

10. Missing data in cost-effectiveness modelling

Gianluca Baio

Department of Statistical Science | University College London

- ✉ g.baio@ucl.ac.uk
- 🌐 <https://gianluca.statistica.it/>
- 🌐 <https://egon.stats.ucl.ac.uk/research/statistics-health-economics/>
- 🌐 <https://github.com/giabaio>
- 🌐 <https://github.com/StatisticsHealthEconomics>
- 🐦 [@gianlubaio](https://twitter.com/gianlubaio)

STAT0019 - Bayesian Methods in Health Economics, UCL

- Background – what's difficult about missing data?
- Missing data mechanisms
- Missing data in health economic evaluation
- Example – the MenSS Trial

References

- *The BUGS Book*, chapter 9.1  [Library](#)  [Book website](#)
- *Missing data in longitudinal studies*  [Library](#)  [Book website](#)
- Gabrio et al (2018)
- Gabrio et al (2020)

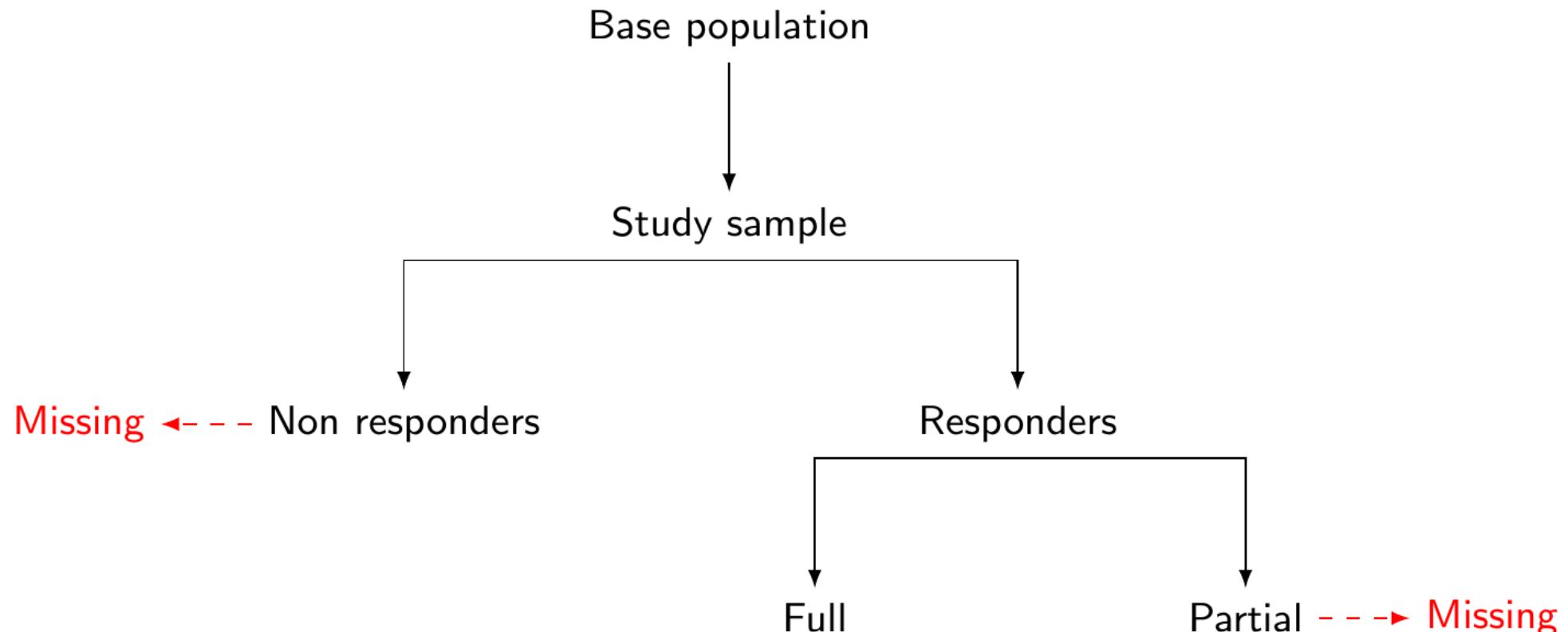
The problems with missing data...

- 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...

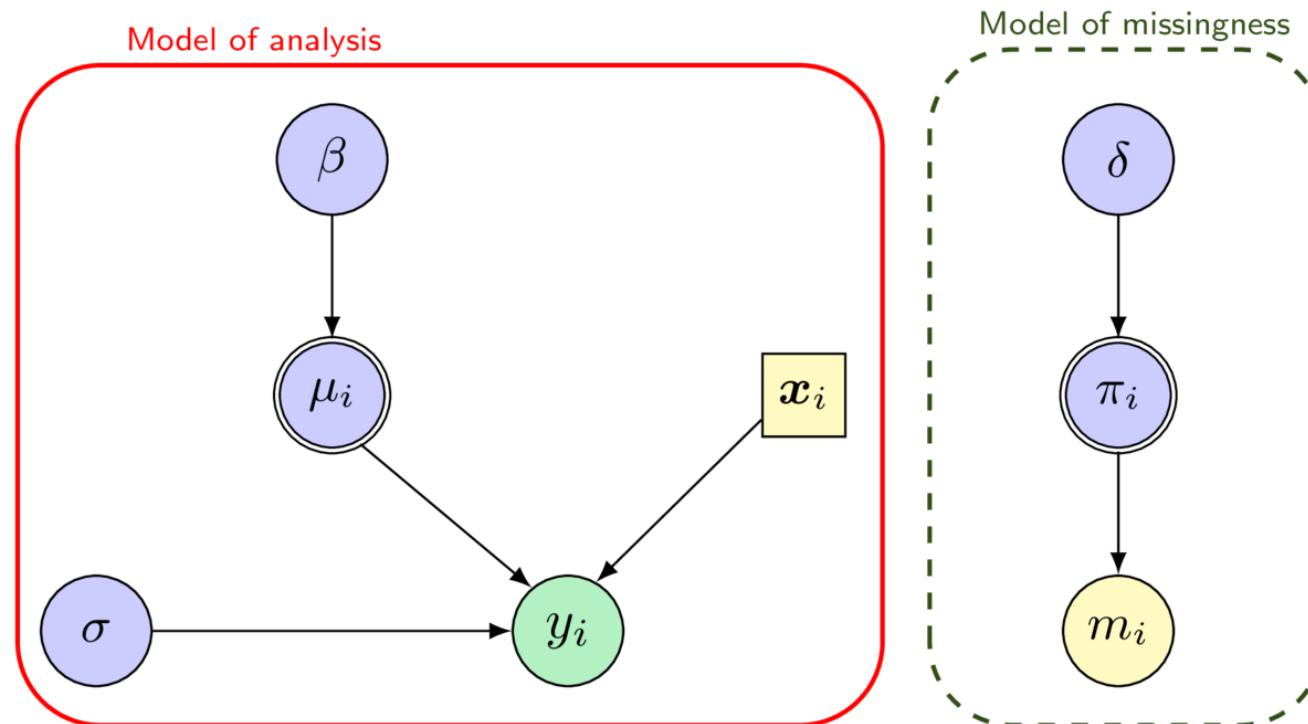
- We plan to observe n_{planned} data points, but end up with a (much) lower number of observations n_{observed}
 - What is the proportion of missing data?
 - Does it matter?...
- We typically don't know **why** the unobserved points are missing and **what** their value might have been
 - Missingness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



Missing data mechanism

Missing Completely at Random (MCAR)



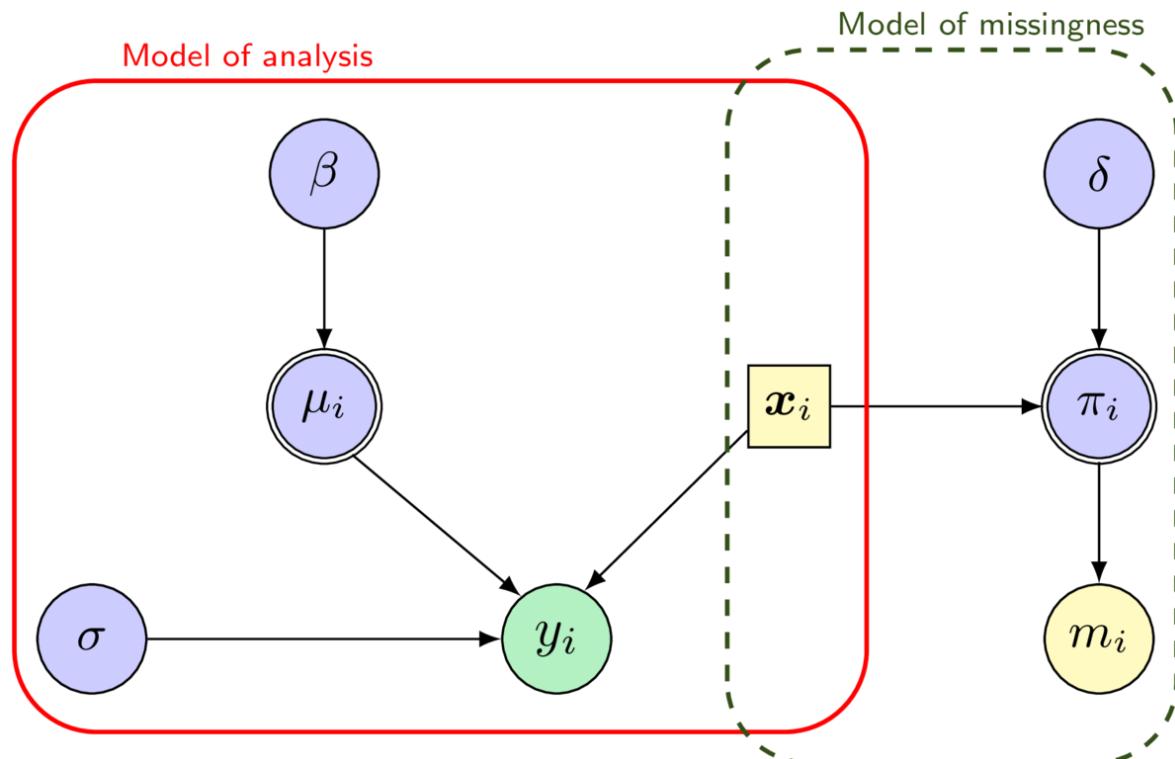
- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

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

NB: things are even more complicated when missing is (also) for the predictors/covariates!

Missing data mechanism

Missing at Random (MAR)



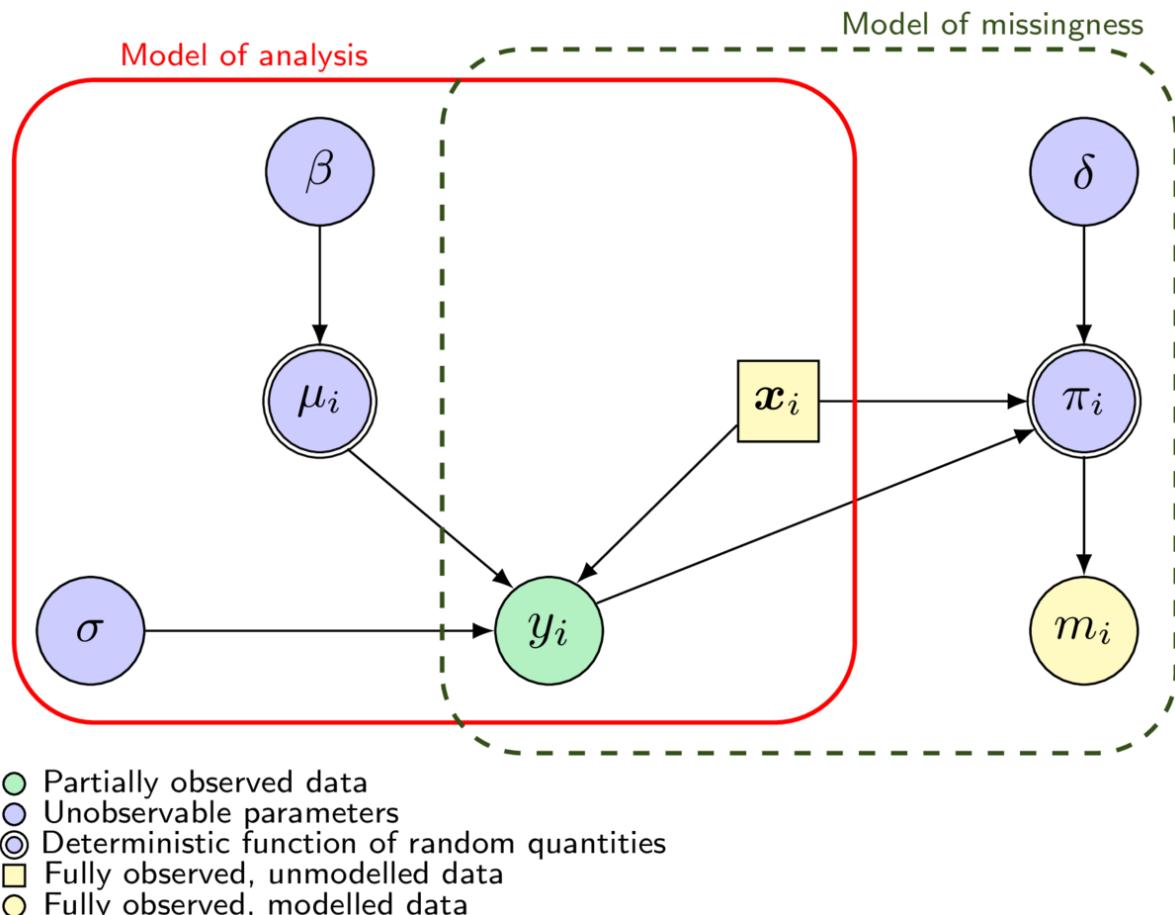
- Partially observed data
- Unobservable parameters
- Deterministic function of random quantities
- Fully observed, unmodelled data
- Fully observed, modelled data

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

NB: things are even more complicated when missing is (also) for the predictors/covariates!

Missing data mechanism

Missing Not at Random (MNAR)



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

NB: things are even more complicated when missing is (also) for the predictors/covariates!

Missing data analysis methods

- Complete Case Analysis
 - Elimination of partially observed cases
 - Simple but reduce efficiency and possibly give 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 low(high) 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 K times to obtain K different imputed datasets
 - Each dataset is analysed and K 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

Code Output

```
> # A very simple simulation study...
>
> # 1. Generates 100 data points for a covariate 'x', assuming x~Normal(mean=2, sd=5)
> x=rnorm(100,2,5)
>
> # 2. Generates 100 data points for the outcome 'y', assuming a linear regression with
> # a. intercept (alpha) = 0.5
> # b. "effect" of x (beta) = 1.5
> # c. error variance = 1 (ie add 'rnorm(100)', which simulates 100 values from Normal(0,1))
> y=.5+1.5*x+rnorm(100)
>
> # 3. Removes 20 values at random and replaces them with 'NA' (missing)
> y[sample(1:100,20)]=NA
>
> # 4. Runs a linear regression model (y on x) and print results (see 'Output' panel...)
> summary(lm(y~x))
```

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

Code Output

Call:

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

Residuals:

Min	1Q	Median	3Q	Max
-2.62167	-0.75375	-0.08869	0.70455	2.23963

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.43660	0.11402	3.829	0.000258 ***
x	1.47617	0.02217	66.579	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.007 on 78 degrees of freedom

(20 observations deleted due to missingness)

Multiple R-squared: 0.9827, Adjusted R-squared: 0.9825

F-statistic: 4433 on 1 and 78 DF, p-value: < 2.2e-16

Advantages

- Clearly 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 give 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 low(high) 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 K times to obtain K different imputed datasets
 - Each dataset is analysed and K 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 a kind of **propensity score** – and this procedure is similar to increasingly popular methods in health economics (e.g. **Multiple Adjusted Indirect Comparisons**)

(Ridiculously simplified!) Example

Suppose we *know* the data to be

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

- The "true" estimation given the complete data set is

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

(Ridiculously simplified!) Example

Suppose we *know* the data to be

Group	Group A	Group B	Group 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 missing and thus the *actual* data are

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

- The CCA of the observed data would give an estimate of

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

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 $w_i = \hat{p}^{-1}$

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 $w_i = \hat{p}^{-1}$
- In this case, **assuming the missingness mechanism is MCAR**
 - $\hat{p}_i^A = 1/3 \Rightarrow w_i^A = 1/\hat{p}_i^A = 3$
 - $\hat{p}_i^B = 3/3 \Rightarrow w_i^B = 1/\hat{p}_i^B = 1$
 - $\hat{p}_i^C = 2/3 \Rightarrow w_i^C = 1/\hat{p}_i^C = 3/2$

- 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 missingness mechanism is MCAR**
 - $\hat{p}_i^A = 1/3 \Rightarrow w_i^A = 1/\hat{p}_i^A = 3$
 - $\hat{p}_i^B = 3/3 \Rightarrow w_i^B = 1/\hat{p}_i^B = 1$
 - $\hat{p}_i^C = 2/3 \Rightarrow w_i^C = 1/\hat{p}_i^C = 3/2$
- 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 give 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 low(high) 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 K times to obtain K different imputed datasets
 - Each dataset is analysed and K 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

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

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 give 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 low(high) 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 K times to obtain K different imputed datasets
 - Each dataset is analysed and K 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

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

Example

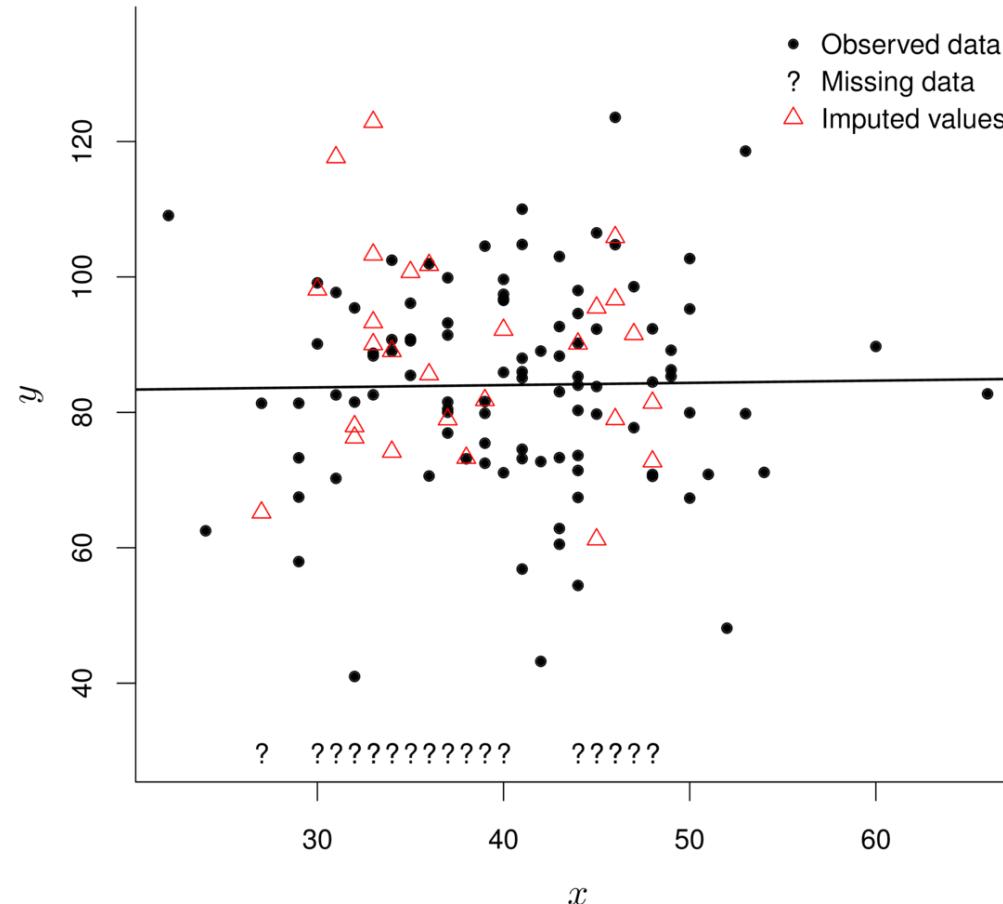
- Suppose we partially observe a continuous variable y on $n_{\text{obs}} < n$ individuals
- Also, we completely observe an additional variable x on the n individuals in the sample
- The simplest way in which we can describe the relationship between y and x is through linear regression

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

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

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

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

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 ("Rubin's rule")

- 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 give 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 low(high) 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 K times to obtain K different imputed datasets
 - Each dataset is analysed and K 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 (see [Lecture 5](#))

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

- Common assumption (explicit in a Bayesian setup): 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 (see [Lecture 5](#))

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

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

- Pattern mixture models**

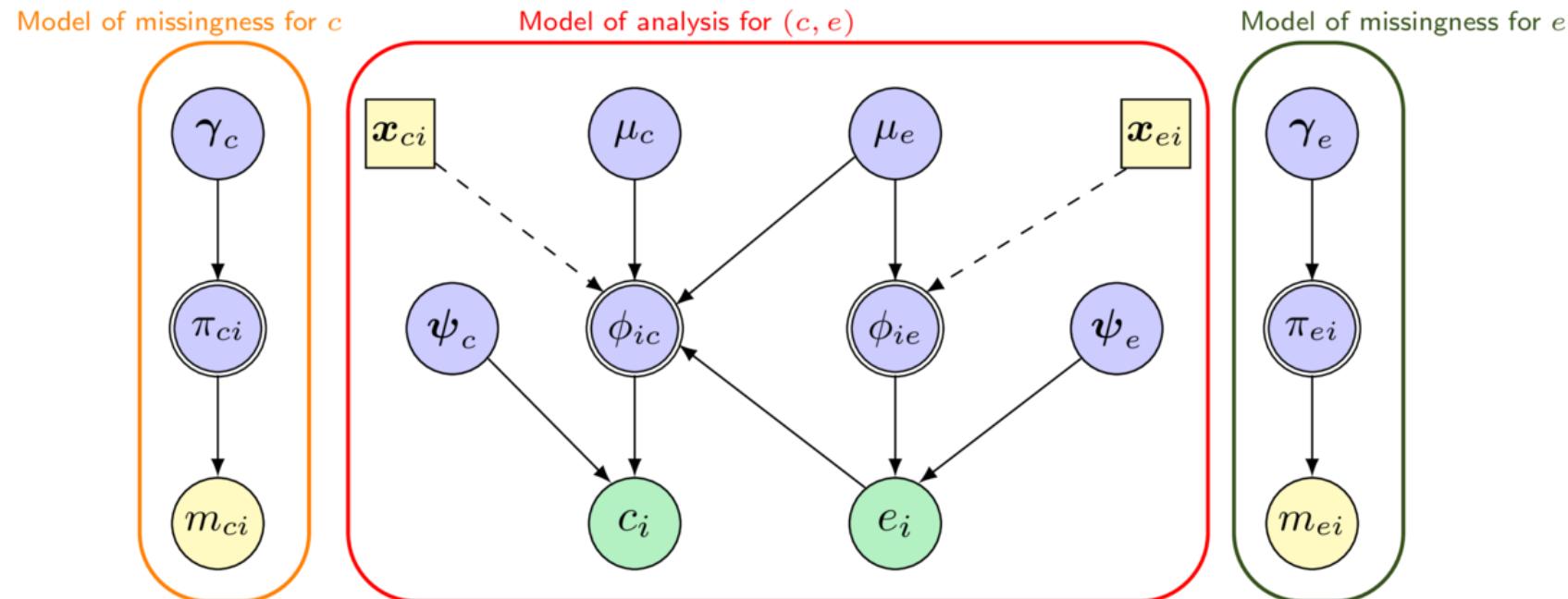
- Needs to model the full possible missingness "patterns" m using a marginal distribution
- Models for data more natural

- Selection models**

- Models directly the marginal distribution of the observable data
- Needs to figure out how the missingness model may be affected by it

- Missing data are complicated in any context
 - But are fairly established in medical/bio-statistical research
- In HTA it's even more complicated...
 - Bivariate outcome, usually correlated
 - Normality not reasonable (skewness)
 - Other features of the data ("spikes")
 - Main objective: decision-making, not inference! (See [Lecture 4](#))

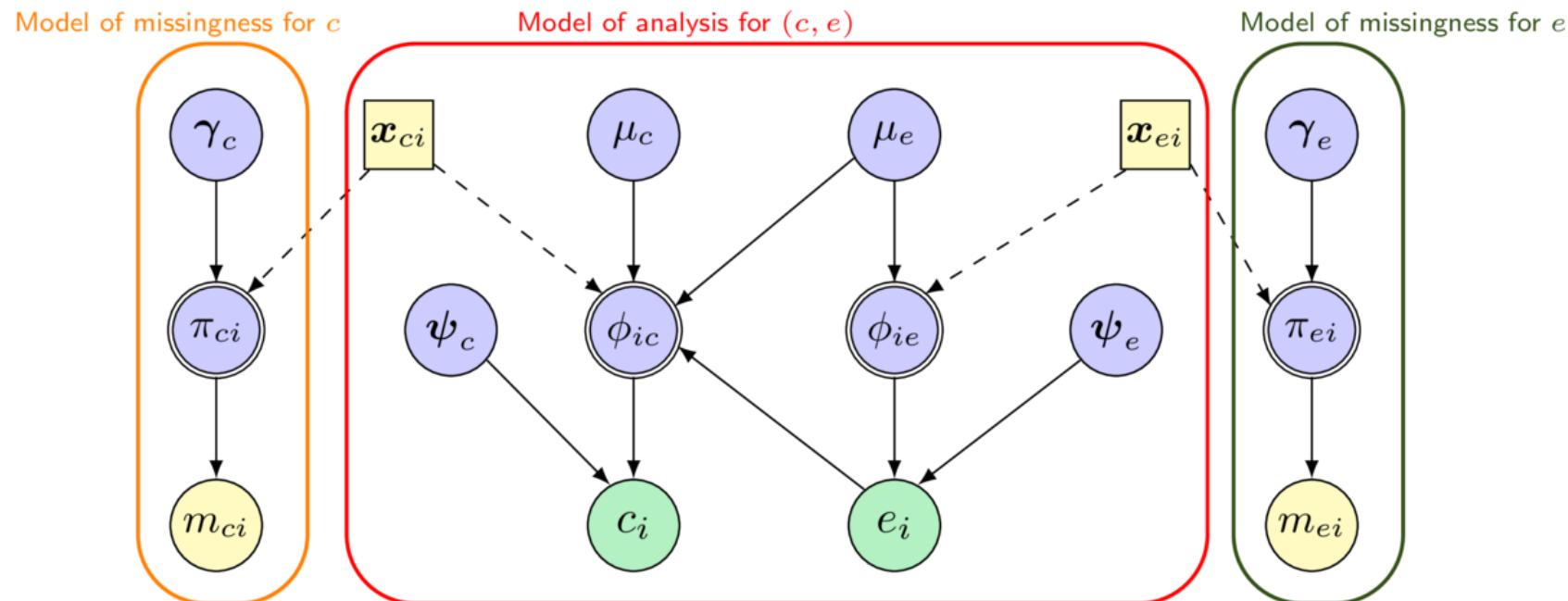
Selection models: MCAR (e, c)



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

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

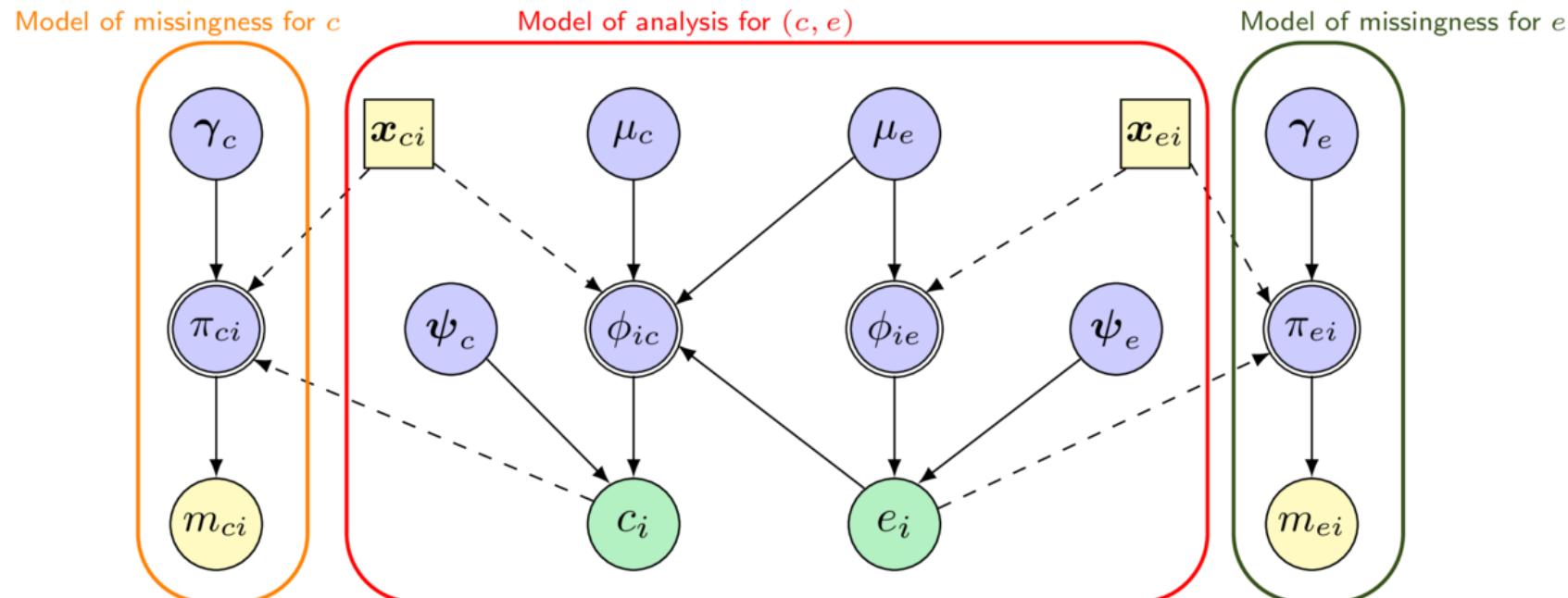
Selection models: MAR (e, c)



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

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

Selection models: MNAR (e, c)



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

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

Motivating example: MenSS trial

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

Motivating example: MenSS trial

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

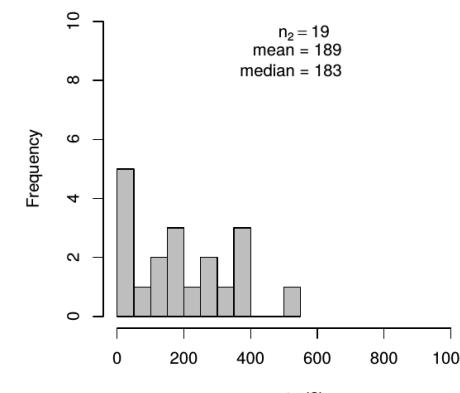
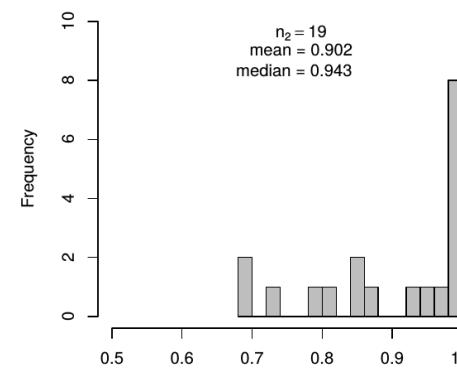
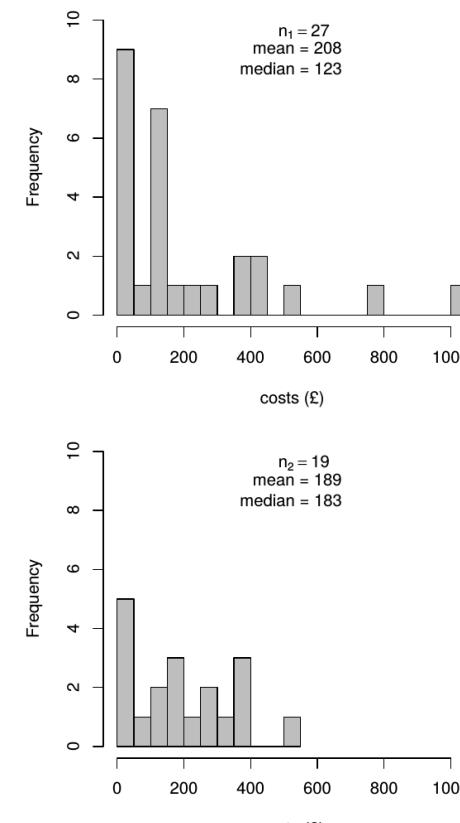
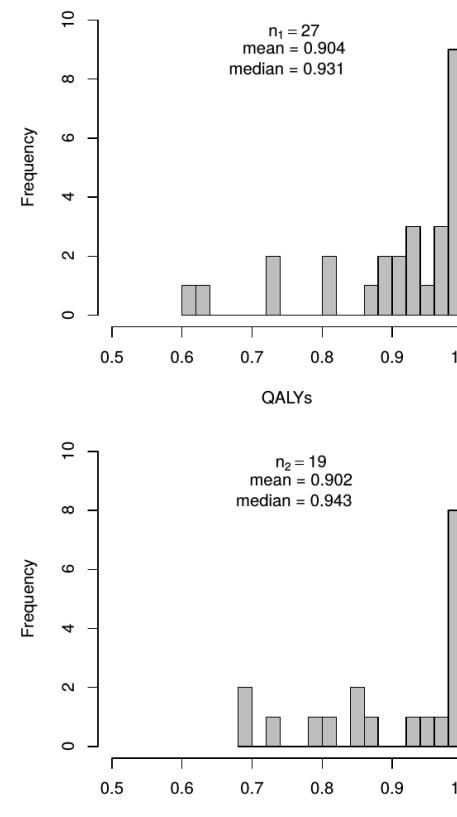
Partially observed data

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%)

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)

Skewness and "structural values"



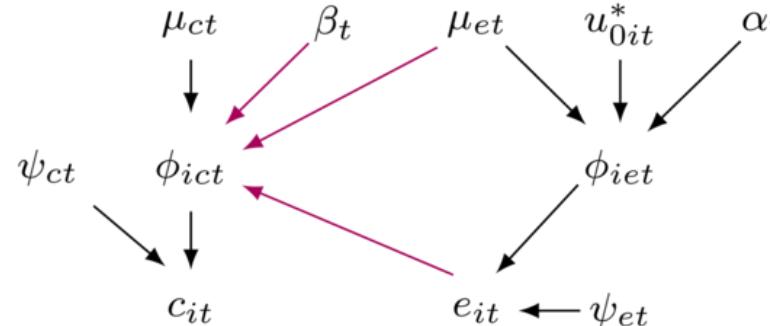
1

Bivariate Normal

- Simpler and closer to "standard" frequentist models
- Accounts for correlation between QALYs and costs

Conditional model for $c | e$

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

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


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

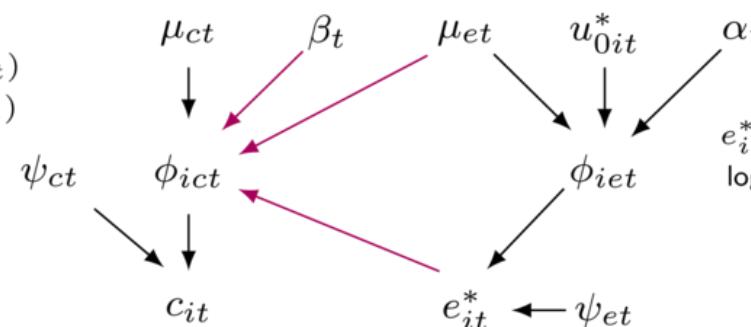
- Simpler and closer to "standard" frequentist models
- Accounts for correlation between QALYs and costs

2

Beta-Gamma

- Account for correlation between outcomes **and** model the relevant ranges: QALYs $\in (0, 1)$ and costs $\in (0, \infty)$
- **But:** needs to rescale the 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

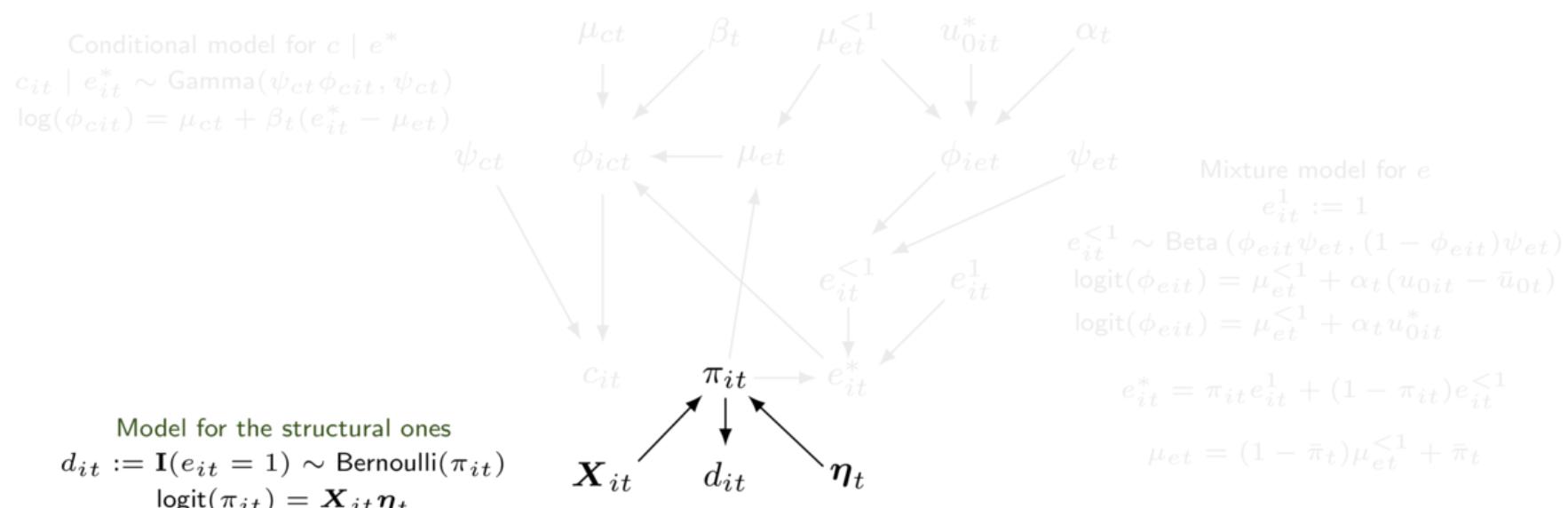
- Simpler and closer to "standard" frequentist models
- Accounts for correlation between QALYs and costs

2 Beta-Gamma

- Account for correlation between outcomes **and** model the relevant ranges: QALYs $\in (0, 1)$ and costs $\in (0, \infty)$
- **But:** needs to rescale the 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 + relevant ranges + structural values
- May expand further to account for partially observed baseline utilities u_{0it} (needs untestable assumptions!)



1 Bivariate Normal

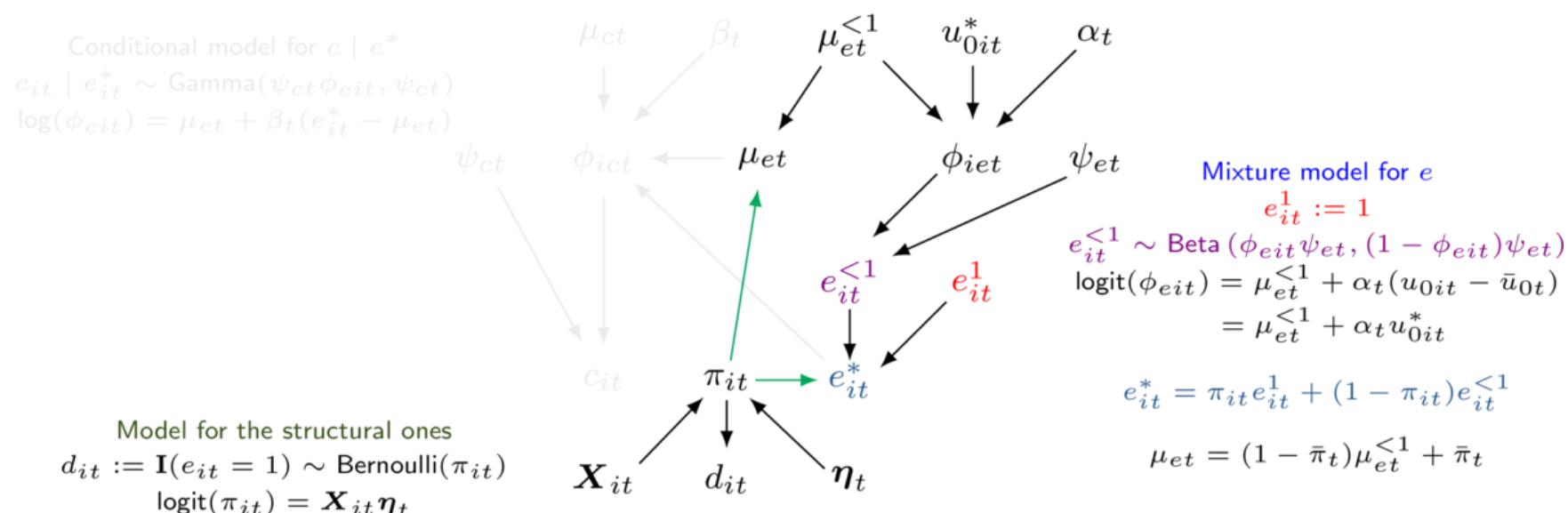
- Simpler and closer to "standard" frequentist models
- Accounts for correlation between QALYs and costs

2 Beta-Gamma

- Account for correlation between outcomes **and** model the relevant ranges: QALYs $\in (0, 1)$ and costs $\in (0, \infty)$
- **But:** needs to rescale the 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 + relevant ranges + structural values
- May expand further to account for partially observed baseline utilities u_{0it} (needs untestable assumptions!)



1 Bivariate Normal

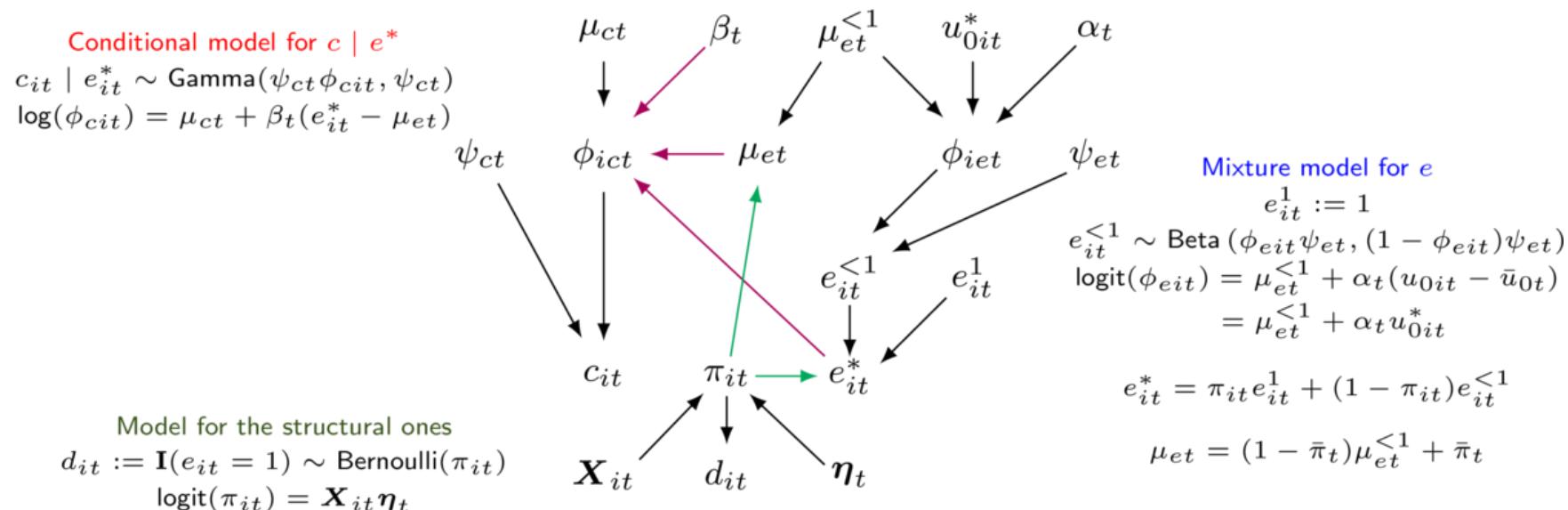
- Simpler and closer to "standard" frequentist models
- Accounts for correlation between QALYs and costs

2 Beta-Gamma

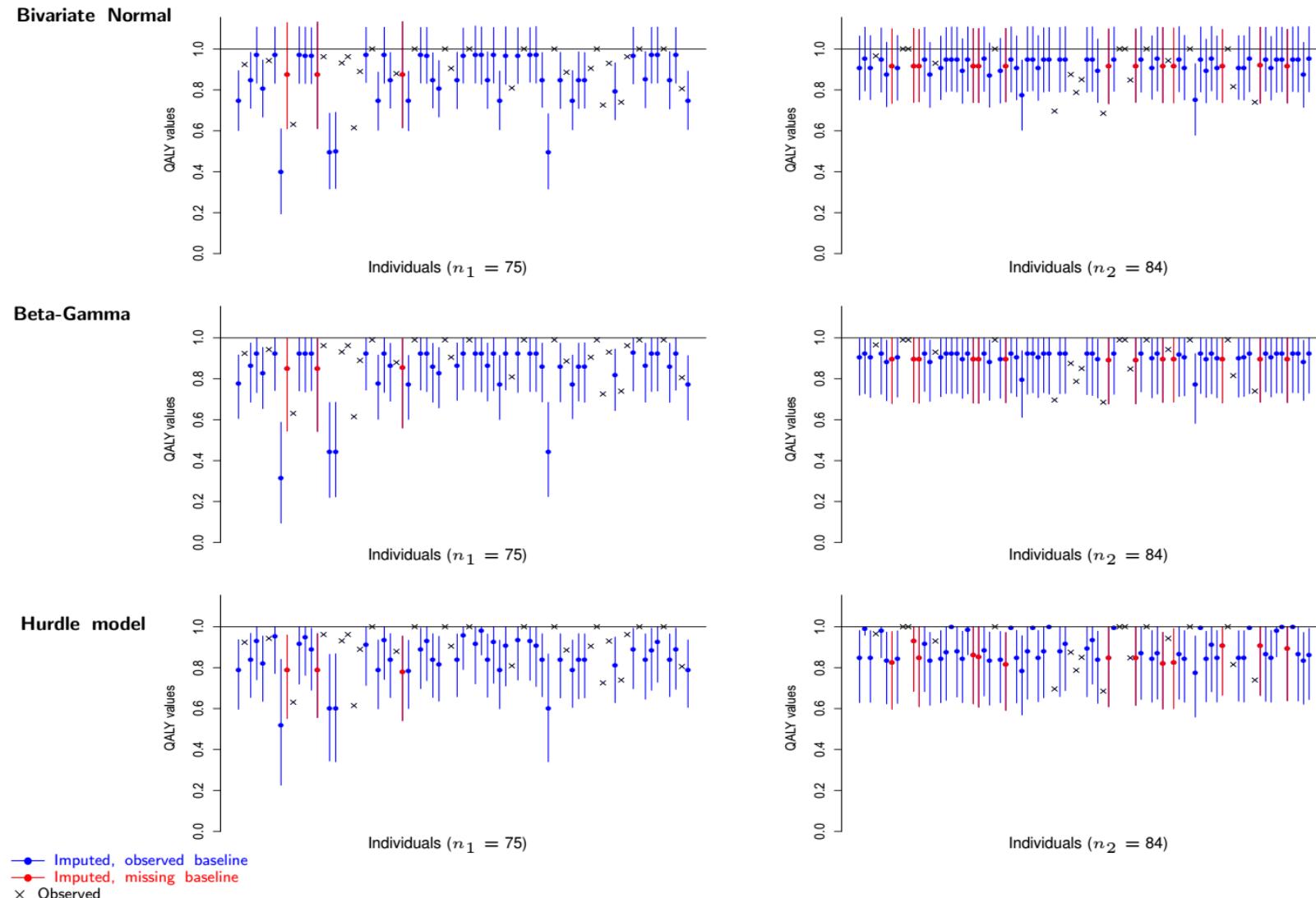
- Account for correlation between outcomes **and** model the relevant ranges: QALYs $\in (0, 1)$ and costs $\in (0, \infty)$
- **But:** needs to rescale the observed data $e_{it}^* = (e_{it} - \epsilon)$ to avoid spikes at 1

3 Hurdle model

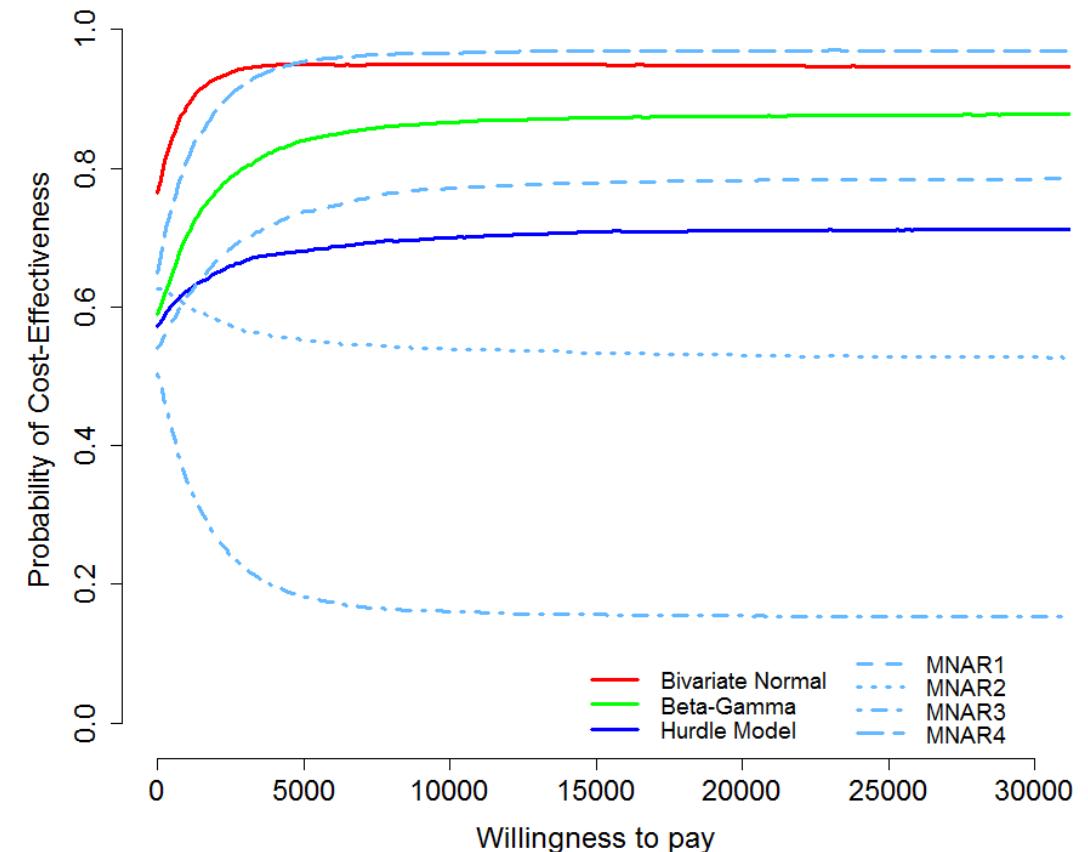
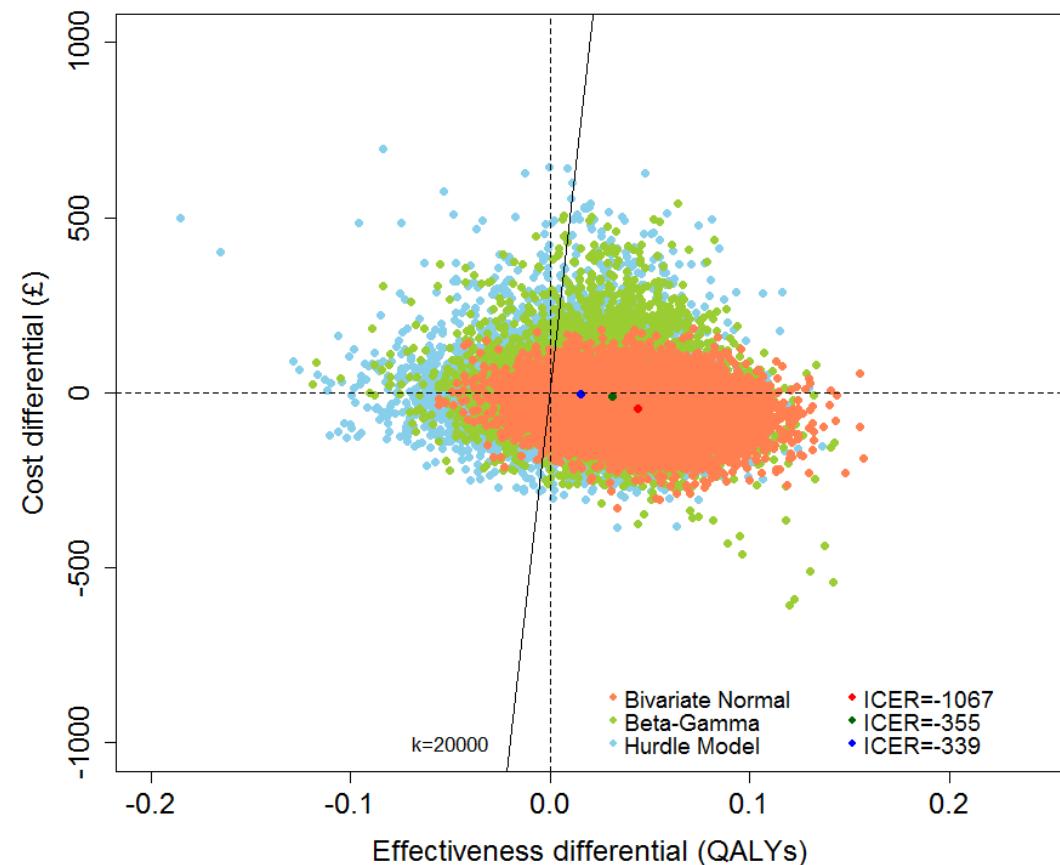
- Model e_{it} as a **mixture** to account for correlation between outcomes + relevant ranges + structural values
- May expand further to account for partially observed baseline utilities u_{0it} (needs untestable assumptions!)



Bayesian multiple imputation (MAR)



Cost-effectiveness analysis



Next lecture