

Métodos para tratar los datos faltantes en la evaluación de tecnologías de la salud

(Thanks/blame to Google Translate)

Gianluca Baio

Andrea Gabrio

University College London
Department of Statistical Science

g.baio@ucl.ac.uk

<http://www.statistica.it/gianluca>
<https://github.com/giabaio>

ucakgab@ucl.ac.uk

<https://sites.google.com/site/agabriostats/>
<https://github.com/AnGabrio>

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

XXXIX Jornadas de Economía de la Salud, Albacete

Tuesday 11 June 2019

Outline

- 1 Introduction to (Bayesian) modelling in HTA
- 2 Introduction to missing data
- 3 Missing data in HTA
- 4 Examples
- 5 Practicals

Part 1

Introduction to (Bayesian) modelling in HTA

[Back to Table of content](#)

Disclaimer...

Before I begin:

Disclaimer...

Before I begin:

- My **personal** view of the world:
Statisticians should be in charge of everything.

Disclaimer...

Before I begin:

- My **personal** view of the world:
Statisticians should be in charge of everything.
- And actually, come to think about it:
Bayesian *Statisticians should be in charge of all Statisticians.*

Disclaimer...

Before I begin:

- My **personal** view of the world:
Statisticians should be in charge of everything.
- And actually, come to think about it:
Bayesian *Statisticians should be in charge of all Statisticians.*
- So I probably will be very annoying throughout the day...¹

¹But luckily no non-Bayesian 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)

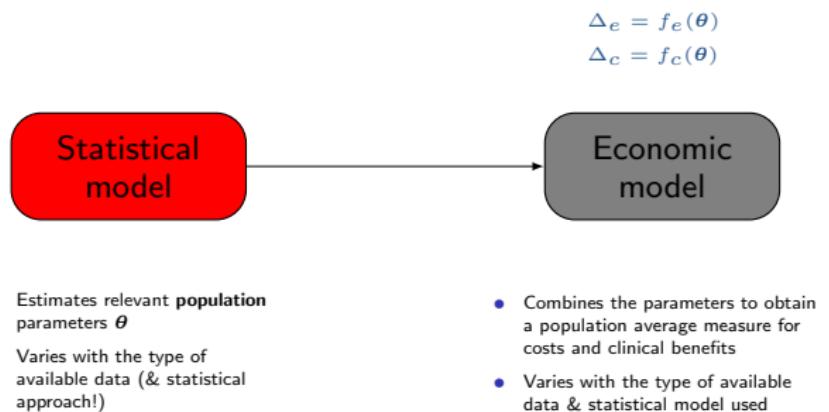
Objective: Combine **costs** & **benefits** of a given intervention into a rational scheme for allocating resources

Statistical model

- Estimates relevant **population** parameters θ
- Varies with the type of available data (& statistical approach!)

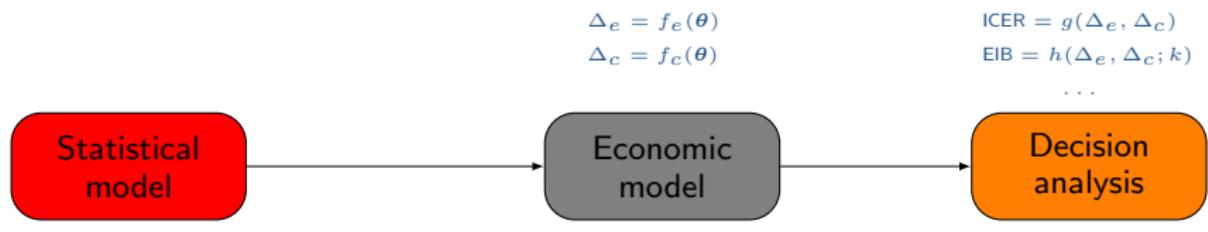
Health technology assessment (HTA)

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



Health technology assessment (HTA)

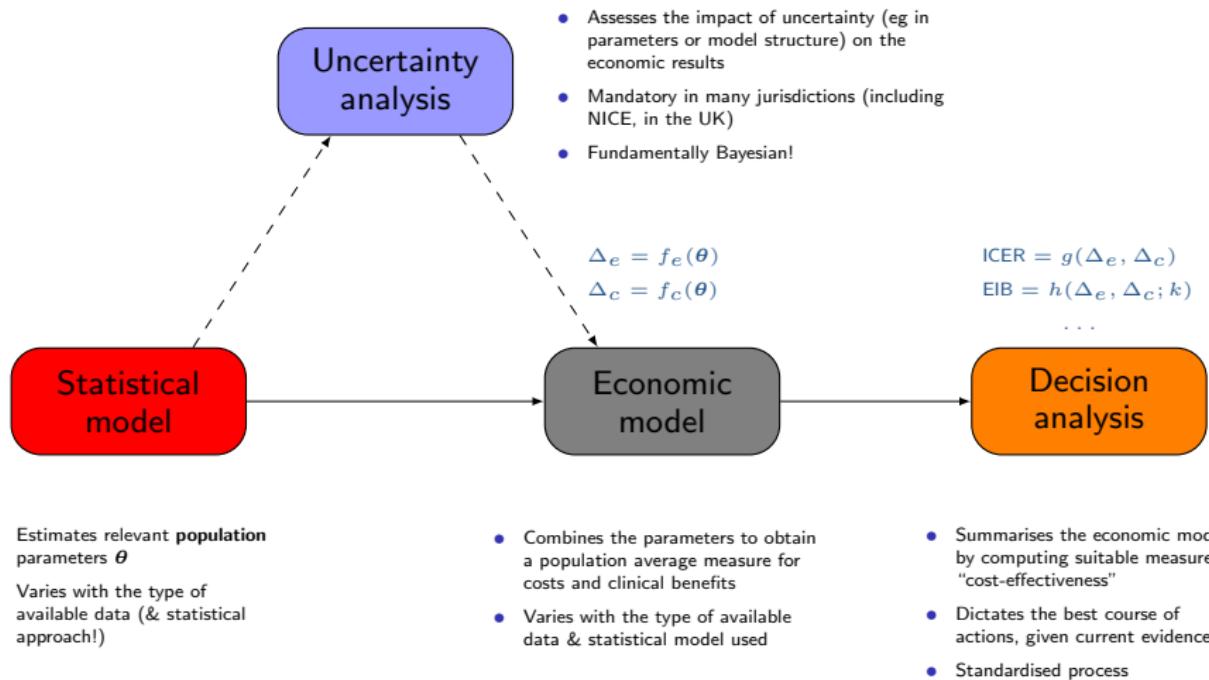
Objective: Combine **costs** & **benefits** of a given intervention into a rational scheme for allocating resources



- Estimates relevant **population** parameters θ
- 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



1. (“Standard”) Statistical modelling

Individual level data

ID	Trt	Demographics			HRQL data			Resource use data			Clinical outcome					
		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	1204	1808	...	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. (“Standard”) Statistical modelling

Individual level data

ID	Trt	Demographics			HRQL data			Resource use data			Clinical outcome					
		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_{J1}
2	1	M	21	...	0.12	0.16	...	0.38	1204	1808	...	877	y_{20}	y_{21}	...	y_{J2}
3	2	F	19	...	0.49	0.55	...	0.88	16	12	...	22	y_{30}	y_{31}	...	y_{J3}
...

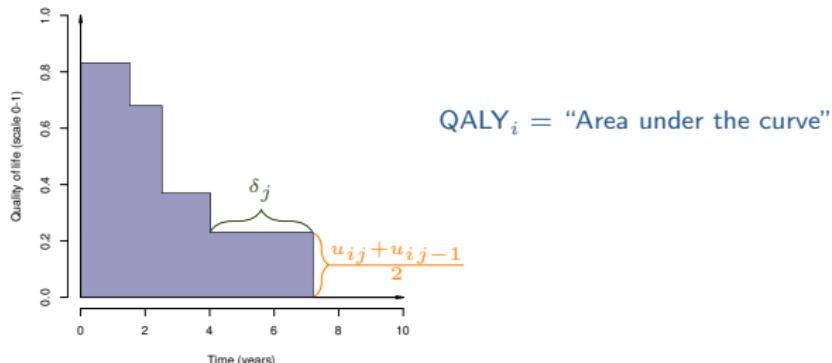
y_{ij} = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...

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, ...)

- Compute individual QALYs and total costs as

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



1. (“Standard”) Statistical modelling

Individual level data

ID	Trt	Demographics			HRQL data			Resource use data			Clinical outcome					
		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	1204	1808	...	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, ...)

- Compute individual QALYs and total costs as

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

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

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

1. (“Standard”) Statistical modelling

Individual level data

ID	Trt	Demographics			HRQL data			Resource use data			Clinical outcome					
		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	1204	1808	...	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, ...)

- Compute individual QALYs and total costs as

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

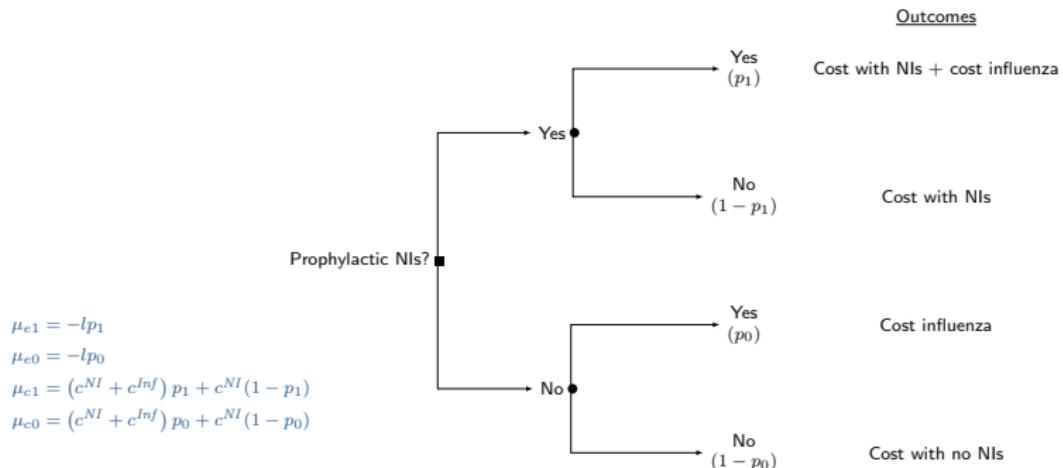
- (Often implicitly) assume normality and linearity and model **independently** individual QALYs and total costs by controlling for baseline values

$$e_i = \alpha_{e0} + \alpha_{e1} u_{0i} + \alpha_{e2} \text{Trt}_i + \varepsilon_{ei} [+ \dots], \quad \varepsilon_{ei} \sim \text{Normal}(0, \sigma_e)$$

$$c_i = \alpha_{c0} + \alpha_{c1} c_{0i} + \alpha_{c2} \text{Trt}_i + \varepsilon_{ci} [+ \dots], \quad \varepsilon_{ci} \sim \text{Normal}(0, \sigma_c)$$

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

④ Build a **population level model** (eg decision tree/Markov model)

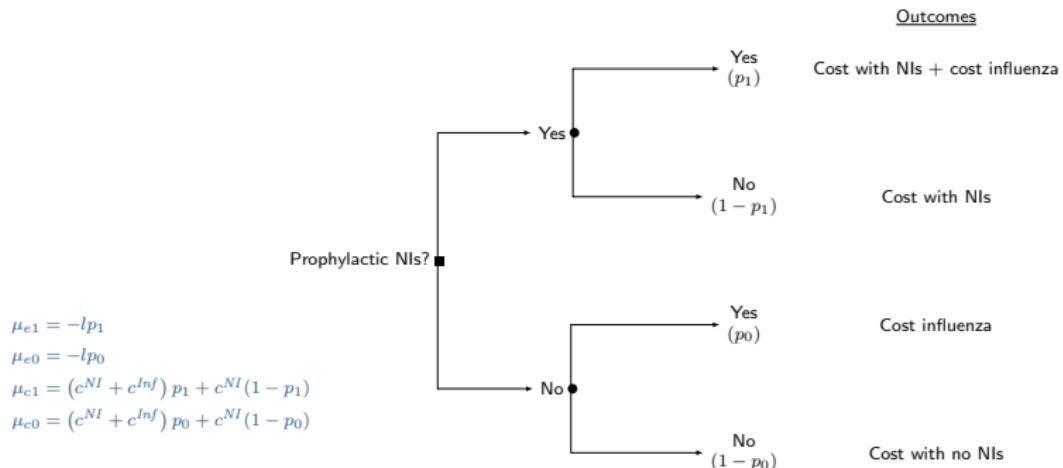


NB: in this case, the “data” are typically represented by summary statistics for the parameters of interest $\theta = (p_0, p_1, l, \dots)$, but may also have access to a combination of ILD and summaries

1. (“Standard”) Statistical modelling

Aggregated level data

- Build a population level model (eg decision tree/Markov model)



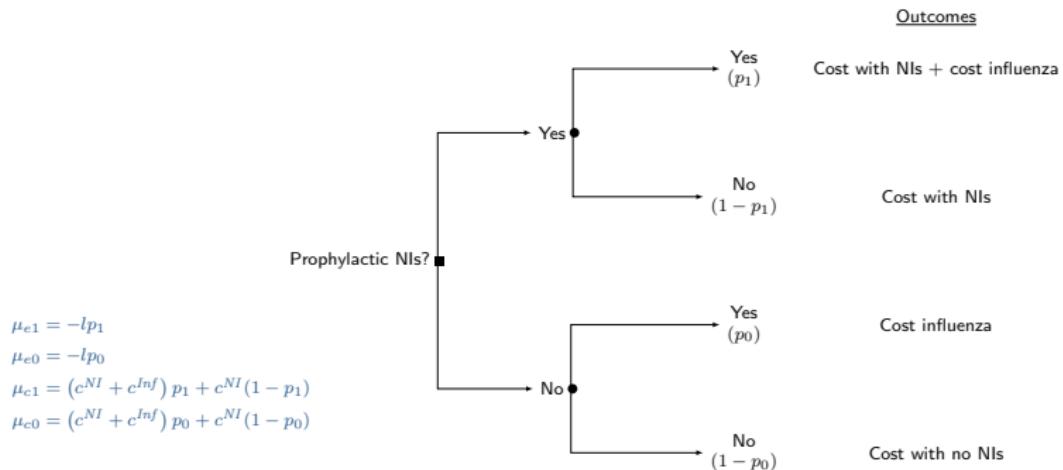
NB: in this case, the “data” are typically represented by summary statistics for the parameters of interest $\theta = (p_0, p_1, l, \dots)$, but may also have access to a combination of ILD and summaries

- Use point estimates for the parameters to build the “base-case” (average) evaluation

1. (“Standard”) Statistical modelling

Aggregated level data

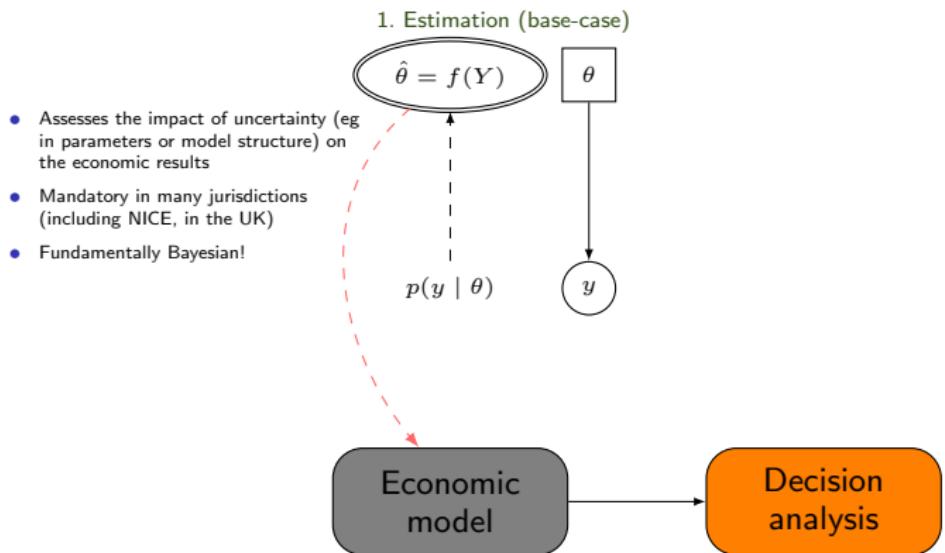
- ① Build a population level model (eg decision tree/Markov model)



NB: in this case, the “data” are typically represented by summary statistics for the parameters of interest $\theta = (p_0, p_1, l, \dots)$, but may also have access to a combination of ILD and summaries

- ② Use point estimates for the parameters to build the “base-case” (average) evaluation
- ③ Use resampling methods (eg bootstrap) to propagate uncertainty in the point estimates and perform uncertainty analysis

“Standard” approach to HTA — “Two-stage”

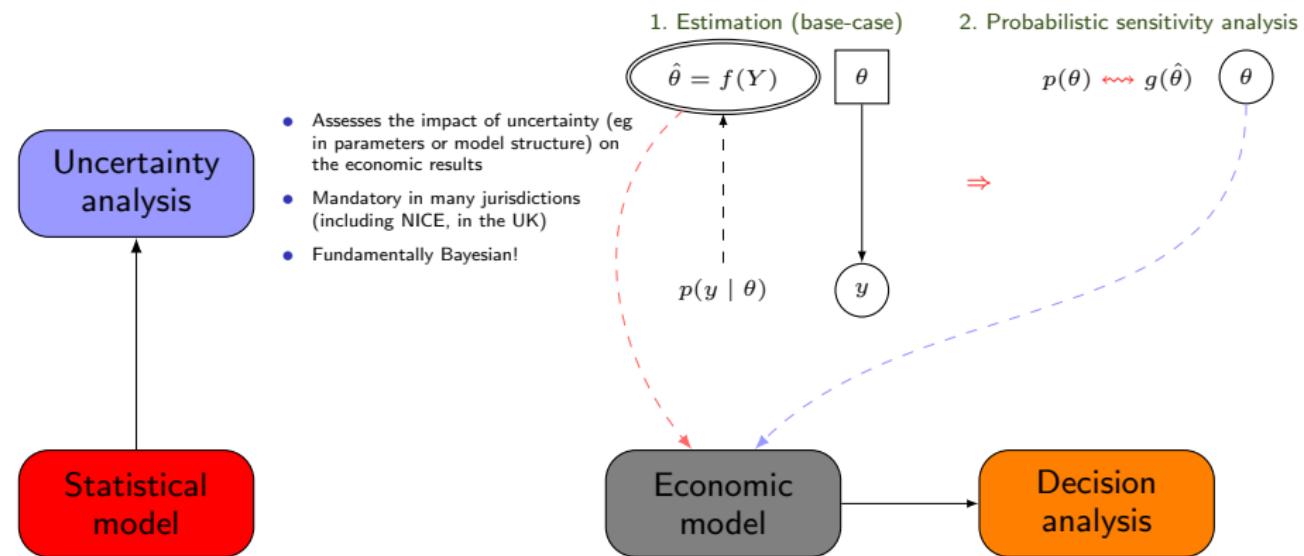


- Estimates relevant population parameters θ
- Varies with the type of available data (& statistical approach!)

- 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”

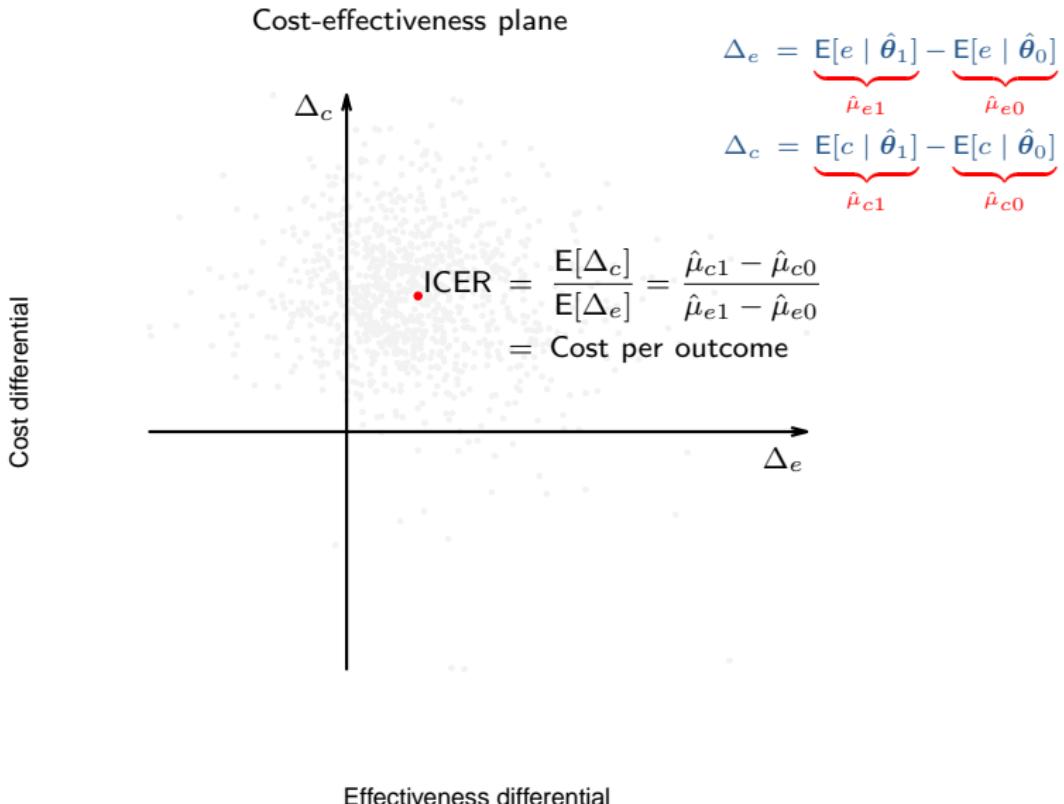


- Estimates relevant population parameters θ
- Varies with the type of available data (& statistical approach!)

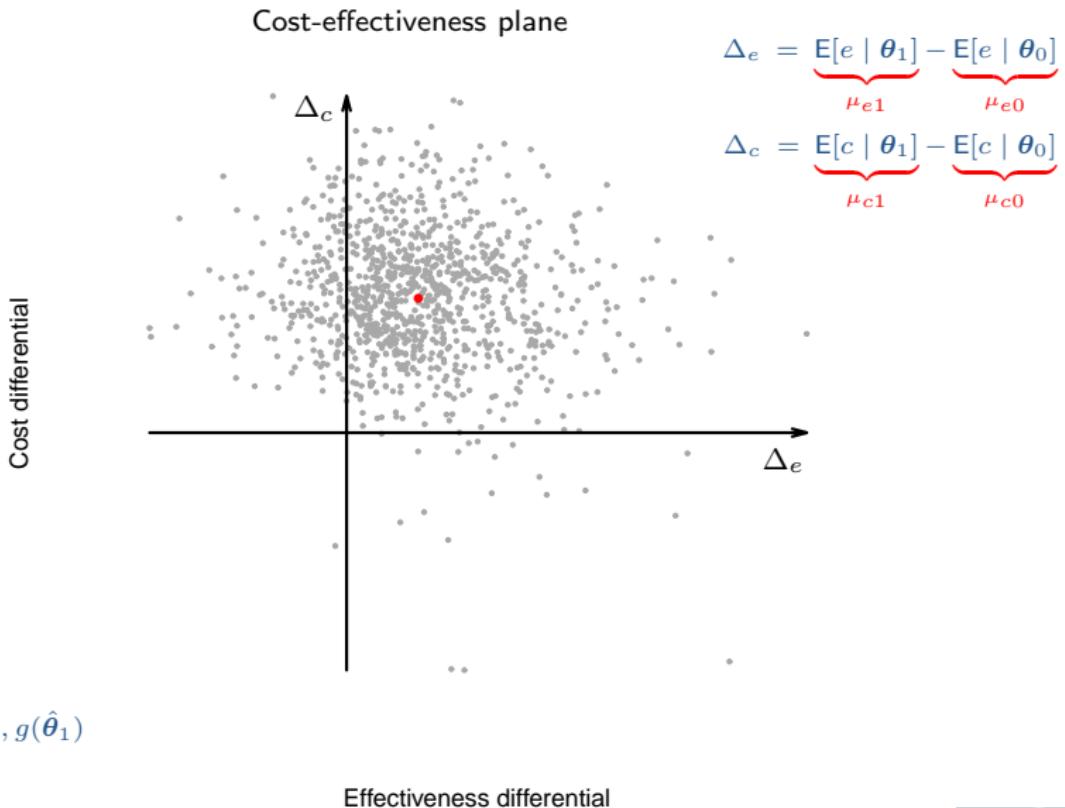
- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- 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)



4. Uncertainty analysis*



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

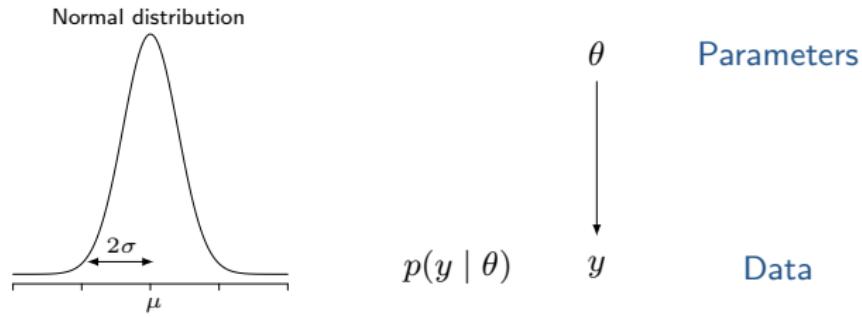
The Sherlock conundrum...



What is statistics all about?

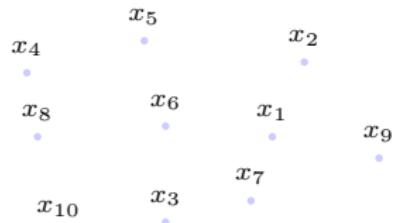
- Typically, we observe some data and we want to use them to learn about some unobservable feature of the general population in which we are interested
- To do this, we use statistical models to describe the probabilistic mechanism by which (**we assume!**) that the data have arisen

Data generating process

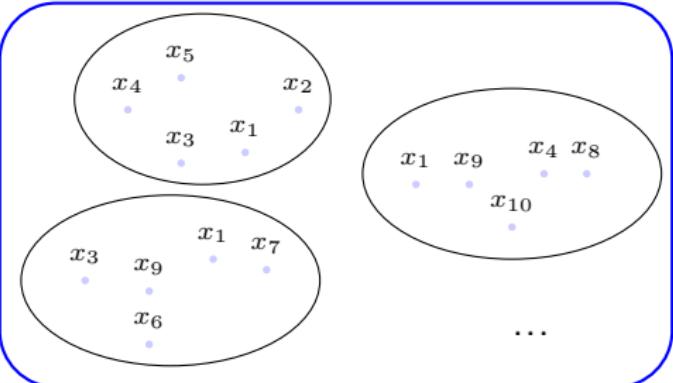


NB: Roman letters (y or x) typically indicate **observable data**, while Greek letters (θ , μ , σ , ...) indicate **population parameters**

The entire population



Some samples of size 5

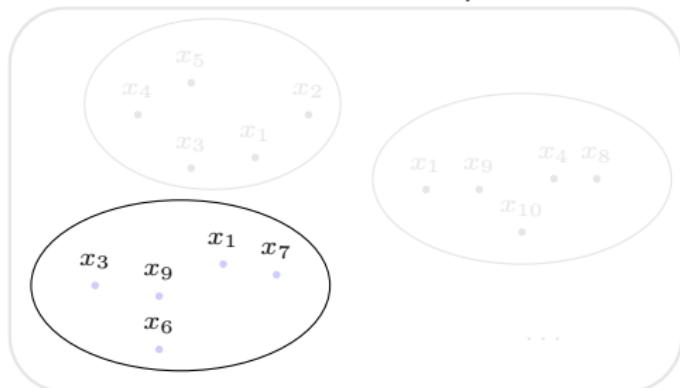


- Population size $N = 10$
- “True” population Mean μ
- “True” Standard deviation σ
- Sample size $n = 5$
- Sample Mean \bar{x}
- Sample Standard deviation s_x

The entire population



The observed sample



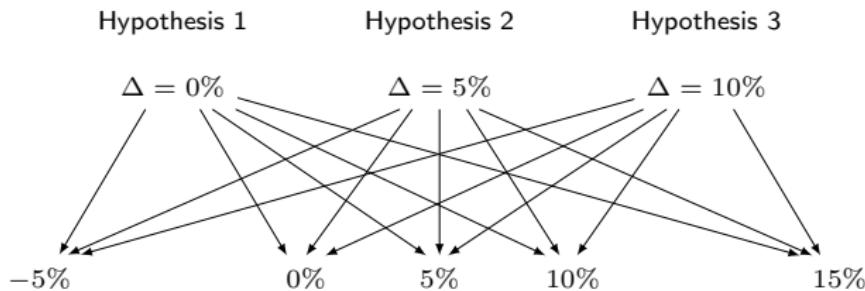
- Population size $N = 10$
- “True” population Mean μ
- “True” Standard deviation σ



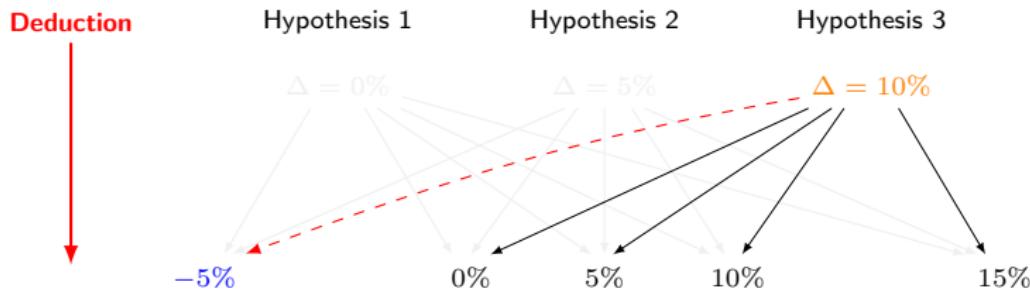
- Sample size $n = 5$
- Sample Mean \bar{x}
- Sample Standard deviation s_x

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

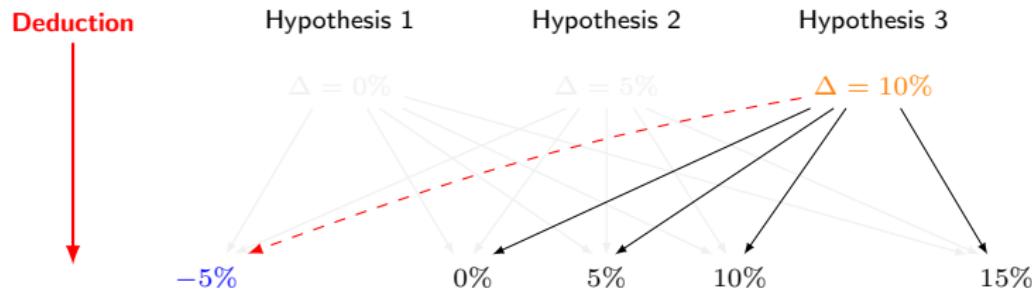


Deductive vs inductive inference



- 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 actually 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 actually observed? If small, then *deduce* weak support of the evidence to the hypothesis
 - Assess $\Pr(\text{Observed data} \mid \text{Hypothesis})$
 - Directly relevant for standard frequentist summaries, eg p-values, Confidence Intervals, etc
 - **NB:** Comparison with data that could have been observed, but haven't!

Confidence intervals

- Drug to cure headaches — “true” probability of success: $\pi = 40/73 \approx 0.55$

Confidence intervals

- Drug to cure headaches — “true” probability of success: $\pi = 40/73 \approx 0.55$
- You get to see data for, say, $n = 10$ individuals, under the “true” **data generating process (DGP)**: $\mathbf{y} = (y_1, \dots, y_{10}) = (0, 0, 1, 1, 0, 1, 0, 1, 0, 1)$
- Can make estimates to infer from sample to population
 - Sample mean: $\bar{y} = \hat{\pi} = \sum_{i=1}^n y_i = \frac{5}{10} = 0.5$
 - Standard error: $se(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}} = 0.16$

Confidence intervals

- Drug to cure headaches — “true” probability of success: $\pi = 40/73 \approx 0.55$
- You get to see data for, say, $n = 10$ individuals, under the “true” **data generating process (DGP)**: $\mathbf{y} = (y_1, \dots, y_{10}) = (0, 0, 1, 1, 0, 1, 0, 1, 0, 1)$
- Can make estimates to infer from sample to population
 - Sample mean: $\bar{y} = \hat{\pi} = \sum_{i=1}^n y_i = \frac{5}{10} = 0.5$
 - Standard error: $se(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}} = 0.16$
- Can compute the interval estimate (using some approximation/theoretical results...)

$$\begin{aligned} 95\% \text{ CI} &\approx [\hat{\pi} - 2se(\hat{\pi}); \hat{\pi} + 2se(\hat{\pi})] \\ &= [0.5 - 0.32; 0.5 + 0.32] = [0.19; 0.81] \end{aligned}$$

Confidence intervals

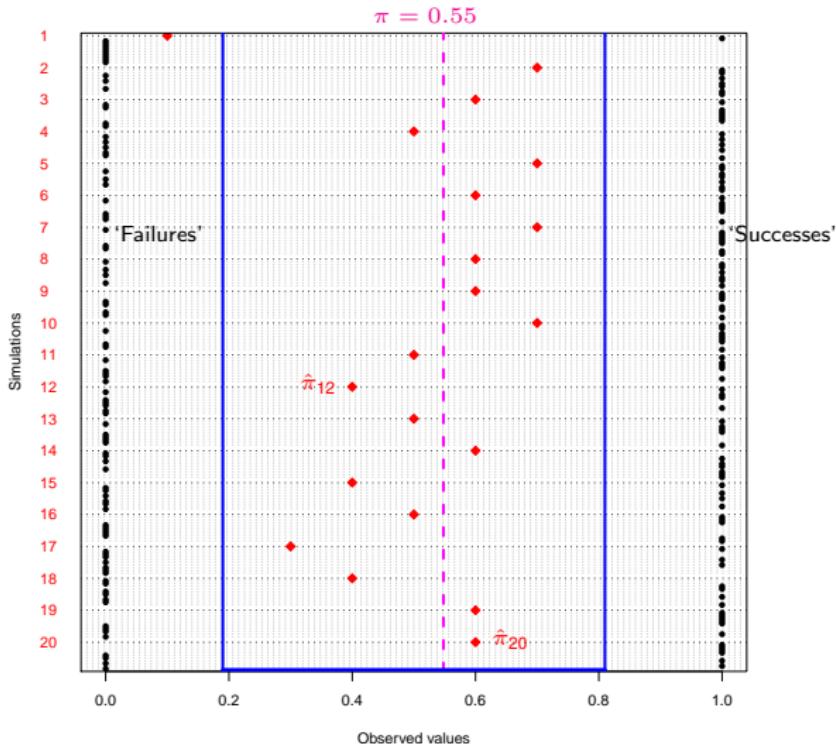
- Drug to cure headaches — “true” probability of success: $\pi = 40/73 \approx 0.55$
- You get to see data for, say, $n = 10$ individuals, under the “true” **data generating process (DGP)**: $\mathbf{y} = (y_1, \dots, y_{10}) = (0, 0, 1, 1, 0, 1, 0, 1, 0, 1)$
- Can make estimates to infer from sample to population
 - Sample mean: $\bar{y} = \hat{\pi} = \sum_{i=1}^n y_i = \frac{5}{10} = 0.5$
 - Standard error: $se(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}} = 0.16$
- Can compute the interval estimate (using some approximation/theoretical results...)

$$\begin{aligned} 95\% \text{ CI} &\approx [\hat{\pi} - 2se(\hat{\pi}); \hat{\pi} + 2se(\hat{\pi})] \\ &= [0.5 - 0.32; 0.5 + 0.32] = [0.19; 0.81] \end{aligned}$$

- Assuming the observed sample is representative of the DGP and using the sample estimates, **if** we were able to replicate the experiment over and over again under the same conditions, 95% of the times, the estimate for the “true” probability of success will be included in the interval [0.19; 0.81]
- **That's how you interpret a 95% Confidence Interval!**

Confidence intervals

- Simulate n_{sim} (e.g. = 20) studies sampling data from a DGP assuming “true” $\pi = 0.5$ (although in fact $\pi = 0.55!$) and $n = 10$
- For each, compute the estimated probability $\hat{\pi}$



Confidence intervals

60% of the time, it works every time!

<https://youtu.be/pjvQFtlNQ-M>



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

<https://youtu.be/Pb0DigCZqL8>

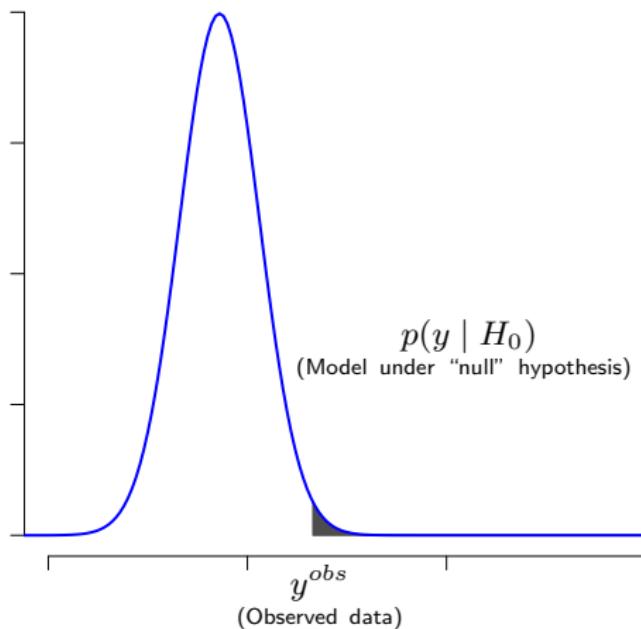


- Designing a study is just as important as analysing it
 - If we don't have "enough" information in the data, we won't be able to detect an underlying signal
- Related to "hypothesis" testing

		"Null" hypothesis H_0	
		True	False
Decision on "Null"	Reject	Type I error α (False positive)	Correct inference (True positive)
	Fail to reject	Correct inference (True negative)	Type II error β (False negative)

- ① Set the Type I error to some low level (**typically**: $\alpha = 0.05$)
 - ② Set the Type II error to some set level (**typically**: $\beta = 0.10$ or $\beta = 0.20$)
 - ③ Define the "clinically relevant outcome" (eg difference in treatment effects), δ
 - ④ Set an estimate of variability in the underlying population
 - ⑤ Use assumptions about sampling variability and determine minimum number of observations to be able to detect δ
- Originally devised to guide quality control of processes

- Interpretation: Under the null hypothesis (ie **IF** it is true), what is the probability of observing something as extreme or even more extreme as the observed data?



- If $p < 0.01 \Rightarrow$ **strong** evidence against H_0
- If $0.01 < p < 0.05 \Rightarrow$ **fairly strong** evidence against H_0
- If $p > 0.05 \Rightarrow$ **little or no** evidence against H_0



Two sides of the same coin?

- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**

Two sides of the same coin?

- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**
- Hypothesis testing (HT)**
 - Considers formally two competing hypotheses — a “null” and an “alternative”
(NB: that determines the treatment effect)
 - Sets the probabilities of error α and β
 - Aims at “rejecting” the null — so it has a binary outcome (yes/no)

Two sides of the same coin?

- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**
- Hypothesis testing (HT)**
 - Considers formally two competing hypotheses — a “null” and an “alternative”
(NB: that determines the treatment effect)
 - Sets the probabilities of error α and β
 - Aims at “rejecting” the null — so it has a binary outcome (yes/no)
- Significance testing (ST, p-values)**
 - Concerned with the sampling distribution of the data under the null hypothesis
 - Measures the strength of the evidence for/against the null, but has no formal involvement of alternative explanations for the observed data

Two sides of the same coin?

- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**
- Hypothesis testing (HT)**
 - Considers formally two competing hypotheses — a “null” and an “alternative”
(NB: that determines the treatment effect)
 - Sets the probabilities of error α and β
 - Aims at “rejecting” the null — so it has a binary outcome (yes/no)
- Significance testing (ST, p-values)**
 - Concerned with the sampling distribution of the data under the null hypothesis
 - Measures the strength of the evidence for/against the null, but has no formal involvement of alternative explanations for the observed data
- $p \neq \alpha$ even if often the **threshold** is set at 0.05 for both!
 - α is **set** by the researcher
 - p is **computed** from the data (as extreme or more extreme than those observed)

Two sides of the same coin?

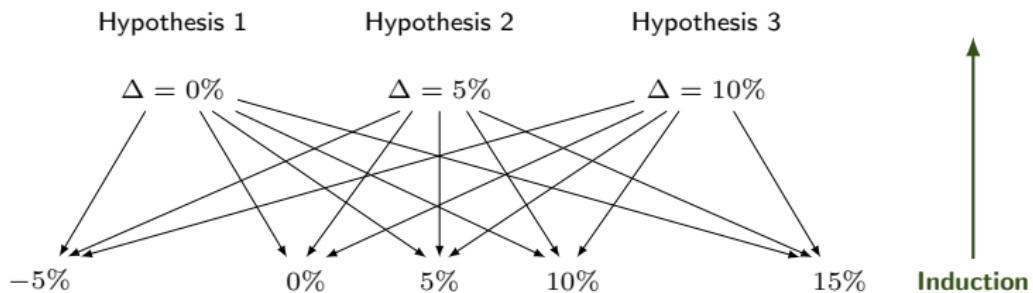
- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**
- **Hypothesis testing** (HT)
 - Considers formally two competing hypotheses — a “null” and an “alternative”
(NB: that determines the treatment effect)
 - Sets the probabilities of error α and β
 - Aims at “rejecting” the null — so it has a binary outcome (yes/no)
- **Significance testing** (ST, p-values)
 - Concerned with the sampling distribution of the data under the null hypothesis
 - Measures the strength of the evidence for/against the null, but has no formal involvement of alternative explanations for the observed data
- $p \neq \alpha$ even if often the **threshold** is set at 0.05 for both!
 - α is **set** by the researcher
 - p is **computed** from the data (as extreme or more extreme than those observed)
- **NB:** Confusingly, experimental studies are **designed** under a HT setting, but **analysed** under a ST setting!
- Increasing recognition of pitfalls in science
 - [http://annals.org/aim/fullarticle/712762/
toward-evidence-based-medical-statistics-1-p-value-fallacy](http://annals.org/aim/fullarticle/712762/toward-evidence-based-medical-statistics-1-p-value-fallacy)
 - <https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf>

Is there another way?...



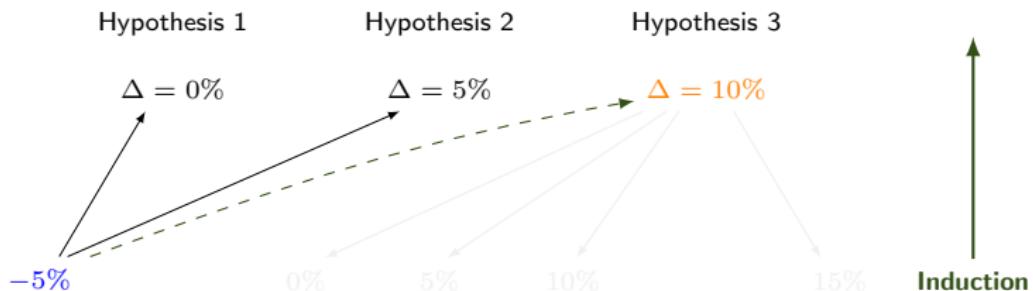
© DAVE GRANLUND.COM
POLITICALCARTOONS.COM

Deductive vs **inductive** inference

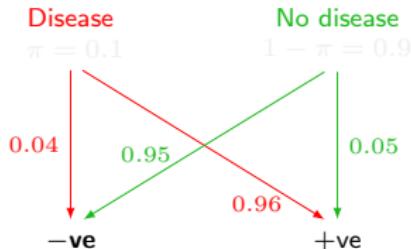


- The **Bayesian** philosophy proceeds fixing the value of the observed data and, **by induction**, makes inference on unobservable hypotheses
 - What is the probability of my hypothesis, given the data I observed? If less than the probability of other competing hypotheses, then weak support of the evidence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

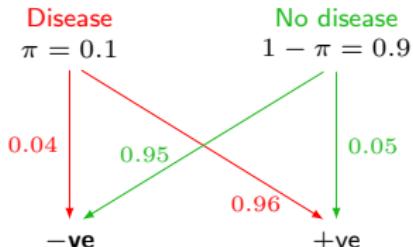
Deductive vs **inductive** inference



- 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!



- Suppose a patient is tested for HIV. The test comes up negative (**-ve**)
- Given the assumptions/model, this indicates **fairly strong** evidence against the hypothesis that the true status is "Disease" — basically, $p = 0.04$



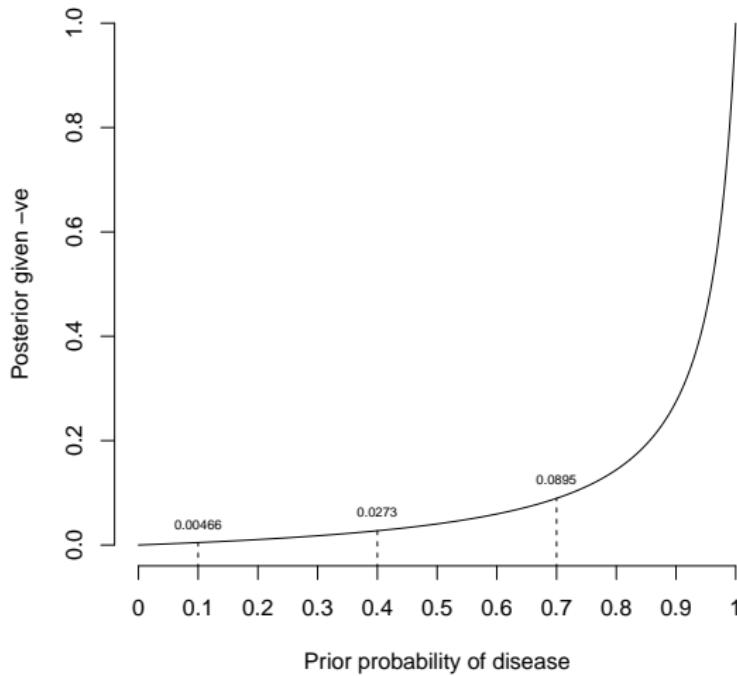
- Suppose a patient is tested for HIV. The test comes up negative (**-ve**)
- Given the assumptions/model, this indicates **fairly strong** evidence against the hypothesis that the true status is "Disease" — basically, $p = 0.04$
- But: how **prevalent** is the disease in the population?
 - We can model our prior knowledge about this and combine this information with the evidence from the data (using **Bayes' theorem**)

$$\Pr(\text{Disease} \mid \text{-ve}) = \frac{\Pr(\text{Disease}) \Pr(\text{-ve} \mid \text{Disease})}{\Pr(\text{-ve})}$$

- Update uncertainty given the evidence provided by the data

Bayesian inference — prior vs posterior

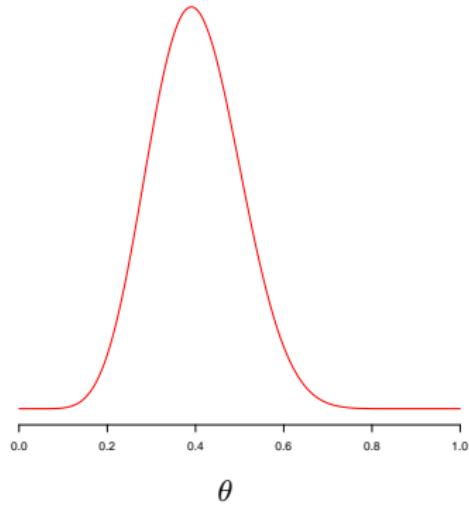
- The evidence **from the data alone**, tells us that the observed result is extremely unlikely under the hypothesis of “Disease”
- This is strongly dependent on the **context**, as provided by the prior knowledge/epistemic uncertainty, though!



Existing knowledge

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

$$p(\theta)$$



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

Existing knowledge

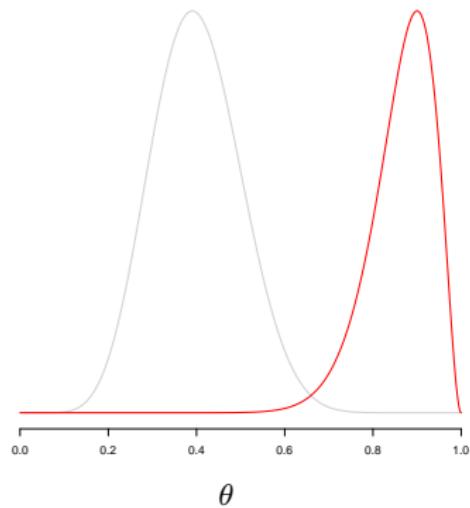
- Population registries
- Observational studies
- Small/pilot RCTs
- Expert options

$$p(\theta)$$

Current data

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

$$p(y | \theta)$$



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

Existing knowledge

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

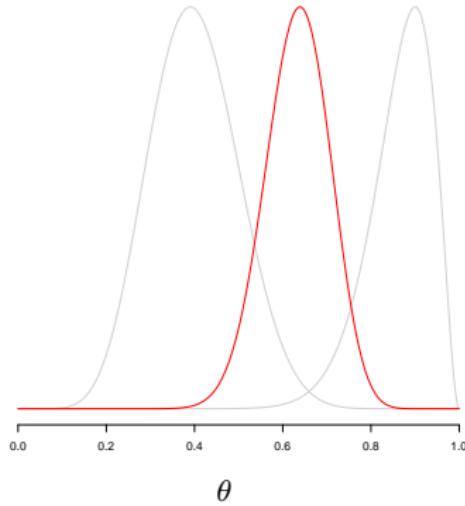
$p(\theta)$

Updated knowledge

$p(\theta | y)$

Current data

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



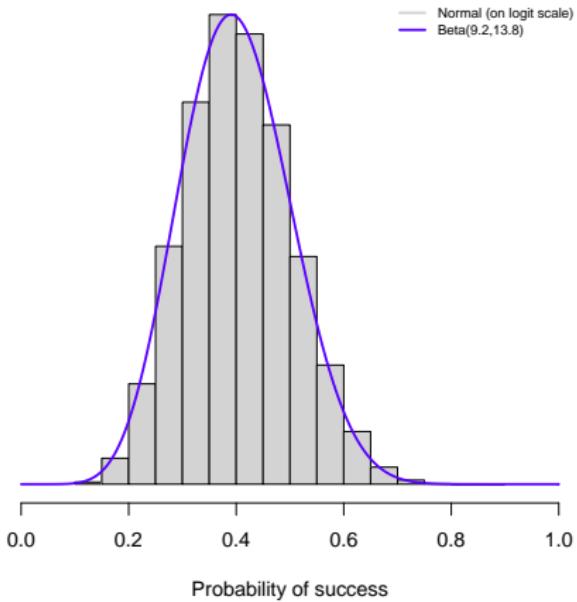
Update knowledge to describe revised “state of science”

Prior information vs prior distribution

- The main objective of Bayesian inference is to **quantify the level of uncertainty**
 - For unobservable quantities (eg parameters)
 - For unobserved quantities (eg future data)
- Only one “language” — probability distribution to describe **any** type of uncertainty

Prior information vs prior distribution

- The main objective of Bayesian inference is to **quantify the level of uncertainty**
 - For unobservable quantities (eg parameters)
 - For unobserved quantities (eg future data)
- Only one “language” — probability distribution to describe **any** type of uncertainty
- There are very many ways of **encoding** the *qualitative* knowledge into a probability distribution



“But how can I form a prior? I don't know anything about this parameter!...”

"But how can I form a prior? I don't know anything about this parameter!..."



"But how can I form a prior? I don't know anything about this parameter!..."

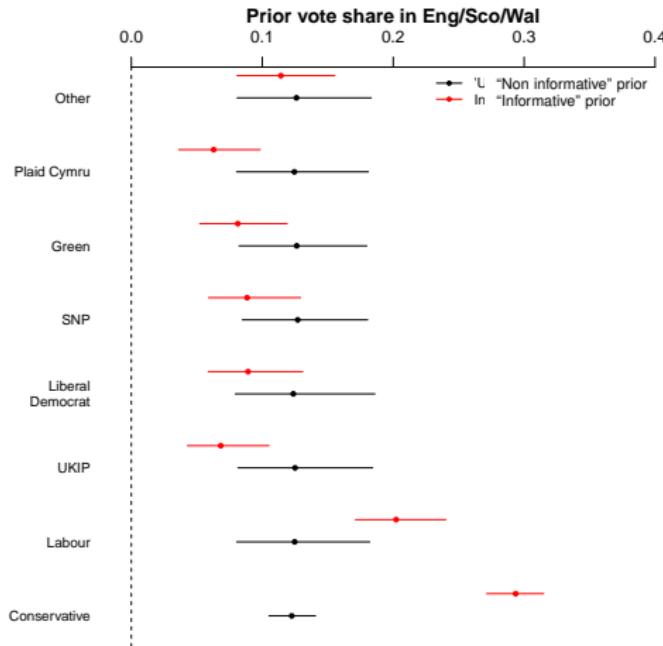


"But how can I form a prior? I don't know anything about this parameter!..."

- Predicting the output of the 2017 UK General Election using poll data
 - Data: number of people out of the N_i respondents in poll i intending to vote for party p (multinomial counts)
 - **Objective of estimation:** $(\pi_1, \dots, \pi_P) =$ population vote share for each party
 - Can model $\pi_p = (\phi_p / \sum \phi_p)$ and $\log(\phi_p) = \alpha_p + \beta_p X_p$

"But how can I form a prior? I don't know anything about this parameter!..."

- Predicting the output of the 2017 UK General Election using poll data
 - Data: number of people out of the N_i respondents in poll i intending to vote for party p (multinomial counts)
 - **Objective of estimation:** $(\pi_1, \dots, \pi_P) =$ population vote share for each party
 - Can model $\pi_p = (\phi_p / \sum \phi_p)$ and $\log(\phi_p) = \alpha_p + \beta_p X_p$



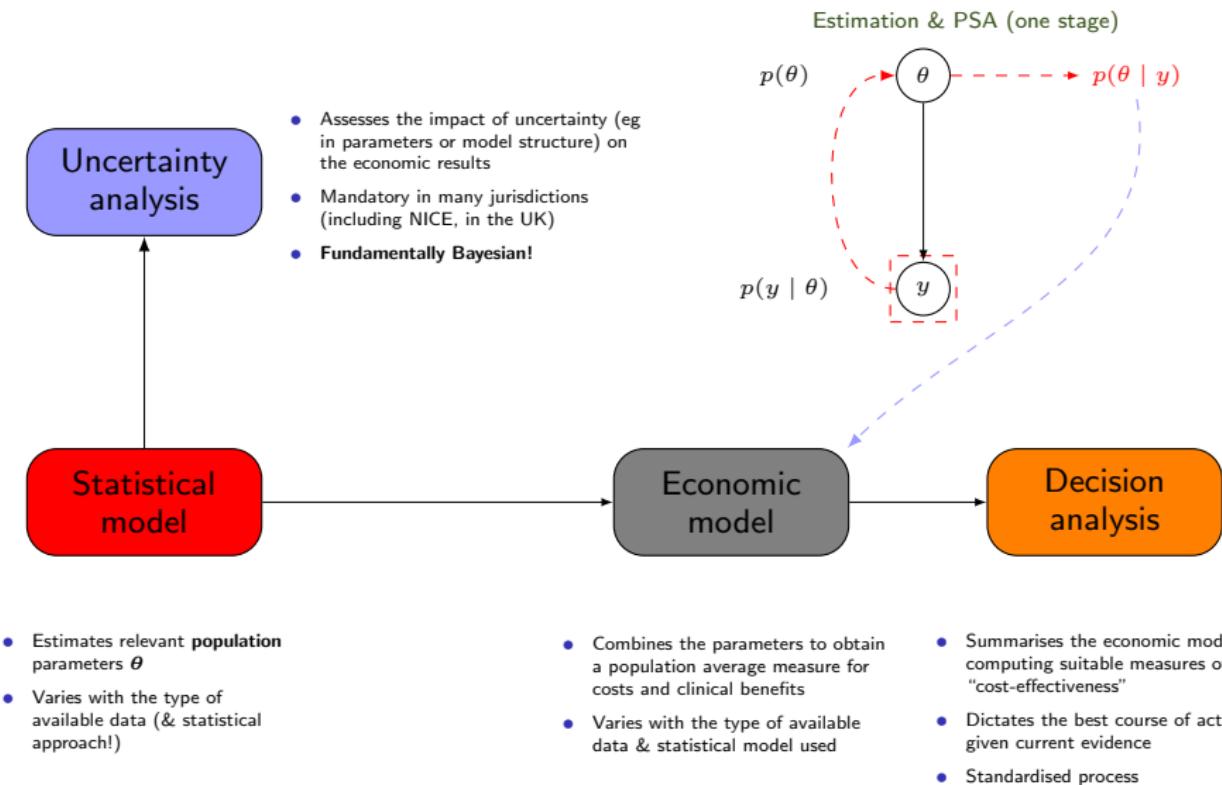
- In artificially simplified modelling structures, Bayesian computations are just as easy as “standard” statistical models
 - Thomas Bayes (1763)
 - ★ Set up (what we now call) a Binomial model for number of “successes” out of a set number of “trials”
 - ★ Applied to billiard balls — <https://youtu.be/-e8w0caascM>
 - Pierre-Simon Laplace (1786)
 - ★ Analysed data on christening in Paris from 1745 to 1770 using (what we now consider) a Bayesian model
 - ★ Concludes that he was “morally certain” that $\text{Pr}(\text{new born is boy} \mid \text{data}) \geq 0.5$ — divine providence to account for the fact that males died at higher rates...

Bayesian computation

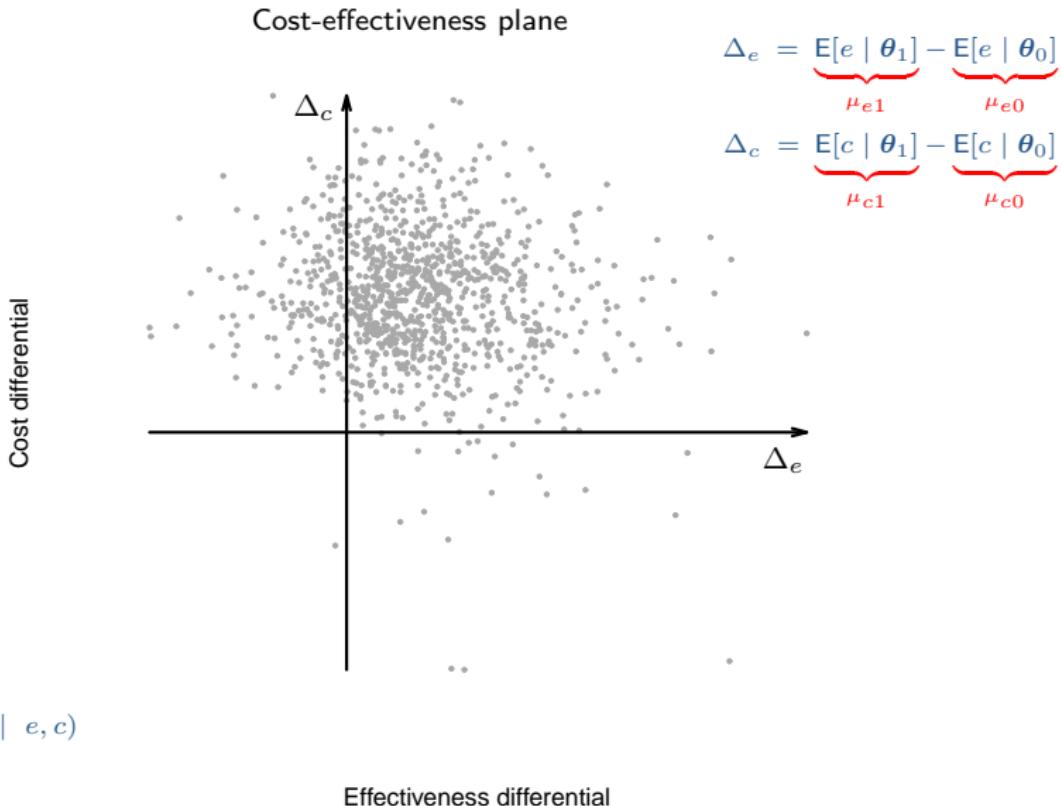
- In artificially simplified modelling structures, Bayesian computations are just as easy as “standard” statistical models
 - Thomas Bayes (1763)
 - ★ Set up (what we now call) a Binomial model for number of “successes” out of a set number of “trials”
 - ★ Applied to billiard balls — <https://youtu.be/-e8w0caascM>
 - Pierre-Simon Laplace (1786)
 - ★ Analysed data on christening in Paris from 1745 to 1770 using (what we now consider) a Bayesian model
 - ★ Concludes that he was “morally certain” that $\text{Pr}(\text{new born is boy} \mid \text{data}) \geq 0.5$ — divine providence to account for the fact that males died at higher rates...
- But they can become very complicated in realistic models
 - Alan Turing (1940s) — breaking the Enigma code
 - ★ Using “prior” information and guess a stretch of letters in an Enigma message, measure their belief in the validity of these guesses and add more clues as they arrived
 - ★ <https://youtu.be/oTltncUkckQ>

- In artificially simplified modelling structures, Bayesian computations are just as easy as “standard” statistical models
 - Thomas Bayes (1763)
 - ★ Set up (what we now call) a Binomial model for number of “successes” out of a set number of “trials”
 - ★ Applied to billiard balls — <https://youtu.be/-e8wOcaascM>
 - Pierre-Simon Laplace (1786)
 - ★ Analysed data on christening in Paris from 1745 to 1770 using (what we now consider) a Bayesian model
 - ★ Concludes that he was “morally certain” that $\text{Pr}(\text{new born is boy} \mid \text{data}) \geq 0.5$ — divine providence to account for the fact that males died at higher rates...
- But they can become very complicated in realistic models
 - Alan Turing (1940s) — breaking the Enigma code
 - ★ Using “prior” information and guess a stretch of letters in an Enigma message, measure their belief in the validity of these guesses and add more clues as they arrived
 - ★ <https://youtu.be/oTltncUkckQ>
- 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
 - Can still be computationally intensive — but variants of “vanilla” implementations can be made **very** efficient

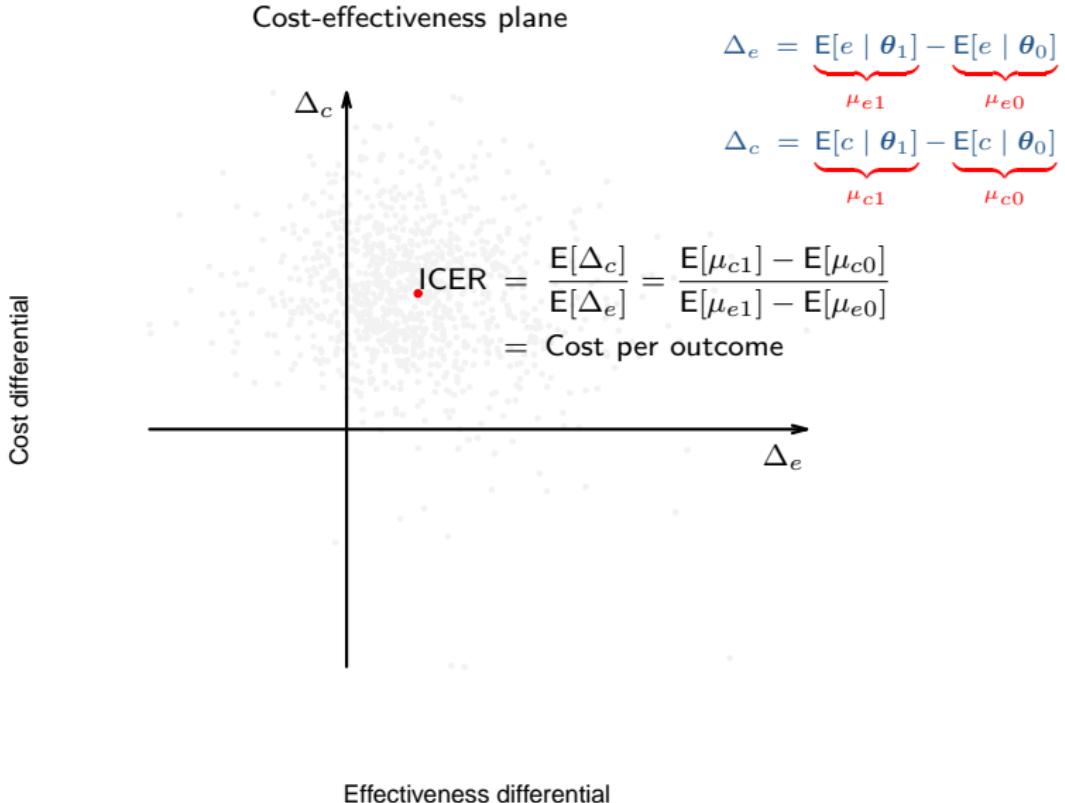
Bayesian approach to HTA



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

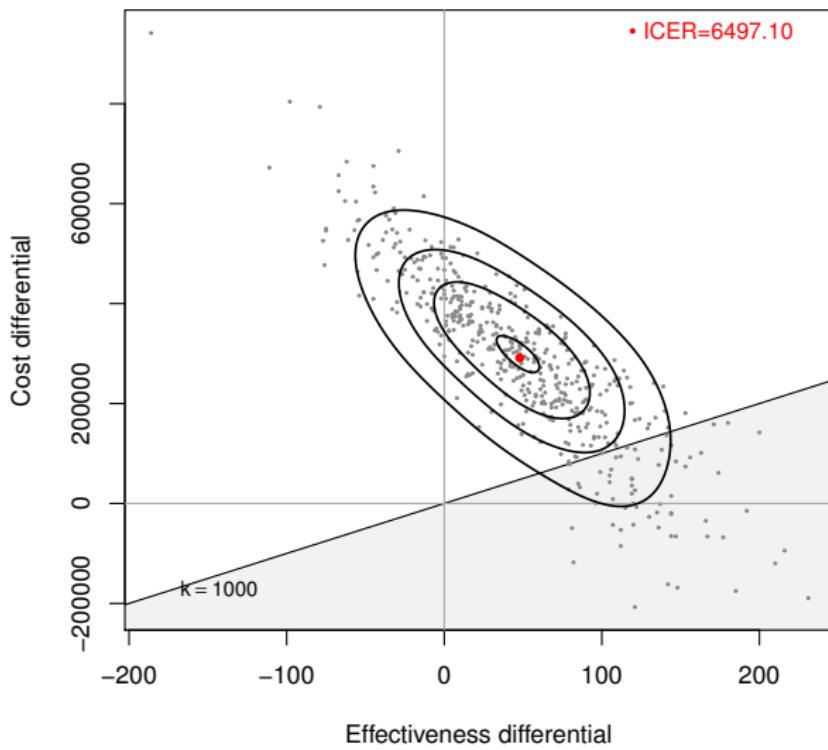


*Induced by $p(\theta | e, c)$



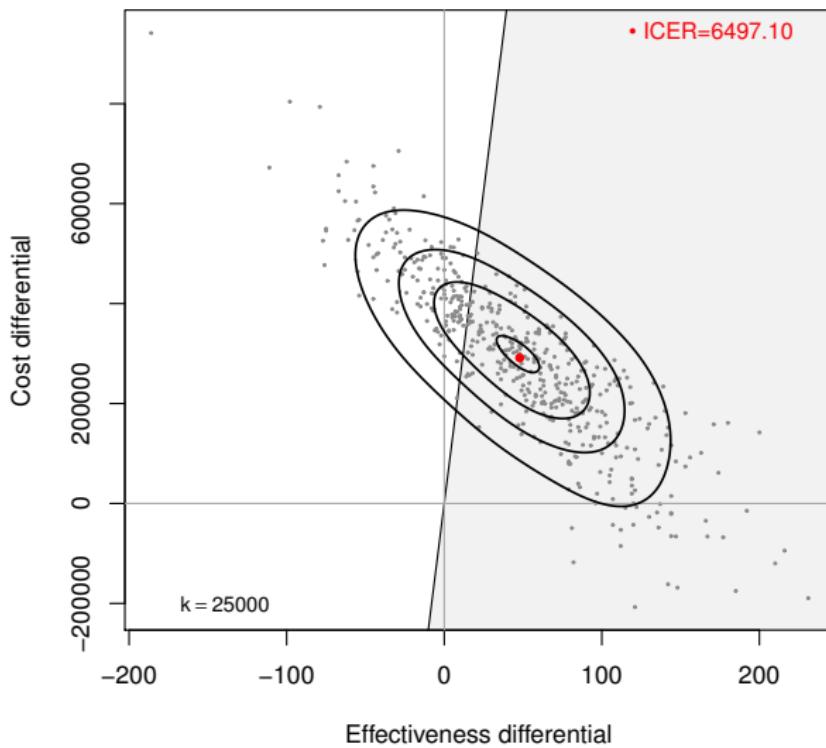
Uncertainty (Probabilistic Sensitivity) Analysis

Cost effectiveness plane
New Chemotherapy vs Old Chemotherapy

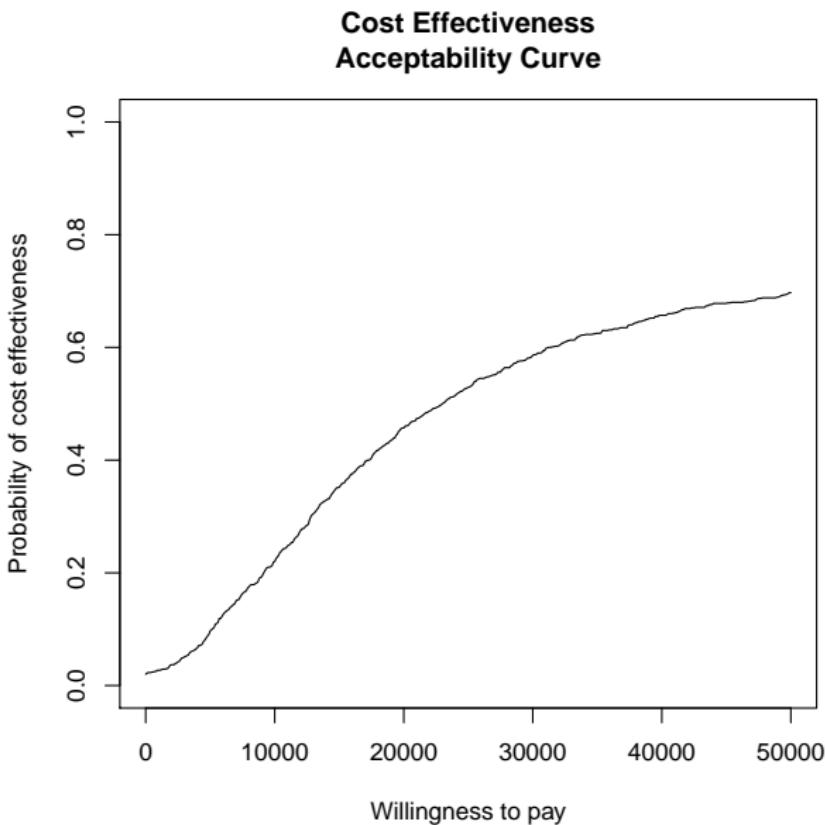


Uncertainty (Probabilistic Sensitivity) Analysis

Cost effectiveness plane
New Chemotherapy vs Old Chemotherapy

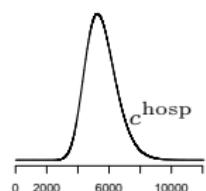
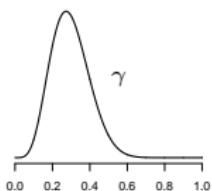
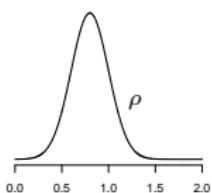
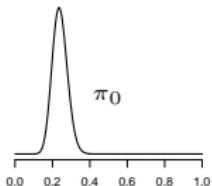


Uncertainty (Probabilistic Sensitivity) Analysis



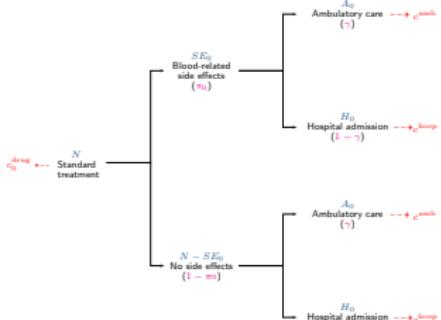
- Potential correlation between costs & clinical benefits [Both ILD and ALD]
 - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research \Rightarrow are associated with higher unit costs
 - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
 - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**
- Joint/marginal normality not realistic [Mainly ILD]
 - Costs usually skewed and benefits may be bounded in [0; 1]
 - Can use transformation (e.g. logs) — but care is needed when back transforming to the natural scale
 - Should use more suitable models (e.g. Beta, Gamma or log-Normal) — generally easier under a Bayesian framework
- ... 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!

Parameters

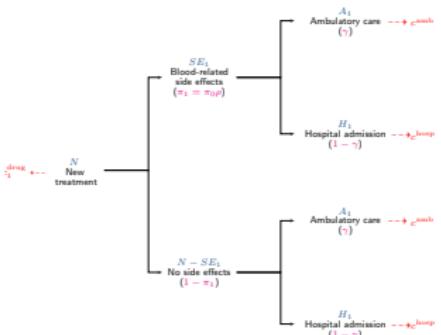


Model structure

Old chemotherapy



New chemotherapy



Decision analysis

Old chemotherapy

Benefits	Costs

Benefits	Costs

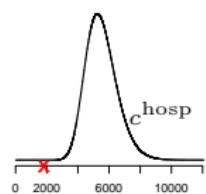
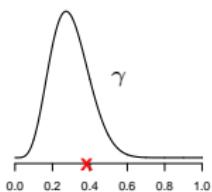
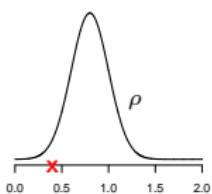
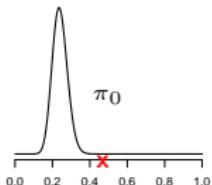
New chemotherapy

Benefits	Costs

PSA & correlation

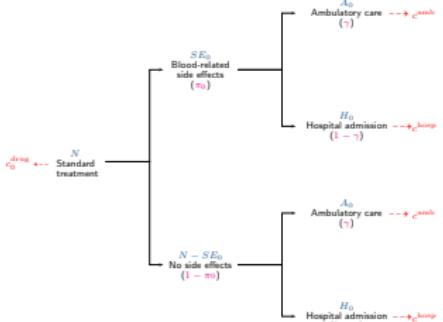
$p(\theta \mid e, c)$ vs $g_i(\theta_i)$

Parameters

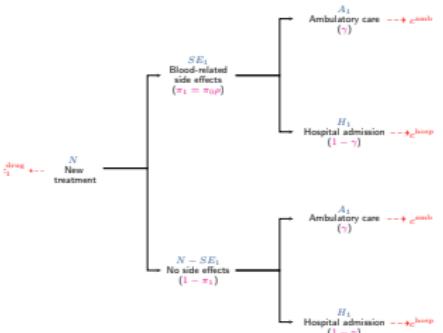


Model structure

Old chemotherapy



New chemotherapy



Decision analysis

Old chemotherapy

Benefits	Costs
741	670 382.1

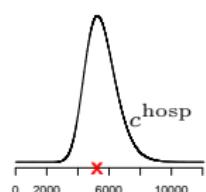
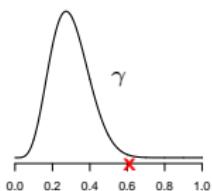
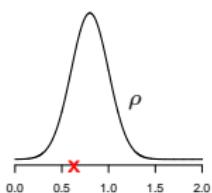
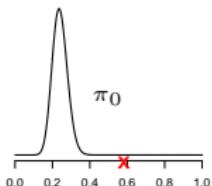
New chemotherapy

Benefits	Costs
732	1 131 978

PSA & correlation

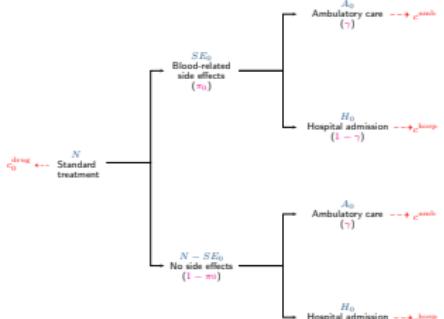
$$[p(\theta | e, c) \text{ vs } g_i(\theta_i)]$$

Parameters

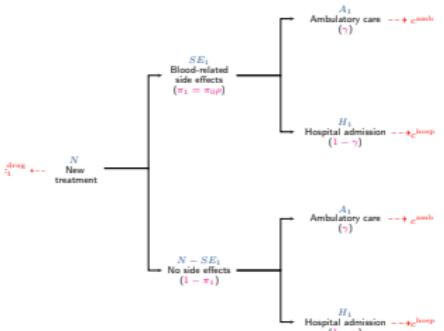


Model structure

Old chemotherapy



New chemotherapy



Decision analysis

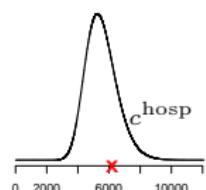
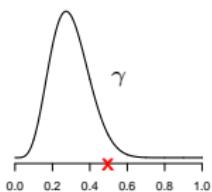
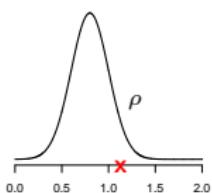
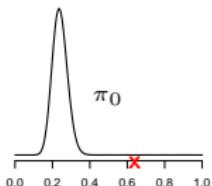
Old chemotherapy

Benefits	Costs
741	670 382.1
699	871 273.3

New chemotherapy

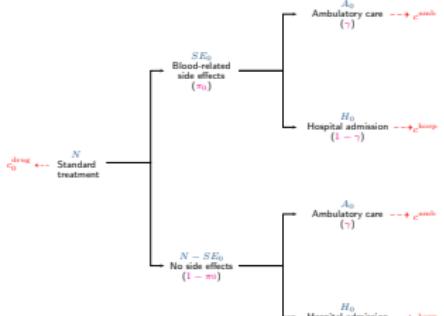
Benefits	Costs
732	1 131 978
664	1 325 654

Parameters

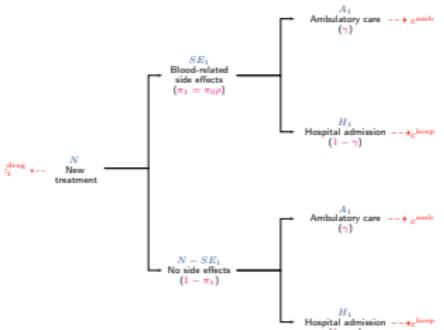


Model structure

Old chemotherapy



New chemotherapy



Decision analysis

Old chemotherapy	
Benefits	Costs
741	670 382.1
699	871 273.3
...	...
726	425 822.2
716.2	790 381.2

New chemotherapy	
Benefits	Costs
732	1 131 978
664	1 325 654
...	...
811	766 411.4
774.5	1 066 849.8

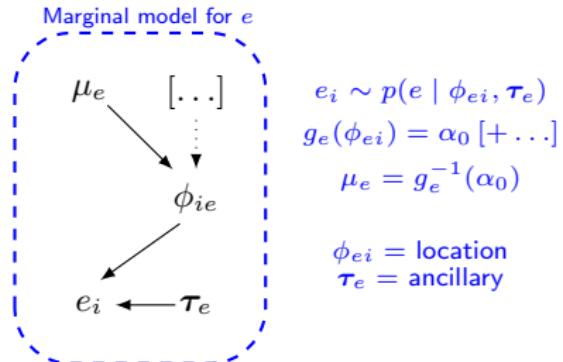
$$\text{ICER} = \frac{276\,468.6}{58.3} = 6\,497.1$$

Bayesians do it better...

- Potential correlation between costs & clinical benefits [Both ILD and ALD]
 - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research ⇒ are associated with higher unit costs
 - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
 - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**
- Joint/marginal normality not realistic [Mainly ILD]
 - Costs usually skewed and benefits may be bounded in $[0; 1]$
 - Can use transformation (e.g. logs) — but care is needed when back transforming to the natural scale
 - Should use more suitable models (e.g. Beta, Gamma or log-Normal) — **generally easier under a Bayesian framework**
- ... 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

- In general, can represent the joint distribution as $p(e, c) = p(e)p(c | e) = p(c)p(e | c)$

- In general, can represent the joint distribution as $p(e, c) = p(e)p(c | e) = p(c)p(e | c)$



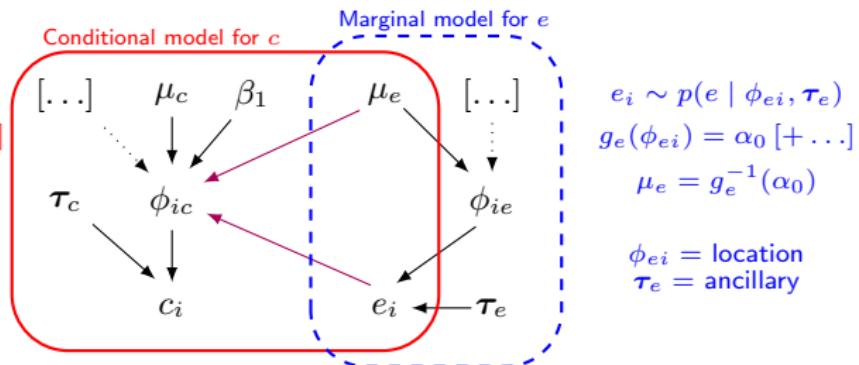
- In general, can represent the joint distribution as $p(e, c) = p(e)p(c | e) = p(c)p(e | c)$

$$c_i \sim p(c | e, \phi_{ci}, \tau_c)$$

$$g_c(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots]$$

$$\mu_c = g_c^{-1}(\beta_0)$$

ϕ_{ci} = location
 τ_c = ancillary



$$e_i \sim p(e | \phi_{ei}, \tau_e)$$

$$g_e(\phi_{ei}) = \alpha_0 [+ \dots]$$

$$\mu_e = g_e^{-1}(\alpha_0)$$

ϕ_{ei} = location
 τ_e = ancillary

- In general, can represent the joint distribution as $p(e, c) = p(e)p(c | e) = p(c)p(e | c)$

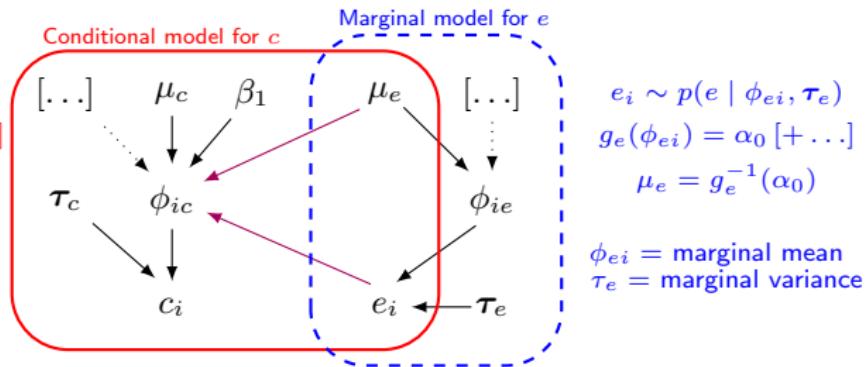
$$c_i \sim p(c | e, \phi_{ci}, \tau_c)$$

$$g_c(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots]$$

$$\mu_c = g_c^{-1}(\beta_0)$$

ϕ_{ci} = conditional mean

τ_c = conditional variance



- For example:

$$e_i \sim \text{Normal}(\phi_{ei}, \tau_e), \quad \phi_{ei} = \alpha_0 [+ \dots], \quad \mu_e = \alpha_0$$

$$c_i | e_i \sim \text{Normal}(\phi_{ci}, \tau_c), \quad \phi_{ci} = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots], \quad \mu_c = \beta_0$$

- In general, can represent the joint distribution as $p(e, c) = p(e)p(c | e) = p(c)p(e | c)$

$$c_i \sim p(c | e, \phi_{ci}, \tau_c)$$

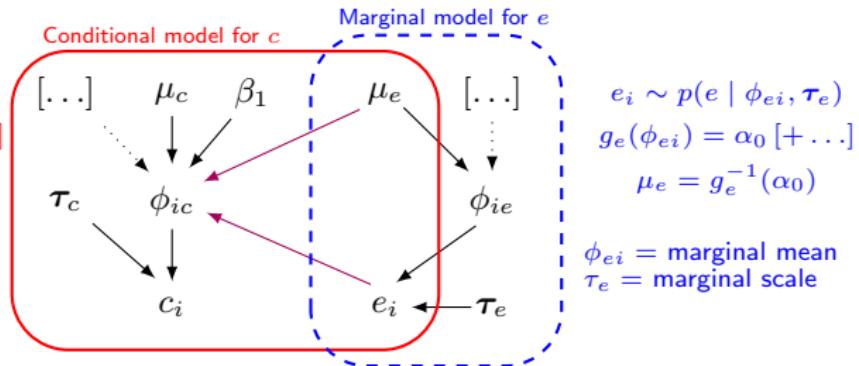
$$g_c(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots]$$

$$\mu_c = g_c^{-1}(\beta_0)$$

ϕ_{ci} = conditional mean

τ_c = shape

τ_c/ϕ_{ci} = rate



- For example:

$$e_i \sim \text{Beta}(\phi_{ei}\tau_e, (1 - \phi_{ei})\tau_e), \quad \text{logit}(\phi_{ei}) = \alpha_0 [+ \dots], \quad \mu_e = \frac{\exp(\alpha_0)}{1 + \exp(\alpha_0)}$$

$$c_i | e_i \sim \text{Gamma}(\tau_c, \tau_c/\phi_{ci}), \quad \log(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots], \quad \mu_c = \exp(\beta_0)$$

- In general, can represent the joint distribution as $p(e, c) = p(e)p(c | e) = p(c)p(e | c)$

$$c_i \sim p(c | e, \phi_{ci}, \tau_c)$$

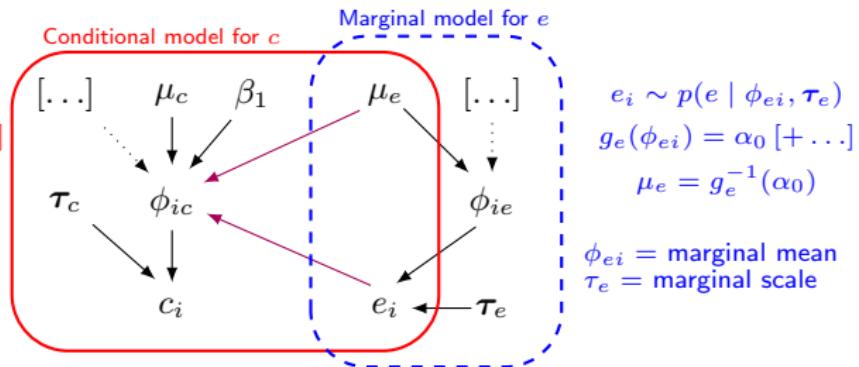
$$g_c(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots]$$

$$\mu_c = g_c^{-1}(\beta_0)$$

ϕ_{ci} = conditional mean

τ_c = shape

τ_c/ϕ_{ci} = rate



- For example:

$$e_i \sim \text{Beta}(\phi_{ei}\tau_e, (1 - \phi_{ei})\tau_e), \quad \text{logit}(\phi_{ei}) = \alpha_0 [+ \dots], \quad \mu_e = \frac{\exp(\alpha_0)}{1 + \exp(\alpha_0)}$$

$$c_i | e_i \sim \text{Gamma}(\tau_c, \tau_c/\phi_{ci}), \quad \log(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots], \quad \mu_c = \exp(\beta_0)$$

- Combining “modules” and fully characterising uncertainty about deterministic functions of random quantities is relatively straightforward using MCMC

- In general, can represent the joint distribution as $p(e, c) = p(e)p(c | e) = p(c)p(e | c)$

$$c_i \sim p(c | e, \phi_{ci}, \tau_c)$$

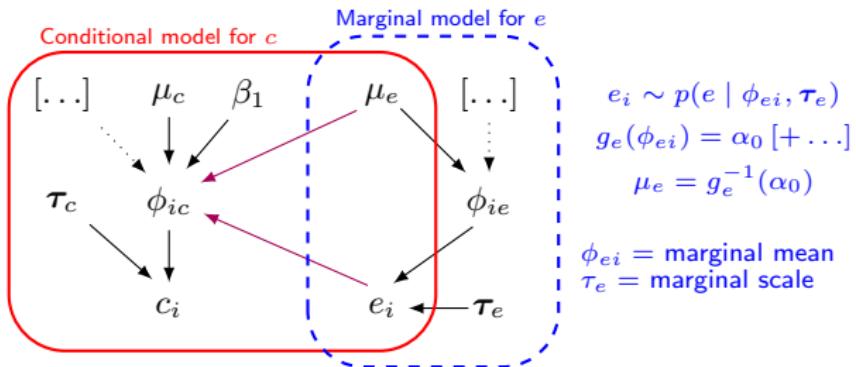
$$g_c(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots]$$

$$\mu_c = g_c^{-1}(\beta_0)$$

ϕ_{ci} = conditional mean

τ_c = shape

τ_c/ϕ_{ci} = rate



- For example:

$$e_i \sim \text{Beta}(\phi_{ei}\tau_e, (1 - \phi_{ei})\tau_e), \quad \text{logit}(\phi_{ei}) = \alpha_0 [+ \dots], \quad \mu_e = \frac{\exp(\alpha_0)}{1 + \exp(\alpha_0)}$$

$$c_i | e_i \sim \text{Gamma}(\tau_c, \tau_c/\phi_{ci}), \quad \log(\phi_{ci}) = \beta_0 + \beta_1(e_i - \mu_e) [+ \dots], \quad \mu_c = \exp(\beta_0)$$

- Combining “modules” and fully characterising uncertainty about deterministic functions of random quantities is relatively straightforward using MCMC
- Prior information can help stabilise inference (especially with sparse data!), eg
 - Cancer patients are unlikely to survive as long as the general population
 - ORs are unlikely to be greater than ± 5

Bayesians do it better...

- Potential correlation between costs & clinical benefits [Both ILD and ALD]
 - Strong positive correlation — effective treatments are innovative and result from intensive and lengthy research ⇒ are associated with higher unit costs
 - Negative correlation — more effective treatments may reduce total care pathway costs e.g. by reducing hospitalisations, side effects, etc.
 - Because of the way in which standard models are set up, bootstrapping generally only approximates the underlying level of correlation — **MCMC does a better job!**
- Joint/marginal normality not realistic [Mainly ILD]
 - Costs usually skewed and benefits may be bounded in [0; 1]
 - Can use transformation (e.g. logs) — but care is needed when back transforming to the natural scale
 - Should use more suitable models (e.g. Beta, Gamma or log-Normal) — generally easier under a Bayesian framework
- ... 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

Introduction to missing data

[Back to Table of content](#)

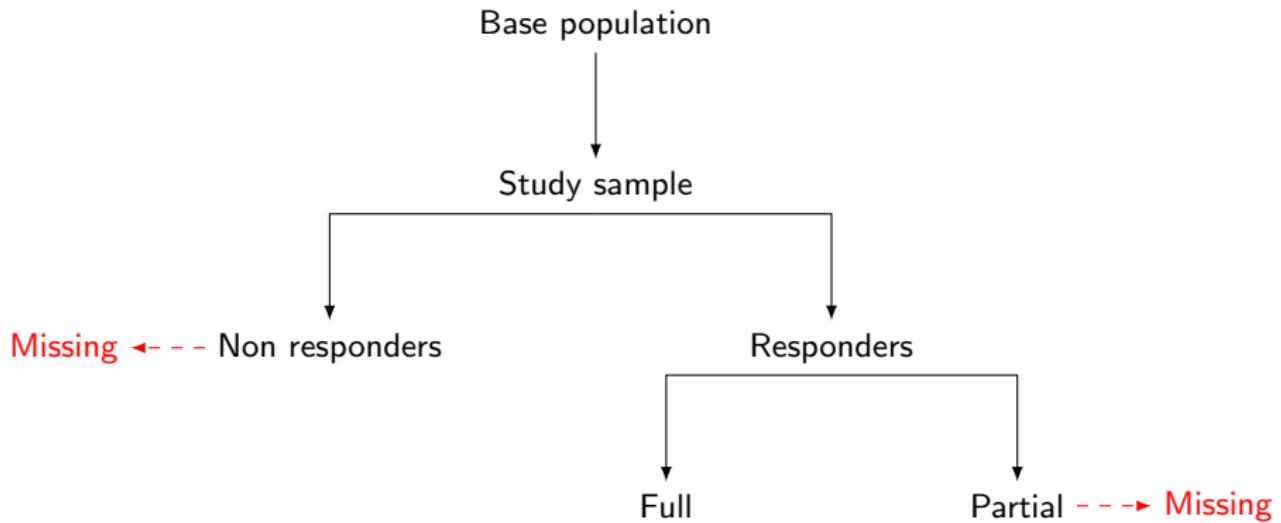
The problems with missing data...

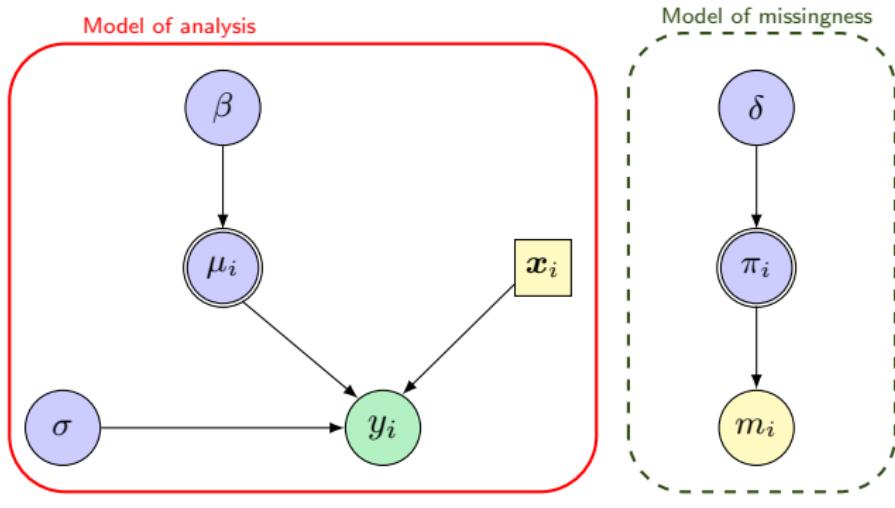
- We plan to observe n_{planned} data points, but end up with a (much) lower number of observations n_{observed}
 - What is the proportion of missing data? Does it matter?...
- We typically don't know **why** the unobserved points are missing and **what** their value might have been
 - Missingness can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

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!

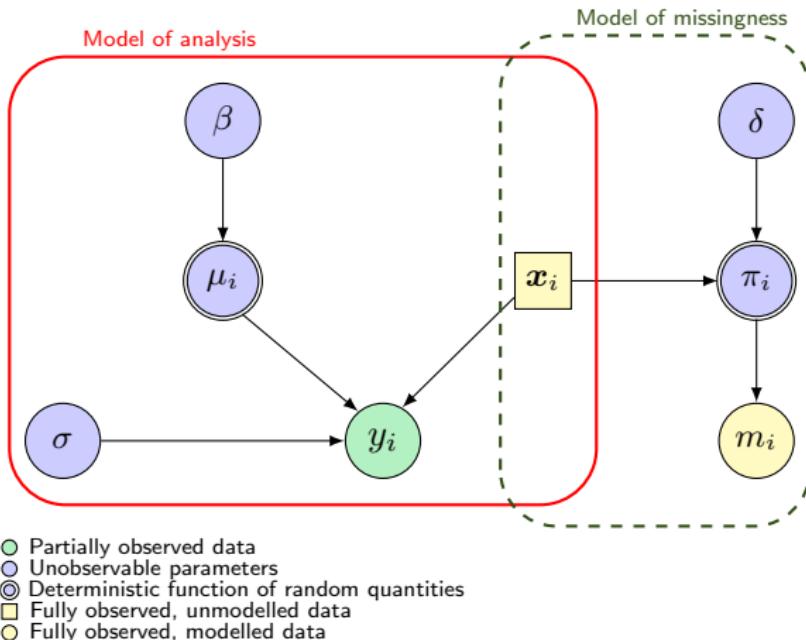
Types of missing data



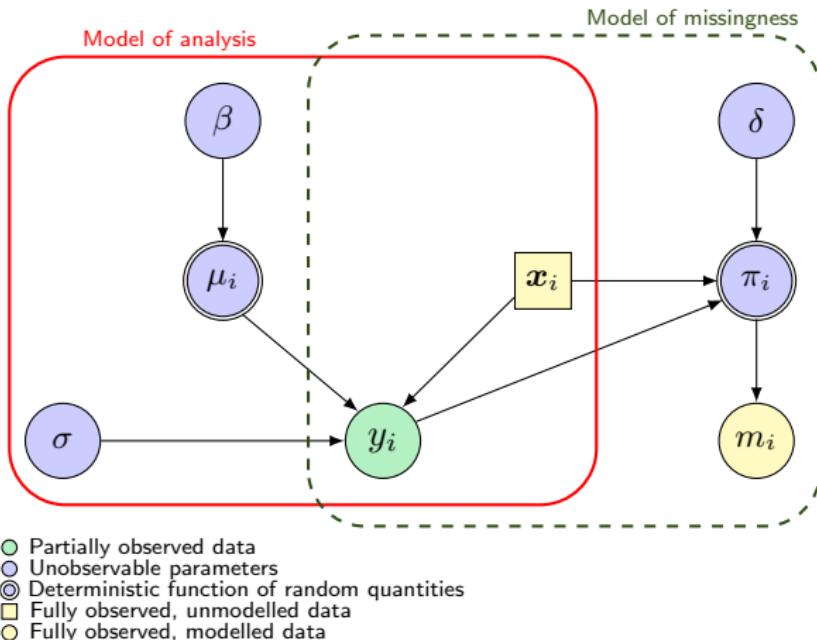


- Green circle: Partially observed data
- Purple circle: Unobservable parameters
- Grey circle: Deterministic function of random quantities
- Yellow square: Fully observed, unmodelled data
- Yellow circle: Fully observed, modelled data

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



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



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

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

Complete case analysis (CCA)

- Essentially swipes the problem under the rug and restricts the analysis to only "valid" cases (ie those completely observed on **all** the variables under consideration)
- Typically, common statistical packages take this approach by default

Call:

```
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.06260	-0.82538	-0.08366	0.84571	2.33579

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)		
(Intercept)	0.99599	0.12418	8.02	8.59e-12 ***		
x	1.01662	0.06796	14.96	< 2e-16 ***		

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’	0.1 ‘ ’	1

Residual standard error: 1.108 on 78 degrees of freedom

(20 observations deleted due to missingness)

Multiple R-squared: 0.7416, Adjusted R-squared: 0.7382

F-statistic: 223.8 on 1 and 78 DF, p-value: < 2.2e-16

Advantages

- Clearly, it is very easy to implement
- If data are MCAR or MAR (and the model is correctly specified!), the results are unbiased

Disadvantages

- The sample size is reduced, due to the fact that units subject to missingness are dropped out
 - If missing data affect more than one variable, this has an even bigger impact
- Consequently, the estimates will be associated with (artificially) larger standard errors
 - **NB:** this is a feature of analyses that are characterised by missing data
 - Intuitively, it makes sense: we do not have all the possible relevant information, and thus our estimation will be less precise than we would obtain, had we been able to observe all the relevant data!
- Does not account for the missing generating mechanism
- Can affect sample size calculations

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

- The basic idea is to estimate, for each unit $i = 1, \dots, n_{\text{obs}}$, the probability of being observed, \hat{p}_i
- Then, each unit is weighted by the inverse of this value, $w_i = \hat{p}_i^{-1}$
 - Thus units that are observed and actually have a large chance of being observed are “normal” — since we do expect to see them
 - Hence they will be multiplied by a small factor
 - Conversely, units that are observed but are associated with a small chance of being available are assumed to be representative of “underrepresented” cases
 - Hence they will be multiplied by a larger factor to “balance” the sample
- This is effective particularly (but not necessarily only!) in stratified cases
- **NB:** \hat{p}_i is some kind of **propensity score** — and this procedure is similar to increasingly popular methods in health economics (e.g. **Multiple Adjusted Indirect Comparisons**)

Inverse Probability Weighting

(Ridiculously simplified!) Example

- Suppose we *know* the data to be

Group	A			B			C		
Response	0	1	2	1	2	3	2	3	4

- The “true” estimation given the complete data set is

$$\hat{\mu} = \frac{18}{9} = 2$$

- But actually, some of the values are partially observed and thus the data are

Group	A			B			C		
Response	?	1	?	1	2	3	2	?	4

- The CCA of the observed data would give an estimation

$$\hat{\mu} = \frac{13}{6} \neq 2$$

Inverse Probability Weighting

- We had originally planned 3 observations in group A
- However, because of missingness, only 1 observation is available
- IPW assumes that the available data will carry information **also** for those that are missing and thus re-weight the observation by $w_i = \hat{p}_i^{-1}$
- In this case, assuming the data are MCAR
 - $p_i^A = 1/3 \Rightarrow w_i^A = 1/p_i^A = 3$
 - $p_i^B = 3/3 \Rightarrow w_i^B = 1/p_i^B = 1$
 - $p_i^C = 2/3 \Rightarrow w_i^C = 1/p_i^C = 3/2$
- The IPW estimator can be computed as

$$\begin{aligned}\hat{\mu}_{\text{IPW}} &= \frac{\sum_i y_i w_i}{\sum_i n_i w_i} \\ &= \frac{1 \times 3 + (1+2+3) \times 1 + (2+4) \times 1.5}{(1 \times 3) + (3 \times 1) + (2 \times 1.5)} \\ &= \frac{18}{9} = 2\end{aligned}$$

Advantages

- Not too complex — only requires that we are able to estimate the weights (i.e. the probability of being observed)
- IPW estimates are generally *consistent*
 - With large enough sample size, they produce an unbiased estimation
 - This does not necessarily apply for moderate to small sample sizes

Disadvantages

- We need to estimate the probability of being observed — this might not be so easy!
- Even when we can estimate the “inclusion” probabilities, these estimations might be highly unstable
 - When p_i is close to 0, then the resulting weight tends to ∞
 - If the inclusion probabilities are not estimated with precision, neither will the weights thus rendering the procedure unreliable

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

Single (deterministic) imputation

- Another approach to handling missing data is to **impute** the missing values, using available information
- The most straightforward way of doing this is to substitute missing values with the mean of the observed values

Individual	Response (y)	Imputed (y_{Imp})
1	5.67	5.67
2	3.28	3.28
3	8.14	8.14
4	4.40	4.40
5	?	7.14
6	9.18	9.18
7	?	7.14
8	?	7.14
9	12.19	12.19
10	?	7.14

$$\begin{aligned} \text{Mean } \bar{y}_{\text{obs}} &= \frac{1}{n_{\text{obs}}} \sum_{i=1}^{n_{\text{obs}}} y_i & \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i \\ &= 7.14 & &= 7.14 \end{aligned}$$

Advantages

- Again, very easy to use

Disadvantages

- If observed data are on an ordinal scale, strictly speaking not applicable
- Assumes no underlying individual variability around missing data
 - All missing data points are treated as if they came from the exact same generating process
 - This goes even beyond (= it's worse than) the assumption of MCAR!

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

Example

- Suppose we partially observe a continuous variable y on $n_{\text{obs}} < n$ individuals
- Also, we completely observe an additional variable x on the n individuals in the sample
- The simplest way in which we can describe the relationship between y and x is through linear regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{with } \varepsilon_i \sim \text{Normal}(0, \sigma^2)$$

Example

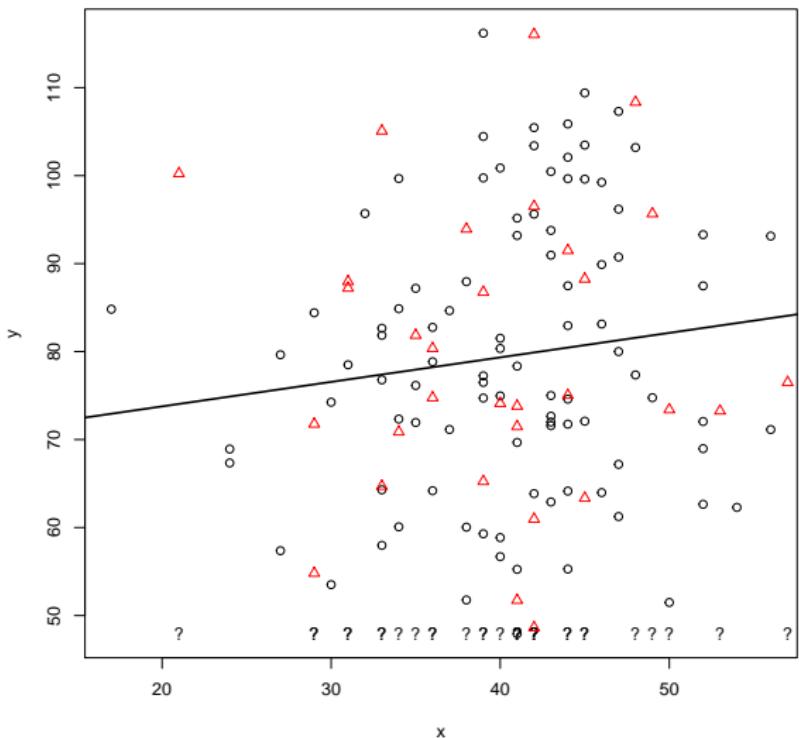
- Suppose we partially observe a continuous variable y on $n_{\text{obs}} < n$ individuals
- Also, we completely observe an additional variable x on the n individuals in the sample
- The simplest way in which we can describe the relationship between y and x is through linear regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{with } \varepsilon_i \sim \text{Normal}(0, \sigma^2)$$

- We can proceed in the following way
 - ① Fit the regression model to the complete cases and estimate the parameters $\theta = (\beta_0, \beta_1, \sigma^2)$
 - ② Build a linear predictor for the missing data points $\mu_i^{\text{mis}} = \hat{\beta}_0 + \hat{\beta}_1 x_i$
 - ③ Simulate a random value by the induced distribution of the missing data

$$y_i^{\text{mis}} \sim \text{Normal}(\mu_i^{\text{mis}}, \hat{\sigma}^2)$$

Multiple imputation



- If we had infinite data on which to base our estimation of the parameters $\theta = (\beta_0, \beta_1, \sigma^2)$ for the imputation model, then a single draw would be sufficient to fully represent our uncertainty in the missingness patterns (provided that the imputation model is correctly specified!)

- If we had infinite data on which to base our estimation of the parameters $\theta = (\beta_0, \beta_1, \sigma^2)$ for the imputation model, then a single draw would be sufficient to fully represent our uncertainty in the missingness patterns (provided that the imputation model is correctly specified!)
- However, in general that is **not** the case
 - The dataset is typically small to moderate
 - We are only using the complete cases anyway to estimate θ
 - The imputation model is possibly mis-specified
- Consequently, the resulting linear predictor and the estimation of the individual variance are likely to be with limited precision

- If we had infinite data on which to base our estimation of the parameters $\theta = (\beta_0, \beta_1, \sigma^2)$ for the imputation model, then a single draw would be sufficient to fully represent our uncertainty in the missingness patterns (provided that the imputation model is correctly specified!)
- However, in general that is **not** the case
 - The dataset is typically small to moderate
 - We are only using the complete cases anyway to estimate θ
 - The imputation model is possibly mis-specified
- Consequently, the resulting linear predictor and the estimation of the individual variance are likely to be with limited precision
- To deal with this...
 - Simulate multiple instances of the possible value that the missing observations might have had (if we had been able to observe them), by considering a number K of draws from their distribution

① Imputation

- Impute (fill in) the missing entries of the incomplete data sets K (typically 5-10) times, by simulating from a given model (e.g. linear regression, in the previous case)
- This step results in K complete data sets

① Imputation

- Impute (fill in) the missing entries of the incomplete data sets K (typically 5-10) times, by simulating from a given model (e.g. linear regression, in the previous case)
- This step results in K complete data sets

② Analysis

- Analyse each of the K completed datasets using standard methods
- This amounts to computing the estimate \bar{y}_k with variance σ_k^2

Multiple imputation

① Imputation

- Impute (fill in) the missing entries of the incomplete data sets K (typically 5-10) times, by simulating from a given model (e.g. linear regression, in the previous case)
- This step results in K complete data sets

② Analysis

- Analyse each of the K completed datasets using standard methods
- This amounts to computing the estimate \bar{y}_k with variance σ_k^2

③ Pooling

- Combine the K analysis results into a single final result

$$\bar{y}_{\text{MI}} = \frac{1}{K} \sum_{k=1}^K \bar{y}_k$$

with variance

$$\sigma_{\text{MI}}^2 = \left(1 + \frac{1}{K}\right) \underbrace{\left[\frac{1}{K-1} \sum_{k=1}^K (\bar{y}_k - \bar{y}_{\text{MI}})^2 \right]}_{\text{"between imputation"}} + \underbrace{\left[\frac{1}{K} \sum_{k=1}^K \sigma_k^2 \right]}_{\text{"within imputation"}}$$

Advantages

- (Generally) valid under MCAR and MAR assumptions
- Makes use of the whole dataset
- Can be extended to MNAR, although models become more complex and untestable assumptions are necessary

Disadvantages

- Leads to biased results if the imputation model is completely mis-specified ("congeniality")
- Can be computationally intensive



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

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

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

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

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

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

- 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 HTA

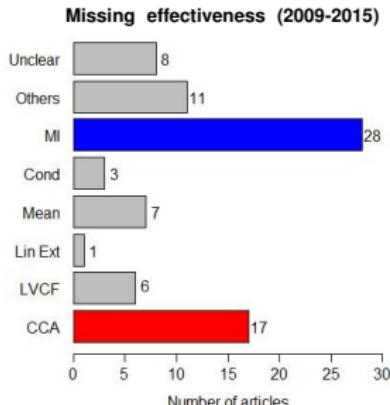
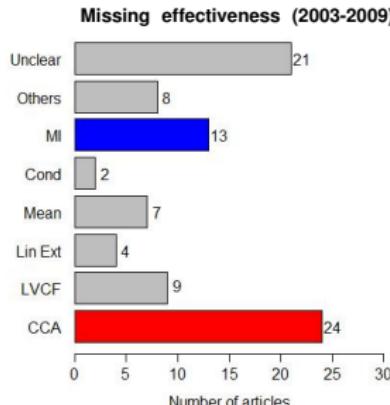
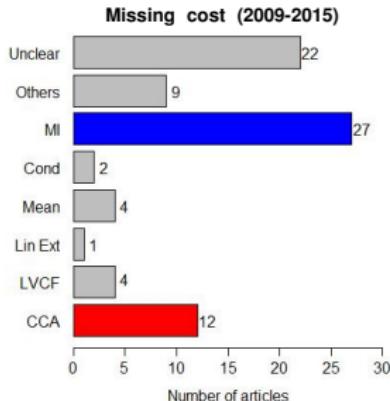
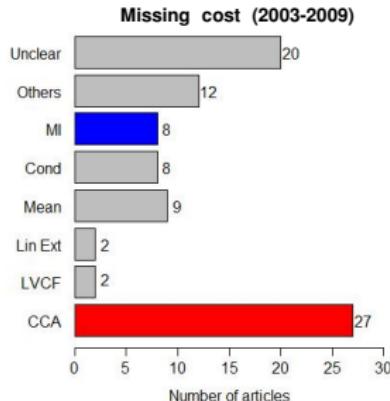
[Back to Table of content](#)

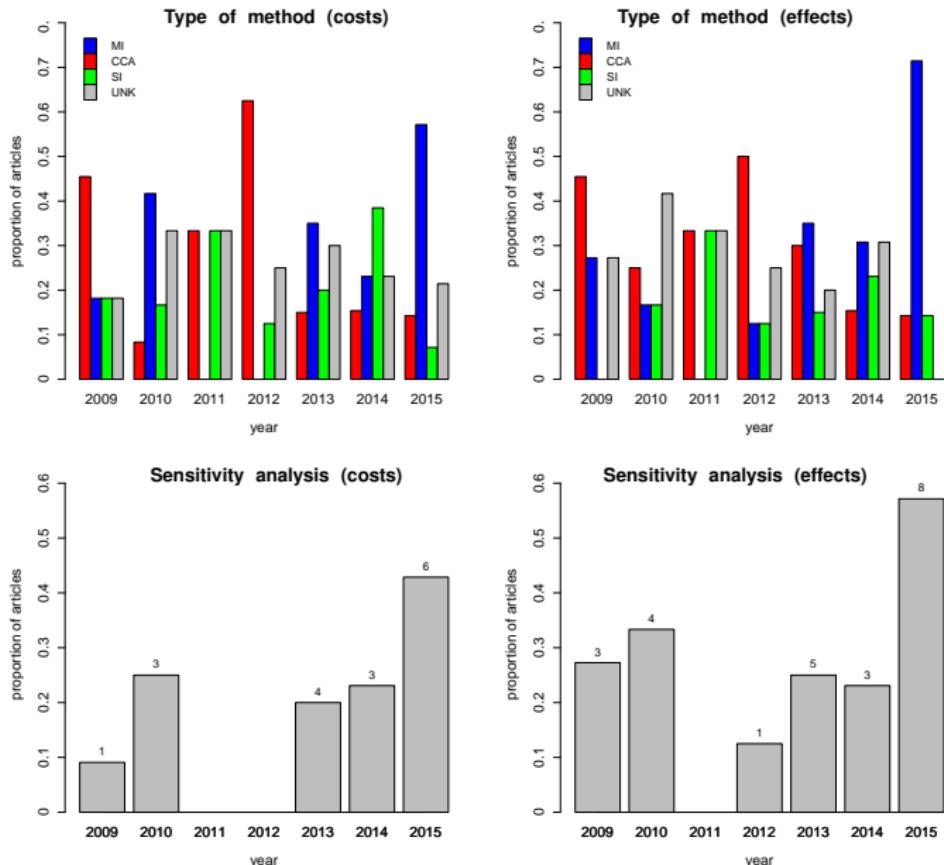
- Decisions in HTA are often informed by within-trial CEA
- Missing data occur frequently in RCTs in both effectiveness and cost outcomes

- 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





Key points:

Key points:

- Change from more to less restrictive methods as default choices (**CCA** → **MI**)
- The number of sensitivity analyses has increased

Key points:

- Change from more to less restrictive methods as default choices (**CCA** → **MI**)
- The number of sensitivity analyses has increased

But ...

- **Transparent information** about the assumptions on missing data is rarely provided
- **Choice of the methods** is often guided by their ease of implementation in standard software packages rather than methodological reasons
- Analyses are almost exclusively based on **MAR**

Problems with HTA data

The analysis of partially-observed data is always complicated

The analysis of partially-observed data is always complicated

- Missingness assumptions are **untestable**
- Cannot check model fit to unobserved data
- Have to accept inherent uncertainty in our analysis

Problems with HTA data

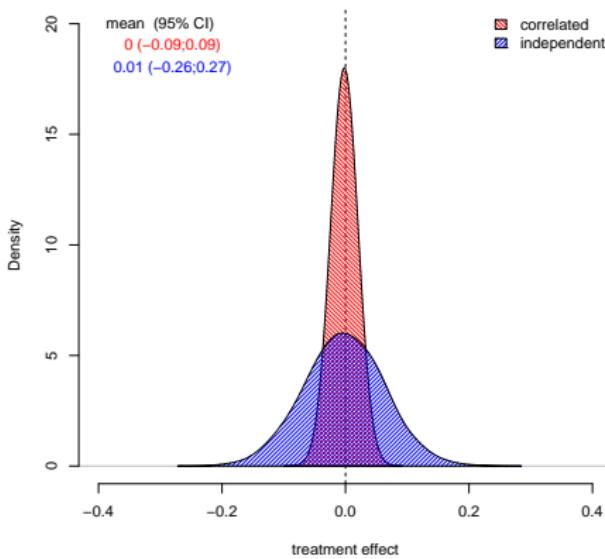
The analysis of partially-observed data is always complicated

- Missingness assumptions are **untestable**
- Cannot check model fit to unobserved data
- Have to accept inherent uncertainty in our analysis

In HTA it's even more complicated ...

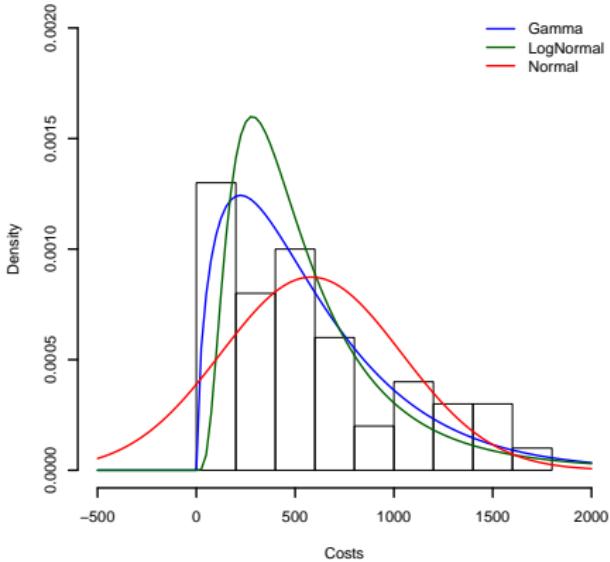
- Bivariate outcome, usually **correlated**
- Normality not reasonable (**skewed** data) and small sample sizes
- **Spikes** at 1 (utilities) or 0 (costs) may occur
- Main objective: **decision-making**, not inference

Problems with HTA data – Correlation



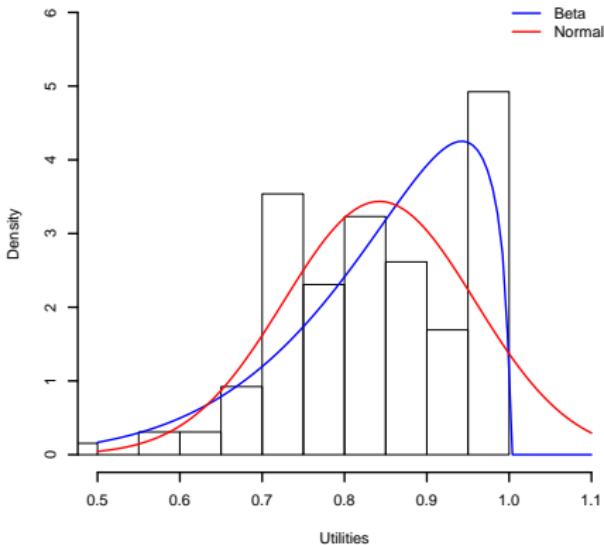
- Ignoring correlation may lead to erroneous statistical inferences due to loss of information (e.g. larger confidence intervals)
- Important in HTA where the assessment of uncertainty is key for decision-making

Problems with HTA data – Skewness



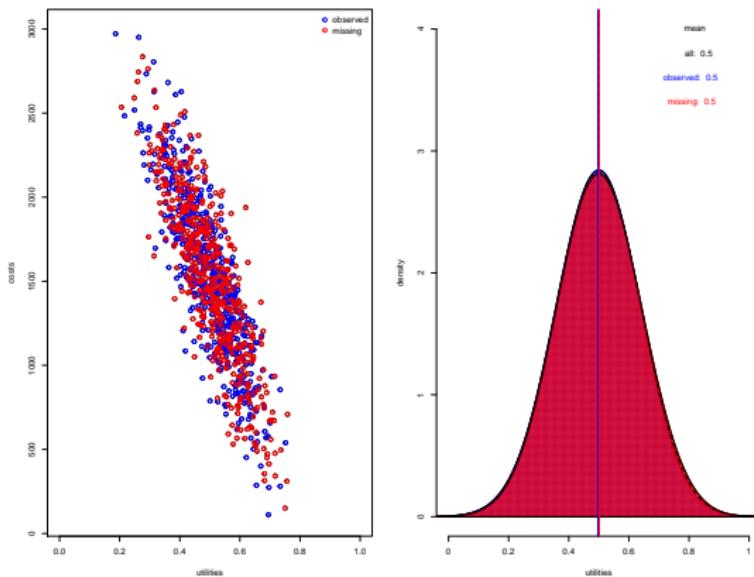
- Alternative parametric distributions have better fit to skewed data (e.g. **Gamma** or **LogNormal**)
- When the sample size is not large, inferences based on the **Normal** distribution may be incorrect

Problems with HTA data – Spikes



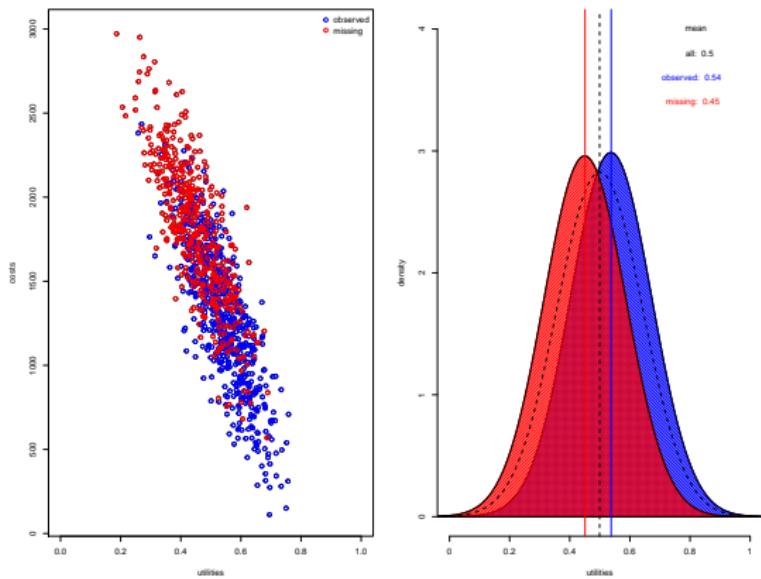
- Spikes often induce high skewness which may be difficult to capture using standard distributions
- The **Beta** distribution has a better fit than the **Normal** but is not defined at 1

Missing data in HTA – MCAR



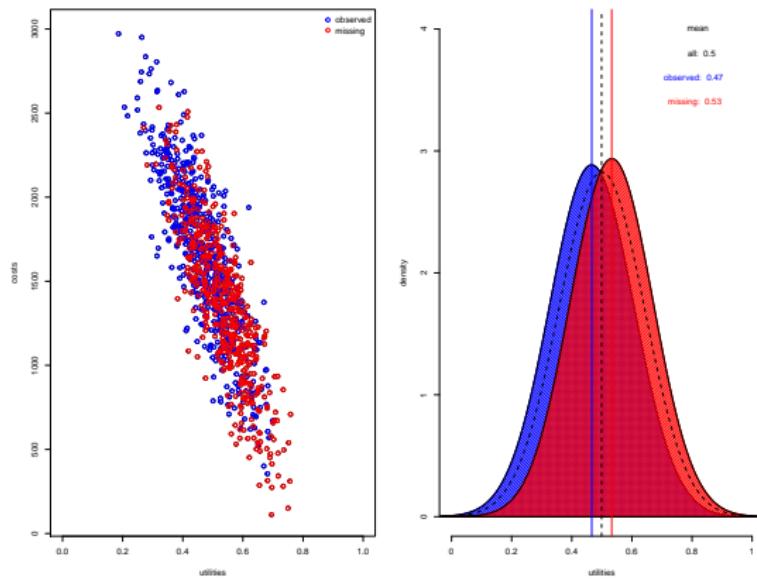
- **Missingness** is completely random
- Inferences on the **observed** data are valid

Missing data in HTA – MAR



- **Missingness** depend on costs (subjects with higher costs have a higher chance to be missing)
- Inferences on the **observed** data are valid if costs are used as predictors of missingness

Missing data in HTA – MNAR



- **Missingness** depend on utilities (subjects with higher utilities have a higher chance to be missing)
- Inferences on the **observed** data are not valid

- Handling missing data in HTA 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

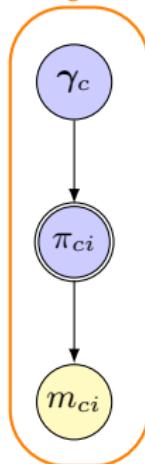
- Handling missing data in HTA can be challenging because:
 - Outcomes (costs and QALYs) are typically correlated and non-normally distributed
 - Limited information is often available about the reasons of missingness
 - MAR may be difficult to justify and cannot be checked from the data
- These features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

- Handling missing data in HTA 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**
- Two general types of models which allow to perform **sensitivity analysis** under MNAR are:
 - **Selection models** – model for the missingness mechanism
 - **Pattern mixture models** – model for each missingness pattern

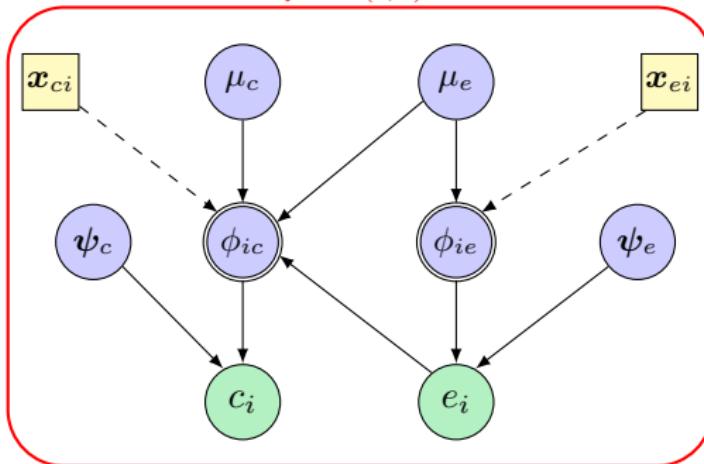
Missing data in HTA – Selection models

MCAR (e, c)

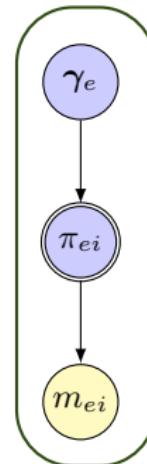
Model of missingness for c



Model of analysis for (c, e)



Model of missingness for e



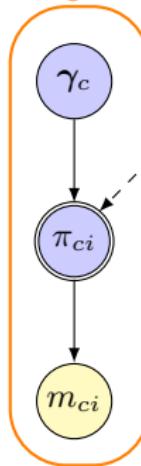
- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

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

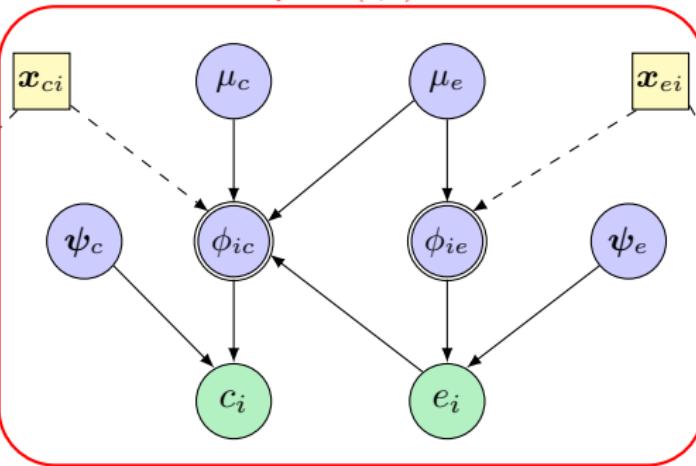
Missing data in HTA – Selection models

MAR (e, c)

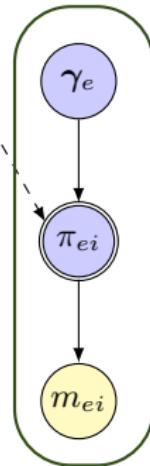
Model of missingness for c



Model of analysis for (c, e)



Model of missingness for e



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

- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$;
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$;

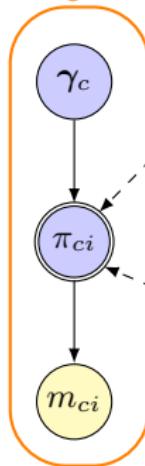
$$\text{logit}(\pi_{ei}) = \gamma_{e0} + \sum_{k=1}^K \gamma_{ek} x_{eik}$$

$$\text{logit}(\pi_{ci}) = \gamma_{c0} + \sum_{h=1}^H \gamma_{ch} x_{cih}$$

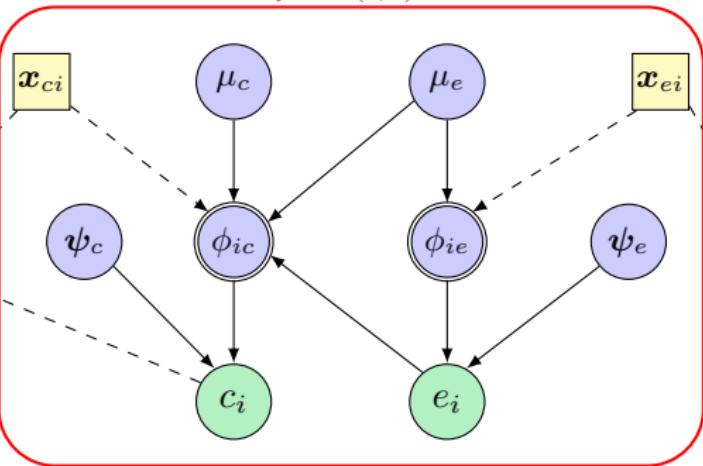
Missing data in HTA – Selection models

MAR e ; MNAR c

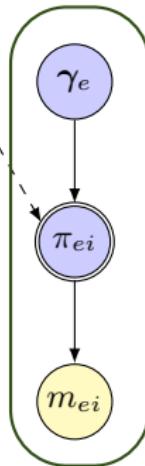
Model of missingness for c



Model of analysis for (c, e)



Model of missingness for e



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

- $m_{ei} \sim \text{Bernoulli}(\pi_{ei});$
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci});$

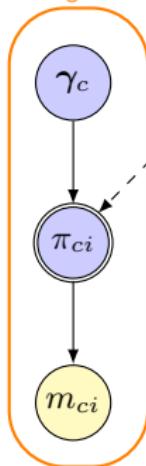
$$\text{logit}(\pi_{ei}) = \gamma_{e0} + \sum_{k=1}^K \gamma_{ek} x_{eik}$$

$$\text{logit}(\pi_{ci}) = \gamma_{c0} + \sum_{h=1}^H \gamma_{ch} x_{cih} + \gamma_{cH+1} c_i$$

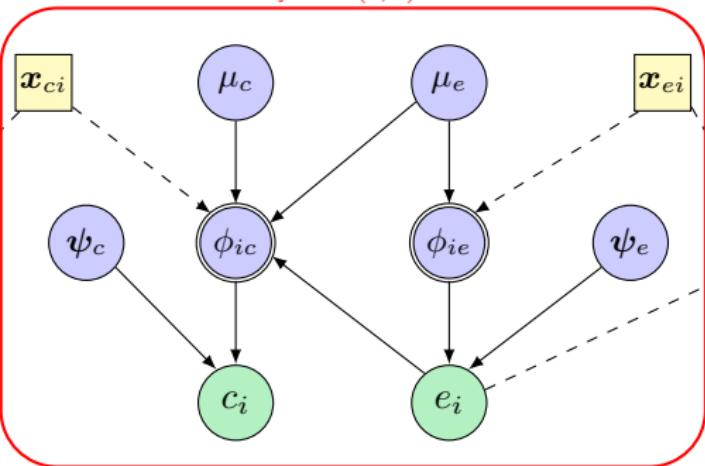
Missing data in HTA – Selection models

MNAR e ; MAR c

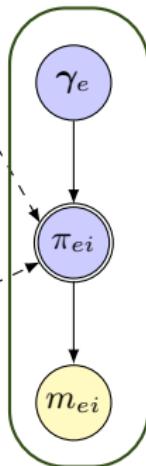
Model of missingness for c



Model of analysis for (c, e)



Model of missingness for e



- Green circle: Partially observed data
- Purple circle: Unobservable parameters
- Grey circle: Deterministic function of random quantities
- Yellow square: Fully observed, unmodelled data
- Yellow circle: Fully observed, modelled data

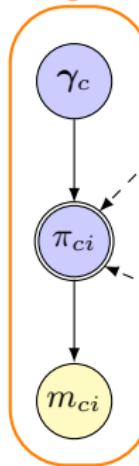
- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$;
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$;

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

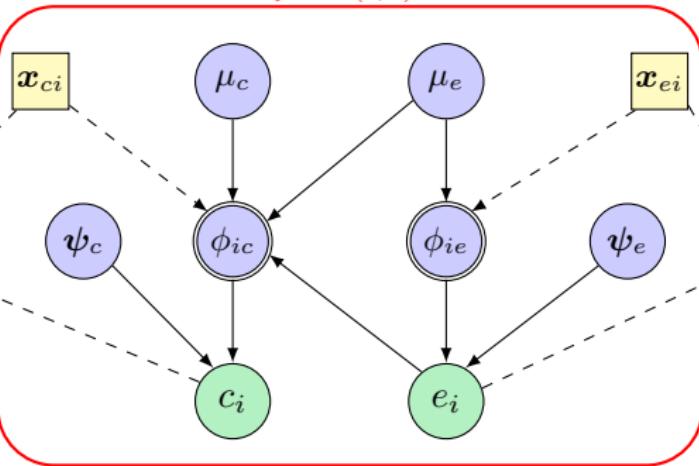
Missing data in HTA – Selection models

MNAR (e, c)

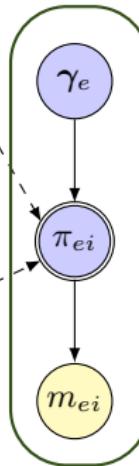
Model of missingness for c



Model of analysis for (c, e)



Model of missingness for e



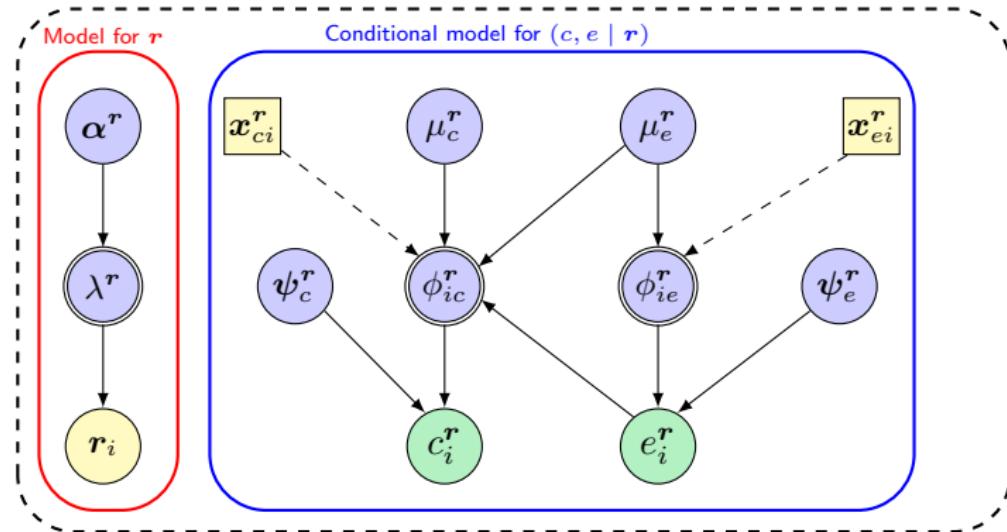
- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

- $m_{ei} \sim \text{Bernoulli}(\pi_{ei})$;
- $m_{ci} \sim \text{Bernoulli}(\pi_{ci})$;

$$\begin{aligned}\text{logit}(\pi_{ei}) &= \gamma_{e0} + \sum_{k=1}^K \gamma_{ek} x_{eik} + \gamma_{eK+1} e_i \\ \text{logit}(\pi_{ci}) &= \gamma_{c0} + \sum_{h=1}^H \gamma_{ch} x_{cih} + \gamma_{cH+1} c_i\end{aligned}$$

Missing data in HTA – Pattern mixture models

MAR (e, c)



● Partially observed data

● Unobservable parameters

● Deterministic function of random quantities

● Fully observed, unmodelled data

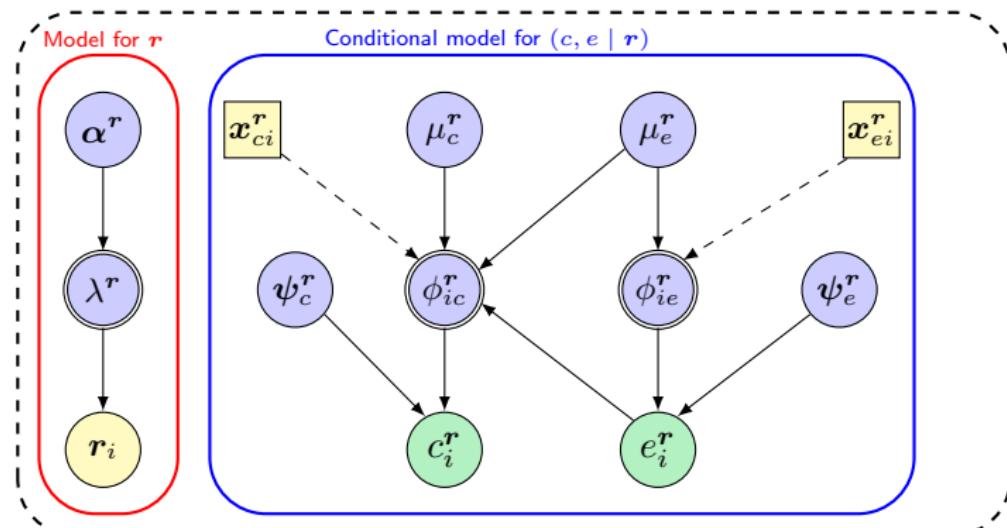
● Fully observed, modelled data

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

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

Missing data in HTA – Pattern mixture models

MAR e ; MNAR c



● Partially observed data

● Unobservable parameters

● Deterministic function of random quantities

● Fully observed, unmodelled data

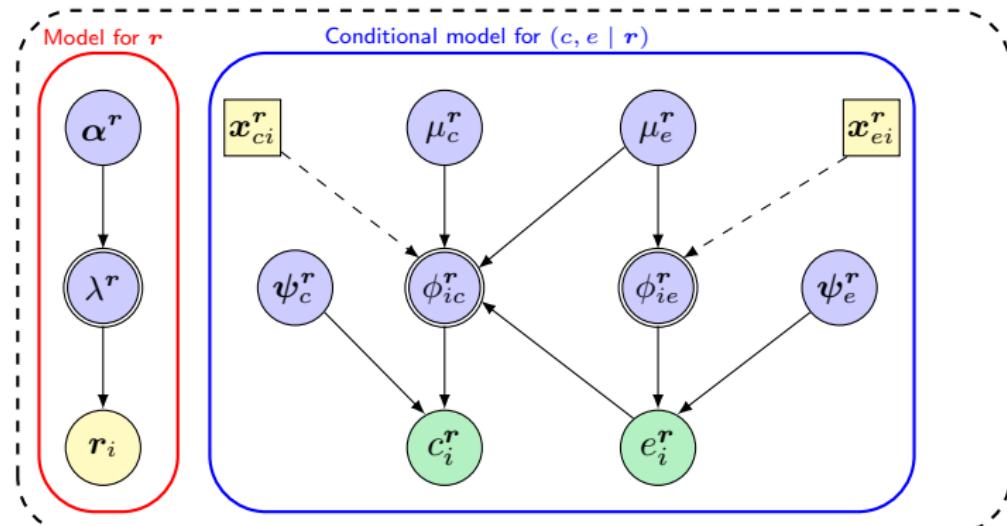
● Fully observed, modelled data

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

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

Missing data in HTA – Pattern mixture models

MNAR e ; MAR c



● Partially observed data

● Unobservable parameters

● Deterministic function of random quantities

● Fully observed, unmodelled data

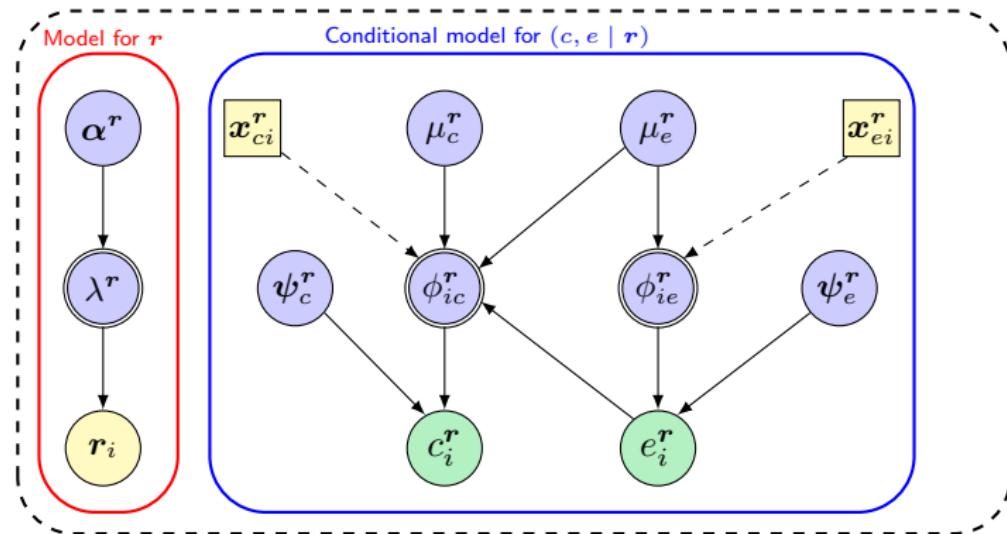
● Fully observed, modelled data

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

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

Missing data in HTA – Pattern mixture models

MNAR (e, c)



● Partially observed data

● Unobservable parameters

● Deterministic function of random quantities

● Fully observed, unmodelled data

● Fully observed, modelled data

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

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

- The steps involved in drawing inference from incomplete data are:
 - Specification of a full data model
 - Specification of the priors
 - Sampling from the posterior of full-data parameters, given the observed data $(e_i^{obs}, c_i^{obs}), (m_{ei}, m_{ci})$ and $(\boldsymbol{x}_{ci}, \boldsymbol{x}_{ei})$

- The steps involved in drawing inference from incomplete data are:
 - Specification of a full data model
 - Specification of the priors
 - Sampling from the posterior of full-data parameters, given the observed data (e_i^{obs}, c_i^{obs}) , (m_{ei}, m_{ci}) and $(\boldsymbol{x}_{ci}, \boldsymbol{x}_{ei})$
- The identification of the model requires unverifiable assumptions about missingness
 - Under an ignorability assumption (MAR), the model is fully-identified from the observed data alone
 - When ignorability is believed not to be a suitable assumption, nonignorable models allow missing data indicators to depend on missing responses themselves (MNAR)
 - These models allow one to parameterise the conditional dependence between (m_{ei}, m_{ci}) and (e_i^{mis}, c_i^{mis})
 - Inference depends on some combination of (a) **unverifiable parametric assumptions** and (b) **informative prior distributions**

- Selection models are attractive because:
 - The analysis model $p(e_i, c_i)$ can be specified directly
 - The model factorisation enables an easy characterisation of the missing data mechanism and appeals directly to the Rubin's taxonomy (MCAR,MAR,MNAR)

- Selection models are attractive because:
 - The analysis model $p(e_i, c_i)$ can be specified directly
 - The model factorisation enables an easy characterisation of the missing data mechanism and appeals directly to the Rubin's taxonomy (MCAR,MAR,MNAR)
- There are two types of **sensitivity analysis** for selection models:
 - **Assumption sensitivity** varies the distribution and form of the analysis and/or missingness models
 - **Parameter sensitivity** varies the prior distributions on the parameters linking the unobserved data to the missingness probabilities
- Inferences may strongly depend on the parametric specification used – **difficult to check**

- Pattern mixture have the advantage that:

- In many cases one can find parameters indexing the distribution of missing data that are not identified by observed data – **sensitivity parameters**
 - Use these parameters to identify the model without relying on distributional assumptions about the full data
 - This makes inferences about full data more transparent in that the source of information about the posterior is clear

- Pattern mixture have the advantage that:
 - In many cases one can find parameters indexing the distribution of missing data that are not identified by observed data – **sensitivity parameters**
 - Use these parameters to identify the model without relying on distributional assumptions about the full data
 - This makes inferences about full data more transparent in that the source of information about the posterior is clear
- **Sensitivity analysis** for pattern mixture models typically requires:
 - The identification of the model through **identifying restrictions**, using the parameters identified from the observed data to define a benchmark scenario
 - The specification of informative priors on the **sensitivity parameters** to explore departures from the benchmark

Missing data **in HTA** – Conclusions

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

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

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

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

Examples

[Back to Table of content](#)

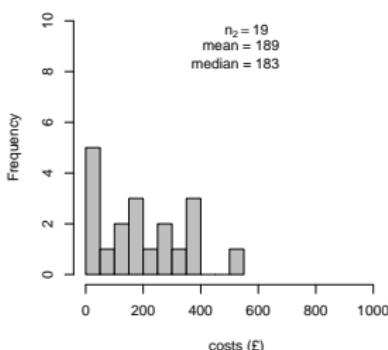
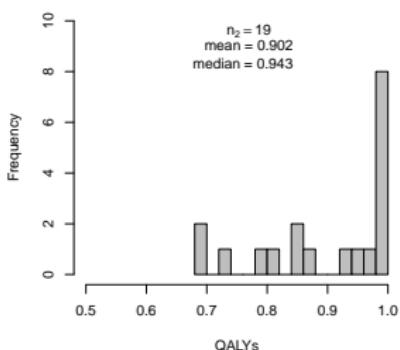
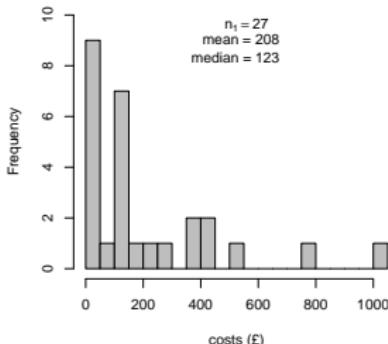
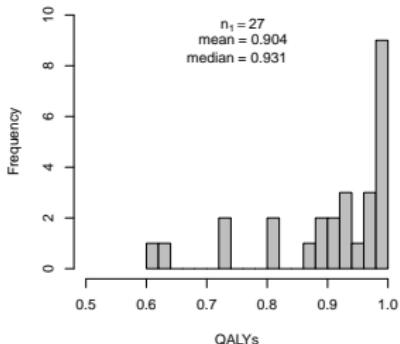
Motivating example: MenSS trial

- 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)

- 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)



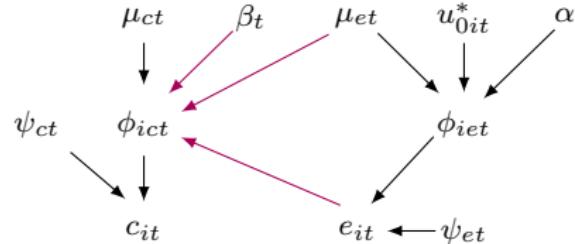
① Bivariate Normal

- 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

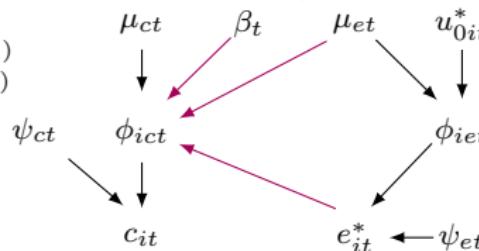
② Beta-Gamma

- 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 $c | e^*$

$$c_{it} | e_{it}^* \sim \text{Gamma}(\psi_{ct}\phi_{cit}, \psi_{ct})$$

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



Marginal model for e^*

$$e_{it}^* \sim \text{Beta}(\phi_{eit}\psi_{et}, (1 - \phi_{eit})\psi_{et})$$

$$\text{logit}(\phi_{eit}) = \mu_{et} + \alpha_t(u_{0it} - \bar{u}_{0t})$$

$$= \mu_{et} + \alpha_t u_{0it}^*$$

① Bivariate Normal

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

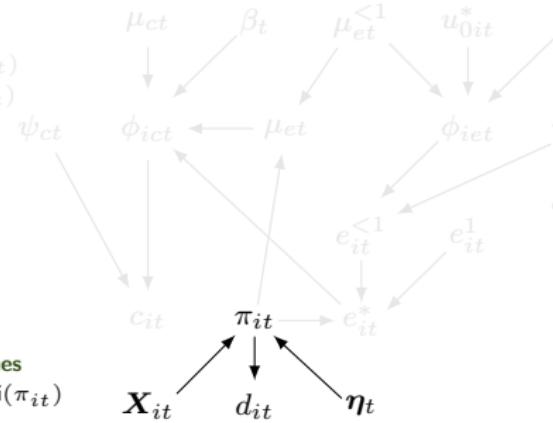
② Beta-Gamma

- 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

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



Mixture model for e
 $e_{it}^1 := 1$
 $e_{it}^{<1>} \sim \text{Beta}(\phi_{eit}\psi_{et}, (1 - \phi_{eit})\psi_{et})$
 $\text{logit}(\phi_{eit}) = \mu_{et}^{<1>} + \alpha_t(u_{0it} - \bar{u}_{0t})$
 $\text{logit}(\phi_{eit}) = \mu_{et}^{<1>} + \alpha_t u_{0it}^*$

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

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

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

① Bivariate Normal

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

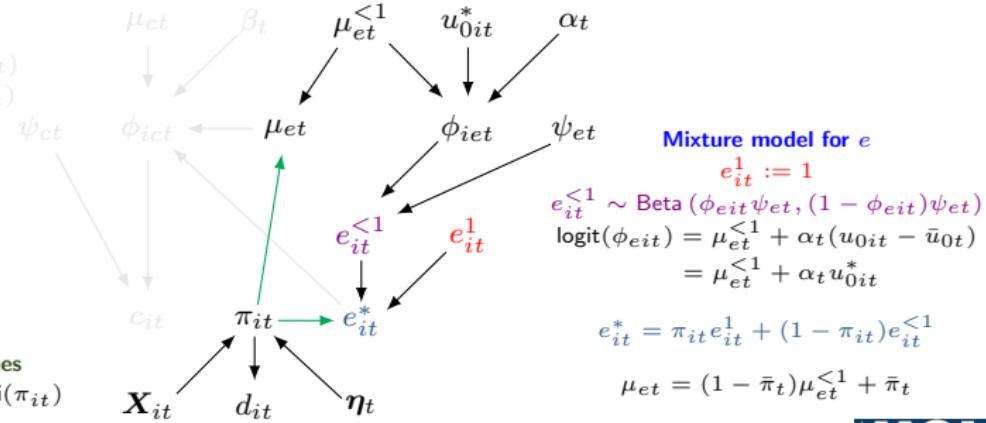
② Beta-Gamma

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



Model for the structural ones

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

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

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

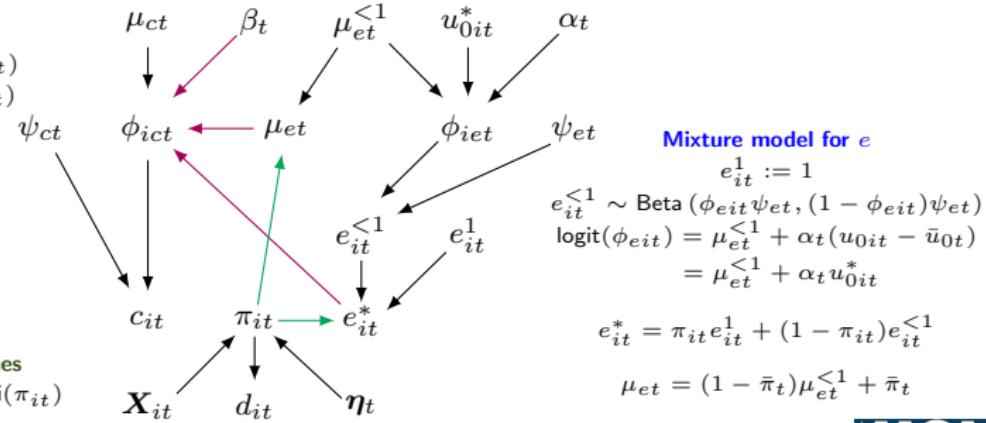
③ Hurdle model

- 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 $c_{it} | e_{it}^*$

$$c_{it} | e_{it}^* \sim \text{Gamma}(\psi_{ct}\phi_{cit}, \psi_{ct})$$

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



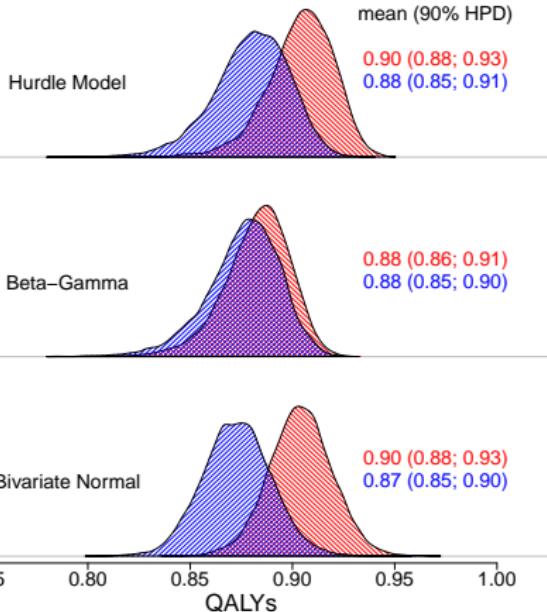
Model for the structural ones

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

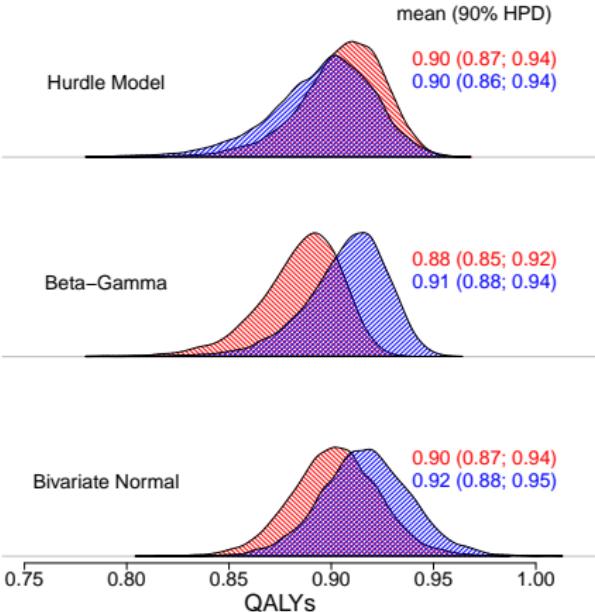
$$\text{logit}(\pi_{it}) = X_{it} \eta_t$$

Results: QALYs

control

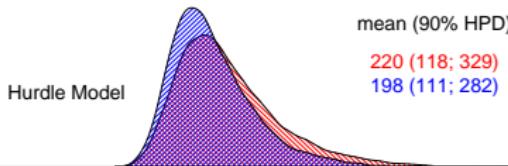


intervention

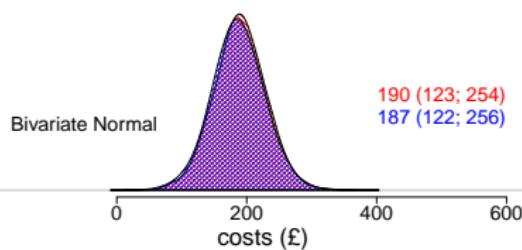
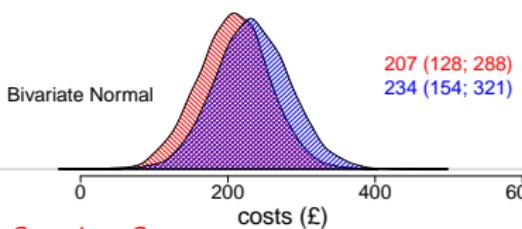
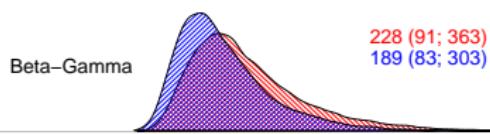
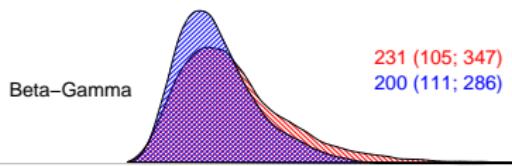
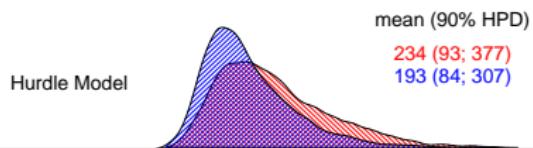


Results: Costs

control



intervention

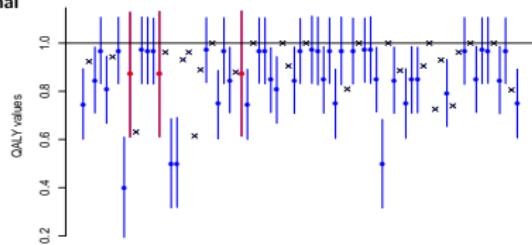


Complete Cases

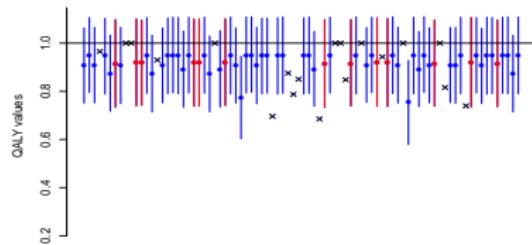
All cases (Missing At Random)

Bayesian multiple imputation (under MAR)

Bivariate Normal



Individuals ($n_1 = 75$)

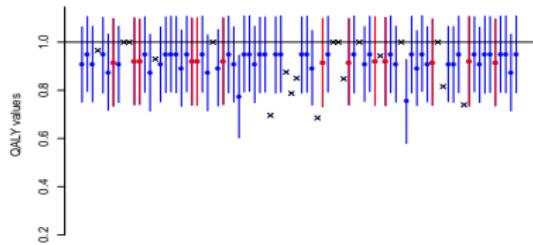
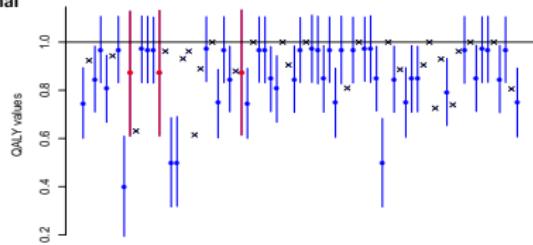


Individuals ($n_2 = 84$)

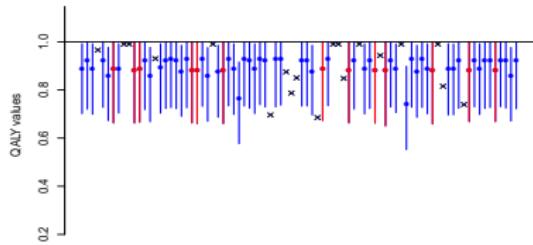
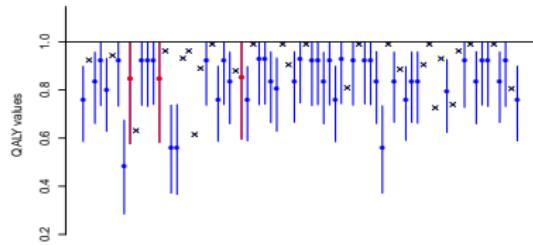
—●— Imputed, observed baseline
—●— Imputed, missing baseline
× Observed

Bayesian multiple imputation (under MAR)

Bivariate Normal



Beta-Gamma



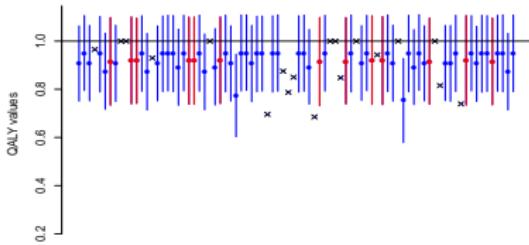
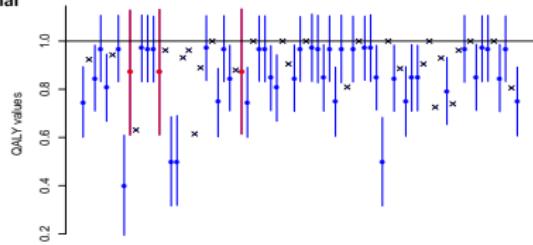
Individuals ($n_1 = 75$)

Individuals ($n_2 = 84$)

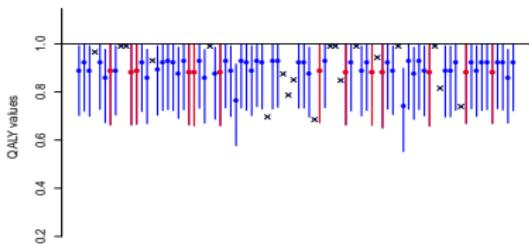
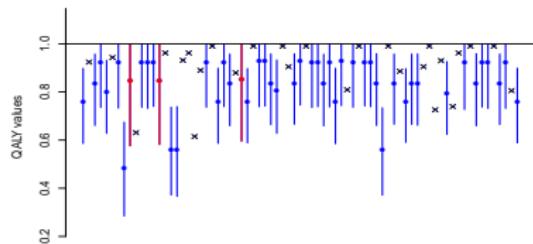
- Imputed, observed baseline
- Imputed, missing baseline
- × Observed

Bayesian multiple imputation (under MAR)

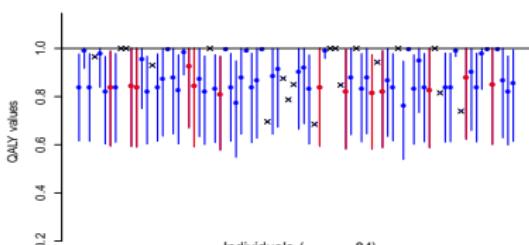
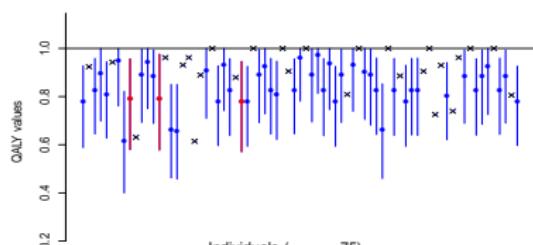
Bivariate Normal



Beta-Gamma



Hurdle model



● Imputed, observed baseline
○ Imputed, missing baseline
✖ Observed

- We observe $n_{01} = 13$ and $n_{02} = 22$ individuals with $u_{0it} = 1$ and $u_{j it} = \text{NA}$, 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 $u_{j it} = \text{NA}$, 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 $u_{j it} = \text{NA}$, 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 $u_{j it} = \text{NA}$, 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 $u_{j it} = \text{NA}$, 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

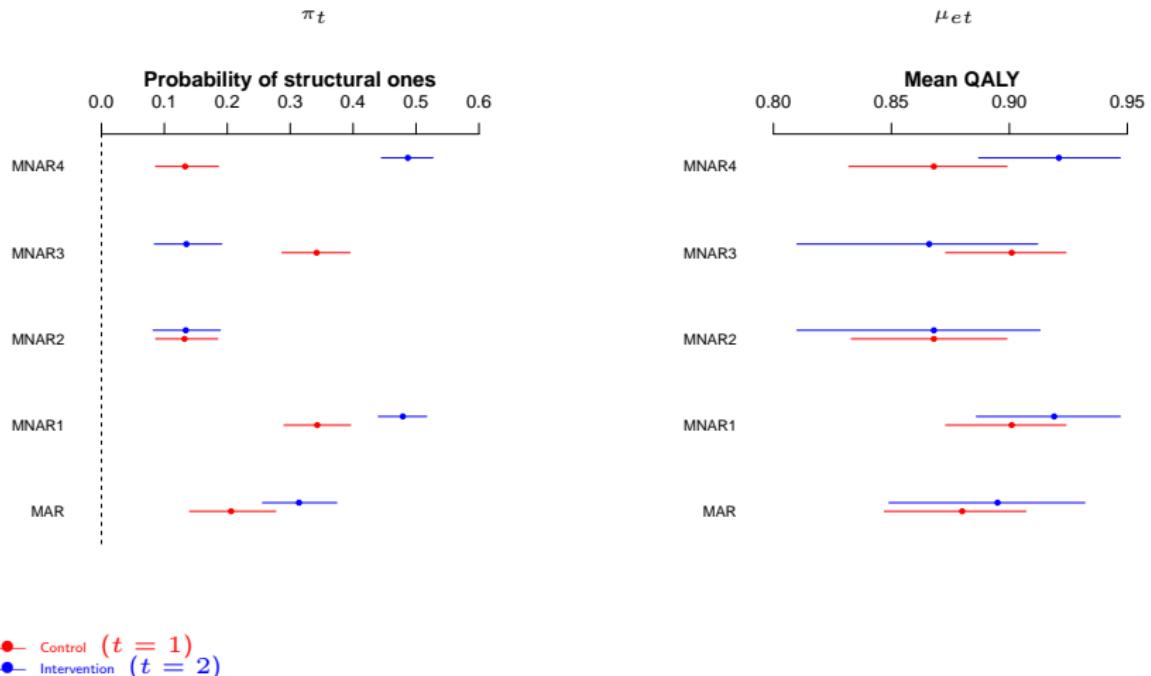
MNAR1. Set $d_{it} = 1$ for all individuals with unit observed baseline utility

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

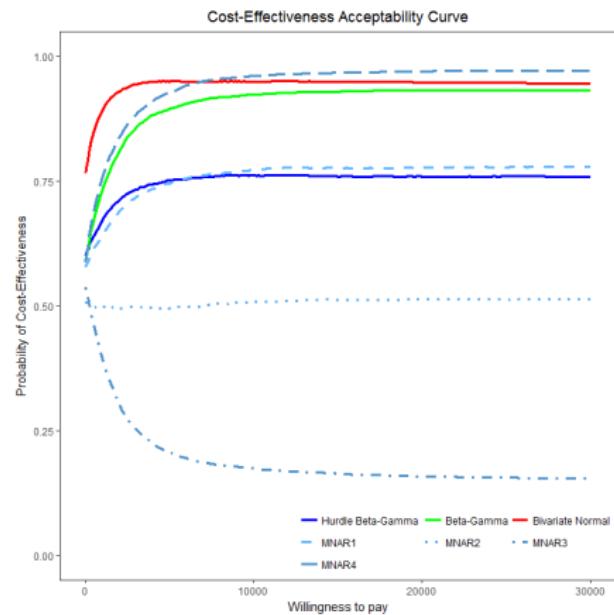
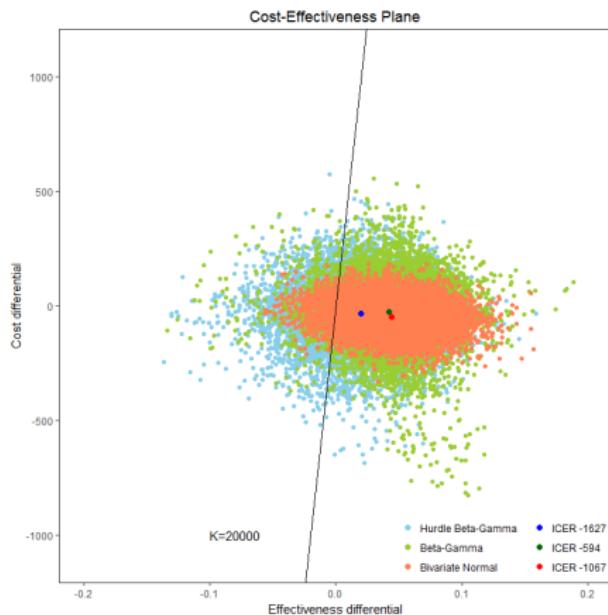
MNAR3. Set $d_{it} = 1$ for the $n_{01} = 13$ individuals with $u_{0i1} = 1$ and $d_{it} = 0$ for the $n_{02} = 22$ individuals with $u_{0i2} = 1$

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 $u_{0i2} = 1$

Results — MNAR



Cost-effectiveness analysis



- HTA 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

Extension: a longitudinal model to handle missingness

- QALYs and total costs are cumulative measures derived from longitudinal data collected over the trial follow-up

- QALYs and total costs are cumulative measures derived from longitudinal data collected over the trial follow-up
 - Each component may have a different missing data pattern
 - Missing data at one timepoint implies that the aggregate variable is also missing

- QALYs and total costs are cumulative measures derived from longitudinal data collected over the trial follow-up
 - Each component may have a different missing data pattern
 - Missing data at one timepoint implies that the aggregate variable is also missing
- Outcome variables can be dealt with at various levels of aggregation:

- QALYs and total costs are cumulative measures derived from longitudinal data collected over the trial follow-up
 - Each component may have a different missing data pattern
 - Missing data at one timepoint implies that the aggregate variable is also missing
- Outcome variables can be dealt with at various levels of aggregation:
 - The typical approach focuses on aggregated data (e_i, c_i) – **cross-sectional**
 - Alternative approach focuses on disaggregated data (u_{ij}, c_{ij}) – **longitudinal**
- When there are missing data, longitudinal models are typically more efficient

Extension: a longitudinal model to handle missingness

- Advantages

- Advantages

- Account for time dependence between outcomes $y_{ij} = (u_{ij}, c_{ij})$
- Use all available utility/cost data in each pattern $r_{ij} = (r_{ij}^u, r_{ij}^c)$

Extension: a longitudinal model to handle missingness

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

Extension: a longitudinal model to handle missingness

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

Extension: a longitudinal model to handle missingness

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

Second example: PBS trial

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

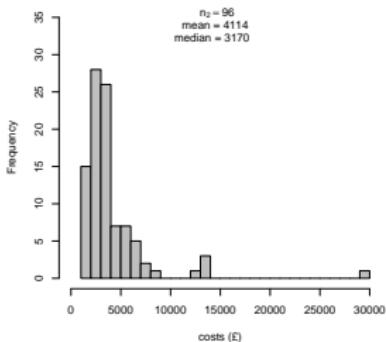
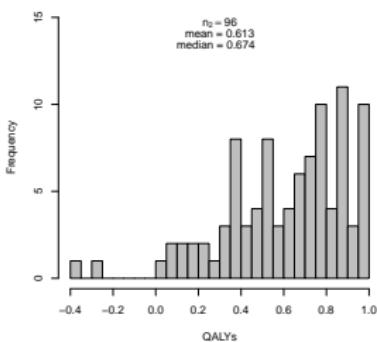
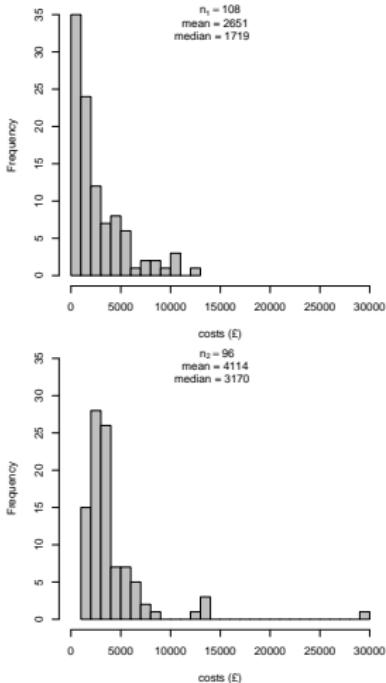
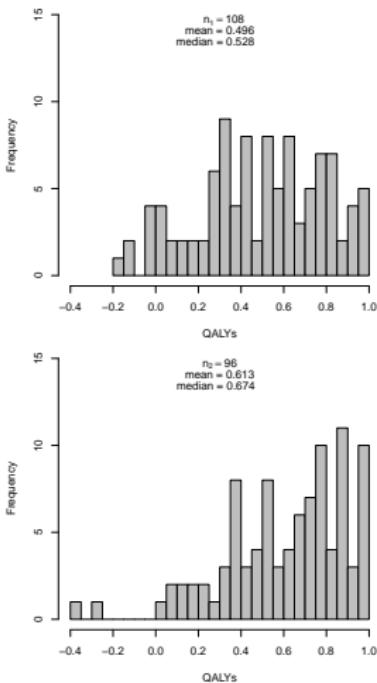
- Multi-centre RCT that evaluates the cost-effectiveness of a new multicomponent intervention for individuals suffering from intellectual disability and challenging behaviour
- Both utilities (EQ-5D) and costs (clinic records) are partially-observed

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

Second example: PBS trial

QALYs and total costs

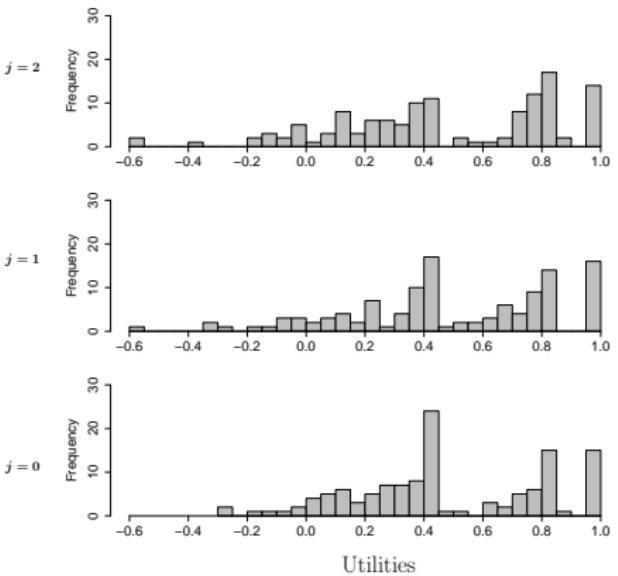
- Multi-centre RCT that evaluates the cost-effectiveness of a new multicomponent intervention for individuals suffering from intellectual disability and challenging behaviour
- Both utilities (EQ-5D) and costs (clinic records) are partially-observed



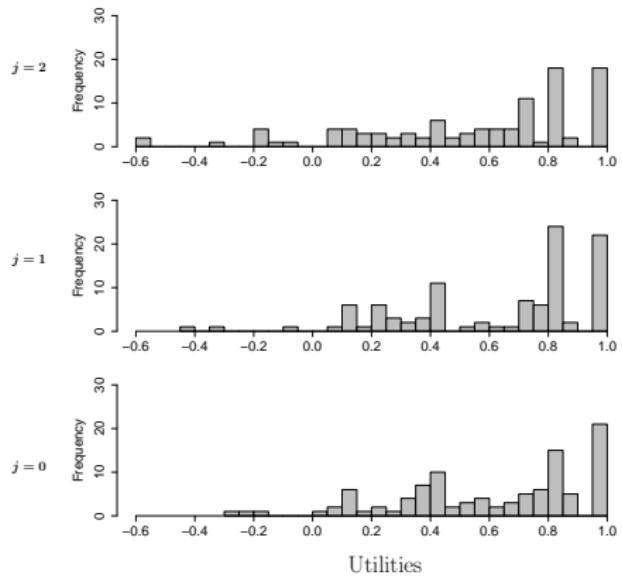
Second example: PBS trial

utilities

Control



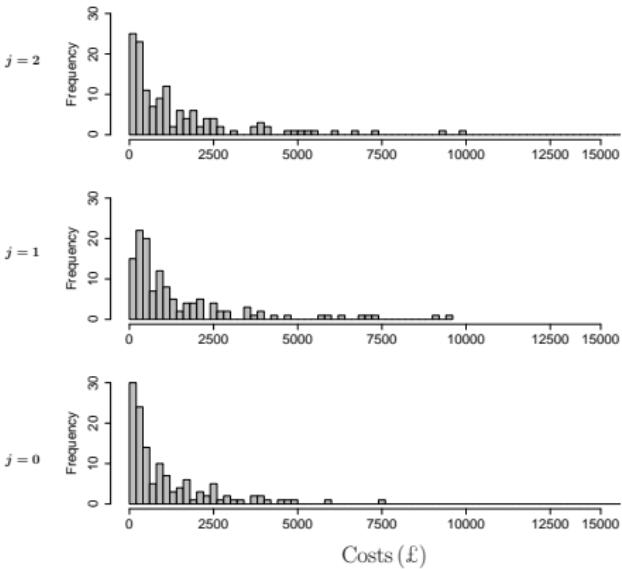
Intervention



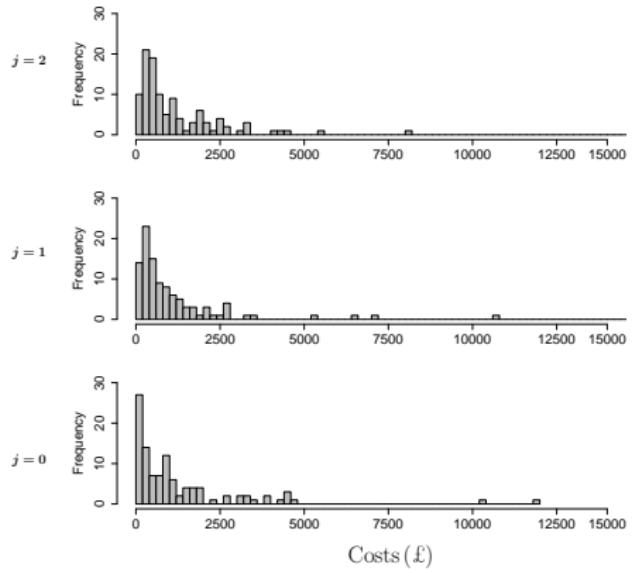
Second example: PBS trial

costs

Control



Intervention

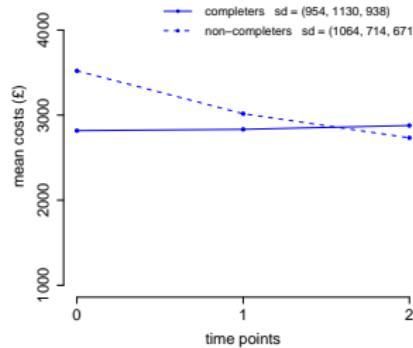
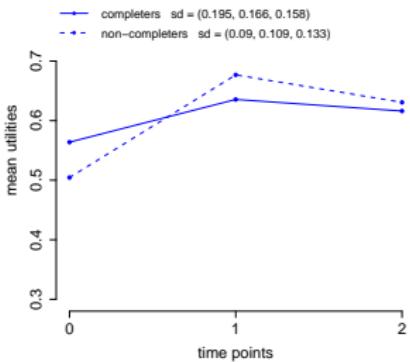
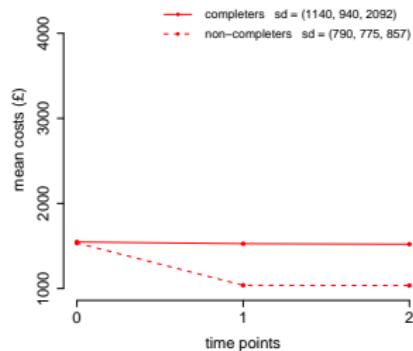
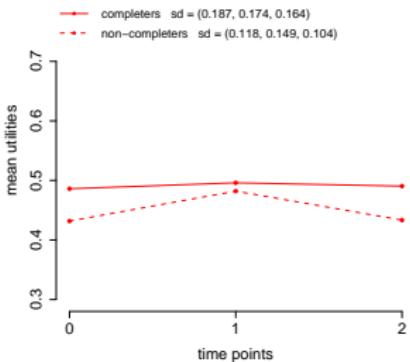


Control ($t = 1$)							n_1^r	Intervention ($t = 2$)							n_2^r
	u_0	c_0	u_1	c_1	u_2	c_2		u_0	c_0	u_1	c_1	u_2	c_2		
r	1	1	1	1	1	1	108	1	1	1	1	1	1	1	96
mean	0.678	1546	0.684	1527	0.680	1520		0.726	2818	0.771	2833	0.759	2878		
r	0	1	1	1	1	1	7	0	1	1	1	1	1	1	5
mean	—	1310	0.704	1440	0.644	1858		—	2573	0.780	2939	0.849	2113		
r	1	1	0	1	1	1	4	1	1	0	1	1	1	1	1
mean	0.709	1620	—	1087	0.737	851		0.467	9649	—	4828	0.259	4930		
r	1	1	1	1	0	1	2	1	1	1	1	0	1	1	1
mean	0.564	640	0.648	512	—	286		0.817	3788	0.884	0	—	0		
r	1	1	0	0	1	1	4	1	1	0	0	1	1	1	1
mean	0.716	2834	—	—	0.634	679		0.501	3608	—	—	0.872	4781		
r	1	1	0	0	0	0	4	1	1	0	0	0	0	0	4
mean	0.434	1528	—	—	—	—		0.760	3086	—	—	—	—		
r	0	1	0	1	1	1	2	0	1	0	1	1	1	1	0
mean	—	595	—	397	0.483	69		—	—	—	—	—	—		
r	1	1	1	1	0	0	2	1	1	1	1	0	0	0	0
mean	0.743	1434	0.705	1606	—	—		—	—	—	—	—	—		
r	1	1	0	1	0	1	3	1	1	0	1	0	1	1	0
mean	0.726	1510	—	432	—	976		—	—	—	—	—	—		

 $\rightarrow r = 1$ $r \neq 1$

Second example: PBS trial

mean profiles



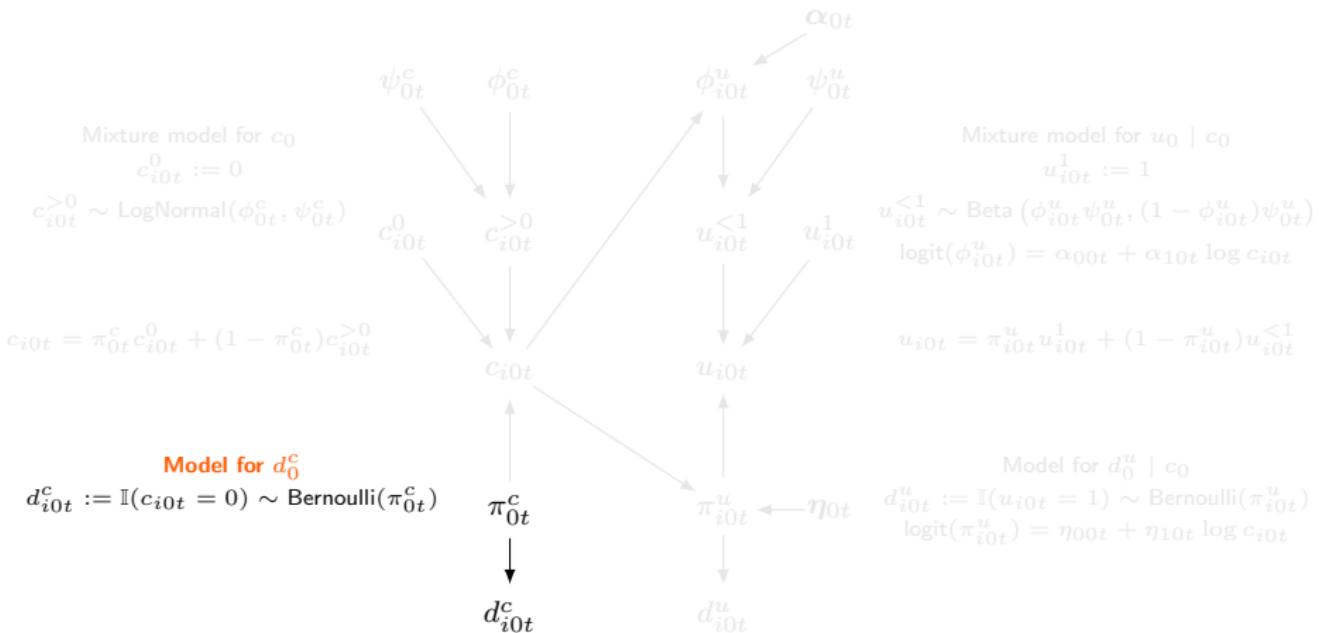
- Fit model to completers $r = 1$ and joint set of all other patterns $r \neq 1$ separately for $t = 1, 2$

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

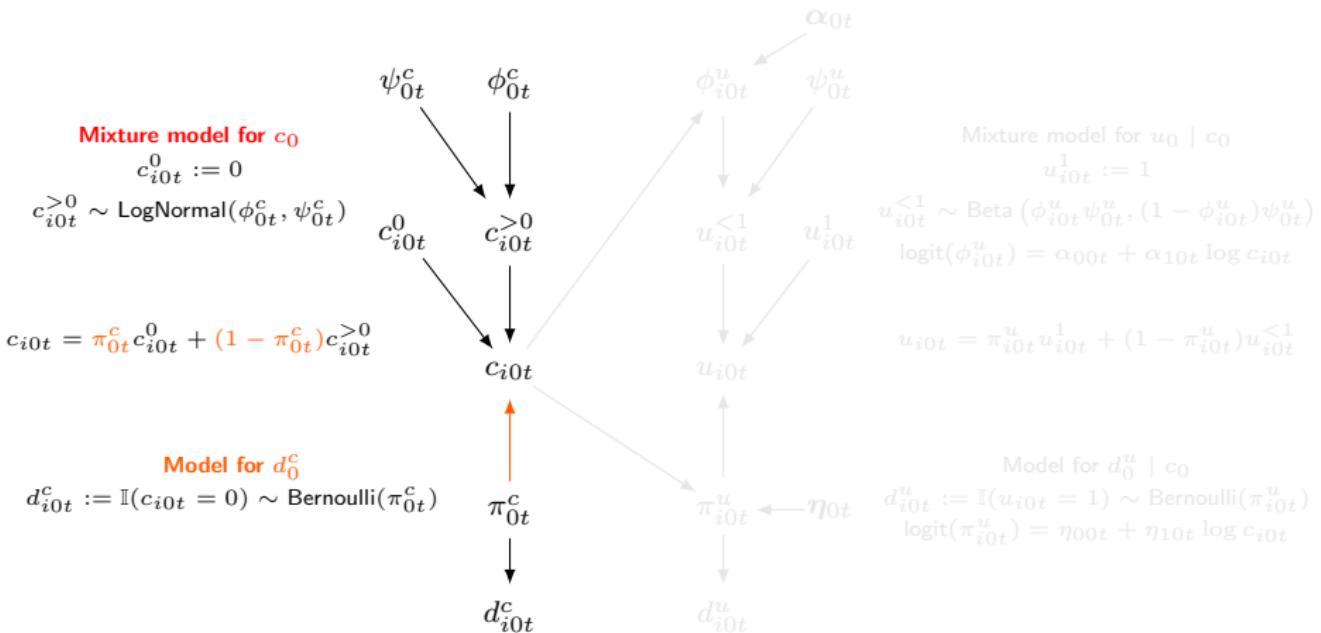
- Fit model to completers $r = 1$ and joint set of all other patterns $r \neq 1$ separately for $t = 1, 2$
- Capture outcome and time dependence through a series of conditional distributions $p(c_{ij} | c_{ij-1}, u_{ij-1})$ and $p(u_{ij} | c_{ij}, u_{ij-1})$
- Account for skewness using **Beta** distributions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

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

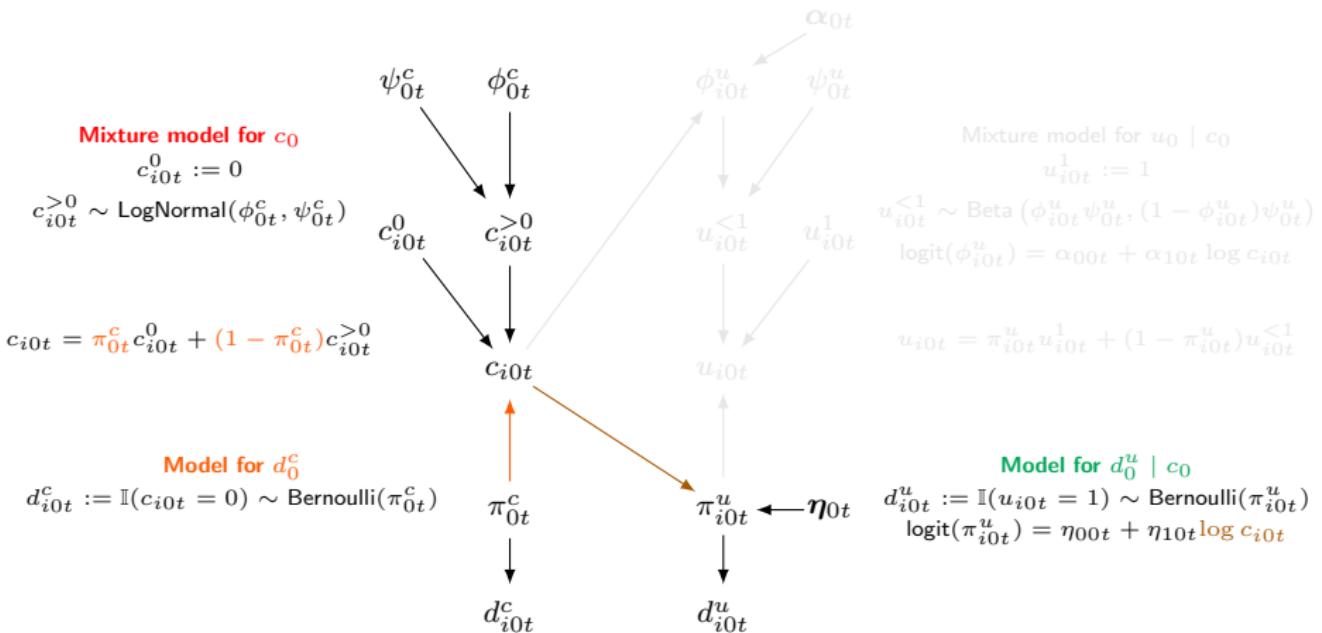
- At $j = 0$



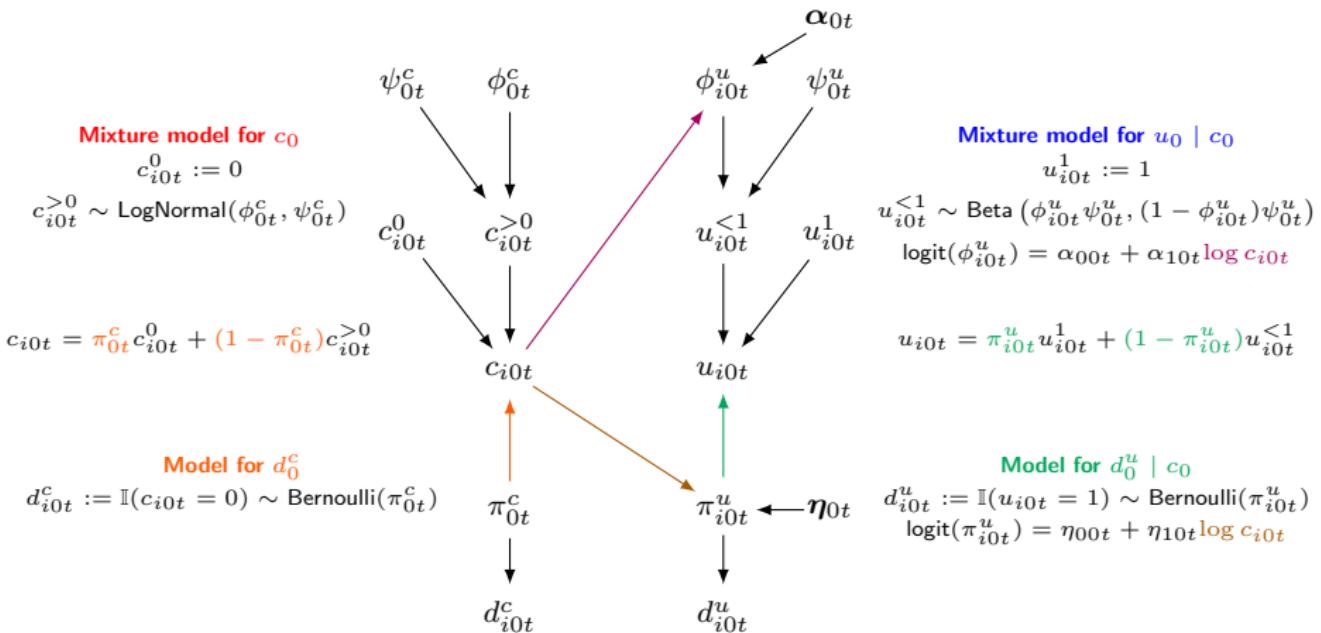
- At $j = 0$



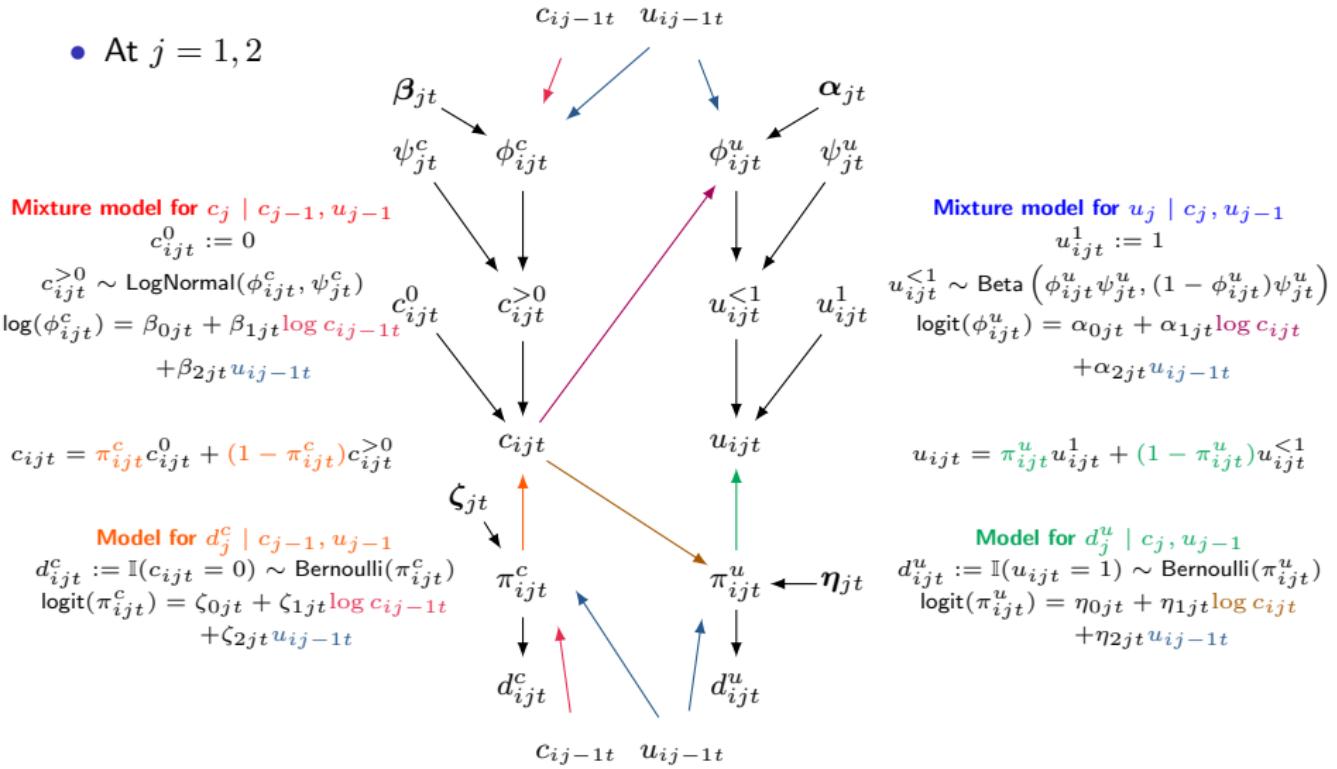
- At $j = 0$



- At $j = 0$



- At $j = 1, 2$



Identifying restrictions and sensitivity parameters

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

Identifying restrictions and sensitivity parameters

- Use Monte Carlo integration to derive the mean estimates $E[\mathbf{y}^r_{obs}]$
- Identify the mean estimates $E[\mathbf{y}^r_{mis}] = E[\mathbf{y}^r_{obs} + \Delta_j]$

Identifying restrictions and sensitivity parameters

- Use Monte Carlo integration to derive the mean estimates $E[\mathbf{y}^r_{obs}]$
- Identify the mean estimates $E[\mathbf{y}^r_{mis}] = E[\mathbf{y}^r_{obs} + \Delta_j]$
- Compute weighted average across r to derive $\mu_{jt} = (\mu_{jt}^u, \mu_{jt}^c)$

Identifying restrictions and sensitivity parameters

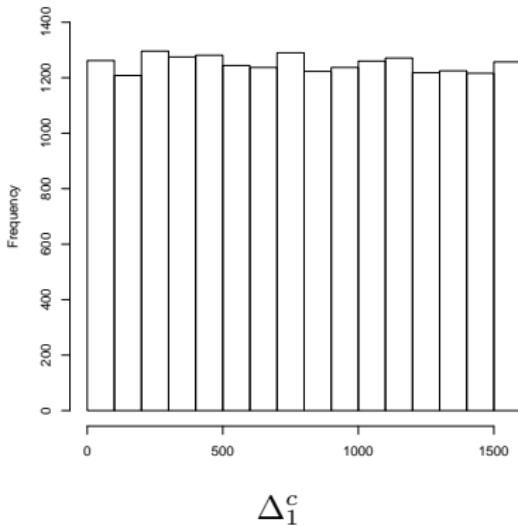
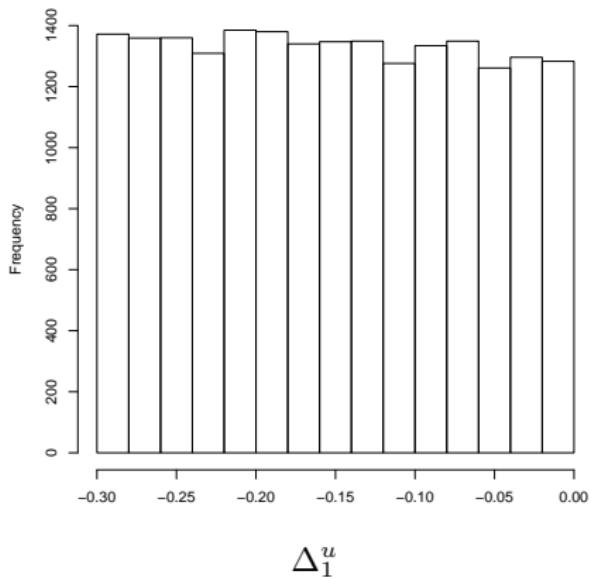
- Use Monte Carlo integration to derive the mean estimates $E[\mathbf{y}^r_{obs}]$
- Identify the mean estimates $E[\mathbf{y}^r_{mis}] = E[\mathbf{y}^r_{obs} + \Delta_j]$
- Compute weighted average across r to derive $\mu_{jt} = (\mu_{jt}^u, \mu_{jt}^c)$
- Set $\Delta_j = \mathbf{0}$ as benchmark assumption

Identifying restrictions and sensitivity parameters

- Use Monte Carlo integration to derive the mean estimates $E[\mathbf{y}^r_{obs}]$
- Identify the mean estimates $E[\mathbf{y}^r_{mis}] = E[\mathbf{y}^r_{obs} + \Delta_j]$
- Compute weighted average across r to derive $\mu_{jt} = (\mu_{jt}^u, \mu_{jt}^c)$
- Set $\Delta_j = \mathbf{0}$ as benchmark assumption
- Specify three alternative priors on $\Delta_j = (\Delta_j^u, \Delta_j^c)$, calibrated based on the variability in the observed data at each time j

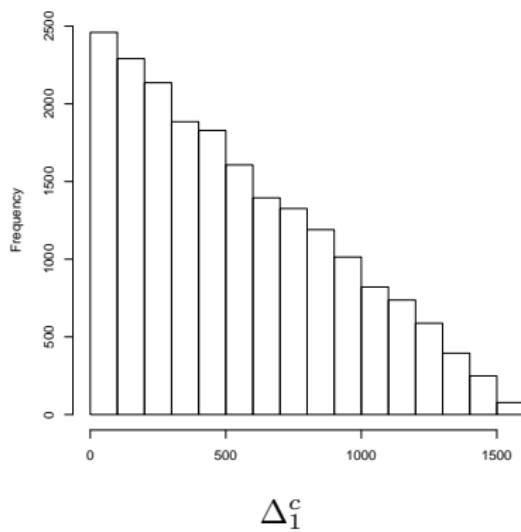
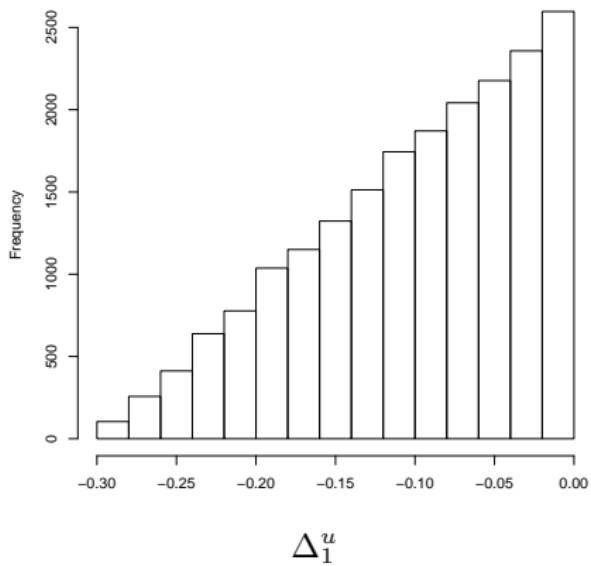
Priors on sensitivity parameters

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



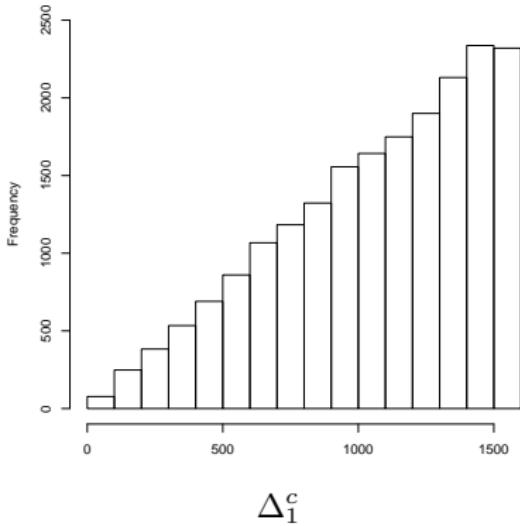
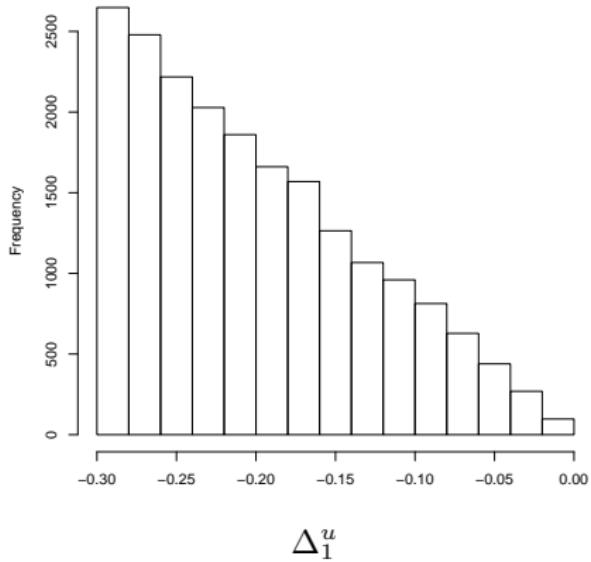
Priors on sensitivity parameters

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



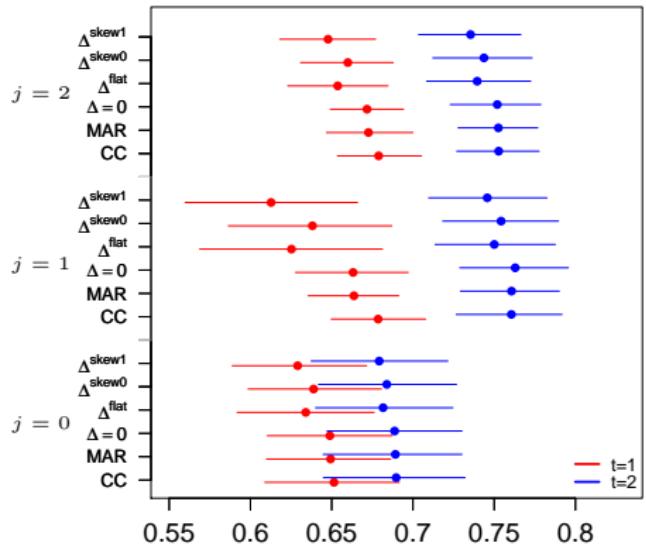
Priors on sensitivity parameters

- Assumption: $\mathbf{u}_{mis} < \mathbf{u}_{obs}$ and $\mathbf{c}_{mis} > \mathbf{c}_{obs}$
- Δ^{skew^1} : Skewed towards values far from 0 on the same range as Δ^{flat}

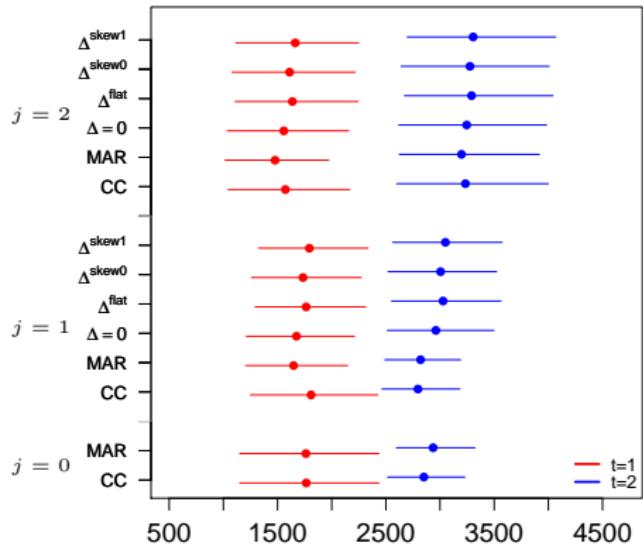


Results: means utilities and costs

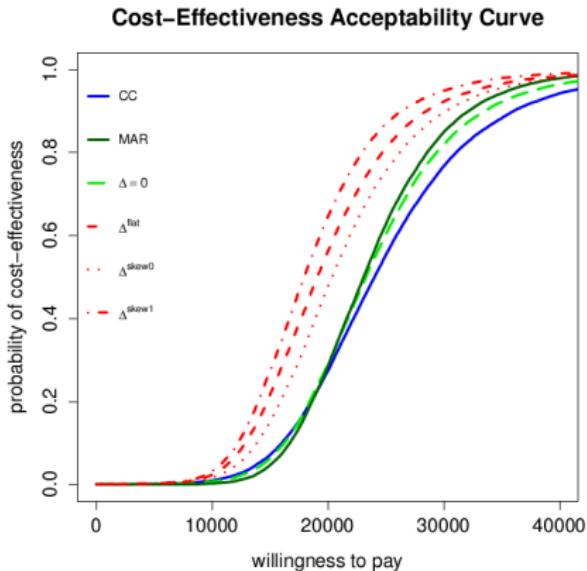
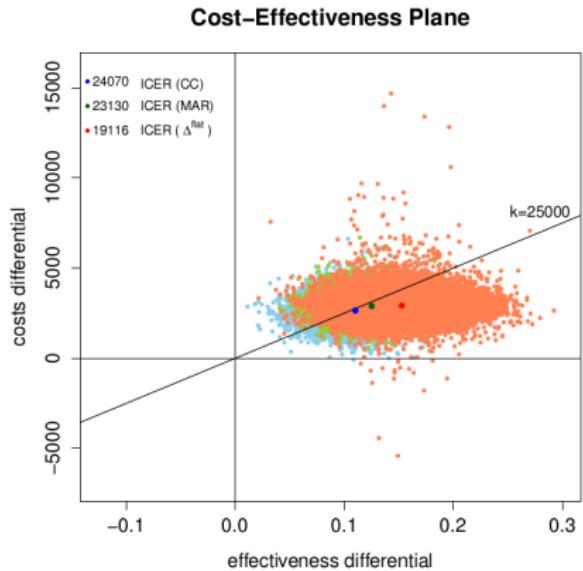
$$\mu_{jt}^u$$



$$\mu_{jt}^c$$



Cost-effectiveness analysis



① Flexibility of the modelling framework

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

① Flexibility of the modelling framework

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

② Extension of standard “imputation methods”

- Performs the estimation and imputation tasks simultaneously
- Fits joint models for missing data in a relatively easy way
- Uses probabilistic approaches that can be implemented in standard software (e.g. OpenBUGS or JAGS)

① Flexibility of the modelling framework

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

② Extension of standard “imputation methods”

- Performs the estimation and imputation tasks simultaneously
- Fits joint models for missing data in a relatively easy way
- Uses probabilistic approaches that can be implemented in standard software (e.g. OpenBUGS or JAGS)

③ Principled incorporation of external evidence through priors

- Crucial for conducting sensitivity analysis to MNAR
- Useful in small/pilot trials where there is limited evidence

Part 5

Practicals

[Back to Table of content](#)

- The focus of these practicals is provide a brief introduction and overview about how to fit Bayesian models using JAGS via R.
- R is a free, open-source software and programming language, which has become one of the dominant software environments for data analysis.
- The latest version of R for Windows, Mac or Linux OS can be downloaded from the CRAN website: <https://www.r-project.org/>
- We also recommend to download and install Rstudio (<https://www.rstudio.com/>), an integrated development environment which provides an "user-friendly" interaction with R (e.g. many drop-down menus, tabs, customisation options)
- There are different R packages which function as frontends for JAGS. These packages make it easy to process the output of Bayesian models and present it in publication-ready form.

- JAGS or **Just Another Gibbs Sampler** is a program for analysis of Bayesian models using Markov Chain Monte Carlo (MCMC) methods
- JAGS is a free software based on the **Bayesian inference Using Gibbs Sampling** (informally BUGS) language at the base of WinBUGS/OpenBUGS but, unlike these programs, it is written in C++ and is platform independent.
- The latest version of JAGS can be dowloaded at
<http://mcmc-jags.sourceforge.net/> and can be installed on multiple OS.
- **R2jags** is an R package that allows to fit JAGS models from within R. In this brief introduction, we will show how to fit JAGS models using the **R2jags** package.

- ① Install the most recent R and Rstudio versions
- ② Install JAGS version 4.2.0 from Martyn Plummer's repository:
<https://sourceforge.net/projects/mcmc-jags/files/JAGS/>
- ③ Install the package **R2jags** from within R or Rstudio, via the package installer or by typing in the command line

```
> install.packages("R2jags", dependencies = TRUE)
```
- ④ The `dependencies = TRUE` option will automatically install all the packages on which the functions in the **R2jags** package rely.

- For an example dataset, we simulate our own data in R. We create a continuous outcome variable y as a function of one predictor x and a disturbance term ϵ . We simulate a dataset with 100 observations

- For an example dataset, we simulate our own data in R. We create a continuous outcome variable y as a function of one predictor x and a disturbance term ϵ . We simulate a dataset with 100 observations
- Create the error term, the predictor and the outcome using a linear form with an intercept β_0 and slope β_1 coefficients

```
> n.sim=100; set.seed(123)
> x=rnorm(n.sim, mean = 5, sd = 2)
> epsilon=rnorm(n.sim, mean = 0, sd = 1)
> beta0=1.5
> beta1=1.2
> y=beta0 + beta1 * x + epsilon
```

- For an example dataset, we simulate our own data in R. We create a continuous outcome variable y as a function of one predictor x and a disturbance term ϵ . We simulate a dataset with 100 observations
- Create the error term, the predictor and the outcome using a linear form with an intercept β_0 and slope β_1 coefficients

```
> n.sim=100; set.seed(123)
> x=rnorm(n.sim, mean = 5, sd = 2)
> epsilon=rnorm(n.sim, mean = 0, sd = 1)
> beta0=1.5
> beta1=1.2
> y=beta0 + beta1 * x + epsilon
```

- Then, we define all the data for JAGS in a list object

```
> datalist=list("y","x","n.sim")
```

- Now, we write the model for JAGS and save it as the text file "basic.mod.txt" in the current working directory

```
> basic.mod= "
+ model {
+ #model
+ for(i in 1:n.sim){
+   y[i] ~ dnorm(mu[i], tau)
+   mu[i] = beta0 + beta1 * x[i]
+ }
+ #priors
+ beta0 ~ dnorm(0, 0.01)
+ beta1 ~ dnorm(0, 0.01)
+ tau ~ dgamma(0.01,0.01)
+ }
```

- Now, we write the model for JAGS and save it as the text file "basic.mod.txt" in the current working directory

```
> basic.mod= "
+ model {
+ #model
+ for(i in 1:n.sim){
+   y[i] ~ dnorm(mu[i], tau)
+   mu[i] = beta0 + beta1 * x[i]
+ }
+ #priors
+ beta0 ~ dnorm(0, 0.01)
+ beta1 ~ dnorm(0, 0.01)
+ tau ~ dgamma(0.01,0.01)
+ }
```

- To write the model as the text file "basic.mod.txt", we type

```
> writeLines(basic.mod, "basic.mod.txt")
```

- Define the parameters whose posterior distributions we are interested in summarising later and set up the initial values for JAGS

```
> params=c("beta0", "beta1")
> inits=function(){list("beta0"=rnorm(1), "beta1"=rnorm(1))}
```

The function creates a list that contains one element for each parameter, which gets assigned a random draw from a normal distribution as a starting value for each chain in the model.

- For simple models like this, it is generally easy to define the initial values for all parameters. However, for more complex models, this may not be immediate and a lot of trial and error may be required.
- JAGS can automatically select the initial values for all parameters in an efficient way even for relatively complex models. This can be achieved by setting `inits=NULL`, which is then passed to the `jags` function in **R2jags**.

- Before using **R2jags** for the first time, you need to load the package, and you may want to set a random seed number for making your estimates replicable

```
> library(R2jags)  
> set.seed(123)
```

- Before using **R2jags** for the first time, you need to load the package, and you may want to set a random seed number for making your estimates replicable

```
> library(R2jags)
> set.seed(123)
```

- Now, we can fit the model in JAGS using the `jags` function in the **R2jags** package and save it in the object `basic.mod`

```
> basic.mod=jags(data = datalist, inits = inits,
+   parameters.to.save = params, n.chains = 2, n.iter = 9000,
+   n.burnin = 1000, model.file = "basic.mod.txt")
## Compiling model graph
## Resolving undeclared variables
## Allocating nodes
## Graph information:
##   Observed stochastic nodes: 100
##   Unobserved stochastic nodes: 3
##   Total graph size: 406
##
## Initializing model
```

- A summary of the posterior estimates and convergence diagnostics for all parameters specified can be seen by typing `print(basic.mod)` or, alternatively,

```
> print(basic.mod$BUGSoutput$summary)
```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
beta0	1.5	0.293	0.95	1.3	1.5	1.7	2.1	1	1600
beta1	1.2	0.053	1.07	1.1	1.2	1.2	1.3	1	920
deviance	278.9	2.508	276.04	277.1	278.2	279.9	285.5	1	2000

- The posterior distribution of each parameter is summarised in terms of
 - The mean, sd and some percentiles
 - Potential scale reduction factor (Rhat) and effective sample size (n.eff)
- The deviance is a goodness of fit statistic and is used in the construction of the "Deviance Information Criterion" (DIC), which is a **relative** measure of model comparison. The DIC of the model can be accessed by typing

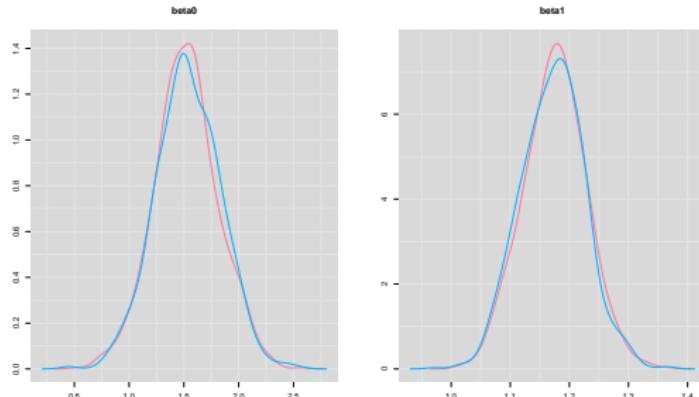
```
> basic.mod$BUGSoutput$DIC  
[1] 282
```

- More diagnostics are available when we convert the model output into an MCMC object using the command

```
> basic.mod.mcmc=as.mcmc(basic.mod)
```

- We can now install and load the **mcmcplots** package to obtain graphical diagnostics and results, e.g. density and trace plots for each parameter.

```
> install.packages("mcmcplots")
> library(mcmcplots)
> denplot(basic.mod.mcmc, parms = c("beta0", "beta1"))
```

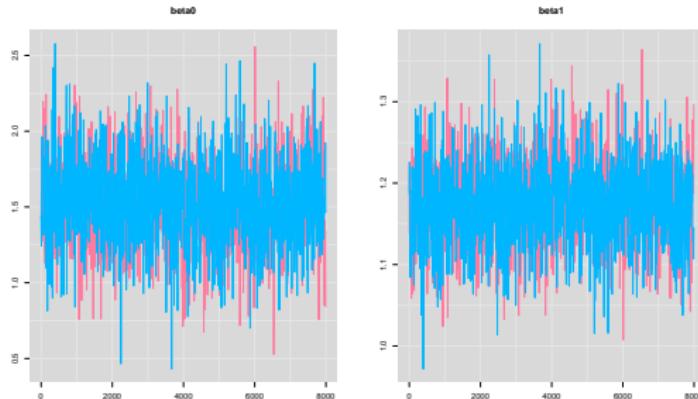


- More diagnostics are available when we convert the model output into an MCMC object using the command

```
> basic.mod.mcmc=as.mcmc(basic.mod)
```

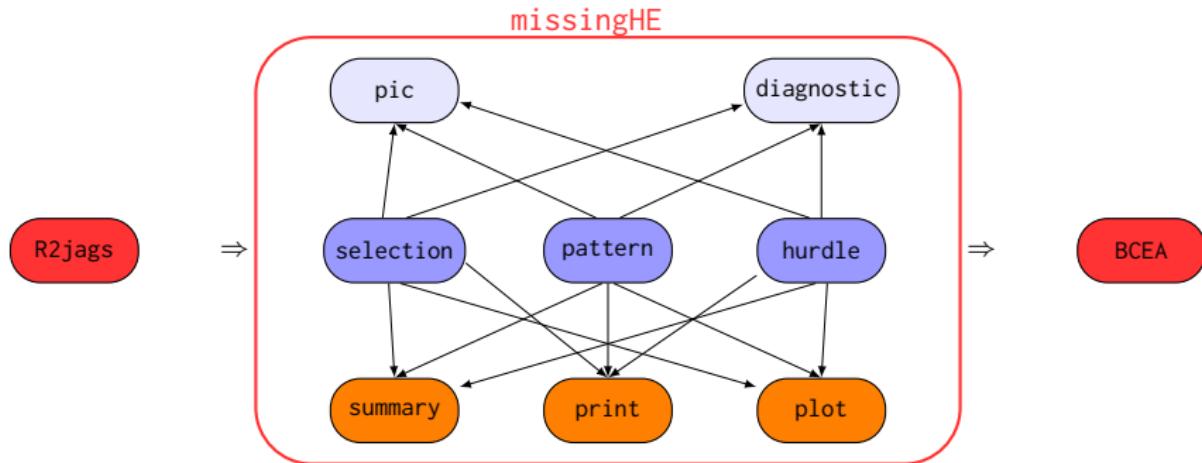
- We can now install and load the **mcmcplots** package to obtain graphical diagnostics and results, e.g. density and trace plots for each parameter.

```
> install.packages("mcmcplots")
> library(mcmcplots)
> traplot(basic.mod.mcmc, parms = c("beta0", "beta1"))
```



missingHE: an R package to deal with missing data in HTA

- Now that you are more familiar with the output of JAGS, we will explore different modelling options for handling missing data in HTA using the **missingHE** package



GitHub repository: <https://github.com/AnGabrio/missingHE>

CRAN repository: <https://cran.r-project.org/web/packages/missingHE>

- Specifies a set of pre-defined JAGS models using the **R2jags** package
- Is linked to the **BCEA** package, which provides summary HTA results
- **missingHE**:
 - Uses the functions hurdle, selection and pattern to implement alternative models under different missingness assumptions
 - Assesses model fit and convergence using the functions pic 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, as well as to summarise the economic results are provided in the handouts
- We also provide HTA data from a pilot RCT, which you can freely explore and analyse using **missingHE**

Muchas gracias!