Bayesian methods for missing data: part 1
Key Concepts

Nicky Best and Alexina Mason

Imperial College London

BAYES 2013, May 21-23, Erasmus University Rotterdam

Outline

- Introduction and motivating examples
- Using Bayesian graphical models to represent different types of missing data processes
- Missing response data
  - ignorable missingness
  - non-ignorable missingness
- Missing covariate data
  - fully Bayesian imputation methods
  - comparison with multiple imputation
- Concluding remarks

Introduction

Introduction

- Missing data are common!
- Usually inadequately handled in both observational and experimental research

Introduction

- Missing data are common!
- Usually inadequately handled in both observational and experimental research
- For example, Wood et al. (2004) reviewed 71 recently published BMJ, JAMA, Lancet and NEJM papers
  - 89% had partly missing outcome data
  - In 37 trials with repeated outcome measures, 46% performed complete case analysis
  - Only 21% reported sensitivity analysis

Introduction

- Missing data are common!
- Usually inadequately handled in both observational and experimental research
- For example, Wood et al. (2004) reviewed 71 recently published BMJ, JAMA, Lancet and NEJM papers
  - 89% had partly missing outcome data
  - In 37 trials with repeated outcome measures, 46% performed complete case analysis
  - Only 21% reported sensitivity analysis
- Sterne et al. (2009) reviewed articles using Multiple Imputation in BMJ, JAMA, Lancet and NEJM from 2002 to 2007
  - 59 articles found, with use doubling over 6 year period
  - However, the reporting was almost always inadequate

Example (1): HAMD antidepressant trial

- 6 centre clinical trial, comparing 3 treatments of depression
- 367 subjects randomised to one of 3 treatments
- Subjects rated on Hamilton depression score (HAMD) on 5 weekly visits
  - week 0 before treatment
  - weeks 1-4 during treatment
- HAMD score takes values 0-50
  - the higher the score, the more severe the depression

Example (1): HAMD antidepressant trial

- 6 centre clinical trial, comparing 3 treatments of depression
- 367 subjects randomised to one of 3 treatments
- Subjects rated on Hamilton depression score (HAMD) on 5 weekly visits
  - week 0 before treatment
  - weeks 1-4 during treatment
- HAMD score takes values 0-50
  - the higher the score, the more severe the depression
- Subjects drop-out from week 2 onwards (246 complete cases)

Example (1): HAMD antidepressant trial

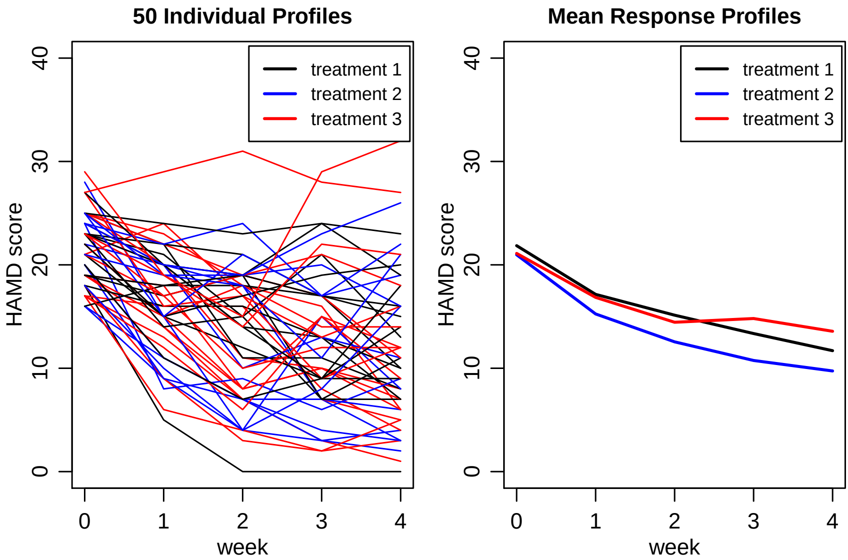
- 6 centre clinical trial, comparing 3 treatments of depression
- 367 subjects randomised to one of 3 treatments
- Subjects rated on Hamilton depression score (HAMD) on 5 weekly visits
  - week 0 before treatment
  - weeks 1-4 during treatment
- HAMD score takes values 0-50
  - the higher the score, the more severe the depression
- Subjects drop-out from week 2 onwards (246 complete cases)
- Data were previously analysed by Diggle and Kenward (1994)

Example (1): HAMD antidepressant trial

- 6 centre clinical trial, comparing 3 treatments of depression
- 367 subjects randomised to one of 3 treatments
- Subjects rated on Hamilton depression score (HAMD) on 5 weekly visits
  - week 0 before treatment
  - weeks 1-4 during treatment
- HAMD score takes values 0-50
  - the higher the score, the more severe the depression
- Subjects drop-out from week 2 onwards (246 complete cases)
- Data were previously analysed by Diggle and Kenward (1994)

Study objective: are there any differences in the effects of the 3 treatments on the change in HAMD score over time?

HAMD example: complete cases



HAMD example: analysis model

- Use the variables
  - y, Hamilton depression (HAMD) score measured at weeks t=0,1,2,3,4
  - x, treatment
- and for simplicity
  - ignore any centre effects
  - assume linear relationships

HAMD example: analysis model

- Use the variables
  - y, Hamilton depression (HAMD) score measured at weeks t=0,1,2,3,4
  - x, treatment
- and for simplicity
  - ignore any centre effects
  - assume linear relationships
- A suitable analysis model might be a hierarchical model with random intercepts and slopes

HAMD example: analysis model

- Use the variables
  - y, Hamilton depression (HAMD) score measured at weeks t=0,1,2,3,4
  - x, treatment
- and for simplicity
  - ignore any centre effects
  - assume linear relationships
- A suitable analysis model might be a hierarchical model with random intercepts and slopes
- We start by just fitting this model to the complete cases (CC)

HAMD example: Complete Case results

Table : posterior mean (95% credible interval) for the contrasts (treatment comparisons) from random effects models fitted to the HAMD data

treatments	complete cases*	
1 v 2	0.50	(-0.03,1.00)
1 v 3	-0.56	(-1.06,-0.04)
2 v 3	-1.06	(-1.56,-0.55)

* individuals with missing scores ignored

CC results suggest that

- treatments 1 and 2 are more effective than treatment 3
- no strong evidence of difference between treatments 1 and 2

HAMD example: Complete Case results

Table : posterior mean (95% credible interval) for the contrasts (treatment comparisons) from random effects models fitted to the HAMD data

treatments	complete cases*	
1 v 2	0.50	(-0.03,1.00)
1 v 3	-0.56	(-1.06,-0.04)
2 v 3	-1.06	(-1.56,-0.55)

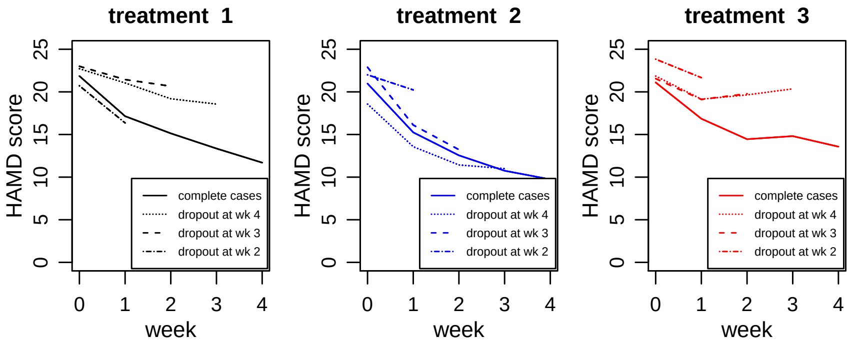
* individuals with missing scores ignored

CC results suggest that

- treatments 1 and 2 are more effective than treatment 3
- no strong evidence of difference between treatments 1 and 2

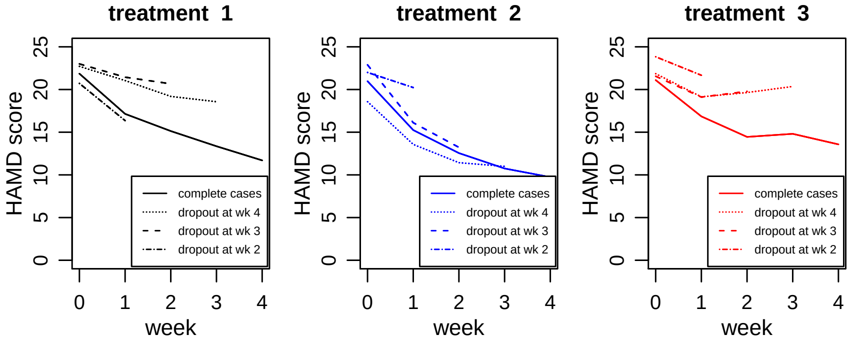
But, takes no account of drop-out

HAMD example: drop out



- Individuals who drop out appear to have somewhat different response profiles to those who remained in the study

HAMD example: drop out



- Individuals who drop out appear to have somewhat different response profiles to those who remained in the study
- Different treatments show slightly different patterns

Example (2): Pollution and low birthweight (LBW)

- Observational study to investigate if there is an association between ambient particulate matter (PM_{10}) concentrations and the risk of term low birth weight

Example (2): Pollution and low birthweight (LBW)

- Observational study to investigate if there is an association between ambient particulate matter (PM_{10}) concentrations and the risk of term low birth weight
- The variables we will use are:

  Y: binary indicator of low birth weight (outcome)

  X: binary indicator of high PM_{10} concentrations (exposure of interest)

  $\boldsymbol{C}$: mother's age, baby gender, deprivation index (vector of fully observed confounders)

  U: maternal smoking (confounder with some missing values)

Example (2): Pollution and low birthweight (LBW)

- Observational study to investigate if there is an association between ambient particulate matter (PM_{10}) concentrations and the risk of term low birth weight
- The variables we will use are:

  Y: binary indicator of low birth weight (outcome)

  X: binary indicator of high PM_{10} concentrations (exposure of interest)

  $\boldsymbol{C}$: mother's age, baby gender, deprivation index (vector of fully observed confounders)

  U: maternal smoking (confounder with some missing values)

- We have data on 8969 births, but only 931 have an observed value for smoking
  - 90% of individuals will be discarded if we use complete case (CC) analysis

LBW example: CC results

- Fit standard logistic regression of Y on X, C and U

		Odds ratio (95% interval) CC (N=931)	
X	High PM_{10}	2.36	(0.96,4.92)
C	Mother's age		
	≤ 25	0.89	(0.32,1.93)
	$25 - 29^{\star}$		1
	$30 - 34$	0.13	(0.00,0.51)
	≥ 35	1.53	(0.39,3.80)
C	Male baby	0.84	(0.34,1.75)
C	Deprivation index	1.74	(1.05,2.90)
U	Smoking	1.86	(0.73,3.89)

* Reference group

- Very wide uncertainty intervals due to excluding 90% of data

Types of missing data

- When dealing with missing data, it is helpful to distinguish between
  - missing responses and missing covariates (regression context)
  - ignorable and non-ignorable missingness mechanisms

Types of missing data

- When dealing with missing data, it is helpful to distinguish between
  - missing responses and missing covariates (regression context)
  - ignorable and non-ignorable missingness mechanisms
- Today, I will focus on missing responses assuming a non-ignorable missingness mechanism
  - Bayesian approach can offer several advantages in this context

Types of missing data

- When dealing with missing data, it is helpful to distinguish between
  - missing responses and missing covariates (regression context)
  - ignorable and non-ignorable missingness mechanisms
- Today, I will focus on missing responses assuming a non-ignorable missingness mechanism
  - Bayesian approach can offer several advantages in this context
- I will also discuss Bayesian methods for handling missing covariates under an ignorable missingness mechanism, and contrast this with multiple imputation (MI)

Graphical Models

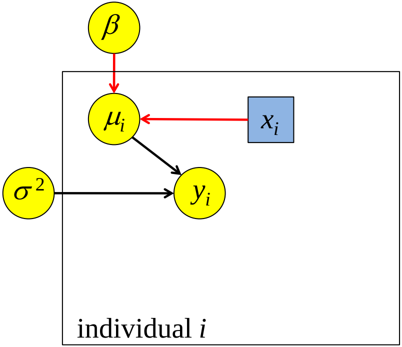
Graphical models to represent different types of missing data

- Graphical models can be a helpful way to visualise different types of missing data and understand their implications for analysis
- More generally, graphical models are a useful tool for building complex Bayesian models

Bayesian graphical models: notation

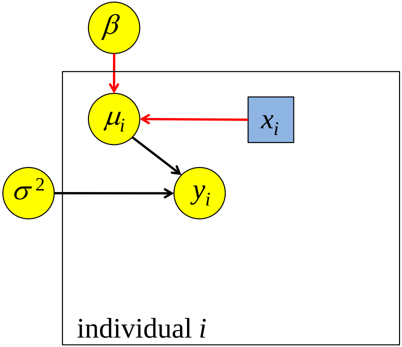
A typical regression model of interest

$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma^2), \quad i = 1, ..., N \\
\mu_i &= \boldsymbol{x}^T \boldsymbol{\beta} \\
\boldsymbol{\beta} &\sim \text{fully specified prior}
\end{aligned}
$$



Bayesian graphical models: notation

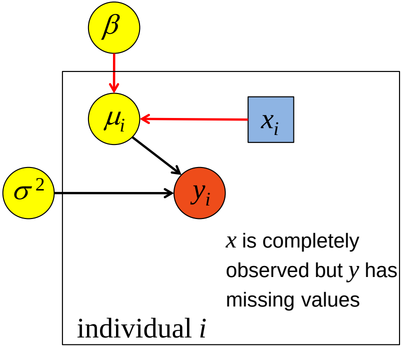
- yellow circles = random variables (data and parameters)
- blue squares = fixed constants (e.g. fully observed covariates)
- black arrows = stochastic dependence
- red arrows = logical dependence
- large rectangles = repeated structures (loops)



Directed Acyclic Graph (DAG) — contains only directed links (arrows) and no cycles

Bayesian graphical models: notation

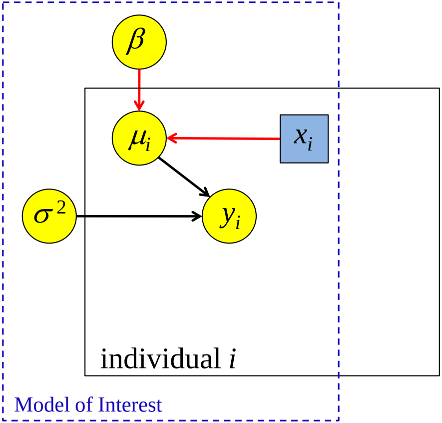
- yellow circles = random variables (data and parameters)
- blue squares = fixed constants (e.g. covariates, denominators)
- black arrows = stochastic dependence
- red arrows = logical dependence
- large rectangles = repeated structures (loops)



- We usually make no distinction in the graph between random variables representing data or parameters
- However, for clarity, we will denote a random variable representing a data node with missing values by an orange circle

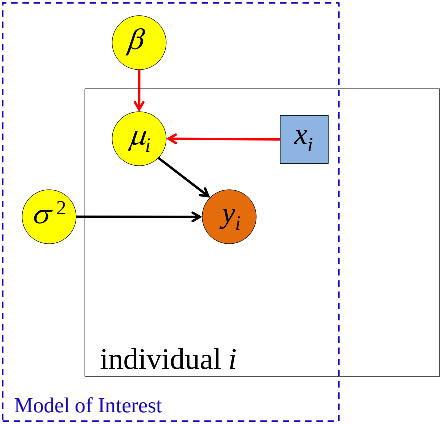
Using DAGs to represent missing data mechanisms

A typical regression model of interest



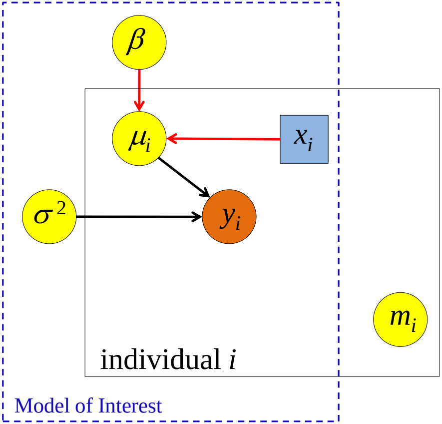
Using DAGs to represent missing data mechanisms

Now suppose x is completely observed, but y has missing values



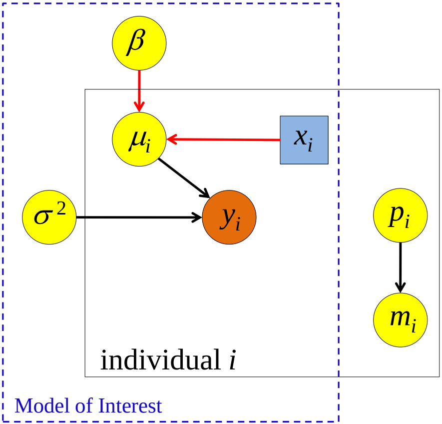
Using DAGs to represent missing data mechanisms

We need to augment the data with a new variable, m_i, that takes value 1 if y_i is missing, and 0 if y_i is observed



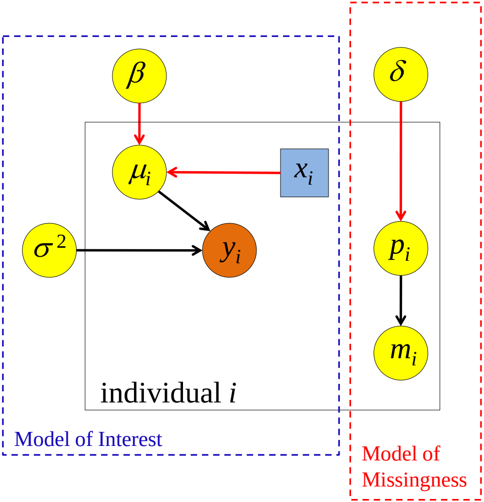
Using DAGs to represent missing data mechanisms

We must then specify a model for the probability, p_i, that $m_i = 1$ (i.e. p_i is the probability that y_i is missing)



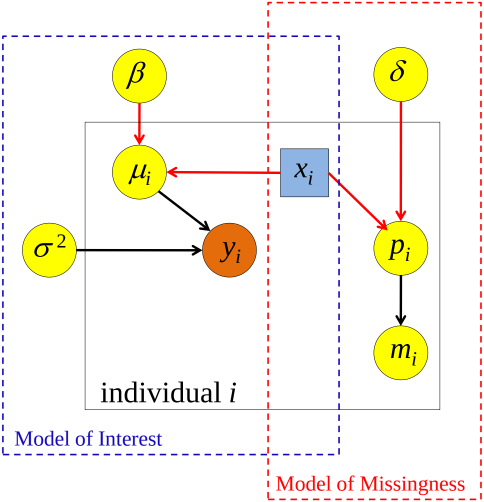
DAG: Missing Completely At Random (MCAR)

e.g. y_i is missing with constant probability δ



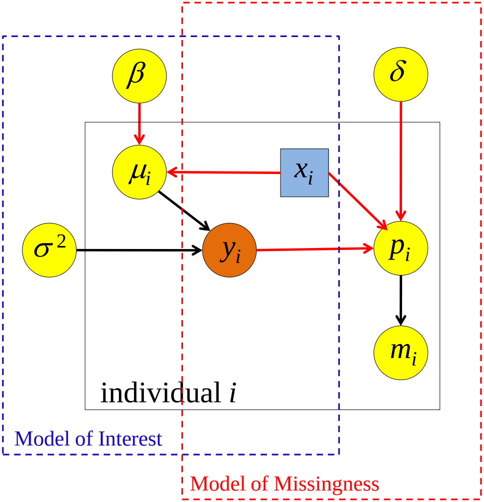
DAG: Missing At Random (MAR)

e.g. y_i is missing with probability that depends on the (observed) covariate value x_i



DAG: Missing Not At Random (MNAR)

e.g. y_i is missing with probability that depends on the (observed) covariate value x_i and on the unobserved value of y_i itself



Joint model for y and m

- The previous DAGs correspond to specifying a joint model (likelihood) for the data of interest and for the missing data indicator:

$$f(y, m|\beta, \sigma^2, \delta, x) = f(y|\beta, \sigma^2, x)f(m|\delta, y, x)$$

Joint model for y and m

- The previous DAGs correspond to specifying a joint model (likelihood) for the data of interest and for the missing data indicator:

$$f(y, m|\beta, \sigma^2, \delta, x) = f(y|\beta, \sigma^2, x) f(m|\delta, y, x)$$

- RHS factorises into analysis model of interest....

Joint model for *y* and *m*

- The previous DAGs correspond to specifying a joint model (likelihood) for the data of interest and for the missing data indicator:

$$f(y, m|\beta, \sigma^2, \delta, x) = f(y|\beta, \sigma^2, x)f(m|\delta, y, x)$$

- RHS factorises into analysis model of interest....
- ..... $\times$ model of missingness

Joint model for *y* and *m*

- The previous DAGs correspond to specifying a joint model (likelihood) for the data of interest and for the missing data indicator:

$$f(y, m|\beta, \sigma^2, \delta, x) = f(y|\beta, \sigma^2, x)f(m|\delta, y, x)$$

- RHS factorises into analysis model of interest....
- ..... $\times$ model of missingness
- This is known as a selection model factorisation

Aside: Pattern mixture factorisation

- Alternatively, we could factorise the joint model as follows:

$$f(y, m|\beta^*, \sigma^{2*}, \delta^*, x) = f(y|m, \beta^*, \sigma^{2*}, x)f(m|\delta^*, x)$$

Aside: Pattern mixture factorisation

- Alternatively, we could factorise the joint model as follows:

$$f(y, m|\beta^*, \sigma^{2*}, \delta^*, x) = f(y|m, \beta^*, \sigma^{2*}, x)f(m|\delta^*, x)$$

- This is known as a pattern mixture model

Aside: Pattern mixture factorisation

- Alternatively, we could factorise the joint model as follows:

$$f(y, m|\beta^*, \sigma^{2*}, \delta^*, x) = f(y|m, \beta^*, \sigma^{2*}, x) f(m|\delta^*, x)$$

- This is known as a pattern mixture model
- Corresponds more directly to what is actually observed (i.e. the distribution of the data within subgroups having different missing data patterns)...

Aside: Pattern mixture factorisation

- Alternatively, we could factorise the joint model as follows:

$$f(y, m|\beta^*, \sigma^{2*}, \delta^*, x) = f(y|m, \beta^*, \sigma^{2*}, x) f(m|\delta^*, x)$$

- This is known as a pattern mixture model
- Corresponds more directly to what is actually observed (i.e. the distribution of the data within subgroups having different missing data patterns)...
- ...but recovering the parameters of the analysis model of interest, $f(y|\beta, \sigma^2, x)$, can be tricky

Aside: Pattern mixture factorisation

- Alternatively, we could factorise the joint model as follows:

$$f(y, m|\beta^*, \sigma^{2*}, \delta^*, x) = f(y|m, \beta^*, \sigma^{2*}, x)f(m|\delta^*, x)$$

- This is known as a pattern mixture model
- Corresponds more directly to what is actually observed (i.e. the distribution of the data within subgroups having different missing data patterns)...
- ...but recovering the parameters of the analysis model of interest, $f(y|\beta, \sigma^2, x)$, can be tricky
- I will focus on the selection model factorisation in this talk

Joint model: integrating out the missing data

- y can be partitioned into $y = (y^{obs}, y^{mis})$

Joint model: integrating out the missing data

- y can be partitioned into $y = (y^{obs}, y^{mis})$
- In order to make inference (Bayesian or MLE) about the model parameters, we need to integrate over the missing data to obtain the observed data likelihood

$$
\begin{aligned}
f(y^{obs}, m|\beta, \sigma^2, \delta, x) &= \int f(y^{obs}, y^{mis}, m|\beta, \sigma^2, \delta, x) dy^{mis} \\
&= \int f(y^{obs}, y^{mis}|\beta, \sigma^2, x) f(m|\delta, y^{obs}, y^{mis}, x) dy^{mis} \qquad (*)
\end{aligned}
$$

Joint model: integrating out the missing data

- y can be partitioned into $y = (y^{obs}, y^{mis})$
- In order to make inference (Bayesian or MLE) about the model parameters, we need to integrate over the missing data to obtain the observed data likelihood

$$\begin{aligned} f(y^{obs}, m|\beta, \sigma^2, \delta, x) &= \int f(y^{obs}, y^{mis}, m|\beta, \sigma^2, \delta, x) dy^{mis} \\ &= \int f(y^{obs}, y^{mis}|\beta, \sigma^2, x) f(m|\delta, y^{obs}, y^{mis}, x) dy^{mis} \qquad (*) \end{aligned}$$

- Under MAR (or MCAR) assumptions, the second term in (*) does not depend on y^{mis}, so the integral can be simplified

$$\begin{aligned} f(y^{obs}, m|\beta, \sigma^2, \delta, x) &= \left\{ \int f(y^{obs}, y^{mis}|\beta, \sigma^2, x) dy^{mis} \right\} f(m|\delta, y^{obs}, x) \\ &= f(y^{obs}|\beta, \sigma^2, x) f(m|\delta, y^{obs}, x) \end{aligned}$$

Joint model: integrating out the missing data

- y can be partitioned into $y = (y^{obs}, y^{mis})$
- In order to make inference (Bayesian or MLE) about the model parameters, we need to integrate over the missing data to obtain the observed data likelihood

$$\begin{aligned} f(y^{obs}, m|\beta, \sigma^2, \delta, x) &= \int f(y^{obs}, y^{mis}, m|\beta, \sigma^2, \delta, x) dy^{mis} \\ &= \int f(y^{obs}, y^{mis}|\beta, \sigma^2, x) f(m|\delta, y^{obs}, y^{mis}, x) dy^{mis} \quad (*) \end{aligned}$$

- Under MAR (or MCAR) assumptions, the second term in (*) does not depend on y^{mis}, so the integral can be simplified

$$\begin{aligned} f(y^{obs}, m|\beta, \sigma^2, \delta, x) &= \left\{ \int f(y^{obs}, y^{mis}|\beta, \sigma^2, x) dy^{mis} \right\} f(m|\delta, y^{obs}, x) \\ &= f(y^{obs}|\beta, \sigma^2, x) f(m|\delta, y^{obs}, x) \end{aligned}$$

$\Rightarrow$ we can ignore the missing data model, $f(m|\delta, y^{obs}, x)$, when making inference about parameters of analysis model

Ignorable/Nonignorable missingness

The missing data mechanism is termed ignorable if

1. the missing data mechanism is MCAR or MAR
2. the parameters of the analysis model (β, σ^2) and the missingness model (δ) are distinct

In the Bayesian setup, an additional condition is

3. the priors on (β, σ^2) and δ are independent

Ignorable/Nonignorable missingness

The missing data mechanism is termed ignorable if

1. the missing data mechanism is MCAR or MAR
2. the parameters of the analysis model (β, σ^2) and the missingness model (δ) are distinct

In the Bayesian setup, an additional condition is

3. the priors on (β, σ^2) and δ are independent

'Ignorable' means we can ignore the model of missingness, but does not necessarily mean we can ignore the missing data!

Ignorable/Nonignorable missingness

The missing data mechanism is termed ignorable if

1. the missing data mechanism is MCAR or MAR
2. the parameters of the analysis model (β, σ^2) and the missingness model (δ) are distinct

In the Bayesian setup, an additional condition is

3. the priors on (β, σ^2) and δ are independent

'Ignorable' means we can ignore the model of missingness, but does not necessarily mean we can ignore the missing data!

However if the data mechanism is nonignorable, then we cannot ignore the model of missingness

Assumptions

- In contrast with the sampling process, which is often known, the missingness mechanism is usually unknown

Assumptions

- In contrast with the sampling process, which is often known, the missingness mechanism is usually unknown
- Although data alone cannot usually definitively tell us the sampling process
  - with fully observed data, we can usually check the plausibility of any assumptions about the sampling process e.g. using residuals and other diagnostics

Assumptions

- In contrast with the sampling process, which is often known, the missingness mechanism is usually unknown
- Although data alone cannot usually definitively tell us the sampling process
  - with fully observed data, we can usually check the plausibility of any assumptions about the sampling process e.g. using residuals and other diagnostics
- Likewise, the missingness pattern, and its relationship to the observations, cannot definitively identify the missingness mechanism
  - Unfortunately, the assumptions we make about the missingness mechanism cannot be definitively checked from the data at hand

Sensitivity analysis

- The issues surrounding the analysis of data sets with missing values therefore centre on assumptions

Sensitivity analysis

- The issues surrounding the analysis of data sets with missing values therefore centre on assumptions
- We have to
  - decide which assumptions are reasonable and sensible in any given setting - contextual/subject matter information will be central to this
  - ensure that the assumptions are transparent
  - explore the sensitivity of inferences/conclusions to the assumptions

Sensitivity analysis

- The issues surrounding the analysis of data sets with missing values therefore centre on assumptions
- We have to
  - decide which assumptions are reasonable and sensible in any given setting - contextual/subject matter information will be central to this
  - ensure that the assumptions are transparent
  - explore the sensitivity of inferences/conclusions to the assumptions
- See talk by Alexina Mason in Part 2 of this session for detailed example

Bayesian inference in the presence of missing data

- Bayesian approach treats missing data as additional unknown quantities for which a posterior distribution can be estimated
  - no fundamental distinction between missing data and unknown parameters

Bayesian inference in the presence of missing data

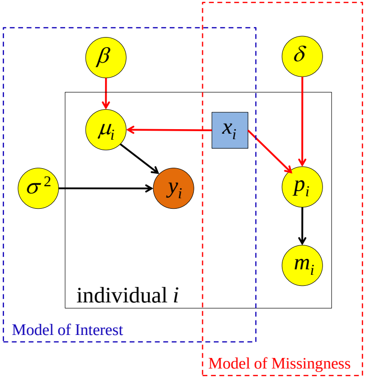
- Bayesian approach treats missing data as additional unknown quantities for which a posterior distribution can be estimated
  - no fundamental distinction between missing data and unknown parameters
- ‘Just’ need to specify appropriate joint model for observed and missing data, the missing data indicator and the model parameters, and estimate in usual way (e.g. using MCMC)

Bayesian inference in the presence of missing data

- Bayesian approach treats missing data as additional unknown quantities for which a posterior distribution can be estimated
  - no fundamental distinction between missing data and unknown parameters
- 'Just' need to specify appropriate joint model for observed and missing data, the missing data indicator and the model parameters, and estimate in usual way (e.g. using MCMC)
- Form of the joint model will depend on
  - whether there are missing values in the response or covariates (or both)
  - whether the missing data mechanism can be assumed to be ignorable or not

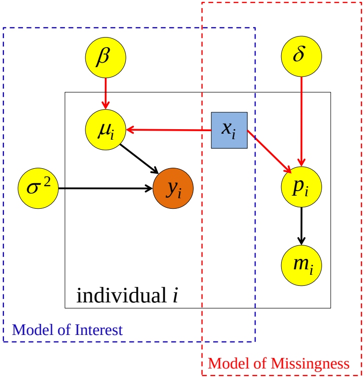
Missing response data

Missing response data
- assuming missing data mechanism is ignorable



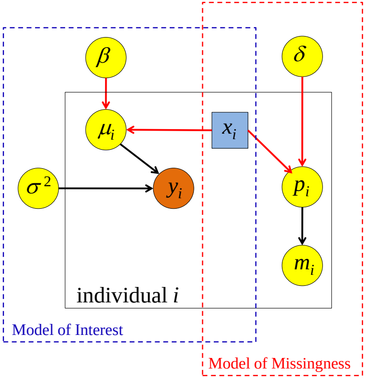
- Model of missingness provides no information about parameters of model of interest, so can be ignored

Missing response data
- assuming missing data mechanism is ignorable



- Model of missingness provides no information about parameters of model of interest, so can be ignored
- Model of interest, $f(y^{obs}, y^{mis}|x, \beta, \sigma^2)$ is just the usual likelihood we would specify for fully observed response y

Missing response data
- assuming missing data mechanism is ignorable



- Model of missingness provides no information about parameters of model of interest, so can be ignored
- Model of interest, $f(y^{obs}, y^{mis}|x, \beta, \sigma^2)$ is just the usual likelihood we would specify for fully observed response y
- Estimating the missing responses y^{mis} is equivalent to posterior prediction from the model fitted to the observed data

HAMD example: ignorable missing data mechanism

Table : posterior mean (95% credible interval) for the contrasts (treatment comparisons) from random effects models fitted to the HAMD data

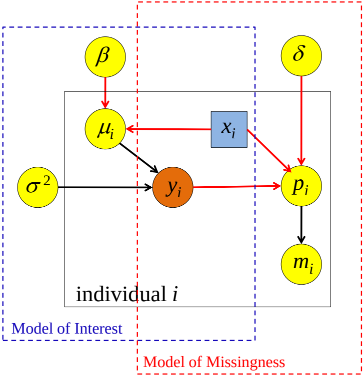
treatments	complete cases*		all cases†	
1 v 2	0.50	(-0.03,1.00)	0.74	(0.25,1.23)
1 v 3	-0.56	(-1.06,-0.04)	-0.51	(-1.01,-0.01)
2 v 3	-1.06	(-1.56,-0.55)	-1.25	(-1.73,-0.77)

* individuals with missing scores ignored
† individuals with missing scores included under the assumption that the missingness mechanism is ignorable

Including all the partially observed cases in the analysis under MAR assumption provides stronger evidence that:

- treatment 2 is more effective than treatment 1
- treatment 2 is more effective than treatment 3

Missing response data
- assuming non-ignorable missing data mechanism



- Inclusion of y (specifically y^{mis}) in the model of missingness
    - changes the missingness assumption from MAR to MNAR
    - provides the link with the analysis model

HAMD example: informative missing data mechanism

- Suppose we think the probability of the HAMD score being missing might be related to the value of that score

HAMD example: informative missing data mechanism

- Suppose we think the probability of the HAMD score being missing might be related to the value of that score
- Then we could model the missing response indicator as follows:

$$
\begin{aligned}
m_{it} &\sim \text{Bernoulli}(p_{it}) \\
\text{logit}(p_{it}) &= \theta + \delta(y_{it} - \bar{y}) \\
\theta, \delta &\sim \text{priors}
\end{aligned}
$$

where $\bar{y}$ is the mean score

HAMD example: informative missing data mechanism

- Suppose we think the probability of the HAMD score being missing might be related to the value of that score
- Then we could model the missing response indicator as follows:

$$\begin{aligned} m_{it} &\sim \text{Bernoulli}(p_{it}) \\ \text{logit}(p_{it}) &= \theta + \delta(y_{it} - \bar{y}) \\ \theta, \delta &\sim \text{priors} \end{aligned}$$

  where $\bar{y}$ is the mean score

- typically, very little information about δ in data
- information depends on parametric model assumptions and error distribution
- advisable to use informative priors (see Alexina Mason's talk)

HAMD Example: MAR v MNAR

Table : posterior mean (95% credible interval) for the contrasts (treatment comparisons) from random effects models fitted to the HAMD data

treatments	complete cases[1]		all cases (mar)[2]		all cases (mnar)[3]	
1 v 2	0.50	(-0.03,1.00)	0.74	(0.25,1.23)	0.75	(0.26,1.24)
1 v 3	-0.56	(-1.06,-0.04)	-0.51	(-1.01,-0.01)	-0.47	(-0.98,0.05)
2 v 3	-1.06	(-1.56,-0.55)	-1.25	(-1.73,-0.77)	-1.22	(-1.70,-0.75)

[1] individuals with missing scores ignored
[2] individuals with missing scores included under the assumption that the missingness mechanism is ignorable
[3] individuals with missing scores included under the assumption that the missingness mechanism is non-ignorable

HAMD Example: MAR v MNAR

Table : posterior mean (95% credible interval) for the contrasts (treatment comparisons) from random effects models fitted to the HAMD data

treatments	complete cases[1]		all cases (mar)[2]		all cases (mnar)[3]	
1 v 2	0.50	(-0.03,1.00)	0.74	(0.25,1.23)	0.75	(0.26,1.24)
1 v 3	-0.56	(-1.06,-0.04)	-0.51	(-1.01,-0.01)	-0.47	(-0.98,0.05)
2 v 3	-1.06	(-1.56,-0.55)	-1.25	(-1.73,-0.77)	-1.22	(-1.70,-0.75)

[1] individuals with missing scores ignored
[2] individuals with missing scores included under the assumption that the missingness mechanism is ignorable
[3] individuals with missing scores included under the assumption that the missingness mechanism is non-ignorable

Allowing for informative missingness with dependence on the current HAMD score:

- has a slight impact on the treatment comparisons
- yields a 95% interval comparing treatments 1 & 3 that includes 0

HAMD Example: Model of missingness parameters

- In a full Bayesian model, it is possible to learn about the parameters of a non-ignorable missingness model (δ)

HAMD Example: Model of missingness parameters

- In a full Bayesian model, it is possible to learn about the parameters of a non-ignorable missingness model (δ)
- However, δ is only identified by the observed data in combination with the model assumptions

HAMD Example: Model of missingness parameters

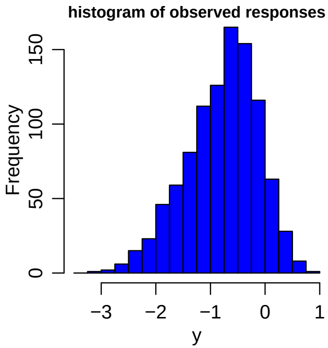
- In a full Bayesian model, it is possible to learn about the parameters of a non-ignorable missingness model (δ)
- However, δ is only identified by the observed data in combination with the model assumptions
- In particular, missing responses are imputed in a way that is consistent with the distributional assumptions in the model of interest

How the distributional assumptions are used

Illustrative example (Daniels & Hogan (2008), Section 8.3.2)

- Consider a cross-sectional setting with
  - a single response
  - no covariates
- Suppose we specify a linear model of missingness,

$$logit(p_i) = \theta_0 + \delta y_i$$



- Assume normal distribution for analysis model, $y_i \sim \text{N}(\mu_i, \sigma^2)$
  - must fill in the right tail $\Rightarrow \delta > 0$
- Assume skew-normal distribution for analysis model
  - $\Rightarrow \delta = 0$

Uncertainty in the analysis model distributional assumptions

- Inference about δ is heavily dependent on the analysis model distributional assumptions about the residuals in combination with the choice and functional form of the covariates

Uncertainty in the analysis model distributional assumptions

- Inference about δ is heavily dependent on the analysis model distributional assumptions about the residuals in combination with the choice and functional form of the covariates
- Unfortunately the analysis model distribution is unverifiable from the observed data when the response is MNAR

Uncertainty in the analysis model distributional assumptions

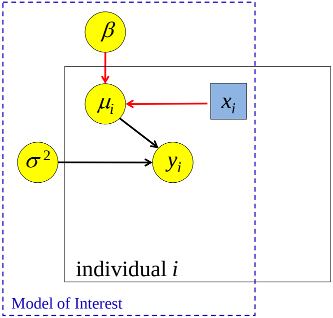
- Inference about δ is heavily dependent on the analysis model distributional assumptions about the residuals in combination with the choice and functional form of the covariates
- Unfortunately the analysis model distribution is unverifiable from the observed data when the response is MNAR
- Different analysis model distributions lead to different results

Uncertainty in the analysis model distributional assumptions

- Inference about δ is heavily dependent on the analysis model distributional assumptions about the residuals in combination with the choice and functional form of the covariates
- Unfortunately the analysis model distribution is unverifiable from the observed data when the response is MNAR
- Different analysis model distributions lead to different results
- Hence sensitivity analysis required to explore impact of different plausible analysis model distributions (see Alexina's talk)

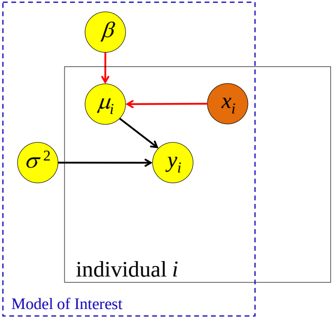
Missing covariate data

Missing covariate data
- assuming missing data mechanism is ignorable



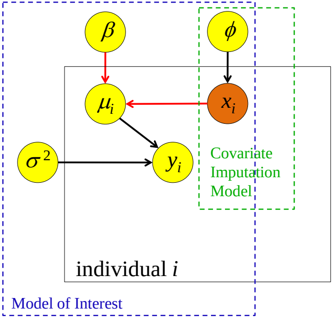
- To include records with missing covariates:

Missing covariate data
- assuming missing data mechanism is ignorable



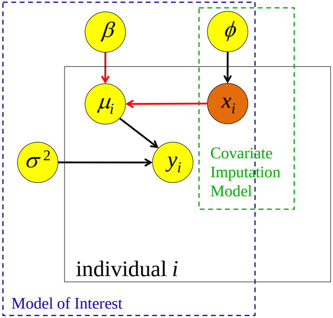
- To include records with missing covariates:
  - we now have to treat covariates as random variables rather than fixed constants

Missing covariate data
- assuming missing data mechanism is ignorable



- To include records with missing covariates:
  - we now have to treat covariates as random variables rather than fixed constants
  - we must build an imputation model to predict their missing values

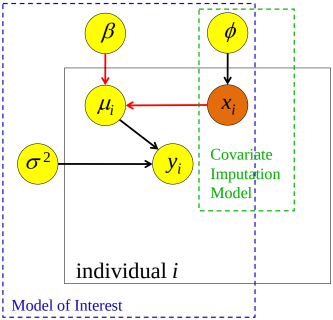
Missing covariate data
- assuming missing data mechanism is ignorable



- To include records with missing covariates:
  - we now have to treat covariates as random variables rather than fixed constants
  - we must build an imputation model to predict their missing values
- Typically this leads to a joint analysis and imputation model of the form

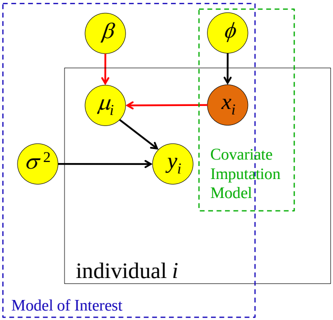
$$f(y, x^{obs}, x^{mis}|\beta, \sigma^2, \phi) = f(y|x^{obs}, x^{mis}, \beta, \sigma^2) f(x^{obs}, x^{mis}|\phi)$$

Missing covariate data
- assuming missing data mechanism is ignorable



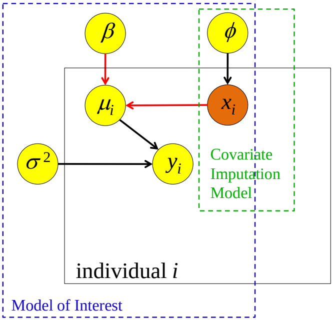
- First term in the joint model, $f(y|x^{obs}, x^{mis}, \beta, \sigma^2)$, is the usual likelihood for the response given fully observed covariates

Missing covariate data
- assuming missing data mechanism is ignorable



- First term in the joint model, $f(y|x^{obs}, x^{mis}, \beta, \sigma^2)$, is the usual likelihood for the response given fully observed covariates
- Second term, $f(x^{obs}, x^{mis}|\phi)$ is a 'prior model' for the covariates, e.g.
  - joint prior distribution, say MVN
  - regression model for each variable with missing values

Missing covariate data
- assuming missing data mechanism is ignorable



- First term in the joint model, $f(y|x^{obs}, x^{mis}, \beta, \sigma^2)$, is the usual likelihood for the response given fully observed covariates
- Second term, $f(x^{obs}, x^{mis}|\phi)$ is a 'prior model' for the covariates, e.g.
    - joint prior distribution, say MVN
    - regression model for each variable with missing values
- It is *not* necessary to include response, y, as a predictor in the covariate imputation model, as its association with x is already accounted for by the first term in the joint model factorisation (unlike multiple imputation)

LBW Example: low birth weight data

- Recall study objective: is there an association between PM_{10} concentrations and the risk of full term low birth weight?
- The variables we will use are:
  - Y: binary indicator of low birth weight (outcome)
  - X: binary indicator of high PM_{10} concentrations (exposure of interest)
  - $\boldsymbol{C}$: mother's age, baby gender, deprivation index (vector of measured confounders)
  - U: smoking (partially observed confounder)
- We have data for 8969 individuals, but only 931 (10%) have an observed value for smoking

LBW Example: missingness assumptions

- Assume that *smoking* is MAR
  - probability of smoking being missing does not depend on whether the individual smokes
  - this assumption is reasonable as the missingness is due to the sample design of the underlying datasets

LBW Example: missingness assumptions

- Assume that *smoking* is MAR
  - probability of smoking being missing does not depend on whether the individual smokes
  - this assumption is reasonable as the missingness is due to the sample design of the underlying datasets
- Also assume that the other assumptions for ignorable missingness hold, so we do not need to specify a model for the missingness mechanism

LBW Example: missingness assumptions

- Assume that *smoking* is MAR
  - probability of smoking being missing does not depend on whether the individual smokes
  - this assumption is reasonable as the missingness is due to the sample design of the underlying datasets
- Also assume that the other assumptions for ignorable missingness hold, so we do not need to specify a model for the missingness mechanism
- However, since *smoking* is a covariate, we must specify an imputation model if we wish to include individuals with missing values of *smoking* in our dataset

LBW Example: specification of joint model

- **Analysis model:** logistic regression for outcome, low birth weight

$$
\begin{aligned}
Y_i &\sim Bernoulli(p_i) \\
logit(p_i) &= \beta_0 + \beta_X X_i + \boldsymbol{\beta}_C^T \boldsymbol{C}_i + \beta_U U_i \\
\beta_0, \beta_X, \ldots &\sim \text{Normal}(0, 10000^2)
\end{aligned}
$$

LBW Example: specification of joint model

- **Analysis model:** logistic regression for outcome, low birth weight

$$
\begin{aligned}
Y_i &\sim Bernoulli(p_i) \\
logit(p_i) &= \beta_0 + \beta_X X_i + \boldsymbol{\beta}_C^T \mathbf{C}_i + \beta_U U_i \\
\beta_0, \beta_X, \ldots &\sim \text{Normal}(0, 10000^2)
\end{aligned}
$$

- **Imputation model:** logistic regression for missing covariate, smoking

$$
\begin{aligned}
U_i &\sim Bernoulli(q_i) \\
logit(q_i) &= \phi_0 + \phi_X X_i + \boldsymbol{\phi}_C^T \mathbf{C}_i \\
\phi_0, \phi_X, \ldots &\sim \text{Normal}(0, 10000^2)
\end{aligned}
$$

LBW Example: specification of joint model

- **Analysis model:** logistic regression for outcome, low birth weight

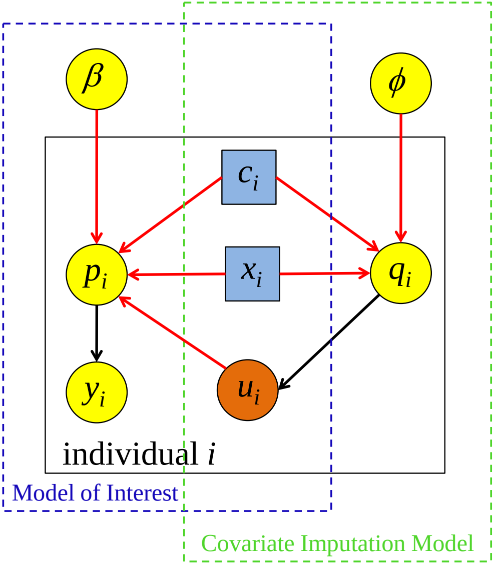
$$
\begin{aligned}
Y_i &\sim Bernoulli(p_i) \\
logit(p_i) &= \beta_0 + \beta_X X_i + \boldsymbol{\beta}_C^T \boldsymbol{C}_i + \beta_U U_i \\
\beta_0, \beta_X, \ldots &\sim \text{Normal}(0, 10000^2)
\end{aligned}
$$

- **Imputation model:** logistic regression for missing covariate, smoking

$$
\begin{aligned}
U_i &\sim Bernoulli(q_i) \\
logit(q_i) &= \phi_0 + \phi_X X_i + \boldsymbol{\phi}_C^T \boldsymbol{C}_i \\
\phi_0, \phi_X, \ldots &\sim \text{Normal}(0, 10000^2)
\end{aligned}
$$

- Unlike multiple imputation, we do not need to include Y as a predictor in the imputation model

LBW example: graphical representation



LBW example: results

		Odds ratio (95% interval)			
		CC (N=931)		All (N=8969)	
X	High PM_{10}	2.36	(0.96,4.92)	1.17	(1.01,1.37)
C	Mother's age				
	≤ 25	0.89	(0.32,1.93)	1.05	(0.74,1.41)
	$25 - 29^\star$		1		1
	$30 - 34$	0.13	(0.00,0.51)	0.80	(0.55,1.14)
	≥ 35	1.53	(0.39,3.80)	1.14	(0.73,1.69)
C	Male baby	0.84	(0.34,1.75)	0.76	(0.58,0.95)
C	Deprivation index	1.74	(1.05,2.90)	1.34	(1.17,1.53)
U	Smoking	1.86	(0.73,3.89)	1.92	(0.80,3.82)

* Reference group

- CC analysis is very uncertain
- Extra records shrink intervals for X coefficient substantially

LBW example: results

		Odds ratio (95% interval)			
		CC (N=931)		All (N=8969)	
X	High PM_{10}	2.36	(0.96,4.92)	1.17	(1.01,1.37)
C	Mother's age				
	≤ 25	0.89	(0.32,1.93)	1.05	(0.74,1.41)
	$25-29^{\star}$		1		1
	$30-34$	0.13	(0.00,0.51)	0.80	(0.55,1.14)
	≥ 35	1.53	(0.39,3.80)	1.14	(0.73,1.69)
C	Male baby	0.84	(0.34,1.75)	0.76	(0.58,0.95)
C	Deprivation index	1.74	(1.05,2.90)	1.34	(1.17,1.53)
U	Smoking	1.86	(0.73,3.89)	1.92	(0.80,3.82)

* Reference group

- Little impact on U coefficient, reflecting uncertainty in imputations

Comments on covariate imputation models

- Covariate imputation model gets more complex if > 1 missing covariates
  - typically need to account for correlation between missing covariates
  - could assume multivariate normality if covariates all continuous
  - for mixed binary, categorical and continuous covariates, could fit latent variable (multivariate probit) model (Chib and Greenberg 1998; BUGS book, Ch. 9)

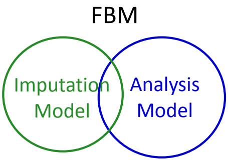
Comments on covariate imputation models

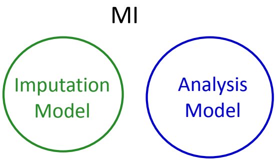
- Covariate imputation model gets more complex if > 1 missing covariates
  - typically need to account for correlation between missing covariates
  - could assume multivariate normality if covariates all continuous
  - for mixed binary, categorical and continuous covariates, could fit latent variable (multivariate probit) model (Chib and Greenberg 1998; BUGS book, Ch. 9)
- If we assume that *smoking* is MNAR, then we must add a third part to the model
  - a model of missingness with a missingness indicator variable for smoking as the response

Multiple Imputation (MI)

- Fully Bayesian Modelling (FBM) is one of a number of ‘statistically principled’ methods for dealing with missing data
- Of the alternatives, standard Multiple Imputation is closest in spirit and has a Bayesian justification
- Multiple imputation was developed by Rubin (1996)
  - Most widely used ‘principled’ method for handling missing data
  - Usually assumes missingness mechanism is MAR (can be used for MNAR but more tricky)
  - Most useful for handling missing **covariates**

Comparison of FBM and MI





- 1 stage procedure
  - Imputation and Analysis Models simultaneously
- imputation model uses joint distribution of all missing variables
- response variable directly informs imputations via feedback from analysis model (congenial)

- 2 stage procedure
  1. fit Imputation Model
  2. fit Analysis Model
- imputation model usually based on a set of univariate conditional distributions (incompatible)
- response variable included as additional predictor in imputation model (uncongenial)

Simulation study to compare FBM and MI

- Generated 1000 simulated data sets with
  - 2 correlated explanatory variables, x and u
  - response, y, dependent on x and u
  - missingness imposed on u

Simulation study to compare FBM and MI

- Generated 1000 simulated data sets with
  - 2 correlated explanatory variables, x and u
  - response, y, dependent on x and u
  - missingness imposed on u
- Each simulated dataset analysed by a series of models to handle missing covariate (GOLD, CC, FBM, MI)
  - correct analysis model used in all cases

Simulation study to compare FBM and MI

- Generated 1000 simulated data sets with
  - 2 correlated explanatory variables, x and u
  - response, y, dependent on x and u
  - missingness imposed on u
- Each simulated dataset analysed by a series of models to handle missing covariate (GOLD, CC, FBM, MI)
  - correct analysis model used in all cases
- Performance (bias, coverage) of models assessed for
  - coefficient for u, β_u, (true value=-2)
  - coefficient for x, β_x, (true value=1)

Simulation study results

- For ‘non-complex’ scenarios (ignorable missingness; non-hierarchical data structure), FBM and MI both perform well (almost unbiased estimates with nominal coverage)

Simulation study results

- For ‘non-complex’ scenarios (ignorable missingness; non-hierarchical data structure), FBM and MI both perform well (almost unbiased estimates with nominal coverage)
- Bigger discrepancies are seen with more complex scenarios
  - hierarchical structure
  - informative missingness

Scenario 1: Hierarchical structure — simulation design

- Data generated with 10 clusters, each with 100 individuals:

$$\begin{pmatrix} x_c \\ u_c \\ \alpha_c \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 & 0.5 & 0.5 \\ 0.5 & 2 & 0.5 \\ 0.5 & 0.5 & 4 \end{pmatrix} \right)$$

$$\begin{pmatrix} x_i \\ u_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} x_c \\ u_c \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$$

$$y_i \sim N(\alpha_c + x_i - 2u_i, 1)$$

c indicates cluster level data; *i* indicates individual level data

Scenario 1: Hierarchical structure — simulation design

- Data generated with 10 clusters, each with 100 individuals:

$$\begin{pmatrix} x_c \\ u_c \\ \alpha_c \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 & 0.5 & 0.5 \\ 0.5 & 2 & 0.5 \\ 0.5 & 0.5 & 4 \end{pmatrix}\right)$$

$$\begin{pmatrix} x_i \\ u_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} x_c \\ u_c \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

$$y_i \sim N(\alpha_c + x_i - 2u_i, 1)$$

c indicates cluster level data; i indicates individual level data

- Impose MAR missingness s.t. u_i is missing with probability p_i

$$logit(p_i) = -0.5 + 0.5y_i$$

Scenario 1: Hierarchical structure — imputation model

Impute $u_i \sim N(\mu_i, \sigma^2)$ where:

- MI: $\mu_i = \gamma_0 + \gamma_1 x_i + \gamma_2 y_i$
- FBM: $\mu_i = \gamma_0 + \gamma_1 x_i$
- FBM (HS: ri): $\mu_i = \gamma_{0,c} + \gamma_1 x_i$
- FBM (HS: ri+rs): $\mu_i = \gamma_{0,c} + \gamma_{1,c} x_i$

Correct analysis model used in all cases

Scenario 1: Hierarchical structure — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-2.00	0.00	0.93	0.14
CC	-1.92	0.08	0.70	0.21
FBM (no HS)	-1.93	0.07	0.67	0.19
FBM (HS: ri)	-2.00	0.00	0.94	0.19
FBM (HS: ri+rs)	-2.00	0.00	0.94	0.19
MI (no HS)	-1.36	0.64	0.00	0.33

Scenario 1: Hierarchical structure — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-2.00	0.00	0.93	0.14
CC	-1.92	0.08	0.70	0.21
FBM (no HS)	-1.93	0.07	0.67	0.19
FBM (HS: ri)	-2.00	0.00	0.94	0.19
FBM (HS: ri+rs)	-2.00	0.00	0.94	0.19
MI (no HS)	-1.36	0.64	0.00	0.33

If hierarchical structure ignored in imputation model

- FBM - slight bias and poor coverage

Scenario 1: Hierarchical structure — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-2.00	0.00	0.93	0.14
CC	-1.92	0.08	0.70	0.21
FBM (no HS)	-1.93	0.07	0.67	0.19
FBM (HS: ri)	-2.00	0.00	0.94	0.19
FBM (HS: ri+rs)	-2.00	0.00	0.94	0.19
MI (no HS)	-1.36	0.64	0.00	0.33

If hierarchical structure ignored in imputation model

- FBM - slight bias and poor coverage
- MI - much worse (no feedback from structure in analysis model)

Scenario 1: Hierarchical structure — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-2.00	0.00	0.93	0.14
CC	-1.92	0.08	0.70	0.21
FBM (no HS)	-1.93	0.07	0.67	0.19
FBM (HS: ri)	-2.00	0.00	0.94	0.19
FBM (HS: ri+rs)	-2.00	0.00	0.94	0.19
MI (no HS)	-1.36	0.64	0.00	0.33

If hierarchical structure incorporated in imputation model

- bias corrected
- nominal coverage rate achieved

Scenario 1: Hierarchical structure — β_x results

	average estimate	bias	coverage rate	interval width
GOLD	1.00	-0.00	0.94	0.14
CC	0.96	-0.04	0.89	0.20
FBM (no HS)	0.85	-0.15	0.21	0.19
FBM (HS: ri)	0.99	-0.01	0.94	0.19
FBM (HS: ri+rs)	0.99	-0.01	0.94	0.19
MI (no HS)	0.53	-0.47	0.01	0.26

Pattern of bias and coverage results similar to β_u

Scenario 2: Informative missingness — simulation design

- Data generated with no hierarchical structure for 100 individuals, as follows:

$$\begin{pmatrix} x \\ u \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$
$$y \sim N(1 + x - 2u, 4^2)$$

Scenario 2: Informative missingness — simulation design

- Data generated with no hierarchical structure for 100 individuals, as follows:

$$\begin{pmatrix} x \\ u \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$
$$y \sim N(1 + x - 2u, 4^2)$$

- Impose MNAR missingness such that u is missing with probability p

$$logit(p) = -2 + 2|u| + 0.5y$$

$\Rightarrow$ u more likely to be missing if it is very small or very large ('v-shaped' missingness)

Scenario 2: Informative missingness — fitted models

FBM models:

- Imputation model: $u_i \sim N(\mu_i, \sigma^2)$; $\mu_i = \gamma_0 + \gamma_1 x_i$
- Covariate missingness: $m_i \sim Bern(p_i)$; $logit\ p_i = ...$
- 4 variants on model for p_i:
  - MAR: no model of covariate missingness
  - MNAR: assumes linear shape (linear)
  - MNAR: allows v-shape (vshape)
  - MNAR: allows v-shape + priors inform signs of slopes (vshape+)

Scenario 2: Informative missingness — fitted models

FBM models:

- Imputation model: $u_i \sim N(\mu_i, \sigma^2)$; $\quad \mu_i = \gamma_0 + \gamma_1 x_i$
- Covariate missingness: $m_i \sim Bern(p_i)$; $\quad logit\ p_i = ...$
- 4 variants on model for p_i:
    - MAR: no model of covariate missingness
    - MNAR: assumes linear shape (linear)
    - MNAR: allows v-shape (vshape)
    - MNAR: allows v-shape + priors inform signs of slopes (vshape+)

MI model:

- Imputation model: $u_i \sim N(\mu_i, \sigma^2)$; $\quad \mu_i = \gamma_0 + \gamma_1 x_i + \gamma_2 y_i$
- Assumes MAR, i.e. no model of covariate missingness
    - most implementations of MI do not readily extend to MNAR
    - ad hoc sensitivity analysis to MNAR possible by inflating or deflating imputations (van Buuren and Groothuis-Oudshoorn, 2011)

Scenario 2: Informative missingness — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-1.99	0.01	0.95	1.68
CC	-1.66	0.34	0.92	2.63
FBM: MAR	-2.25	-0.25	0.93	3.18
FBM: MNAR (linear)	-2.08	-0.08	0.97	3.76
FBM: MNAR (vshape)	-2.06	-0.06	0.96	3.49
FBM: MNAR (vshape+)	-2.02	-0.02	0.96	3.31
MI: MAR	-2.25	-0.25	0.90	3.33

Scenario 2: Informative missingness — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-1.99	0.01	0.95	1.68
CC	-1.66	0.34	0.92	2.63
FBM: MAR	-2.25	-0.25	0.93	3.18
FBM: MNAR (linear)	-2.08	-0.08	0.97	3.76
FBM: MNAR (vshape)	-2.06	-0.06	0.96	3.49
FBM: MNAR (vshape+)	-2.02	-0.02	0.96	3.31
MI: MAR	-2.25	-0.25	0.90	3.33

- MAR results in bias and slightly reduced coverage

Scenario 2: Informative missingness — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-1.99	0.01	0.95	1.68
CC	-1.66	0.34	0.92	2.63
FBM: MAR	-2.25	-0.25	0.93	3.18
FBM: MNAR (linear)	-2.08	-0.08	0.97	3.76
FBM: MNAR (vshape)	-2.06	-0.06	0.96	3.49
FBM: MNAR (vshape+)	-2.02	-0.02	0.96	3.31
MI: MAR	-2.25	-0.25	0.90	3.33

- MAR results in bias and slightly reduced coverage
- improvements if allow MNAR, even if wrong form

Scenario 2: Informative missingness — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-1.99	0.01	0.95	1.68
CC	-1.66	0.34	0.92	2.63
FBM: MAR	-2.25	-0.25	0.93	3.18
FBM: MNAR (linear)	-2.08	-0.08	0.97	3.76
FBM: MNAR (vshape)	-2.06	-0.06	0.96	3.49
FBM: MNAR (vshape+)	-2.02	-0.02	0.96	3.31
MI: MAR	-2.25	-0.25	0.90	3.33

- MAR results in bias and slightly reduced coverage
- improvements if allow MNAR, even if wrong form
- further improvements from correct form

Scenario 2: Informative missingness — β_u results

	average estimate	bias	coverage rate	interval width
GOLD	-1.99	0.01	0.95	1.68
CC	-1.66	0.34	0.92	2.63
FBM: MAR	-2.25	-0.25	0.93	3.18
FBM: MNAR (linear)	-2.08	-0.08	0.97	3.76
FBM: MNAR (vshape)	-2.06	-0.06	0.96	3.49
FBM: MNAR (vshape+)	-2.02	-0.02	0.96	3.31
MI: MAR	-2.25	-0.25	0.90	3.33

- MAR results in bias and slightly reduced coverage
- improvements if allow MNAR, even if wrong form
- further improvements from correct form
- and even better with informative priors

Scenario 2: Informative missingness — β_x results

	average estimate	bias	coverage rate	interval width
GOLD	0.99	-0.01	0.94	1.65
CC	0.70	-0.30	0.91	2.06
FBM: MAR	0.87	-0.13	0.94	1.85
FBM: MNAR (linear)	0.83	-0.17	0.94	1.89
FBM: MNAR (vshape)	0.87	-0.13	0.95	1.91
FBM: MNAR (vshape+)	0.89	-0.11	0.94	1.93
MI: MAR	0.87	-0.13	0.94	1.88

Scenario 2: Informative missingness — β_x results

	average estimate	bias	coverage rate	interval width
GOLD	0.99	-0.01	0.94	1.65
CC	0.70	-0.30	0.91	2.06
FBM: MAR	0.87	-0.13	0.94	1.85
FBM: MNAR (linear)	0.83	-0.17	0.94	1.89
FBM: MNAR (vshape)	0.87	-0.13	0.95	1.91
FBM: MNAR (vshape+)	0.89	-0.11	0.94	1.93
MI: MAR	0.87	-0.13	0.94	1.88

- MAR results in modest bias (FBM and MI)

Scenario 2: Informative missingness — β_x results

	average estimate	bias	coverage rate	interval width
GOLD	0.99	-0.01	0.94	1.65
CC	0.70	-0.30	0.91	2.06
FBM: MAR	0.87	-0.13	0.94	1.85
FBM: MNAR (linear)	0.83	-0.17	0.94	1.89
FBM: MNAR (vshape)	0.87	-0.13	0.95	1.91
FBM: MNAR (vshape+)	0.89	-0.11	0.94	1.93
MI: MAR	0.87	-0.13	0.94	1.88

- MAR results in modest bias (FBM and MI)
- wrong MNAR (linear) slightly worse than MAR

Scenario 2: Informative missingness — β_x results

	average estimate	bias	coverage rate	interval width
GOLD	0.99	-0.01	0.94	1.65
CC	0.70	-0.30	0.91	2.06
FBM: MAR	0.87	-0.13	0.94	1.85
FBM: MNAR (linear)	0.83	-0.17	0.94	1.89
FBM: MNAR (vshape)	0.87	-0.13	0.95	1.91
FBM: MNAR (vshape+)	0.89	-0.11	0.94	1.93
MI: MAR	0.87	-0.13	0.94	1.88

- MAR results in modest bias (FBM and MI)
- wrong MNAR (linear) slightly worse than MAR
- little gain in correct MNAR over MAR

Concluding remarks

Concluding remarks

- Bayesian methods naturally accomodate missing data without requiring new techniques for inference

Concluding remarks

- Bayesian methods naturally accomodate missing data without requiring new techniques for inference
- Bayesian framework is well suited to the process of building complex models, linking smaller sub-models into a coherent joint model

Concluding remarks

- Bayesian methods naturally accomodate missing data without requiring new techniques for inference
- Bayesian framework is well suited to the process of building complex models, linking smaller sub-models into a coherent joint model
- A typical model may consist of 3 parts:
  1. analysis model
  2. covariate imputation model
  3. model of missingness
- Models can become computationally challenging....

Concluding remarks

Covariate imputation

- Full Bayes and MI often produce similar results
- Full Bayes can lead to improved performance with complex data structures

Concluding remarks

Covariate imputation

- Full Bayes and MI often produce similar results
- Full Bayes can lead to improved performance with complex data structures

Non-ignorable missingness

- Typically need informative priors to help identify selection models for informative non-response
- Sensitivity analysis to examine impact of modelling assumptions for non-ignorable missing data mechanisms is essential (see Alexina's talk)

Thank you for your attention!

Funding: ESRC BIAS project (www.bias-project.org.uk)

References and Further Reading

- Daniels, M and Hogan J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* CRC Press.
- Diggle, P and Kenward MG (1994). Informative Drop-out in Longitudinal Data Analysis (with discussion). *JRSSC*, **43**, 49–93.
- Little RJA and Rubin DB (2002). *Statistical Analysis with Missing Data*, 2nd edition, Wiley, New Jersey.
- Mason, A, Richardson, S, Plewis I, and Best, N (2012). Strategy for modelling non-random missing data mechanisms in observational studies using Bayesian methods. *Journal of Official Statistics*. 28:279-302.
- Molitor NT, Best N, Jackson C and Richardson S (2009). Using Bayesian graphical models to model biases in observational studies and to combine multiple datasources: Application to low birth-weight and water disinfection by-products. *Journal of the Royal Statistical Society, Series A*. 172: 615-637.
- Lunn et al (2012). *The BUGS Book.* Chapter 9.1.