MissingHE

An R package to deal with missing data in trial-based health economic evaluations

Andrea Gabrio

University College London
Department of Primary Care and Population Health & Statistical Science

email: ucakgab@ucl.ac.uk

website: https://agabrioblog.onrender.com/

GitHub page: https://github.com/AnGabrio

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

UCL Priment CTU Methodologists' meeting, London

Wednesday 09 December 2019

Outline

Part 1

Introduction to (Bayesian) modelling in economic evaluations

Back to Table of content

Health Economic Evaluation (HEE)

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

Health Economic Evaluation (HEE)

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 Economic Evaluation (HEE)

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

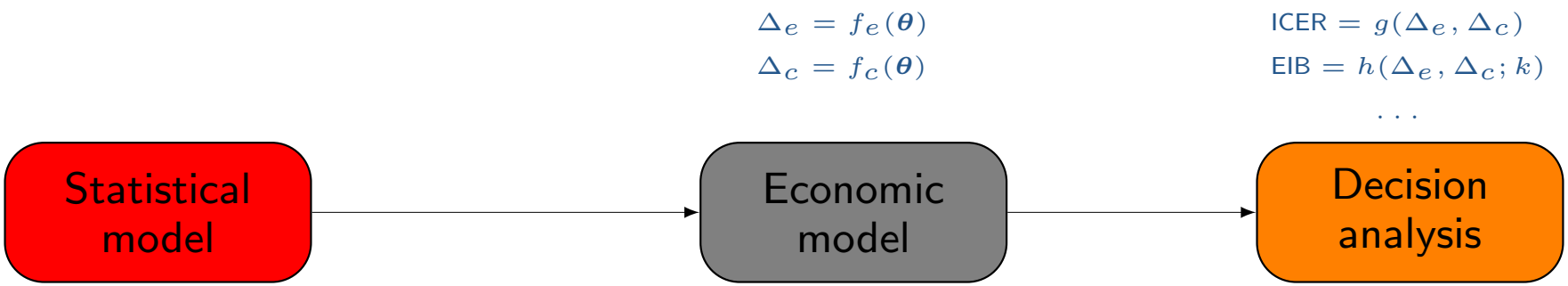
$$\Delta_e = f_e(\boldsymbol{\theta})$$
$$\Delta_c = f_c(\boldsymbol{\theta})$$

- 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 Economic Evaluation (HEE)

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



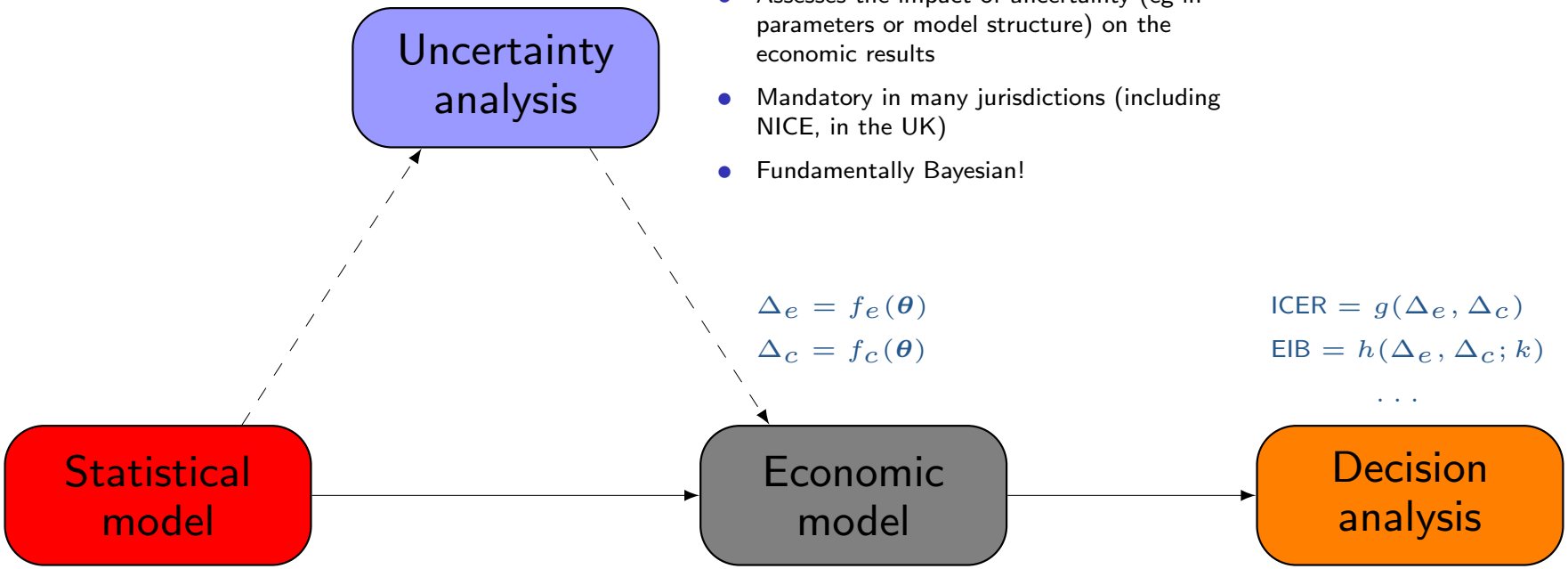
- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

- 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 Economic Evaluation (HEE)

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!

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

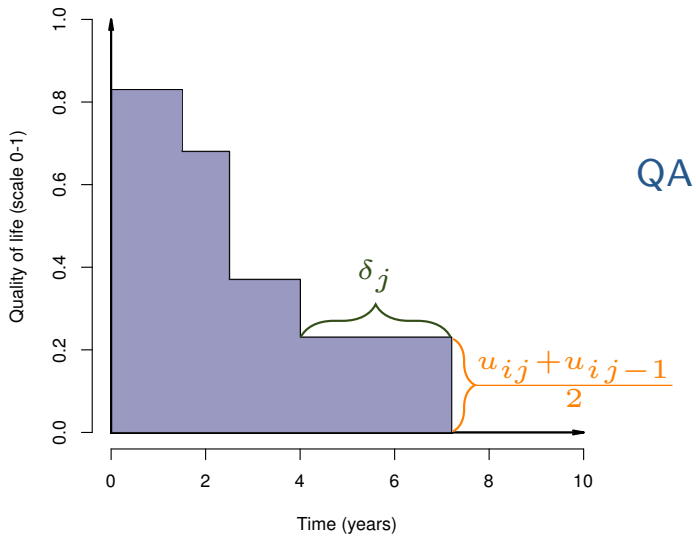
y_{ij} = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
u_{ij} = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
c_{ij} = Use of resources (drugs, hospital, GP appointments, ...)

		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

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

		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

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

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

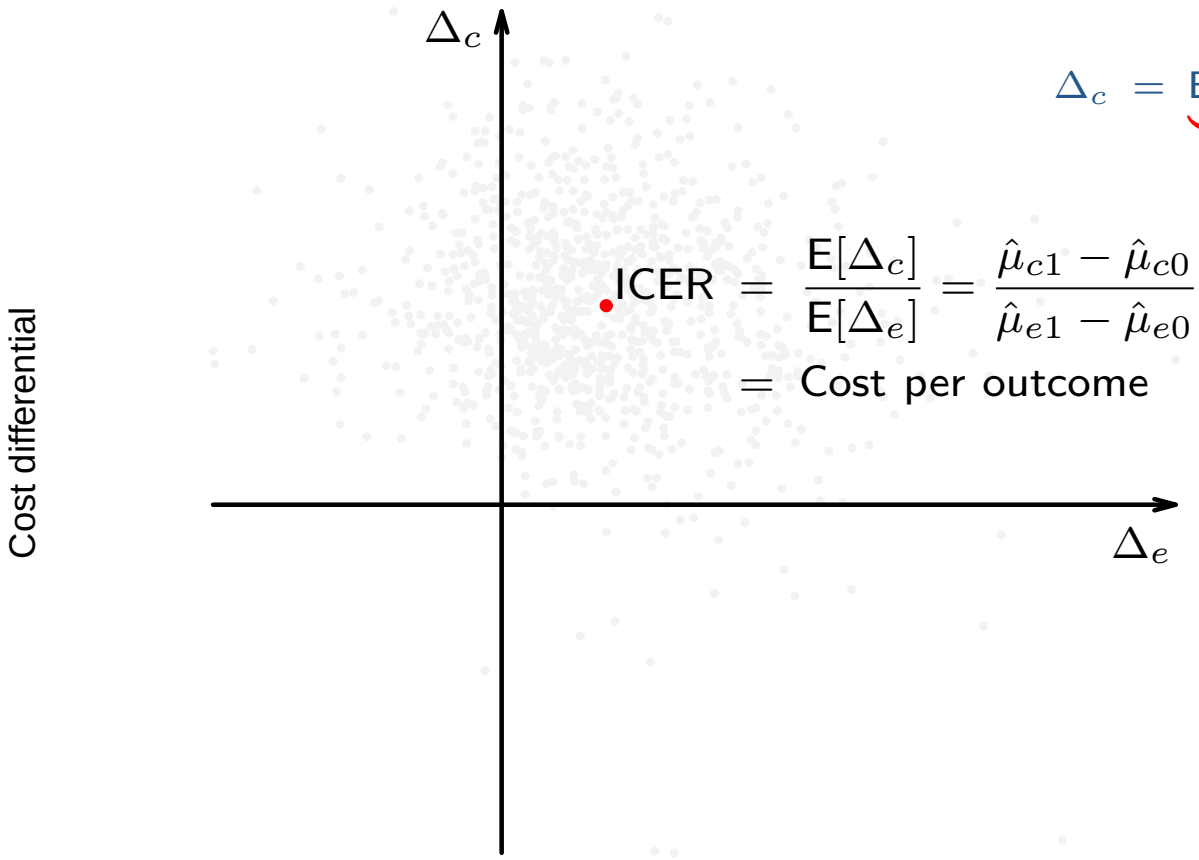
$$\begin{aligned} e_i &= \alpha_{e0} + \alpha_{e1} u_{0i} + \alpha_{e2} \text{Trt}_i + \varepsilon_{ei} \,[+ \ldots], & \varepsilon_{ei} &\sim \text{Normal}(0, \sigma_e) \\ c_i &= \alpha_{c0} + \alpha_{c1} c_{0i} + \alpha_{c2} \text{Trt}_i + \varepsilon_{ci} \,[+ \ldots], & \varepsilon_{ci} &\sim \text{Normal}(0,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

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

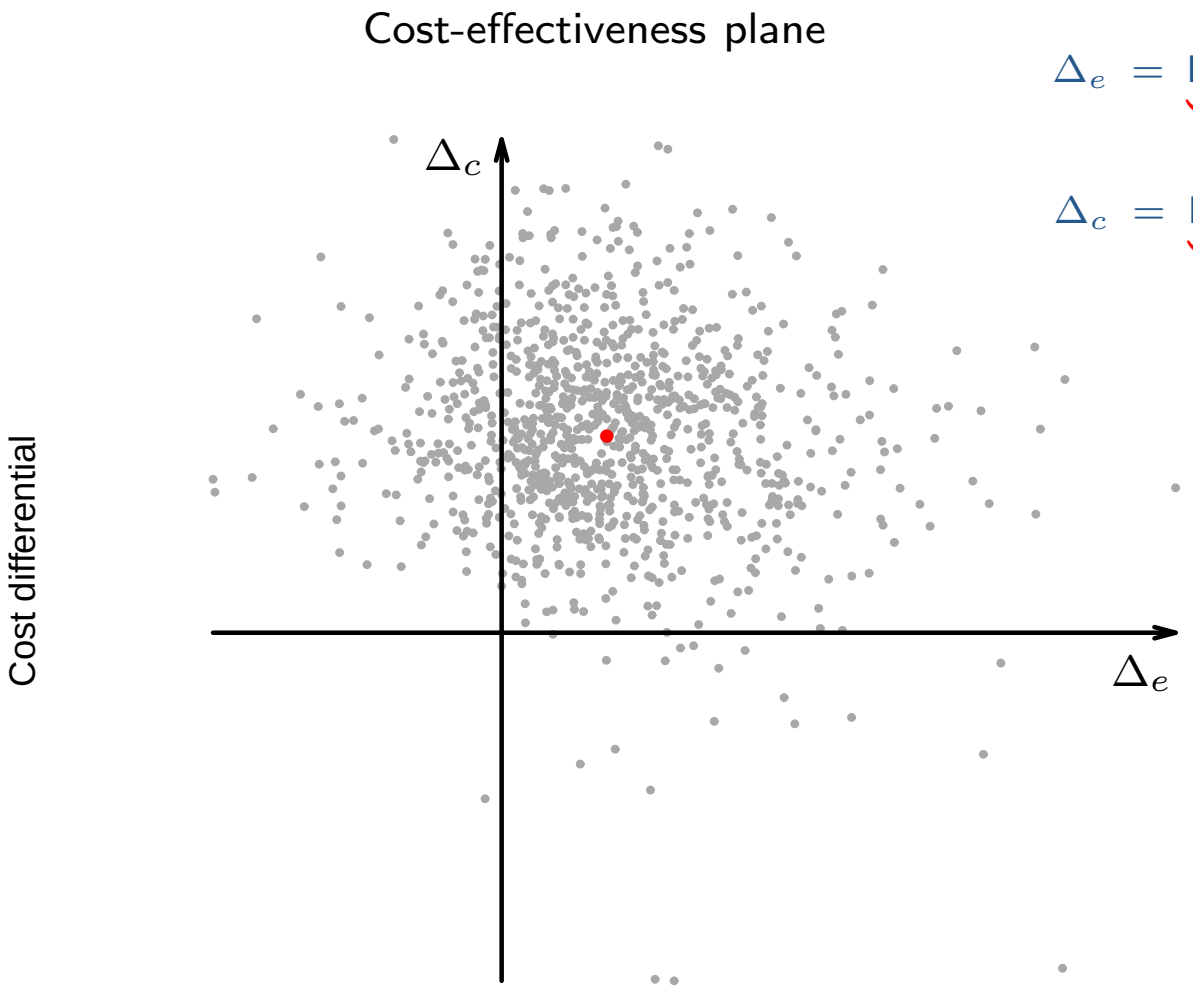
Cost-effectiveness plane

$$\Delta_e = \underbrace{\mathsf{E}[e \mid \hat{\boldsymbol{\theta}}_1]}_{\hat{\mu}_{e1}} - \underbrace{\mathsf{E}[e \mid \hat{\boldsymbol{\theta}}_0]}_{\hat{\mu}_{e0}}$$

$$\Delta_c = \underbrace{\mathsf{E}[c \mid \hat{\boldsymbol{\theta}}_1]}_{\hat{\mu}_{c1}} - \underbrace{\mathsf{E}[c \mid \hat{\boldsymbol{\theta}}_0]}_{\hat{\mu}_{c0}}$$



4. Uncertainty analysis*



$$\Delta_e = \underbrace{\mathrm{E}[e \mid \boldsymbol{\theta}_1]}_{\mu_{e1}} - \underbrace{\mathrm{E}[e \mid \boldsymbol{\theta}_0]}_{\mu_{e0}}$$

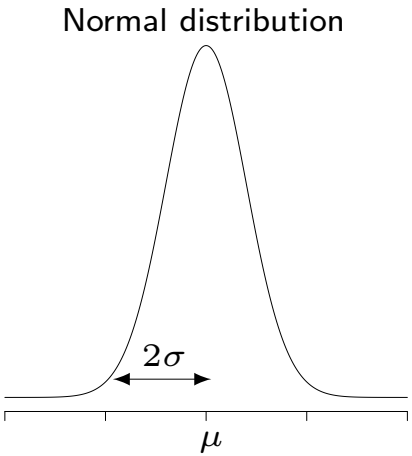
$$\Delta_c = \underbrace{\mathrm{E}[c \mid \boldsymbol{\theta}_1]}_{\mu_{c1}} - \underbrace{\mathrm{E}[c \mid \boldsymbol{\theta}_0]}_{\mu_{c0}}$$

*Induced by $g(\hat{\boldsymbol{\theta}}_0), g(\hat{\boldsymbol{\theta}}_1)$

What is statistics all about?

- Typically, we observe some data and we want to use them to learn about some unobservable feature of the general population in which we are interested
- To do this, we use statistical models to describe the probabilistic mechanism by which (**we assume!**) that the data have arisen

Data generating process



θ Parameters

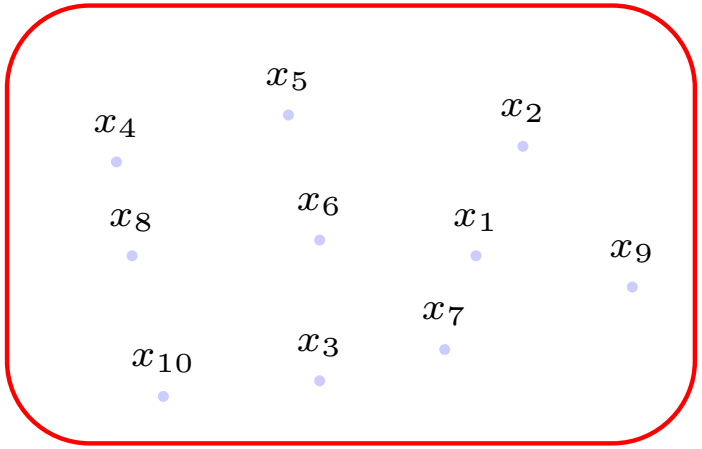
$p(y \mid \theta)$

y Data

NB: Roman letters (y or x) typically indicate **observable data**, while Greek letters (θ, μ, σ, ...) indicate **population parameters**

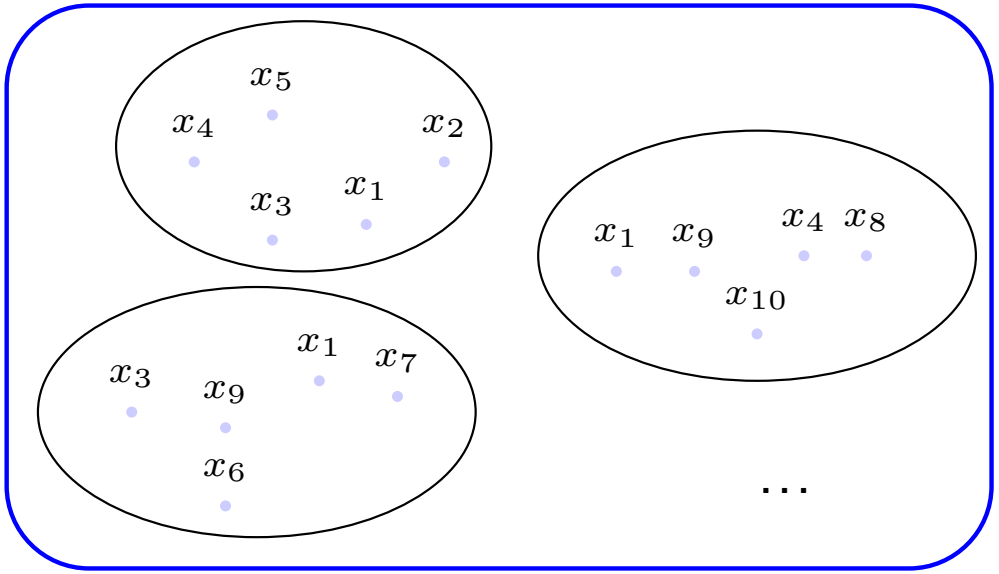
Sampling **variability**

The entire population



$\Rightarrow$

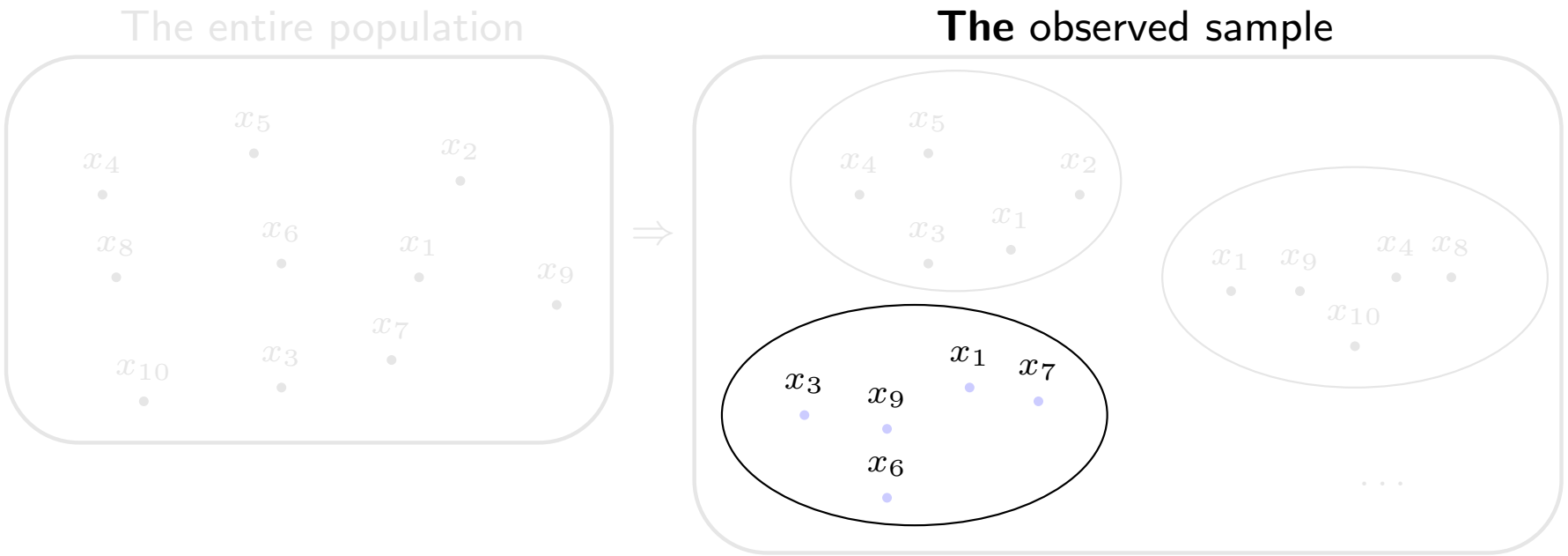
Some samples of size 5



- Population size $N = 10$
- "True" population Mean μ
- "True" Standard deviation σ

- Sample size $n = 5$
- Sample Mean $\bar{x}$
- Sample Standard deviation s_x

Sampling **variability**



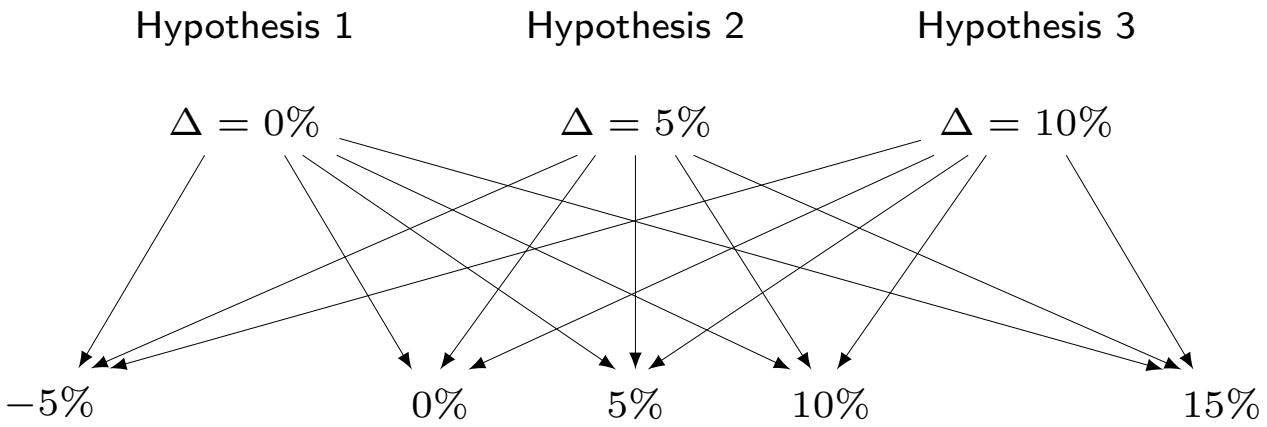
- Population size $N = 10$
- "True" population Mean μ
- "True" Standard deviation σ

$\Leftarrow$

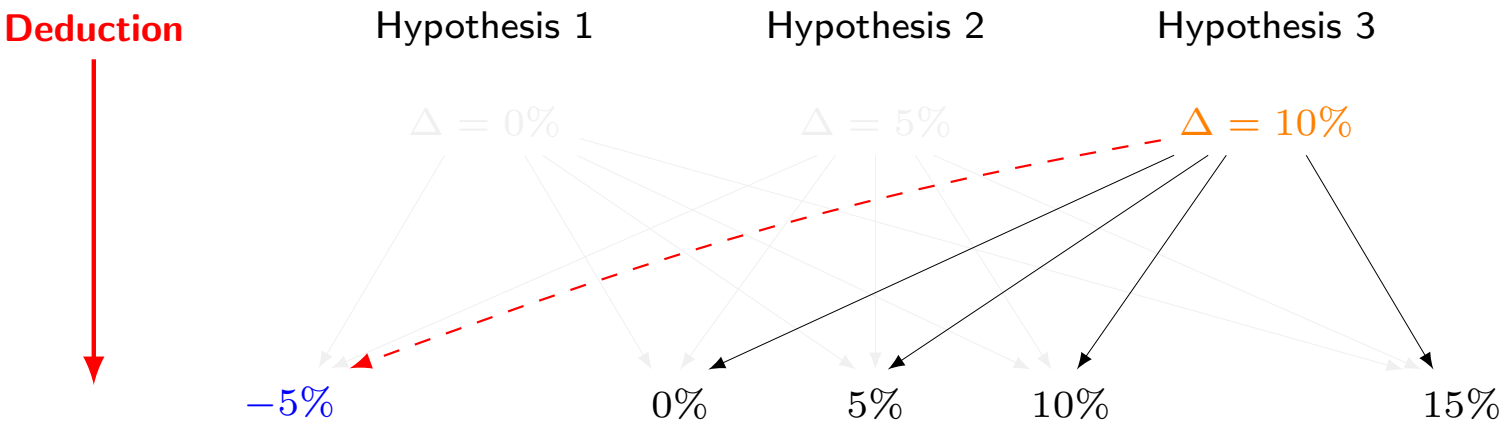
- Sample size $n = 5$
- Sample Mean $\bar{x}$
- Sample Standard deviation s_x

In reality we observe **only one** such sample (out of the many possible — in fact there are **252** different ways of picking **at random** 5 units out of a population of size 10!) and we want to use the information contained in **that** sample to **infer** about the population parameters (e.g. the true mean and standard deviation)

Deductive vs inductive inference

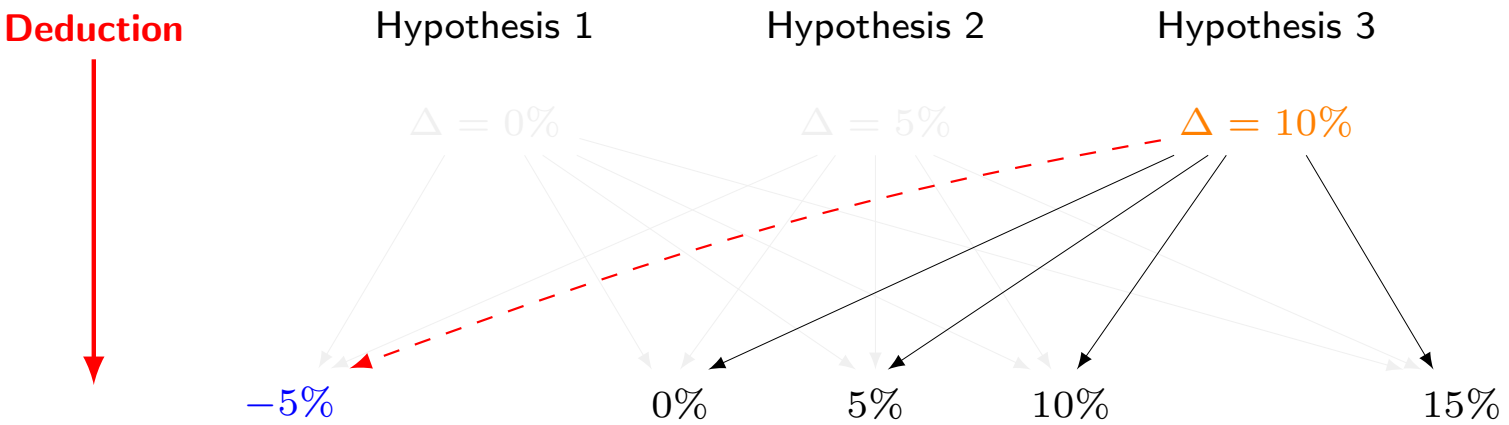


Deductive vs inductive inference



- Standard (frequentist) procedures fix the working hypotheses and, **by deduction**, make inference on the observed data:
  - If my hypothesis is true, what is the probability of randomly selecting the data that I observed? If small, then *deduce* weak support of the evidence to the hypothesis

Deductive vs inductive inference



- Standard (frequentist) procedures fix the working hypotheses and, **by deduction**, make inference on the observed data:
  - If my hypothesis is true, what is the probability of randomly selecting the data that I observed? If small, then *deduce* weak support of the evidence to the hypothesis
  - Assess $\Pr(\text{Observed data} \mid \text{Hypothesis})$
  - Relevant for frequentist summaries, eg p-values, Confidence Intervals, etc
  - **NB**: Comparison with data that could have been observed, but haven't!
  - **If** the observed sample is representative of the DGP and using the sample estimates, **if** we could replicate the experiment under the same conditions, 95% of the times, the estimate for the "true" value will be included in the CI
  - **That's how you interpret a 95% Confidence Interval!**

https://youtu.be/pjvQFtlNQ-M

YouTube GB Search

60% of the time, it works every time....

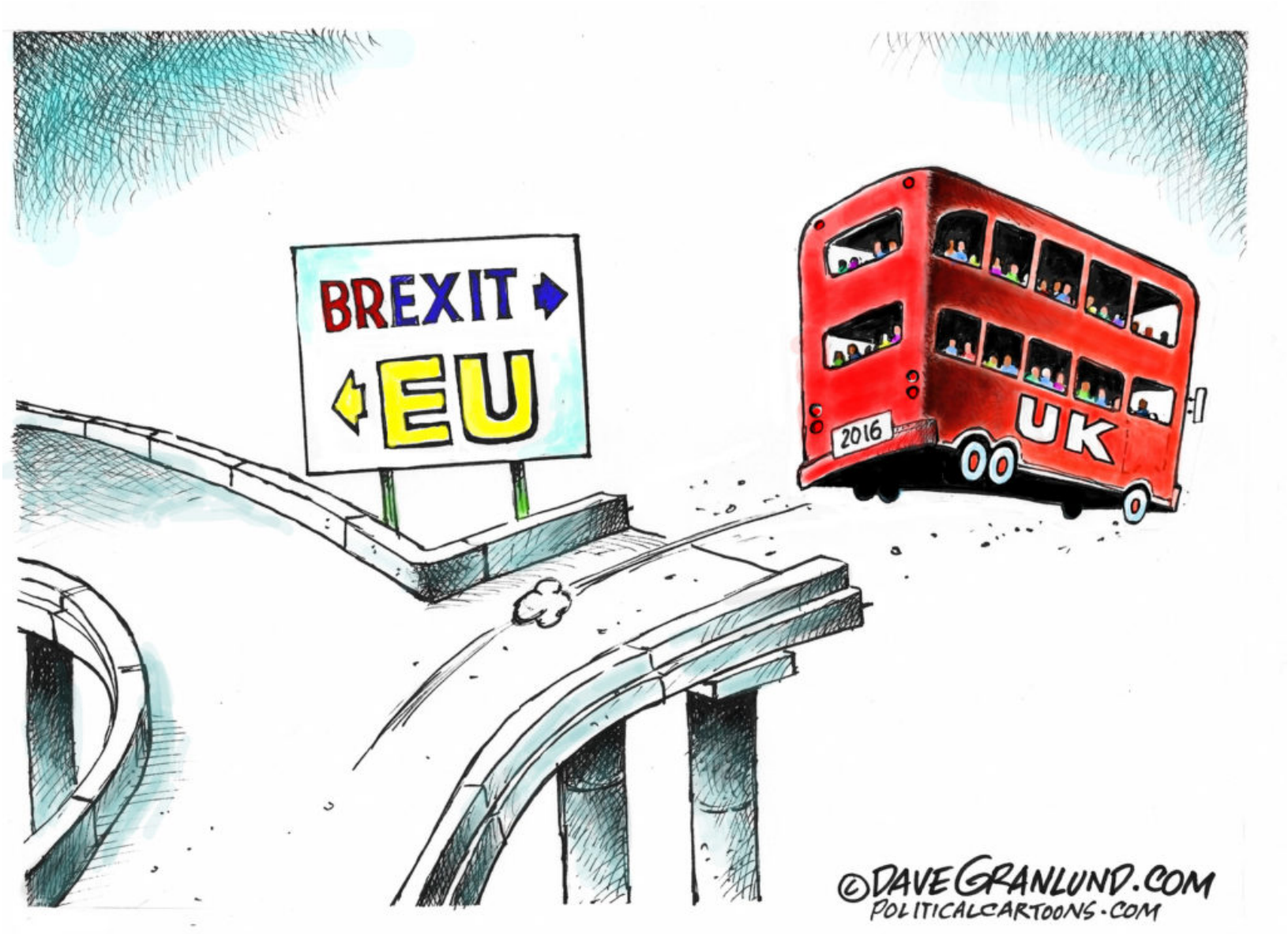
691,867 views

2K 41 SHARE

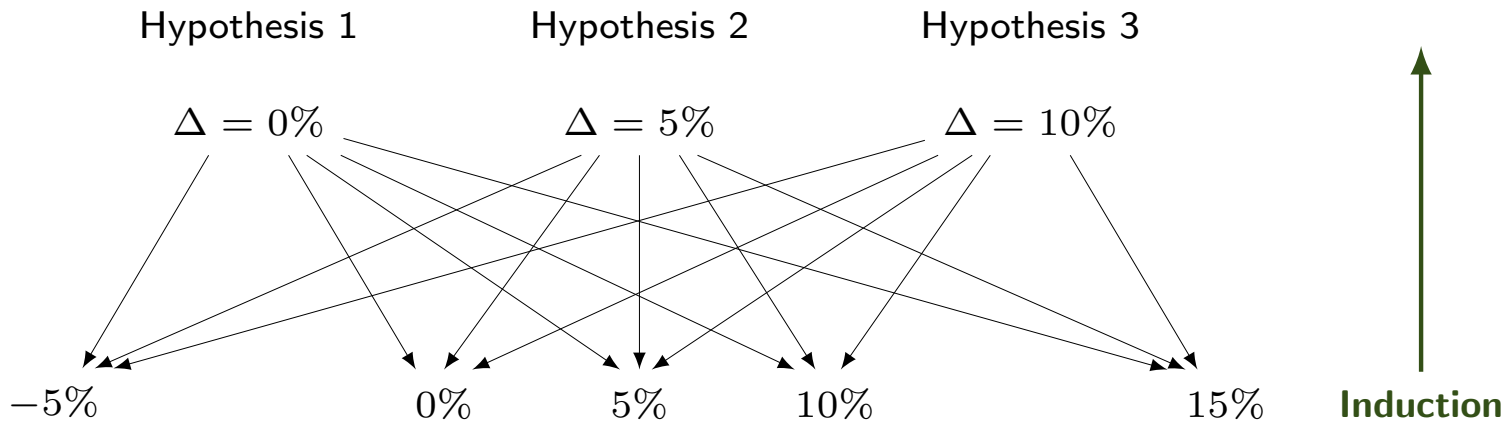
mulletpole
Published on 11 Oct 2010

SUBSCRIBE 949

Is there another way?...

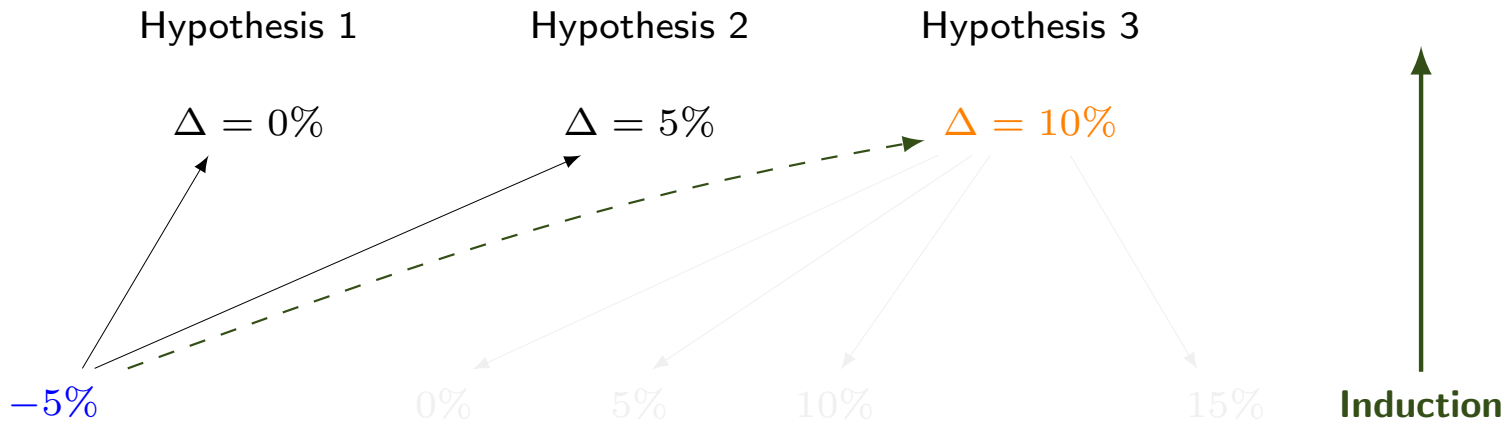


Deductive vs **inductive** inference



- The **Bayesian** philosophy proceeds fixing the value of the observed data and, **by induction**, makes inference on unobservable hypotheses
  - What is the probability of my hypothesis, given the data I observed? If less than the probability of other competing hypotheses, then weak support of the evidence to the hypothesis

Deductive vs **inductive** inference

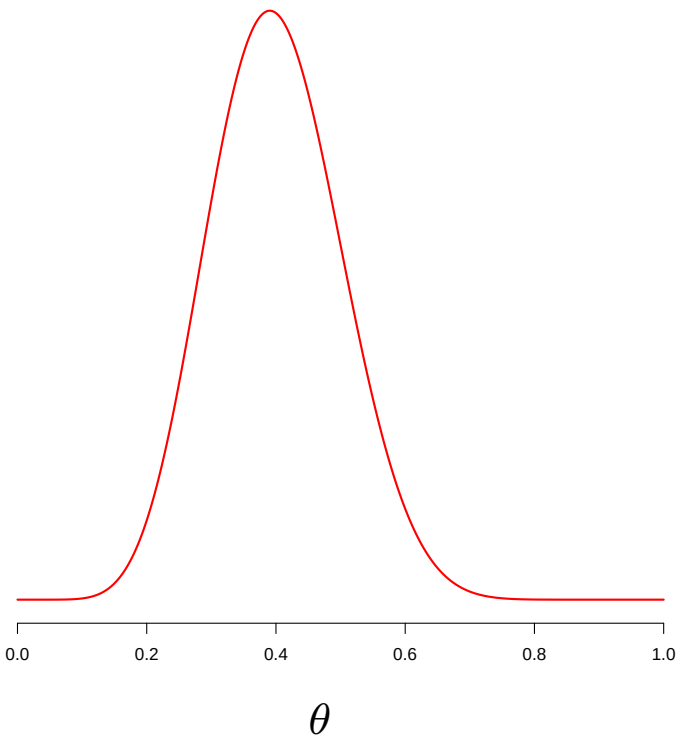


- The **Bayesian** philosophy proceeds fixing the value of the observed data and, **by induction**, makes inference on unobservable hypotheses
  - What is the probability of my hypothesis, given the data I observed? If less than the probability of other competing hypotheses, then weak support of the evidence to the hypothesis
  - Assess $\Pr(\text{Hypothesis} \mid \text{Observed data})$
  - Can express in terms of an **interval** estimate: $\Pr(a \leq \text{parameter} \leq b \mid \text{Data})$
  - **NB**: Unobserved data have no role in the inference!
  - Since the 1990s, rely on computer simulations and a suite of algorithms called **Markov Chain Monte Carlo** (MCMC)
  - Highly generalisable — can throw at it virtually any complexity

Existing knowledge

- Population registries
- Observational studies
- Small/pilot RCTs
- Expert options

$p(\boldsymbol{\theta})$



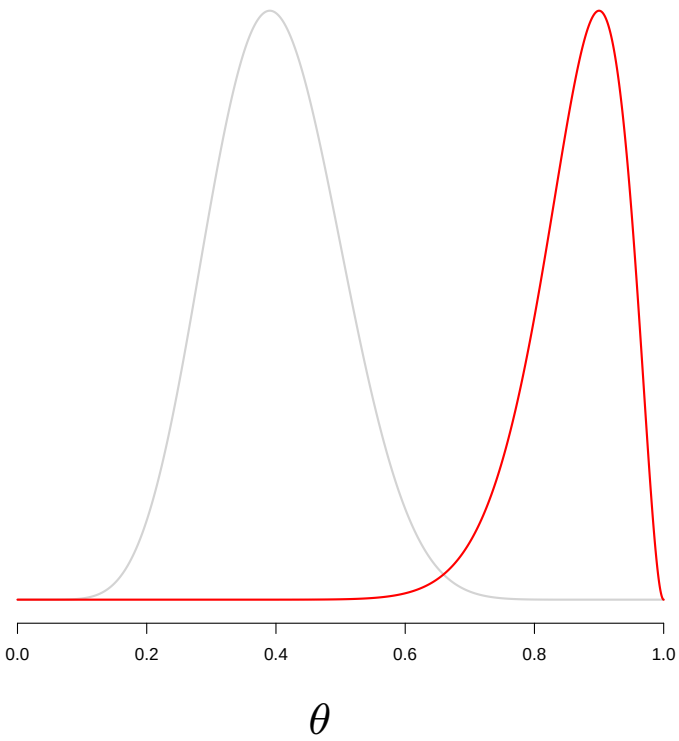
Encode the assumption that a drug has a response rate between 20 and 60%

Existing knowledge
- Population registries
- Observational studies
- Small/pilot RCTs
- Expert options

$p(\boldsymbol{\theta})$

Current data
- Large(r) scale RCT
- Observational study
- Relevant summaries

$p(y \mid \boldsymbol{\theta})$



Observe a study with 150 responders out of 200 patients given the drug

Existing knowledge
- Population registries
- Observational studies
- Small/pilot RCTs
- Expert options

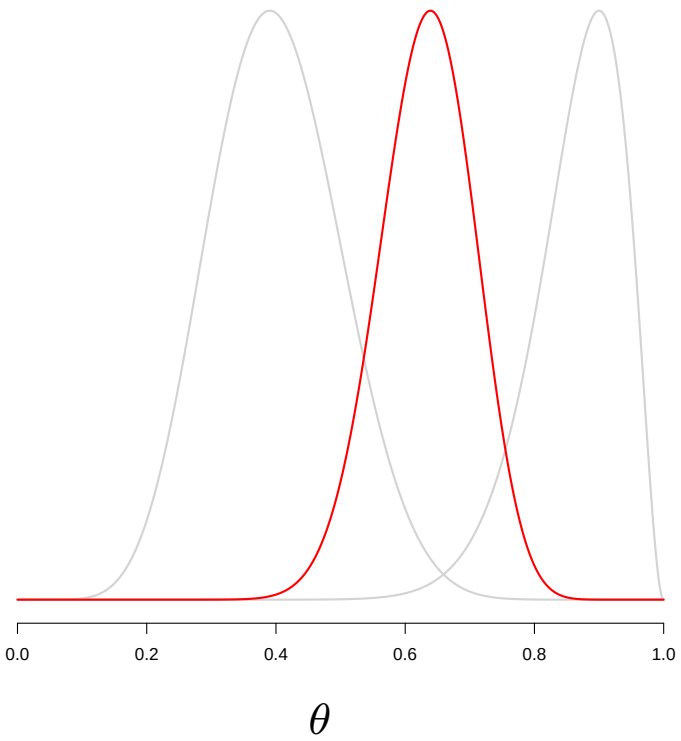
$p(\boldsymbol{\theta})$

Updated knowledge

$p(\boldsymbol{\theta} \mid y)$

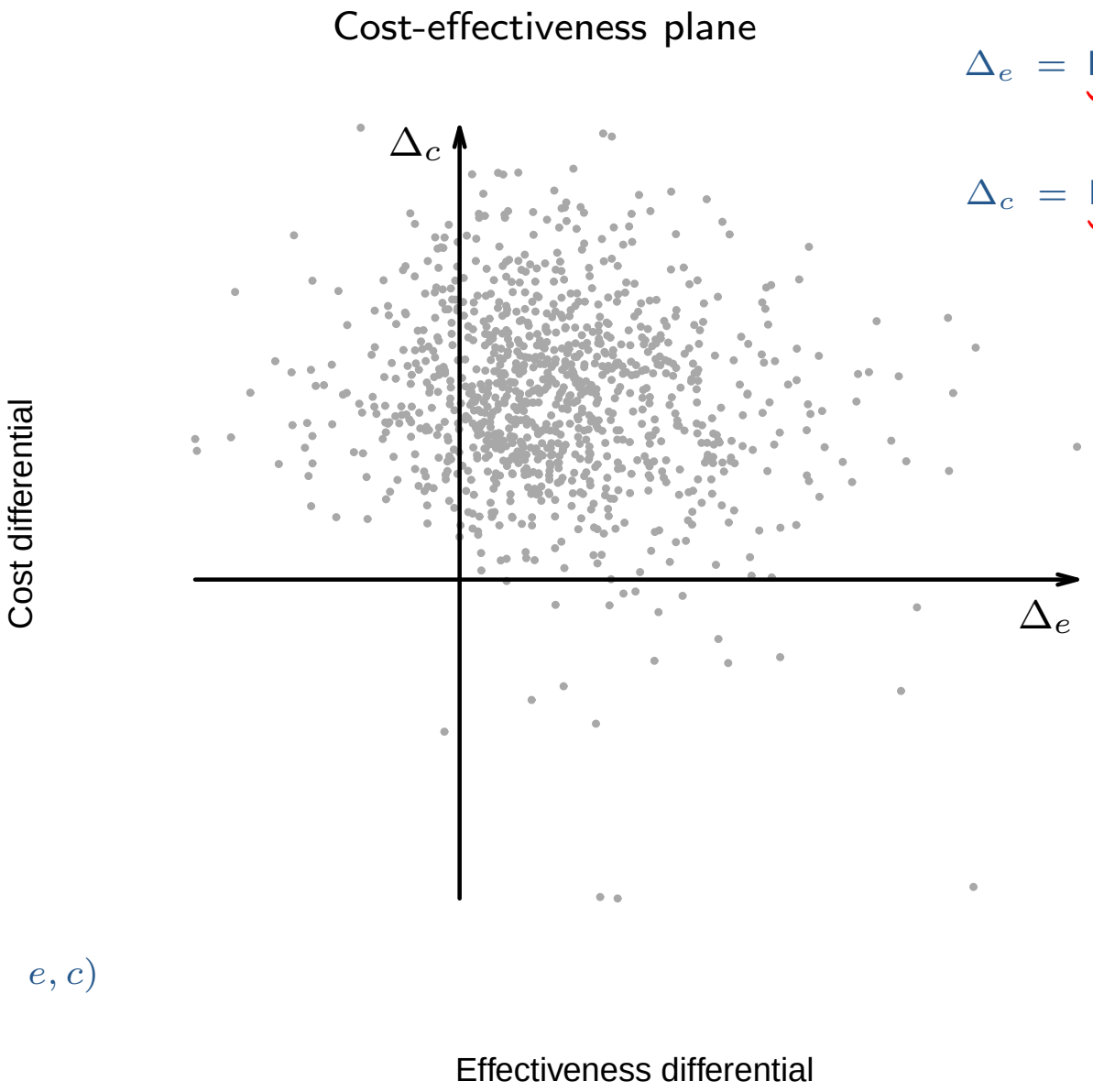
$p(y \mid \boldsymbol{\theta})$

Current data
- Large(r) scale RCT
- Observational study
- Relevant summaries



Update knowledge to describe revised "state of science"

2./3./4. Economic modelling+Decision analysis+Uncertainty analysis*

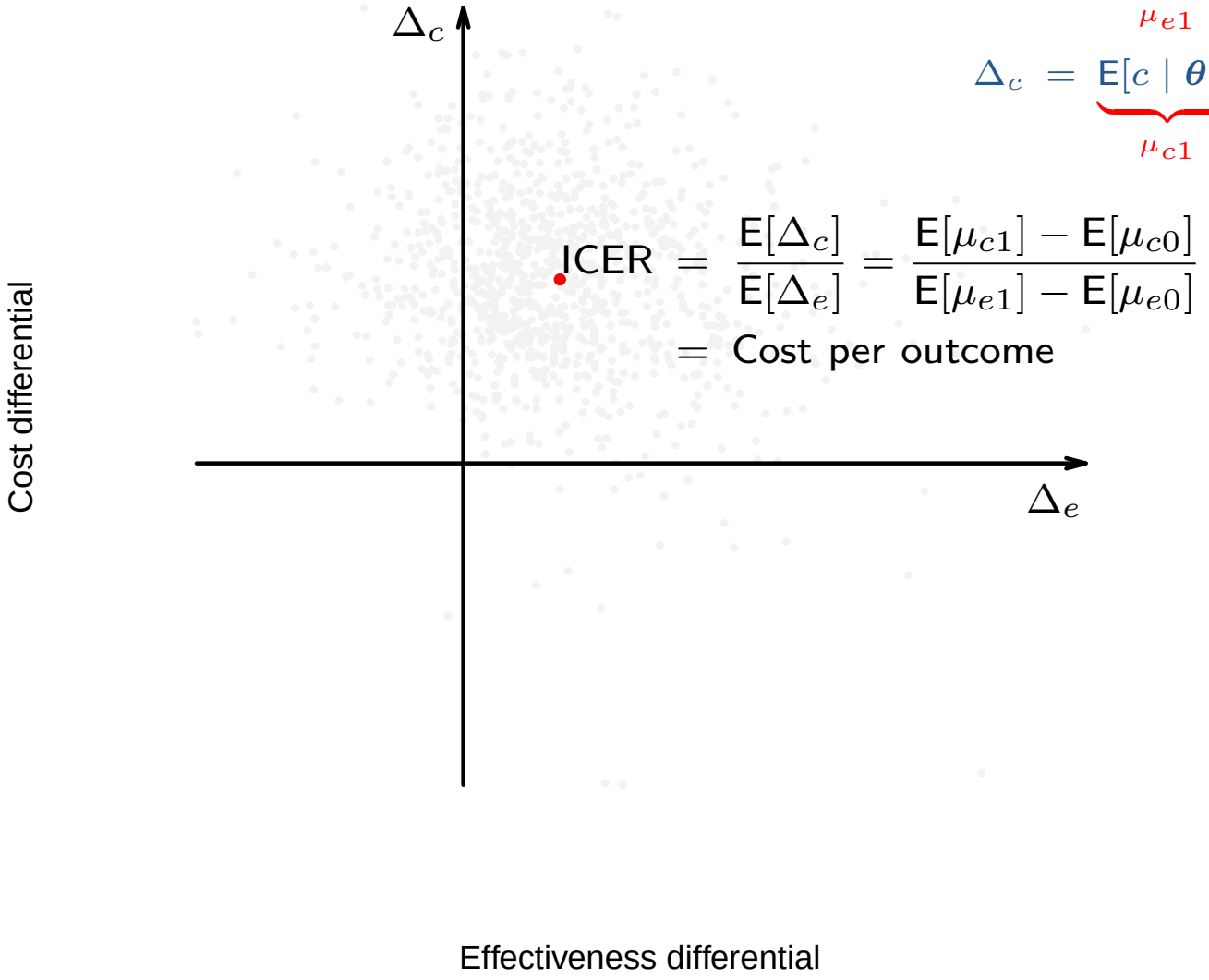


*Induced by $p(\boldsymbol{\theta} \mid e, c)$

Cost-effectiveness plane

$$\Delta_e = \underbrace{\mathsf{E}[e \mid \boldsymbol{\theta}_1]}_{\mu_{e1}} - \underbrace{\mathsf{E}[e \mid \boldsymbol{\theta}_0]}_{\mu_{e0}}$$

$$\Delta_c = \underbrace{\mathsf{E}[c \mid \boldsymbol{\theta}_1]}_{\mu_{c1}} - \underbrace{\mathsf{E}[c \mid \boldsymbol{\theta}_0]}_{\mu_{c0}}$$



Advantages...

- 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!**

Advantages...

- 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**

Advantages...

- 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!**

Part 2

Introduction to missing data

Back to Table of content

Background

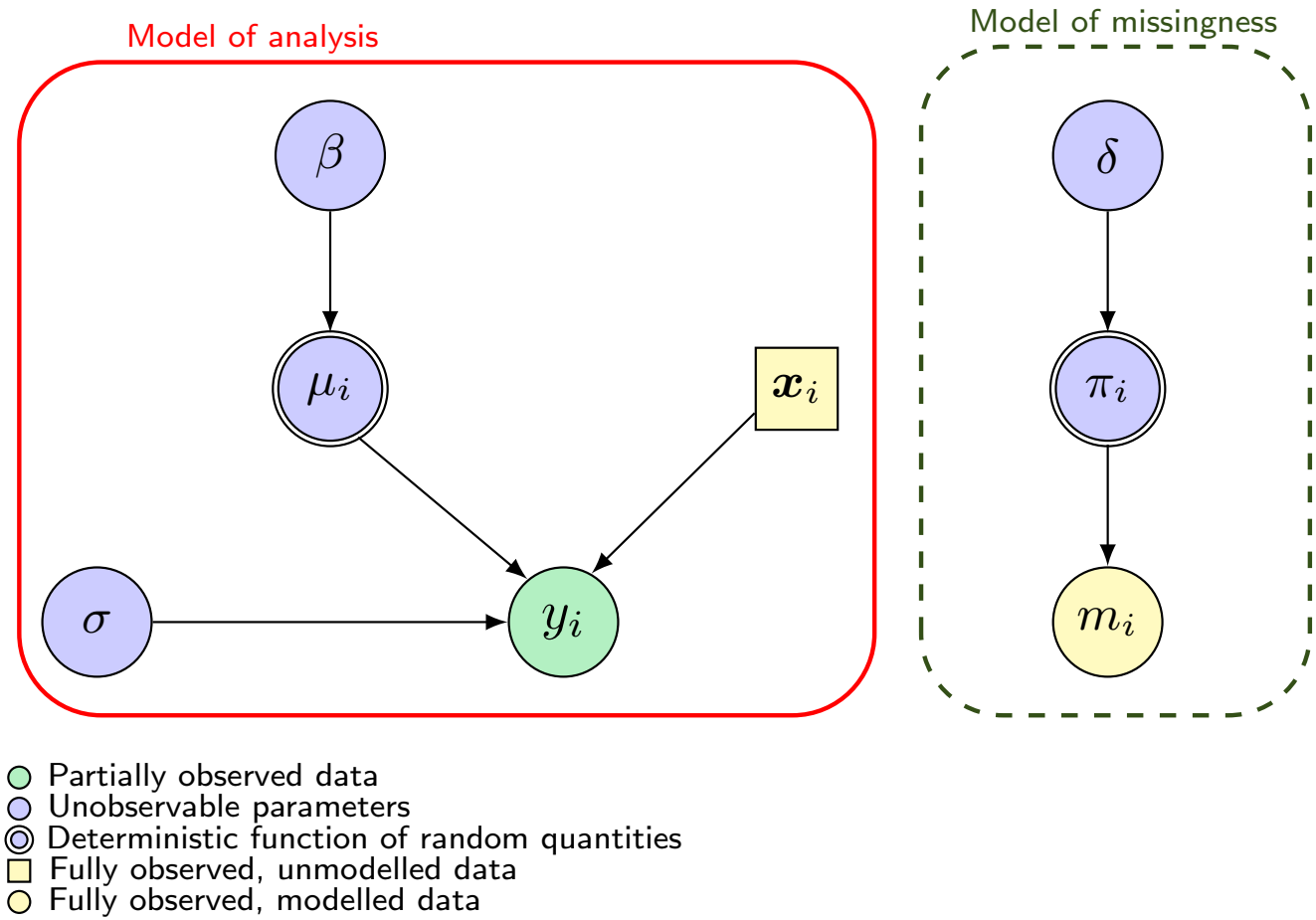
The problems with missing data...

- 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

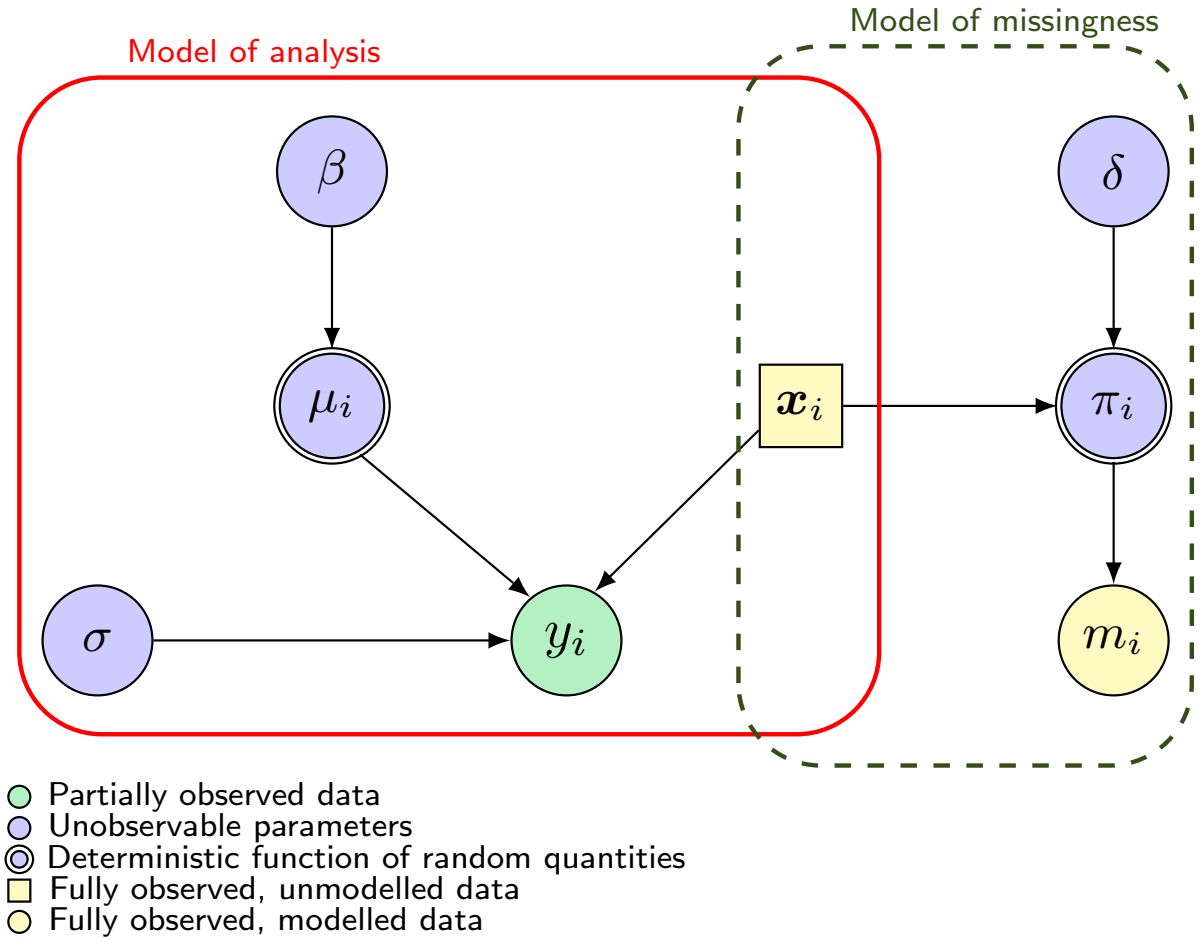
Background

The problems with missing data...

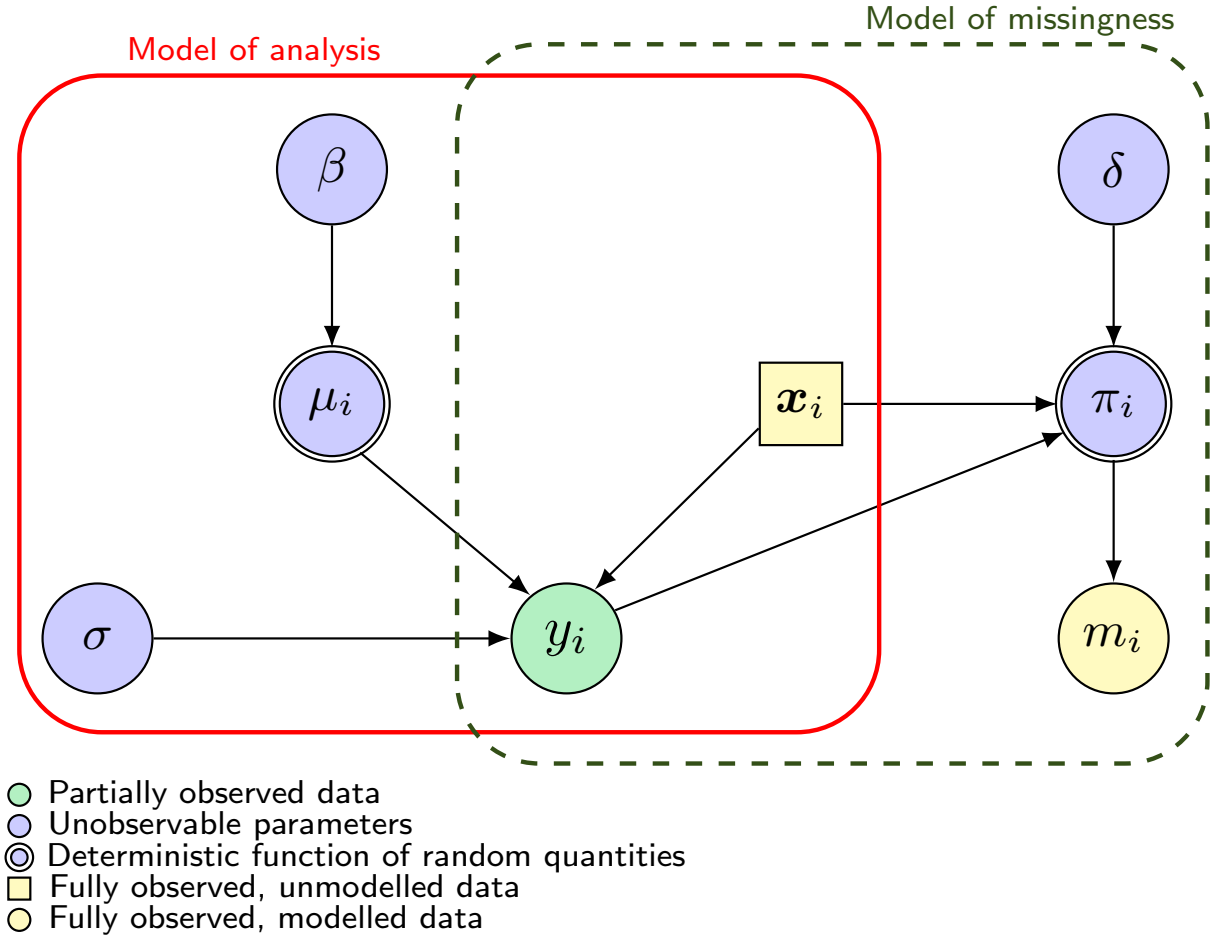
- We plan to observe n_{planned} data points, but end up with a (much) lower number of observations n_{observed}
  - What is the proportion of missing data? Does it matter?...
- We typically don't know **why** the unobserved points are missing and **what** their value might have been
  - Missingness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



- 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
  - $\boldsymbol{\theta}^{\mathrm{MoA}} = (\beta, \sigma)$
  - $\boldsymbol{\theta}^{\mathrm{MoM}} = \delta$

Missing data **analysis methods**

- Complete Case Analysis
  - Elimination of partially observed cases
  - Simple but reduce efficiency and possibly bias parameter estimates
- Inverse probability weighting
  - Weigh the original data (subject to missingness) to account for the fact that the actual sample size is smaller than originally planned
  - Weigh up(down) units that have a high(low) chance of actually being observed
- Single (deterministic) imputation
  - Imputation of missing data with a single value (mean, median, LVCF)
  - Does not account for the uncertainty in the imputation process
- Multiple (stochastic) imputation (MI)
  - Missing data imputed T times to obtain T different imputed datasets
  - Each dataset is analysed and T sets of estimates are derived
  - Parameter estimates are combined into a single quantity
  - The uncertainty due to imputation is incorporated but the validity relies on the correct specification of the imputation model
- "Full Bayesian"
  - Basically extends MI to model formally the missing mechanism

(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

Part 3

Missing data in HEE

Back to Table of content

State of the art

- Decisions in HEE are often informed by within-trial evidence
- Missing data occur frequently in RCTs in both effectiveness and costs

State of the art

- Decisions in HEE are often informed by within-trial evidence
- Missing data occur frequently in RCTs in both effectiveness and costs

- **CCA** has historically represented the standard approach in health economics

State of the art

- Decisions in HEE are often informed by within-trial evidence
- Missing data occur frequently in RCTs in both effectiveness and costs

- **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

State of the art

- Decisions in HEE are often informed by within-trial evidence
- Missing data occur frequently in RCTs in both effectiveness and costs

- **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 have started to appear in the literature (Gabrio et al. (2017). *PharmacoEconomics Open*, 1(2), 79-97)
  - 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

Missing data **in HEE** – sensitivity analysis

- Handling missing data in HEE can be challenging because:
  - Outcomes (costs and QALYs) are typically **correlated** and **non-normally distributed**
  - Limited information is often available about the reasons of missingness
  - **MAR** may be difficult to justify and cannot be checked from the data
- These features have strong implications for the choice of the **missing data method**

Missing data **in HEE** – sensitivity analysis

- Handling missing data in HEE can be challenging because:
  - Outcomes (costs and QALYs) are typically **correlated** and **non-normally distributed**
  - Limited information is often available about the reasons of missingness
  - **MAR** may be difficult to justify and cannot be checked from the data
- These features have strong implications for the choice of the **missing data method**

- It is essential that missingness is addressed in a **principled** way, which entails:
  - A well-defined statistical model for the **complete data**
  - Explicit assumptions about the **missing value mechanism**

Missing data **in HEE** – sensitivity analysis

- Handling missing data in HEE can be challenging because:
  - Outcomes (costs and QALYs) are typically **correlated** and **non-normally distributed**
  - Limited information is often available about the reasons of missingness
  - **MAR** may be difficult to justify and cannot be checked from the data
- These features have strong implications for the choice of the **missing data method**
- It is essential that missingness is addressed in a **principled** way, which entails:
  - A well-defined statistical model for the **complete data**
  - Explicit assumptions about the **missing value mechanism**
- Selection models and Pattern Mixture models offer a convenient framework to conduct **sensitivity analysis** under **MNAR**
  - Both rely on a combination of **unverifiable assumptions** and **informative prior distributions**
  - **Selection models** focus on modelling the missingness mechanism directly
  - **Pattern mixture models** focus on modelling the model by missingness pattern

Missing data **in HEE** – Conclusions

- Often, little or no information is available about missingness in within-trial CEAs

Missing data **in HEE** – Conclusions

- Often, little or no information is available about missingness in within-trial CEAs
- Restrict the analysis to a single missingness scenario (**MAR**) is unlikely to provide a realistic assessment of the cost-effectiveness of the interventions and could mislead decision-makers

Missing data **in HEE** – Conclusions

- Often, little or no information is available about missingness in within-trial CEAs
- Restrict the analysis to a single missingness scenario (**MAR**) is unlikely to provide a realistic assessment of the cost-effectiveness of the interventions and could mislead decision-makers
- Selection and pattern mixture models represent possible choices to perform **sensitivity analysis** to **MNAR**
  - Rely on untestable assumptions about the unobserved data
  - Useful to assess the robustness of the results to a range of **plausible** departures from **MAR**

Missing data **in HEE** – Conclusions

- Often, little or no information is available about missingness in within-trial CEAs
- Restrict the analysis to a single missingness scenario (**MAR**) is unlikely to provide a realistic assessment of the cost-effectiveness of the interventions and could mislead decision-makers
- Selection and pattern mixture models represent possible choices to perform **sensitivity analysis** to **MNAR**
  - Rely on untestable assumptions about the unobserved data
  - Useful to assess the robustness of the results to a range of **plausible** departures from **MAR**
- The Bayesian approach allows the incorporation of **external evidence** into the model, which may be crucial for
  - The selection of the missingness assumptions to explore
  - The assessment and quantification of the impact that missingness uncertainty has on decision-making

Part 4

Case Study

Back to Table of content

Motivating example: MenSS trial

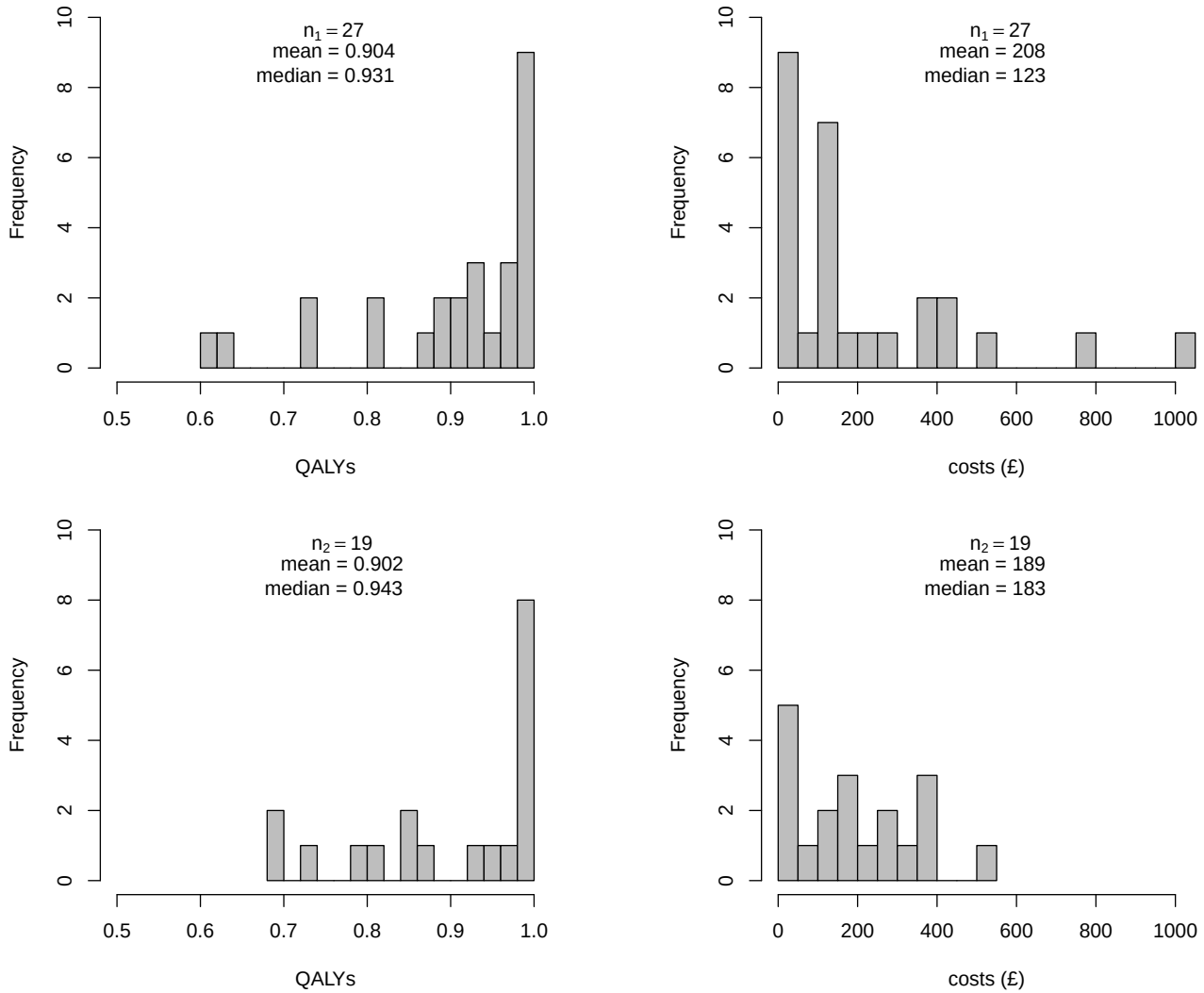
- The MenSS pilot RCT evaluates the cost-effectiveness of a new digital intervention to reduce the incidence of STI in young men with respect to the SOC
  - QALYs calculated from utilities (EQ-5D 3L)
  - Total costs calculated from different components (no baseline)

Motivating example: MenSS trial

- The MenSS pilot RCT evaluates the cost-effectiveness of a new digital intervention to reduce the incidence of STI in young men with respect to the SOC
  - QALYs calculated from utilities (EQ-5D 3L)
  - Total costs calculated from different components (no baseline)

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

- The MenSS pilot RCT evaluates the cost-effectiveness of a new digital intervention to reduce the incidence of STI in young men with respect to the SOC
  - QALYs calculated from utilities (EQ-5D 3L)
  - Total costs calculated from different components (no baseline)

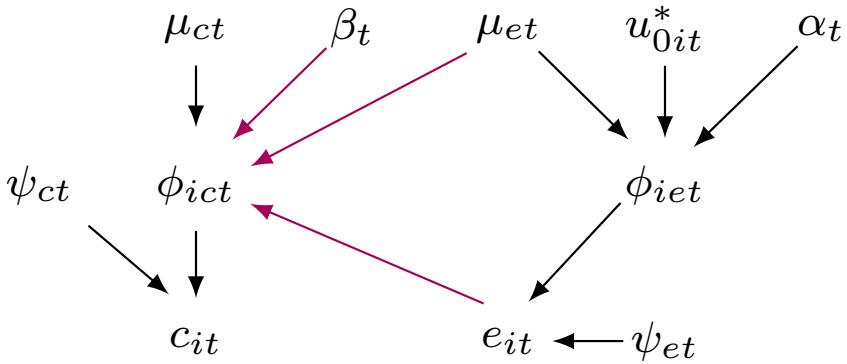


1 Bivariate Normal

- Simpler and closer to "standard" frequentist model
- Account for correlation between QALYs and costs

Conditional model for $c \mid e$

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



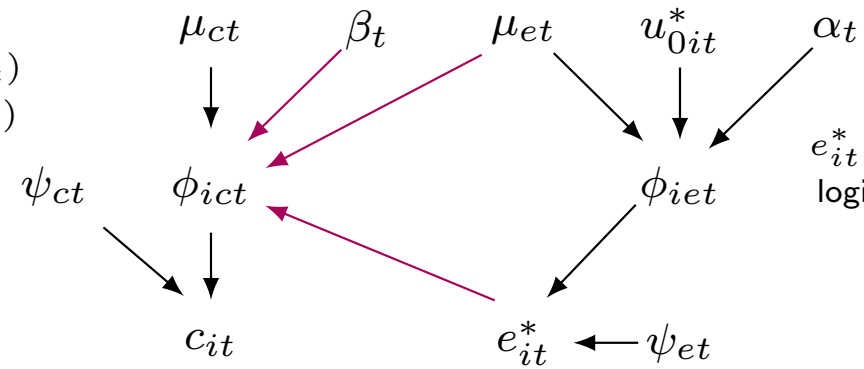
Marginal model for e

$$e_{it} \sim \text{Normal}(\phi_{eit}, \psi_{et})$$
$$\phi_{eit} = \mu_{et} + \alpha_t(u_{0it} - \bar{u}_{0t})$$
$$= \mu_{et} + \alpha_t u^*_{0it}$$

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

Conditional model for $c \mid e^*$

$$c_{it} \mid e^*_{it} \sim \text{Gamma}(\psi_{ct}\phi_{cit}, \psi_{ct})$$
$$\log(\phi_{cit}) = \mu_{ct} + \beta_t(e^*_{it} - \mu_{et})$$



Marginal model for e^*

$$e^*_{it} \sim \text{Beta}\left(\phi_{eit}\psi_{et}, (1-\phi_{eit})\psi_{et}\right)$$
$$\begin{aligned}\text{logit}(\phi_{eit}) &= \mu_{et} + \alpha_t(u_{0it} - \bar{u}_{0t}) \\ &= \mu_{et} + \alpha_t u^*_{0it}\end{aligned}$$

Gabrio et al. (2019). *Statistics in Medicine*, 38(8), 1399-1420

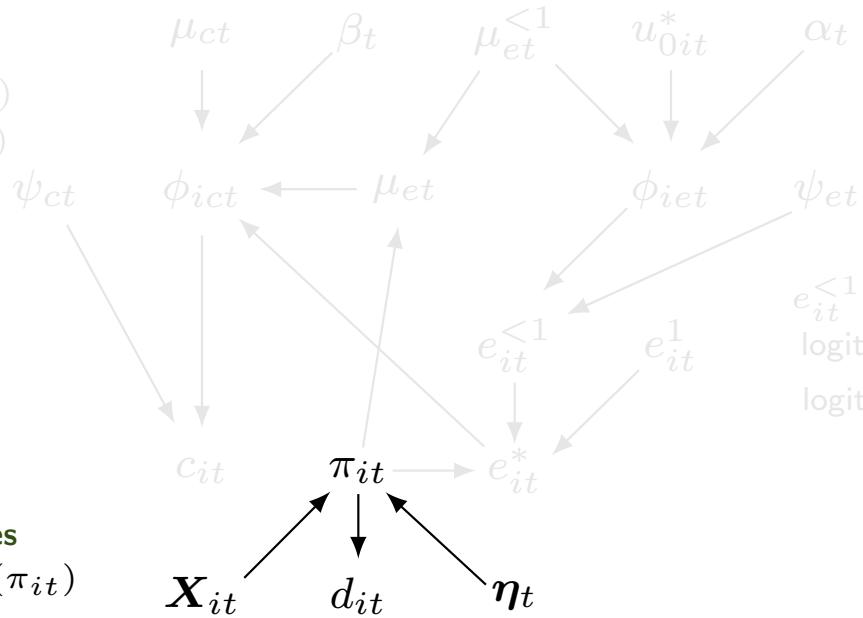
1. **Bivariate Normal**
   - Simpler and closer to "standard" frequentist model
   - Account for correlation between QALYs and costs
2. **Beta-Gamma**
   - Account for correlation between outcomes
   - Model the relevant ranges: QALYs $\in (0,1)$ and costs $\in (0,\infty)$
   - **But**: needs to rescale observed data $e^*_{it} = (e_{it} - \epsilon)$ to avoid spikes at 1
3. **Hurdle model**
   - Model e_{it} as a **mixture** to account for correlation between outcomes, model the relevant ranges and account for structural values
   - May expand to account for partially observed baseline utility u_{0it}

Conditional model for $c \mid e^*$

$$c_{it} \mid e^*_{it} \sim \text{Gamma}(\psi_{ct}\phi_{cit}, \psi_{ct})$$
$$\log(\phi_{cit}) = \mu_{ct} + \beta_t(e^*_{it} - \mu_{et})$$

Mixture model for e

$$e^1_{it} := 1$$
$$e^{<1}_{it} \sim \text{Beta}\left(\phi_{eit}\psi_{et}, (1-\phi_{eit})\psi_{et}\right)$$
$$\text{logit}(\phi_{eit}) = \mu^{<1}_{et} + \alpha_t(u_{0it} - \bar{u}_{0t})$$
$$\text{logit}(\phi_{eit}) = \mu^{<1}_{et} + \alpha_t u^*_{0it}$$
$$e^*_{it} = \pi_{it}e^1_{it} + (1-\pi_{it})e^{<1}_{it}$$
$$\mu_{et} = (1-\bar{\pi}_t)\mu^{<1}_{et} + \bar{\pi}_t$$



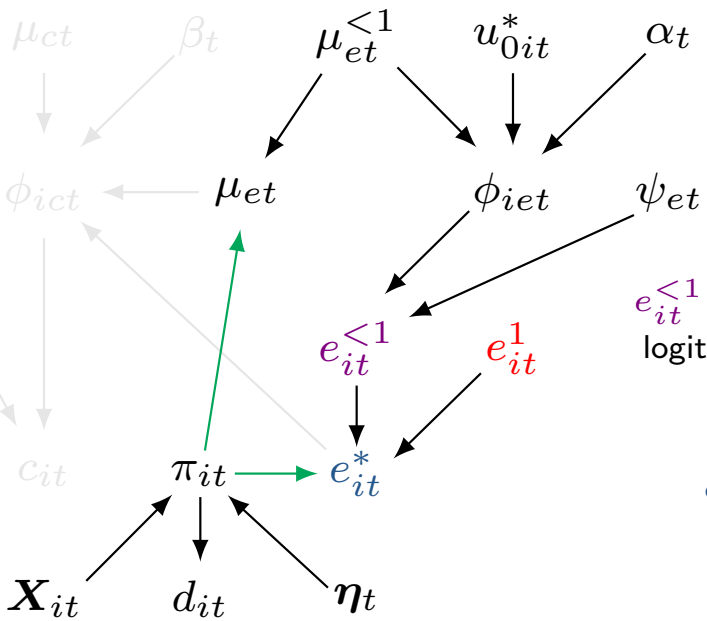
Model for the structural ones

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

1. **Bivariate Normal**
   - Simpler and closer to "standard" frequentist model
   - Account for correlation between QALYs and costs
2. **Beta-Gamma**
   - Account for correlation between outcomes
   - Model the relevant ranges: QALYs $\in (0,1)$ and costs $\in (0,\infty)$
   - **But**: needs to rescale observed data $e^*_{it} = (e_{it} - \epsilon)$ to avoid spikes at 1
3. **Hurdle model**
   - Model e_{it} as a **mixture** to account for correlation between outcomes, model the relevant ranges and account for structural values
   - May expand to account for partially observed baseline utility u_{0it}

Conditional model for $c \mid e^*$

$$c_{it} \mid e^*_{it} \sim \text{Gamma}(\psi_{ct}\phi_{cit}, \psi_{ct})$$
$$\log(\phi_{cit}) = \mu_{ct} + \beta_t(e^*_{it} - \mu_{et})$$



Mixture model for e

$$e^1_{it} := 1$$
$$e^{<1}_{it} \sim \text{Beta}\left(\phi_{eit}\psi_{et}, (1-\phi_{eit})\psi_{et}\right)$$
$$\text{logit}(\phi_{eit}) = \mu^{<1}_{et} + \alpha_t(u_{0it} - \bar{u}_{0t}) = \mu^{<1}_{et} + \alpha_t u^*_{0it}$$

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

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

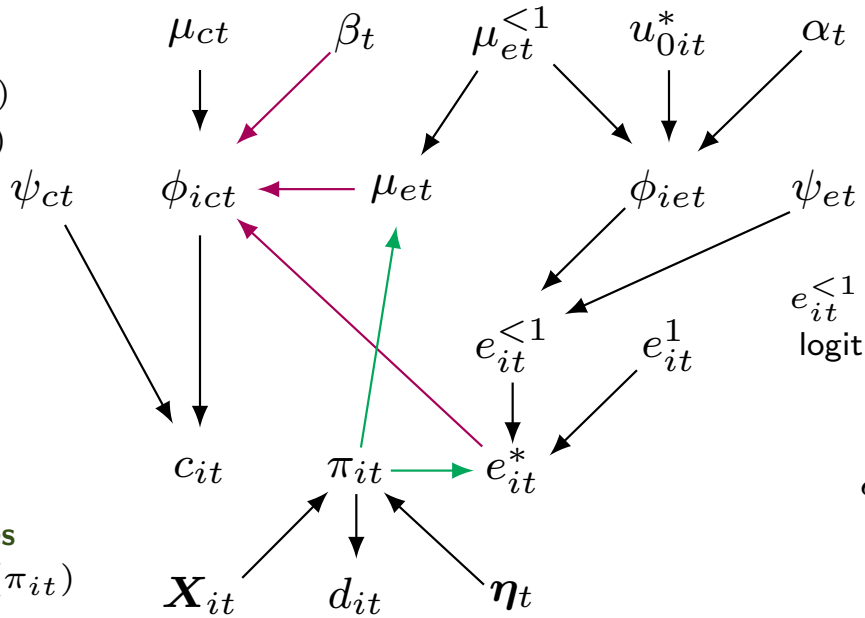
Model for the structural ones

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

1. **Bivariate Normal**
   - Simpler and closer to "standard" frequentist model
   - Account for correlation between QALYs and costs
2. **Beta-Gamma**
   - Account for correlation between outcomes
   - Model the relevant ranges: QALYs $\in (0,1)$ and costs $\in (0,\infty)$
   - **But**: needs to rescale observed data $e^*_{it} = (e_{it} - \epsilon)$ to avoid spikes at 1
3. **Hurdle model**
   - Model e_{it} as a **mixture** to account for correlation between outcomes, model the relevant ranges and account for structural values
   - May expand to account for partially observed baseline utility u_{0it}

Conditional model for $c \mid e^*$

$$c_{it} \mid e^*_{it} \sim \text{Gamma}(\psi_{ct}\phi_{cit}, \psi_{ct})$$
$$\log(\phi_{cit}) = \mu_{ct} + \beta_t(e^*_{it} - \mu_{et})$$



Mixture model for e

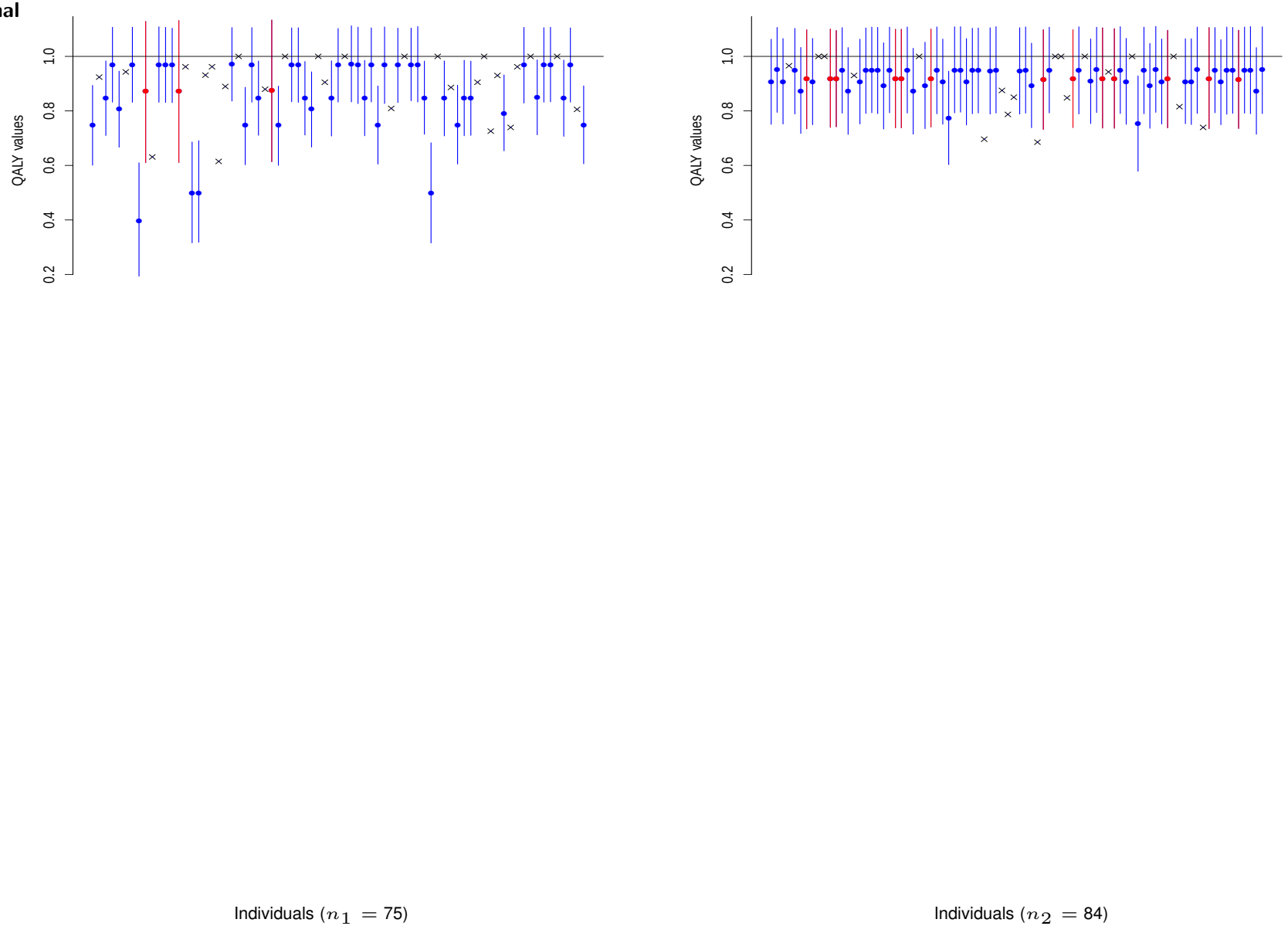
$$e^1_{it} := 1$$
$$e^{<1}_{it} \sim \text{Beta}\left(\phi_{eit}\psi_{et}, (1-\phi_{eit})\psi_{et}\right)$$
$$\text{logit}(\phi_{eit}) = \mu^{<1}_{et} + \alpha_t(u_{0it} - \bar{u}_{0t}) = \mu^{<1}_{et} + \alpha_t u^*_{0it}$$
$$e^*_{it} = \pi_{it}e^1_{it} + (1-\pi_{it})e^{<1}_{it}$$
$$\mu_{et} = (1-\bar{\pi}_t)\mu^{<1}_{et} + \bar{\pi}_t$$

Model for the structural ones

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

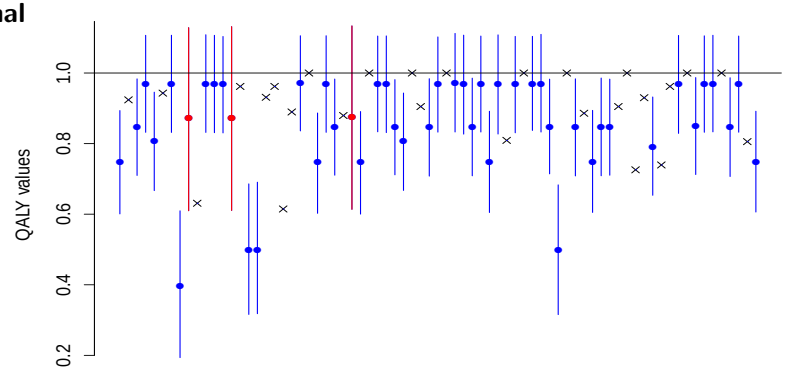
Bayesian multiple imputation (under MAR)

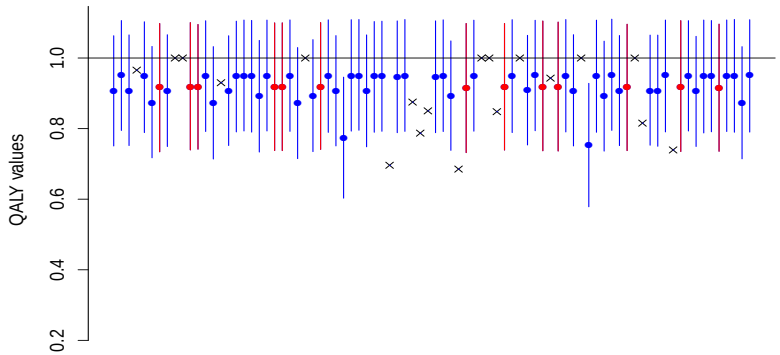
Bivariate Normal



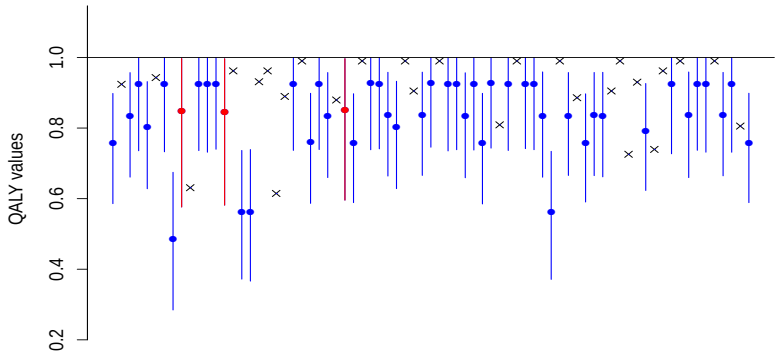
Bayesian multiple imputation (under MAR)

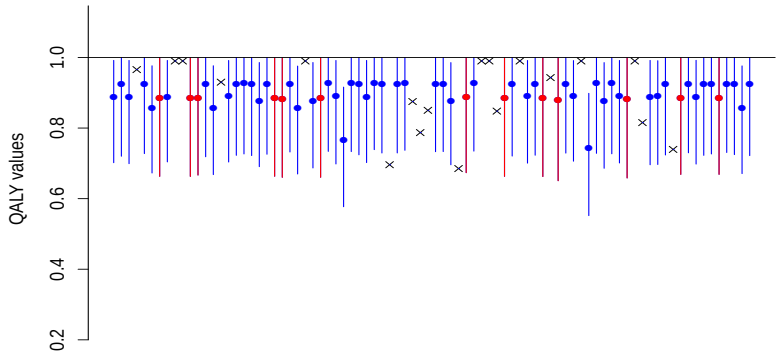
Bivariate Normal





Beta-Gamma





Individuals ($n_1 = 75$)

Individuals ($n_2 = 84$)

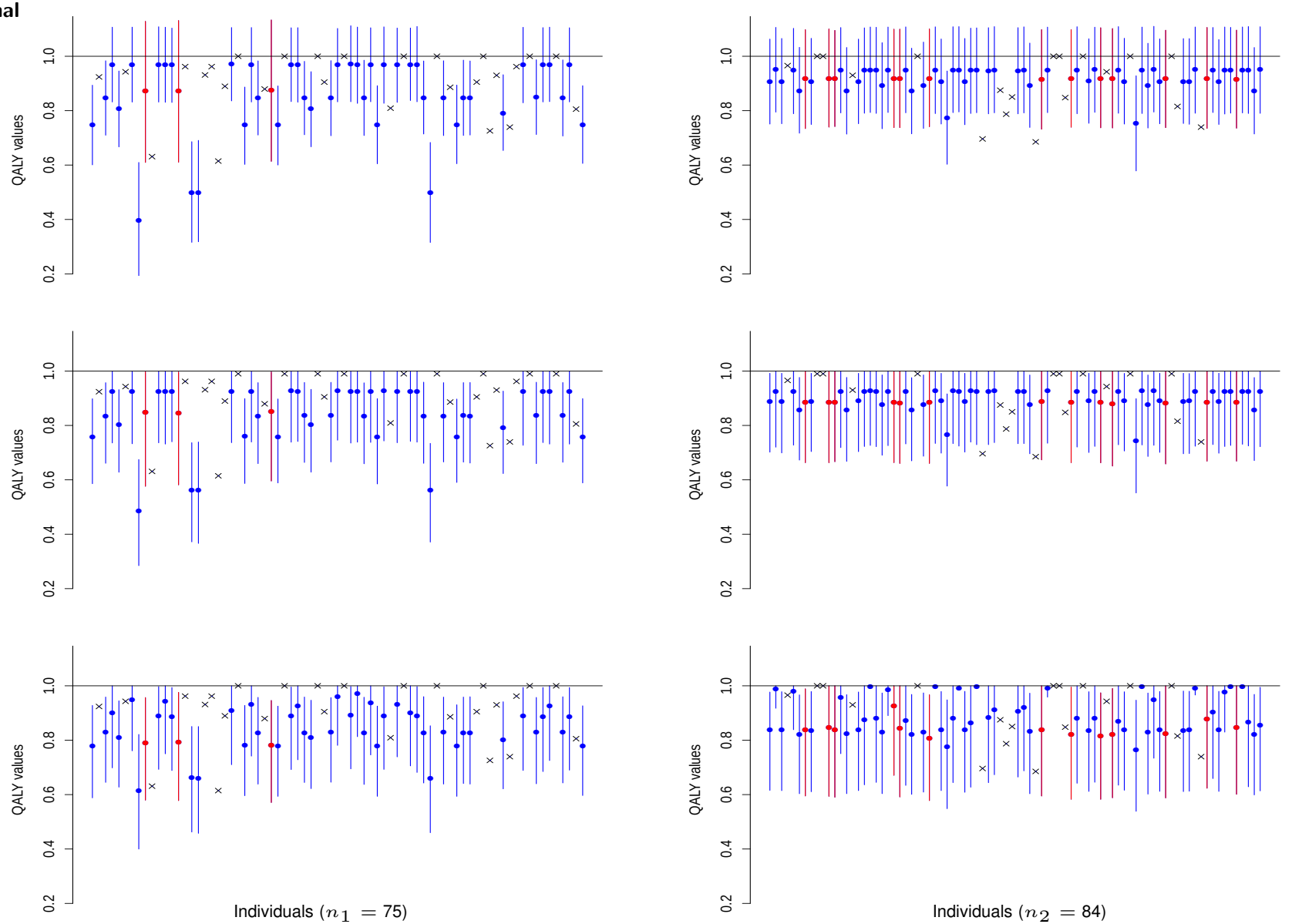
Imputed, observed baseline
Imputed, missing baseline
Observed

Bayesian multiple imputation (under MAR)

Bivariate Normal

Beta-Gamma

Hurdle model



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

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

MNAR1. Set $d_{it} = 1$ for all individuals with unit observed baseline utility

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

MNAR1. Set $d_{it} = 1$ for all individuals with unit observed baseline utility

MNAR2. Set $d_{it} = 0$ for all individuals with unit observed baseline utility

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

MNAR1. Set $d_{it} = 1$ for all individuals with unit observed baseline utility

MNAR2. Set $d_{it} = 0$ for all individuals with unit observed baseline utility

MNAR3. Set $d_{it} = 1$ for the $n_{01} = 13$ individuals with $u_{0i1} = 1$ and $d_{it} = 0$ for the $n_{02} = 22$ individuals with $u_{0i2} = 1$

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

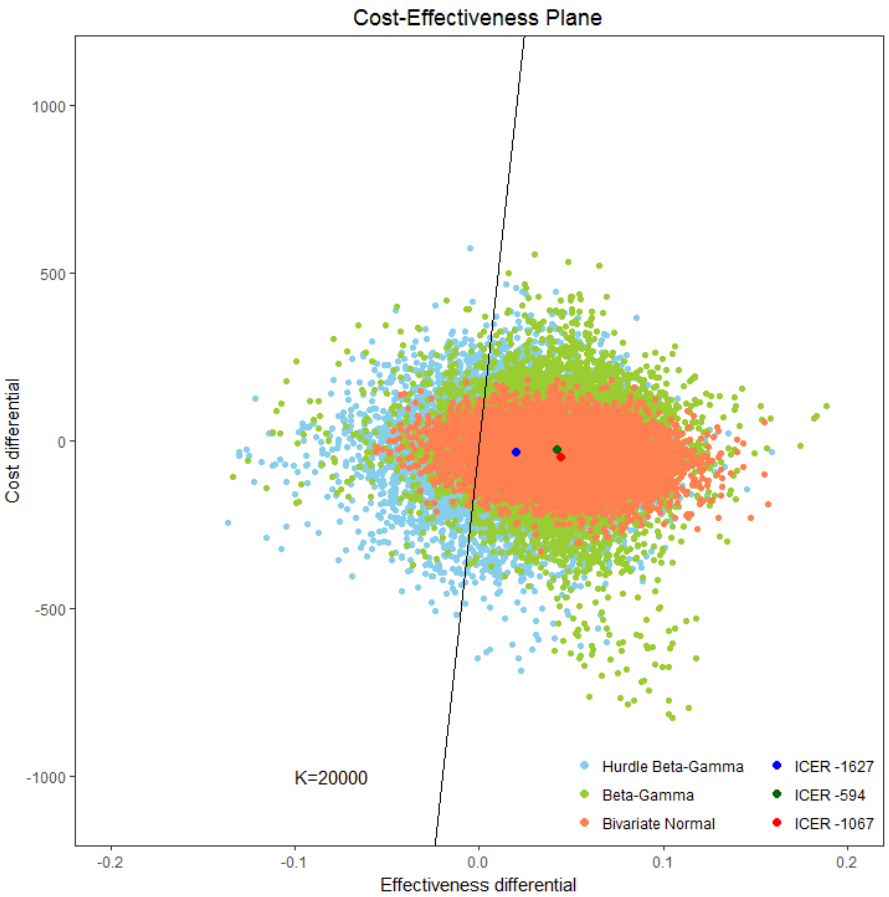
MNAR1. Set $d_{it} = 1$ for all individuals with unit observed baseline utility

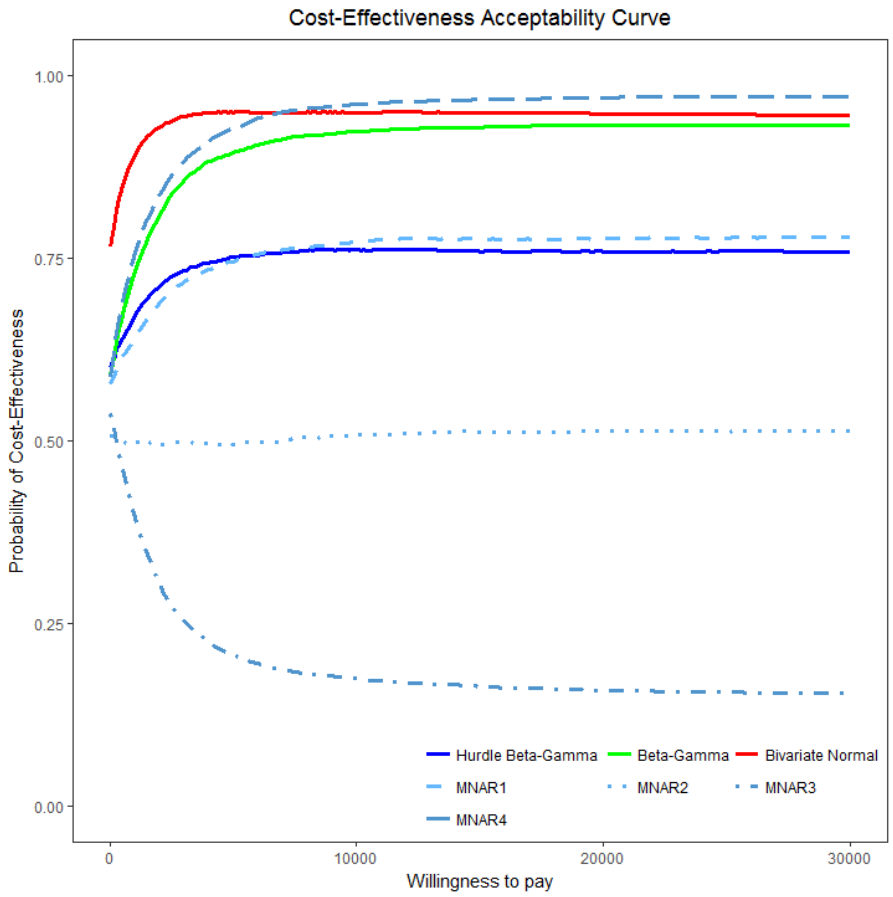
MNAR2. Set $d_{it} = 0$ for all individuals with unit observed baseline utility

MNAR3. Set $d_{it} = 1$ for the $n_{01} = 13$ individuals with $u_{0i1} = 1$ and $d_{it} = 0$ for the $n_{02} = 22$ individuals with $u_{0i2} = 1$

MNAR4. Set $d_{it} = 0$ for the $n_{01} = 13$ individuals with $u_{0i1} = 1$ and $d_{it} = 1$ for the $n_{02} = 22$ individuals with $u_{0i2} = 1$

Cost-effectiveness analysis





MenSS trial: discussion

- HEE data are subject to some complexities that are typically ignored by the "standard" approach, which could lead to biased results
- A Bayesian approach allows to increase model complexity to jointly account for these with relatively little expansion to the basic model
- MAR can be used as reference assumption but plausible MNAR departures should be explored in sensitivity analysis
- Possible to expand the framework to a longitudinal setting to handle missingness more efficiently (Gabrio et al. (2019). *RJSS A*, early view)

Part 5

MissingHE

Back to Table of content

missingHE: an R package to deal with missing data in **HEE**

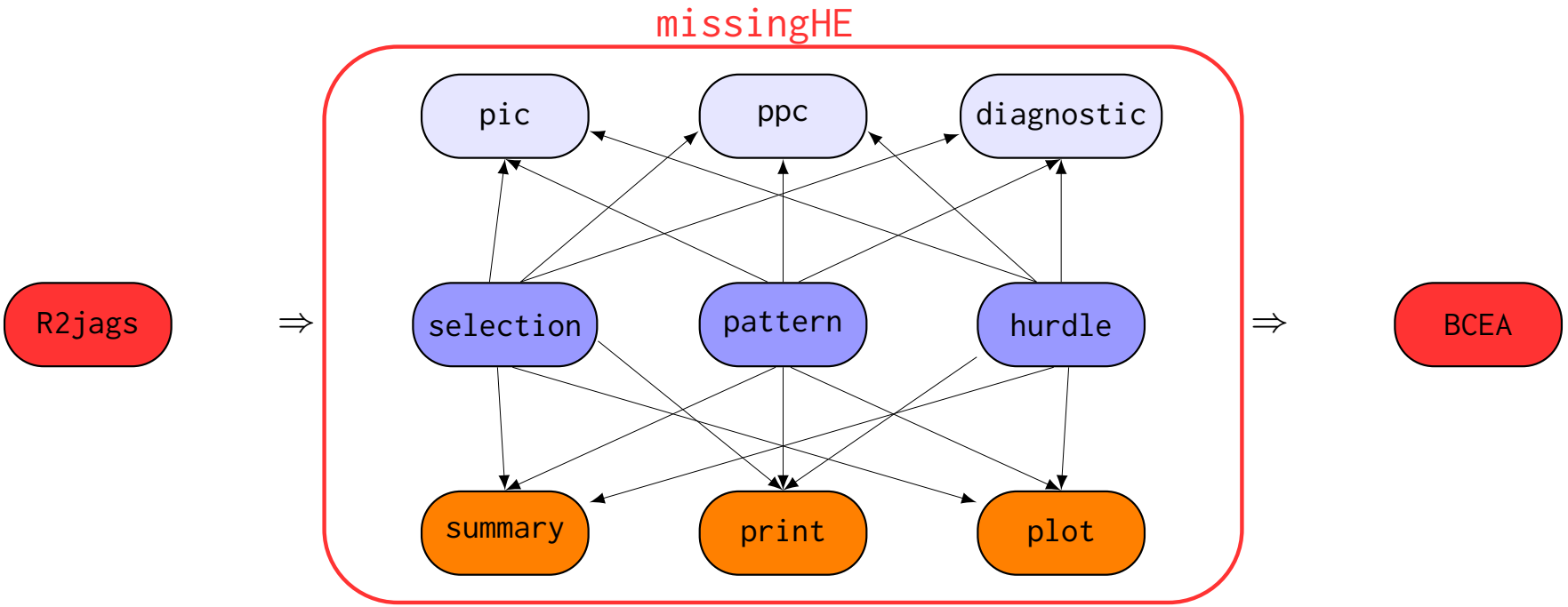
1. Install the most recent R and Rstudio versions
2. Install JAGS from Martyn Plummer's repository: https://sourceforge.net/projects/mcmc-jags/files/JAGS/
3. Install the package **missingHE** from within R or Rstudio, via the package installer or by typing in the command line

```
> install.packages("missingHE", dependencies = TRUE)
```

4. The dependencies = TRUE option will automatically install all the packages on which the functions in the **missingHE** package rely.

missingHE: an R package to deal with missing data in **HEE**

- The **missingHE** package provides different functions to fit Bayesian models for missing data in trial-based HEE



GitHub repository: https://github.com/AnGabrio/missingHE
CRAN repository: https://cran.r-project.org/web/packages/missingHE

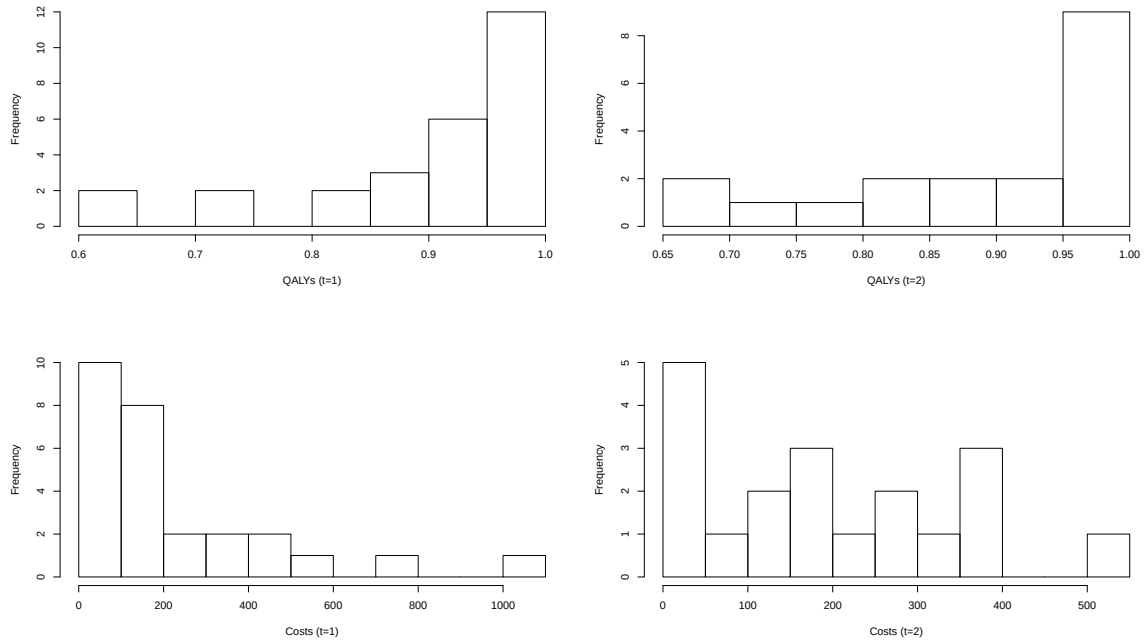
- Load the package using the command `library` or `require`

```
> library("missingHE")
```

- MenSS data already available in the object `MenSS` when loading the package

```
> str(MenSS)
## 'data.frame': 159 obs. of  8 variables:
##  $ id        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ u.0       : num  0.725 0.848 0.848 1 0.796 ...
##  $ e         : num  NA 0.924 NA NA NA 0.943 NA NA NA 0.631 ...
##  $ c         : num  NA 0 NA NA NA 0 NA NA NA 516 ...
##  $ age       : int  23 23 27 27 18 25 48 30 24 27 ...
##  $ ethnicity : Factor w/ 2 levels "0","1": 1 1 2 1 1 2 2 2 2 1 ...
##  $ employment: Factor w/ 2 levels "0","1": 1 2 1 1 1 2 2 2 2 2 ...
##  $ t         : int  1 1 1 1 1 1 1 1 1 1 ...
```

- Check QALYs and costs distribution by treatment arm using the `hist` command



- Example: Fit selection model using Beta and Gamma for QALYs and Costs.
- First need to subtract/add "small" constant to data to avoid 1 and 0 values

```
> MenSS.star=MenSS
> MenSS.star$e=MenSS$e-0.05
> MenSS.star$c=MenSS$c+0.05
```

- Fit selection model using the function `selection` under MNAR for e

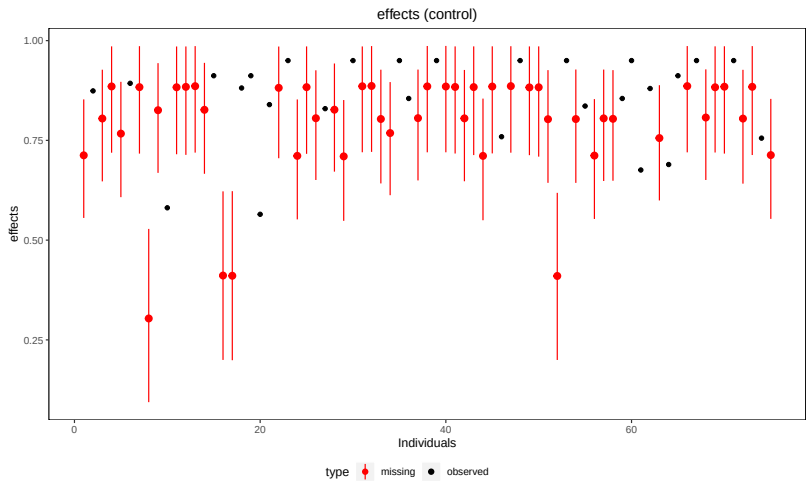
```
> set.seed(123)
> BG.sel=selection(data = MenSS.star, model.eff = e ~ u.0, model.cost = c ~ e,
+   model.me = me ~ e, model.mc = mc ~ age, type = "MNAR", n.chains = 2,
+   n.iter = 10000, n.burnin = 1000, dist_e = "beta", dist_c = "gamma")

## module glm loaded
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 410
##    Unobserved stochastic nodes: 246
##    Total graph size: 4048
##
## Initializing model
```

- A quick inspection of the posterior mean results using the `print` command

```
> print(BG.sel, only.means = TRUE)
##              mean      sd    2.5%   97.5%  Rhat n.eff
## mu_c[1] 257.744 100.284 133.267 505.955 1.004   600
## mu_c[2] 284.191 161.685 108.871 731.672 1.001 18000
## mu_e[1]   0.826   0.018   0.789   0.858 1.001 18000
## mu_e[2]   0.865   0.020   0.819   0.899 1.001  2800
```

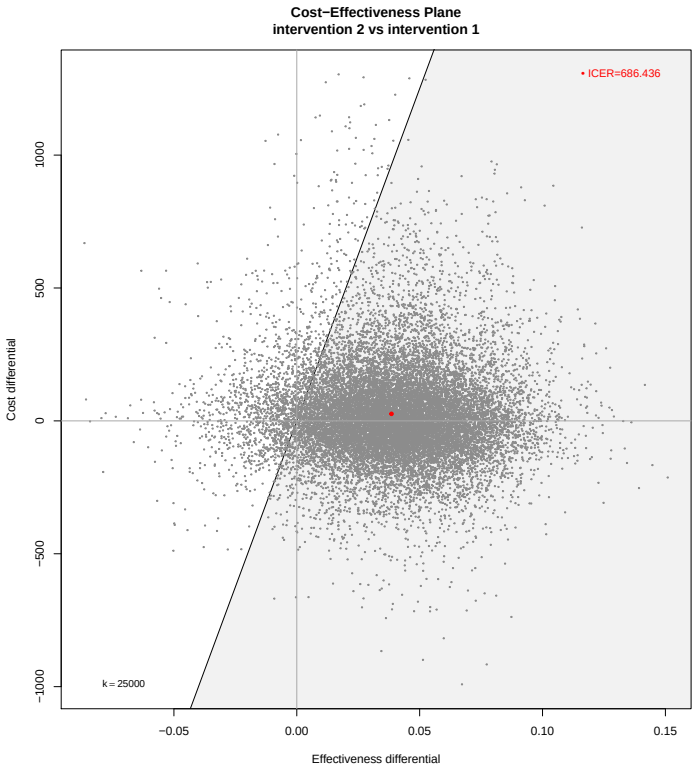
- Next, we use the function `plot` to check the imputed QALYs in the control group for the Beta-Gamma model

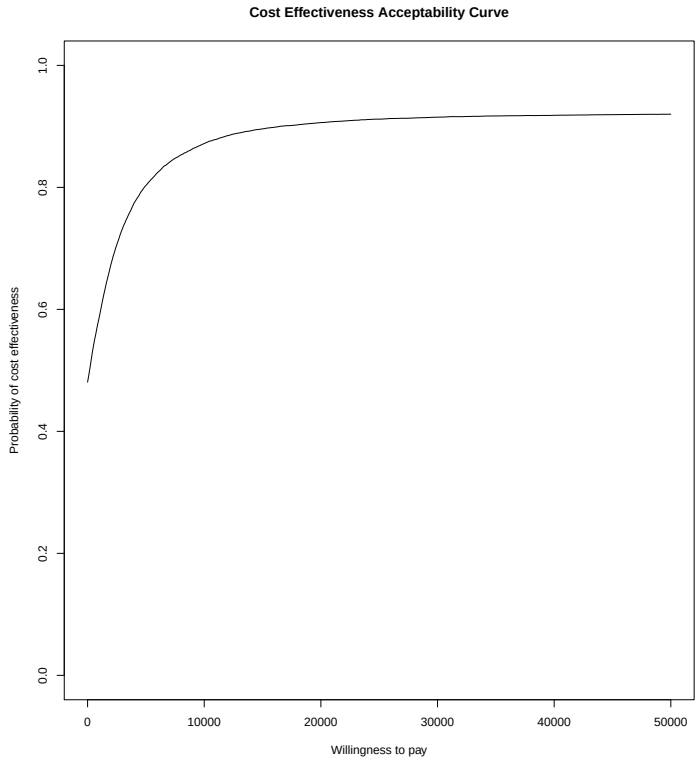
```
> plot(BG.sel, outcome = "effects_arm1")
```



- Conduct the economic evaluation using the function `ceplane.plot` and `ceac.plot` in the package **BCEA** (must be loaded first)

```
> library(BCEA)
> par(mfrow=c(1,2))
> ceplane.plot(BG.sel$cea)
> ceac.plot(BG.sel$cea)
```





missingHE: an R package to deal with missing data in **HEE**

- Specifies a set of pre-defined JAGS models using the **R2jags** package
- Is linked to the **BCEA** package, which provides summary HEE results
- **missingHE**:
  - Uses the functions `hurdle`, `selection` and `pattern` to implement alternative models under different missingness assumptions
  - Assesses model using the functions `pic`, `ppc` and `diagnostic`
  - Summarises the results from the model using the functions `summary`, `print` and `plot`
- Instructions on how to use the functions of **missingHE** to fit and assess different types of models can be accessed by typing `help` on the different functions of the package
- A short course with practicals and solutions is available at `https://github.com/AnGabrio/short-course`
- Code available at `https://github.com/AnGabrio/missingHE`