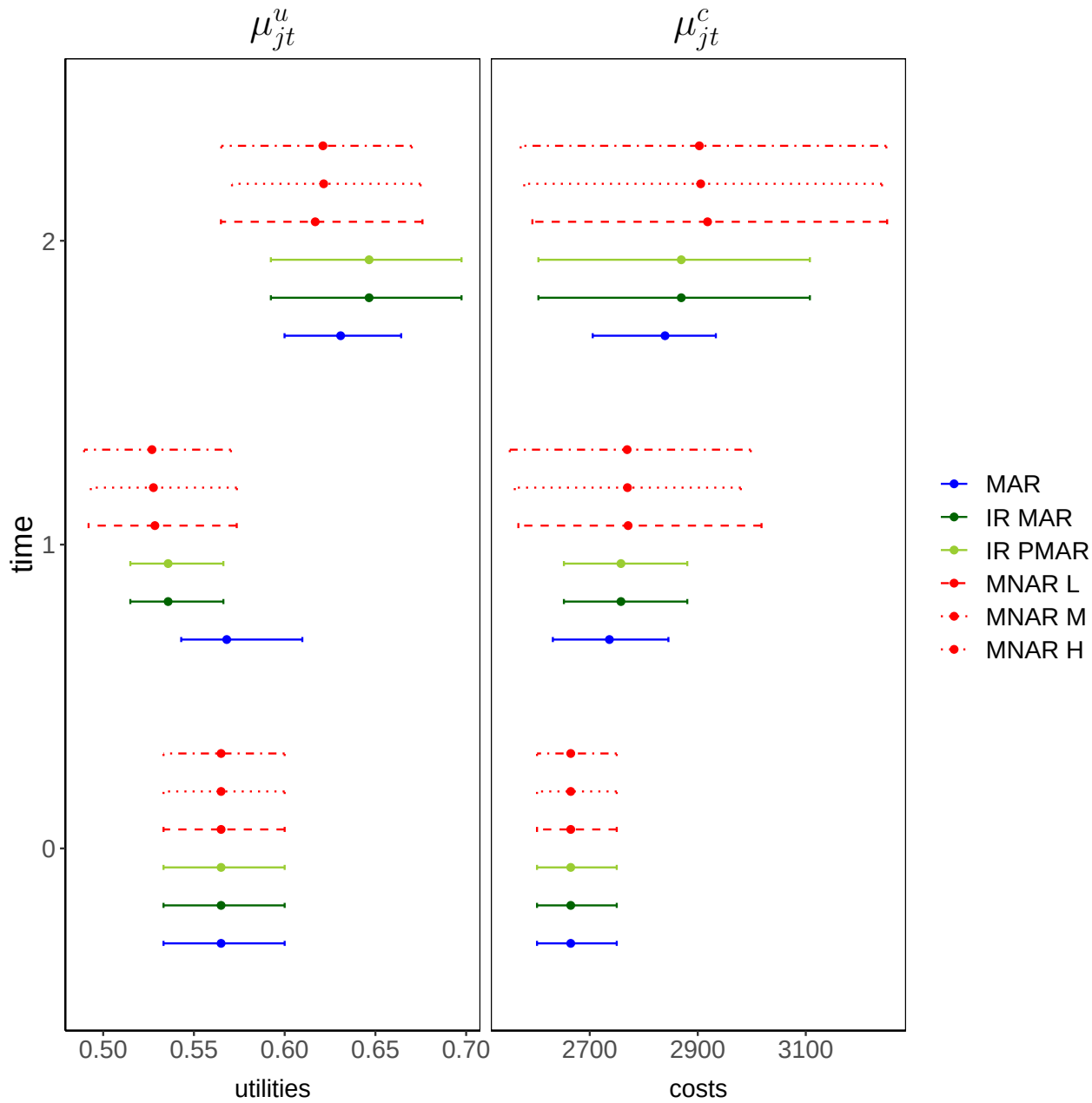
- We also assess the absolute fit of the model using **posterior predictive checks** based on data replications from the posterior predictive $p(\tilde{\boldsymbol{y}}, \tilde{\boldsymbol{r}} \mid \boldsymbol{y}_{\boldsymbol{r}}, \boldsymbol{r}, \boldsymbol{\omega})$.



The PBS study – posterior means and 95% HPD (control)



The PBS study – posterior means and 95% HPD (intervention)



The PBS study – posterior aggregated means and 95% HPD

- Using the posterior of μ_{jt}^u and μ_{jt}^c, we can then calculate the aggregated QALY and total cost means μ_{et} and μ_{ct} over the study period through the formulae:

$$\mu_{et} = \sum_{j=1}^{J} (\mu_{jt}^u + \mu_{j-1t}^u) \frac{\delta_j}{2} \quad \text{and} \quad \mu_{ct} = \sum_{j=1}^{J} \mu_{cjt}$$

The PBS study – posterior aggregated means and 95% HPD

- Using the posterior of μ^u_{jt} and μ^c_{jt}, we can then calculate the aggregated QALY and total cost means μ_{et} and μ_{ct} over the study period through the formulae:

$$\mu_{et} = \sum_{j=1}^{J} (\mu^u_{jt} + \mu^u_{j-1t}) \frac{\delta_j}{2} \quad \text{and} \quad \mu_{ct} = \sum_{j=1}^{J} \mu_{cjt}$$

- We then summarise μ_{et} and μ_{ct} under each missingess scenario

The PBS study – posterior aggregated means and 95% HPD

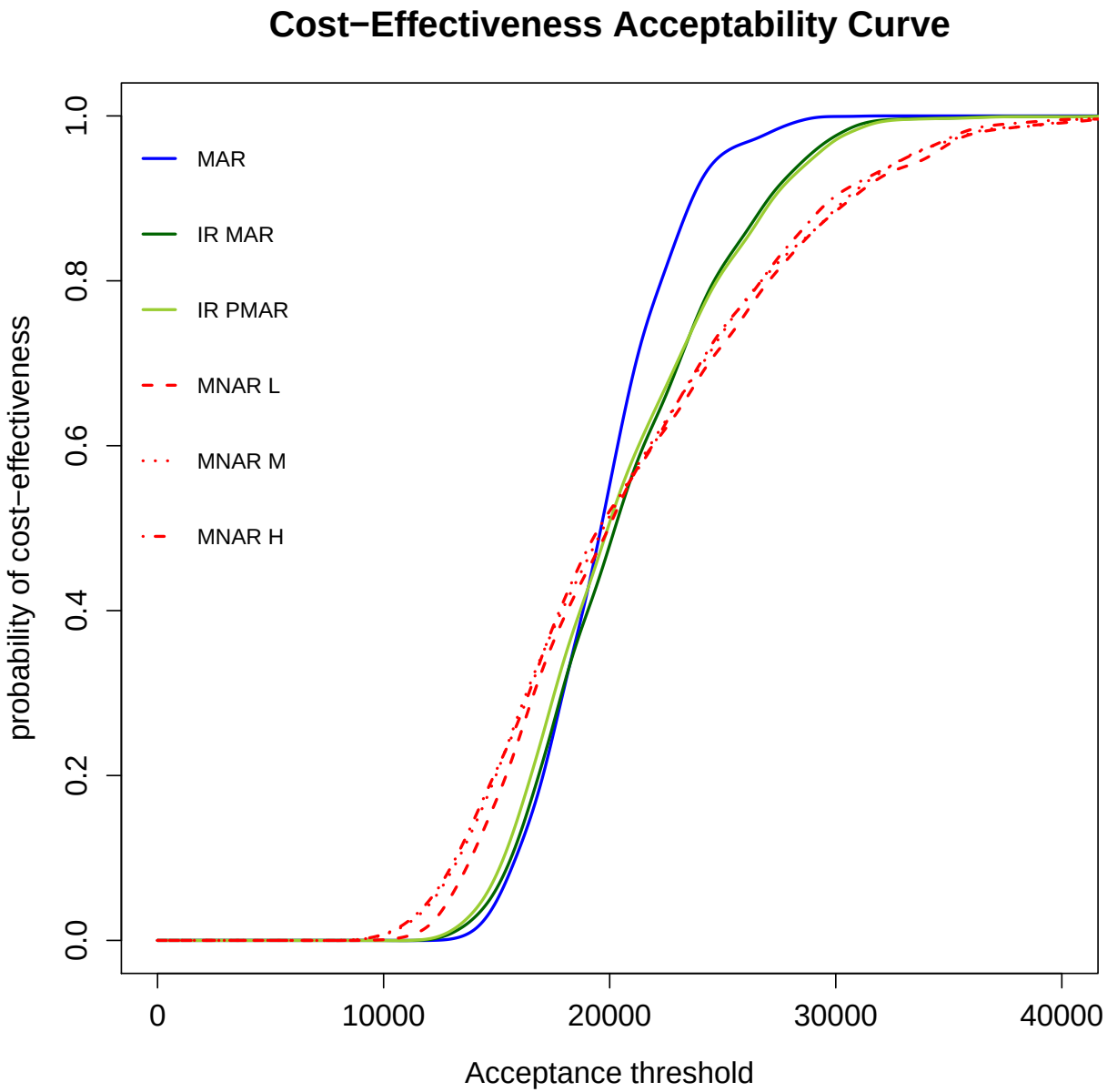
- Using the posterior of μ^u_{jt} and μ^c_{jt}, we can then calculate the aggregated QALY and total cost means μ_{et} and μ_{ct} over the study period through the formulae:

$$\mu_{et} = \sum_{j=1}^{J} (\mu^u_{jt} + \mu^u_{j-1t}) \frac{\delta_j}{2} \quad \text{and} \quad \mu_{ct} = \sum_{j=1}^{J} \mu_{cjt}$$

- We then summarise μ_{et} and μ_{ct} under each missingess scenario

Scenario	μ_{e1}		μ_{e2}		μ_{c1}		μ_{c2}	
	mean	95% CI	mean	95% CI	mean	95% CI	mean	95% CI
MAR	0.50	(0.49;0.50)	0.58	(0.56;0.60)	3894	(3685;4254)	5575	(5351;5742)
IR MAR	0.50	(0.49;0.52)	0.57	(0.55;0.60)	4223	(3972;4591)	5627	(5288;5944)
IR PMAR	0.50	(0.48;0.52)	0.57	(0.55;0.60)	4222	(3973;4592)	5627	(5288;5944)
MNAR L	0.49	(0.47;0.52)	0.56	(0.53;0.58)	4313	(3850;4722)	5689	(5225;6135)
MNAR M	0.49	(0.47;0.52)	0.56	(0.53;0.59)	4311	(3863;4692)	5675	(5188;6143)
MNAR H	0.49	(0.47;0.52)	0.56	(0.53;0.59)	4309	(3875;4695)	5672	(5183;6156)

The PBS study – economic evaluation (CEAC)



Part 5

Conclusions

Back to Table of content

Discussion

1. **Flexibility** of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses a nonparametric approach to account for complexities that otherwise may bias inferences and mislead the economic assessment
   - Can be used to specify complex models in a "relatively easy way"

Discussion

1. **Flexibility** of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses a nonparametric approach to account for complexities that otherwise may bias inferences and mislead the economic assessment
   - Can be used to specify complex models 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`, `JAGS` or `STAN`)

Discussion

1. **Flexibility** of the modelling framework
   - Naturally allows the propagation of uncertainty to the economic model
   - Uses a nonparametric approach to account for complexities that otherwise may bias inferences and mislead the economic assessment
   - Can be used to specify complex models 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`, `JAGS` or `STAN`)
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

Heel erg bedankt!