Bayesian methods for missing data: part 2

Illustration of a General Strategy

Alexina Mason and Nicky Best

Imperial College London

BAYES 2013, May 21-23, Erasmus University Rotterdam

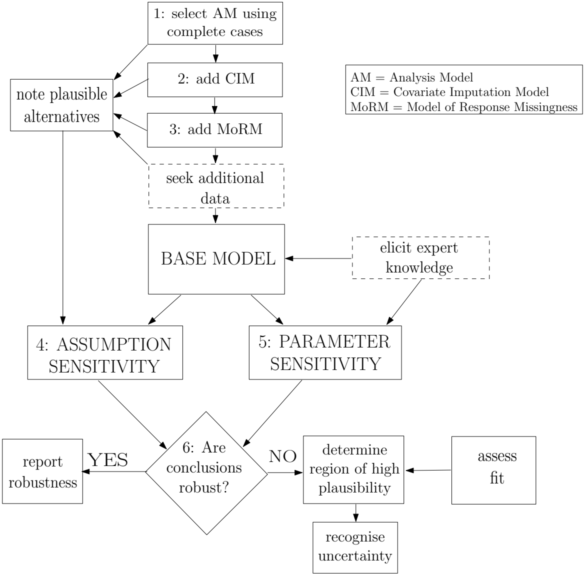
Introduction

- In part 1, we discussed the use of Bayesian joint models for dealing with missing data
- Considering a regression context, key points were:
  - subjects with missing responses can be modelled assuming ignorable missingness using just the analysis model
  - a missingness indicator must be modelled to allow for a non-ignorable missingness mechanism
  - a covariate imputation model must be built to include subjects with missing covariates
- In part 2, we will:
  - demonstrate how these ideas can be incorporated into a general strategy for modelling missing data
  - focus on sensitivity analysis
  - use the HAMD data as an illustrative example throughout

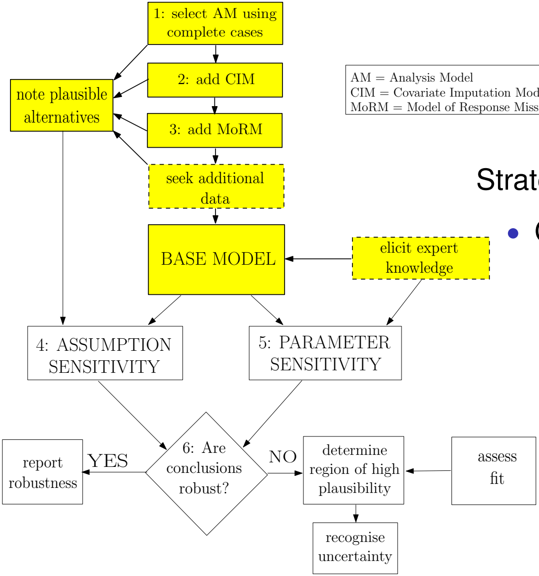
Strategy Overview

- The strategy (Mason et al., 2012b) consists of two parts
  - constructing a base model
  - assessing conclusions from this base model against a selection of well chosen sensitivity analyses
- It allows
  - the uncertainty from the missing data to be taken into account
  - additional sources of information to be utilised
- It can be implemented using currently available software, e.g. WinBUGS

Schematic Diagram



Schematic Diagram: constructing a base model

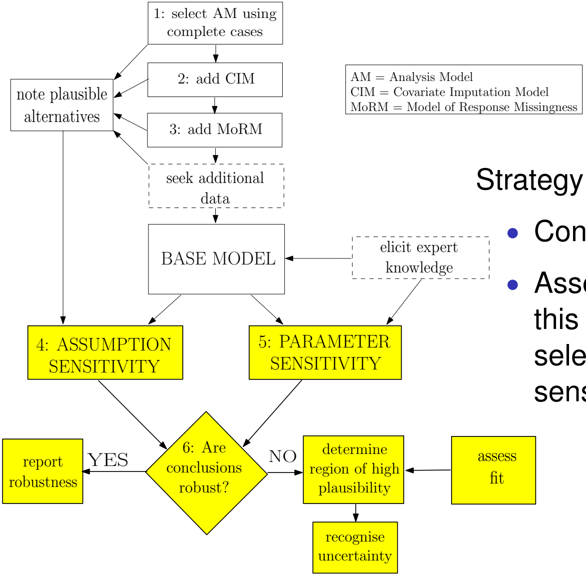


AM = Analysis Model
CIM = Covariate Imputation Model
MoRM = Model of Response Missingness

Strategy consists of two parts:

- Constructing a base model

Schematic Diagram: sensitivity analysis

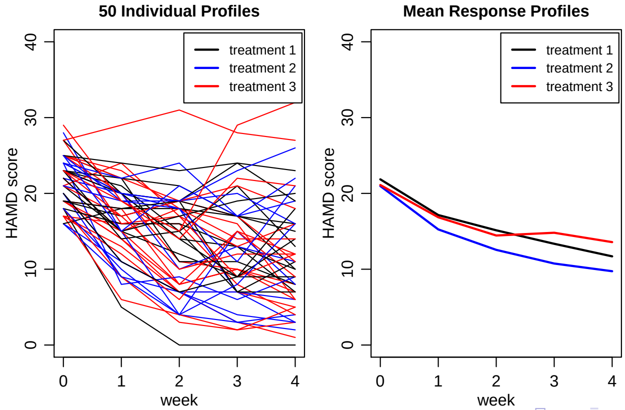


Strategy consists of two parts:

- Constructing a base model
- Assessing conclusions from this base model against a selection of well chosen sensitivity analyses

Illustrative Example: HAMD revisited

- Antidepressant clinical trial, comparing 3 treatments
- Subjects rated on HAMD score on 5 weekly visits
- Objective is to compare the effects of the 3 treatments on the improvement in HAMD score over time



Before the strategy: step 0

- The strategy consists of a series of model building steps
- Before starting, the missingness should be explored to determine
    - which steps are required?
    - are any other modifications needed?
- In particular
    - which variables have missing values?
    - what is the extent and pattern of missingness?
    - what are plausible explanations for the missingness?

HAMD example: step 0

Which variables have missing values?

- HAMD score (model response) missing in weeks 3-5
- No covariate missingness $\Rightarrow$ CIM not needed (omit step 2)

HAMD example: step 0

Which variables have missing values?

- HAMD score (model response) missing in weeks 3-5
- No covariate missingness $\Rightarrow$ CIM not needed (omit step 2)

What is the extent and pattern of missingness?

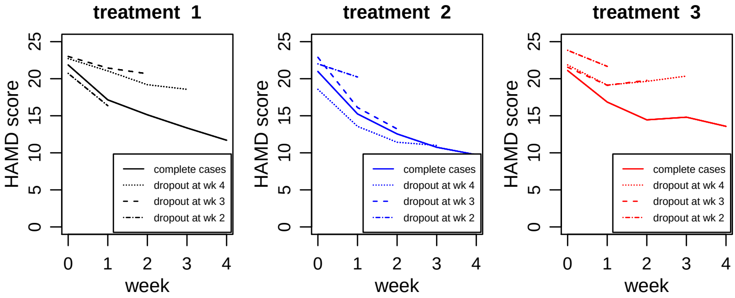
Percentage of missingness by treatment and week

	treat. 1	treat. 2	treat. 3	all treatments
week 2	11.7	22.0	9.3	14.2
week 3	19.2	29.7	16.3	21.5
week 4	36.7	35.6	27.1	33.0

- level and pattern of missingness inconsistent across treatments

HAMD example: step 0 continued

What is the extent and pattern of missingness? (continued)



- individuals have different profiles if they dropped out rather than remained in the study
- the treatments show different patterns

HAMD example: step 0 continued II

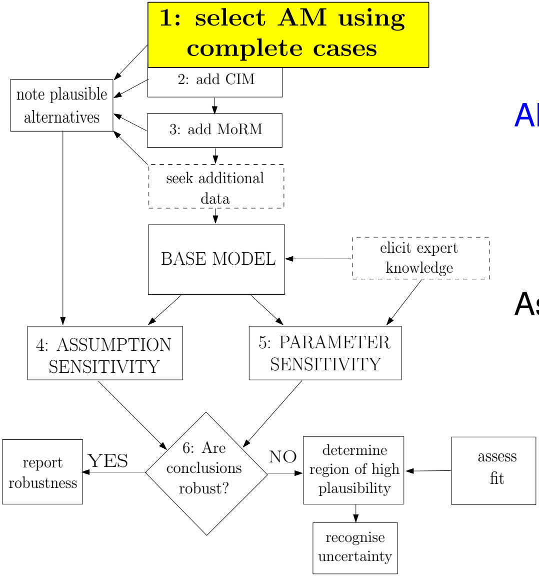
What are plausible explanations for the missingness?

- patients for whom the treatment is successful and get better may decide not to continue in the study
- patients not showing any improvement or feeling worse, may seek alternative treatment and drop-out of the study
- in either case, informative missingness
  $\Rightarrow$ MoRM needed (step 3)

Part 1 (steps 1-3): constructing a base model

- This part involves building a joint model as follows:
  1. choose an analysis model
  2. add a covariate imputation model
  3. add a model of response missingness
- Optionally, the amount of available information can be increased by incorporating data from other sources and/or expert knowledge
- The strategy
  - allows informative missingness in the response
  - but assumes that the covariates are MAR
- However, it can be adapted to reflect alternative assumptions

HAMD example: analysis model (step 1)

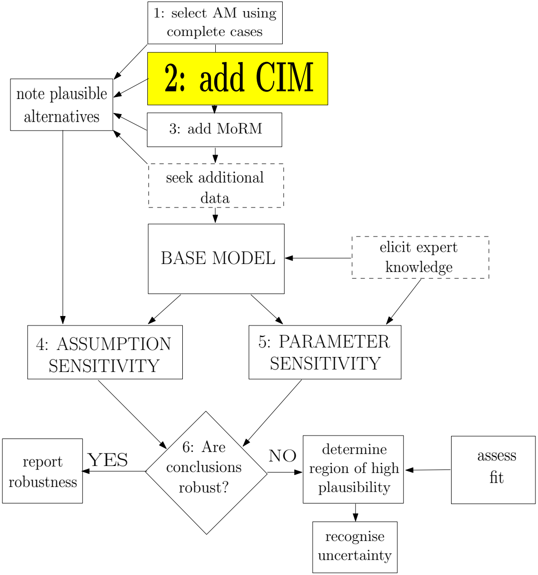


AM = Analysis Model

As discussed in part 1

- a hierarchical model with random intercepts and random slopes is reasonable

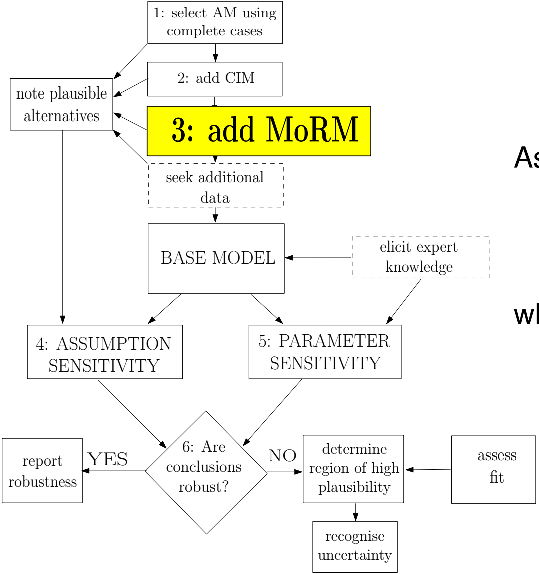
HAMD example: covariate imputation model (step 2)



CIM = Covariate Imputation Model

- No missing covariates in this example, so not required
- If data includes missing covariates, set up CIM to produce realistic imputations at this stage
- See part 1 for details
- Without a CIM, records with missing covariates cannot be included

HAMD ex.: model of response missingness (step 3)



MoRM = Model of Response Missingness

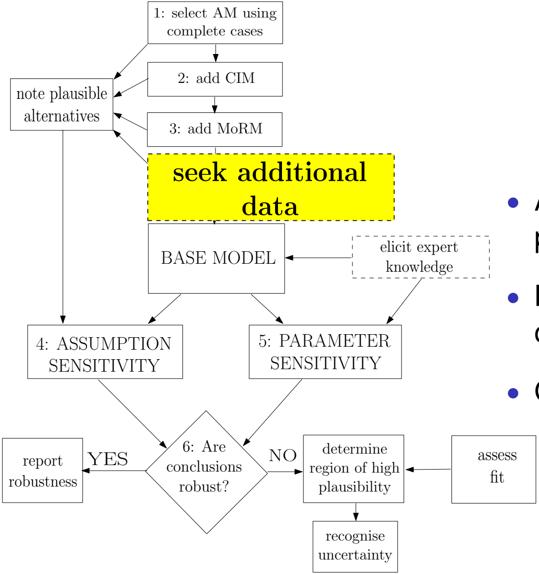
As discussed in part 1 use

$$m_{iw} \sim \text{Bernoulli}(p_{iw})$$
$$\text{logit}(p_{iw}) = \theta_0 + \delta(y_{iw} - \bar{y})$$

where $\bar{y}$ is mean of observed ys

- Allows informative missingness in the response
- Dependence is on current HAMD score

Optional step: seek additional data

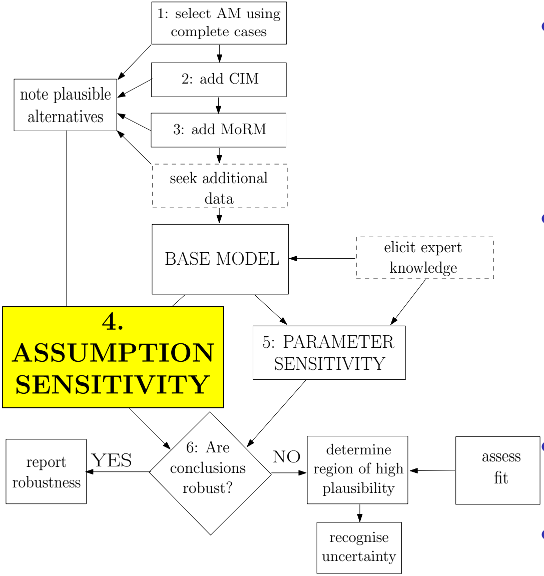


- Additional data can help with parameter estimation
- Most useful with missing covariates
- Omitted for HAMD example

Part 2 (steps 4-6): sensitivity analysis

- Sensitivity analysis is essential because assumptions are untestable from the data
- There are many possible options, and the appropriate choice is problem dependent
- We propose two types of sensitivity analysis:
  1. an assumption sensitivity
  2. a parameter sensitivity

Step 4 - assumption sensitivity



- Assumption sensitivity forms alternative models by changing the assumptions in the different base sub-models
- Key assumptions include:
  - AM error distribution
  - transformation of the AM response
  - functional form of the MoRM
- Stage 1: change single aspect to assess effect
- Stage 2: combine several changes

HAMD example: assumption sensitivity (step 4)

There are many options, including but not limited to the following

- Analysis model:
  - use a t_4 rather than normal error distribution
  - use an autoregressive model, AR(1), rather than random effects
  - include centre effects
  - allow for non-linearity by including a quadratic term
- Model of response missingness:
  - allow dependence on change in HAMD score
  - allow dependence on treatment
  - allow for non-linear relationship (piece-wise linear)

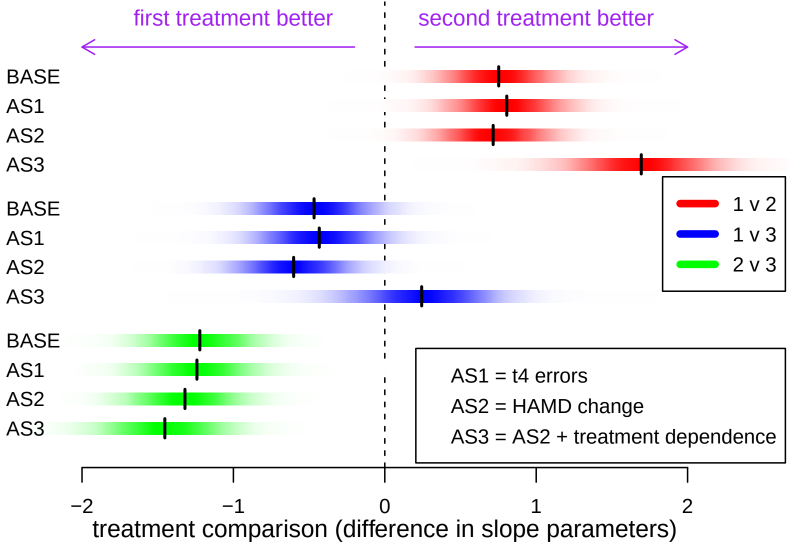
HAMD example: assumption sensitivity (step 4)

There are many options, including but not limited to the following

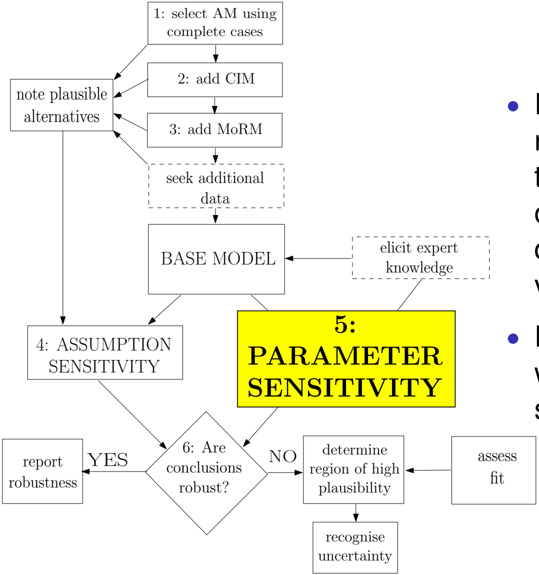
- Analysis model:
  - use a t_4 rather than normal error distribution
  - use an autoregressive model, AR(1), rather than random effects
  - include centre effects
  - allow for non-linearity by including a quadratic term
- Model of response missingness:
  - allow dependence on change in HAMD score
  - allow dependence on treatment
  - allow for non-linear relationship (piece-wise linear)

Example: results from assumption sensitivity

AS3: δ_{T1} 0.33 (0.01,0.62); δ_{T2} -0.41 (-0.63,-0.20); δ_{T3} -0.23 (-0.42,-0.02)



Step 5 - parameter sensitivity



- Parameter sensitivity involves running the base model with the MoRM parameters controlling the extent of the departure from MAR fixed to values in a plausible range
- Expert knowledge can help with setting the parameter sensitivity range

HAMD example: parameter sensitivity (step 5)

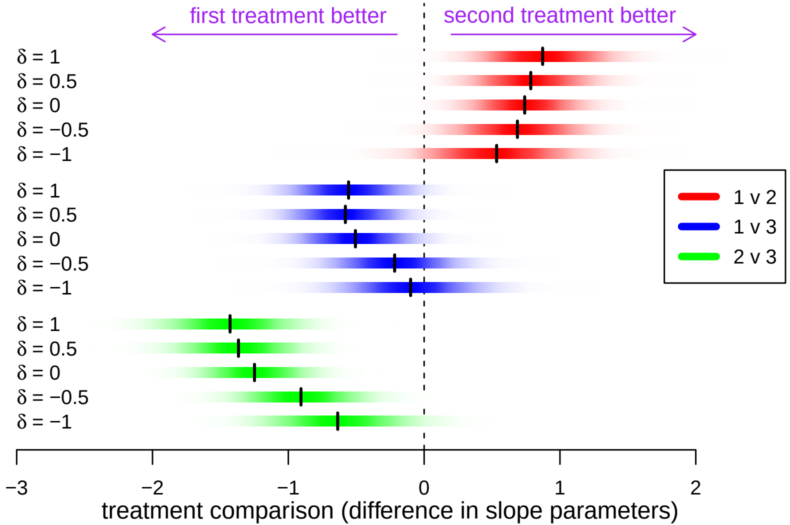
MoRM equation for Base Case

$$\text{logit}(p_{iw}) = \theta_0 + \delta(y_{iw} - \bar{y})$$

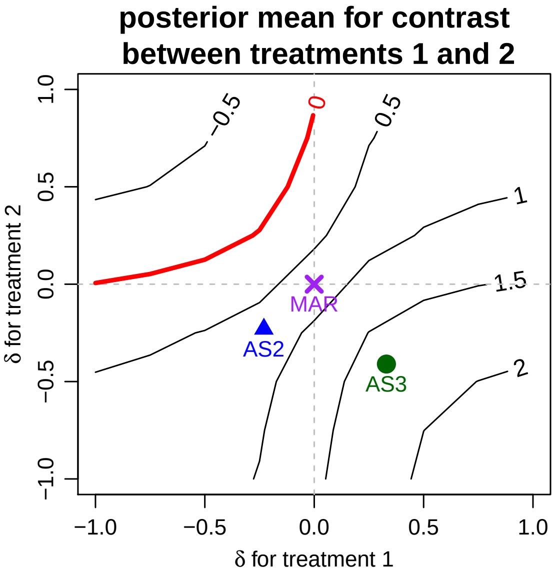
- δ estimated as 0.08 (0.04,0.11) in Base Case
- The value of δ controls the degree of departure from MAR missingness
- δ is difficult to estimate for a model with vague prior
- Run a series of models with δ fixed using point prior
  - 5 variants: values $\{-1, -0.5, 0, 0.5, 1\}$
  - δ corresponds to the log odds ratio of a missing response per point increase in HAMD score
  - range (-1,1) corresponds to assuming odds of non-response per unit increment in HAMD score ranges from ≈ 3 to $\frac{1}{3}$
  - $\delta = 0$ variant is equivalent to assuming the response is MAR

HAMD example: results from parameter sensitivity

MoRM equation: $\text{logit}(p_{iw}) = \theta_0 + \delta(y_{iw} - \bar{y})$

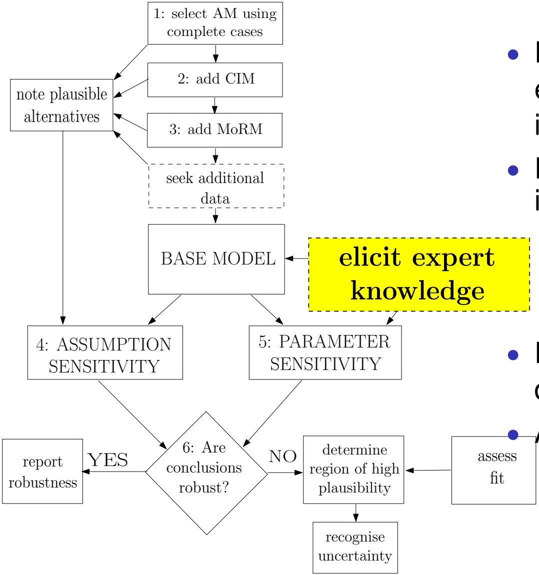


HAMD example: results from parameter sensitivity II



- Second parameter sensitivity based on AS3 (separate δ for each treatment)
- Investigate difference between treatments 1 and 2
- Fix δ_1 and δ_2 to values in range (-1,1)
- Fix $\delta_3 = -0.2$ (value suggested by AS3)

Optional step: elicit expert knowledge



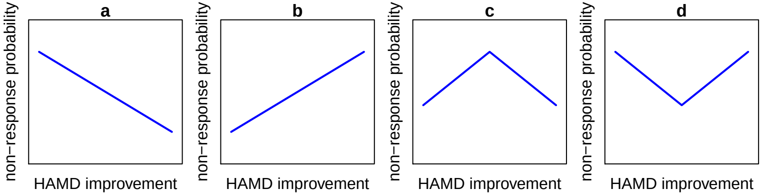
- Expert knowledge can be elicited and incorporated using informative priors
- Focus on parameters not well identified by the data
  - particularly those associated with the degree of departure from MAR
- Eliciting priors on parameters directly is difficult
- A better strategy is
  - elicit information about the probability of response
  - convert to informative priors

HAMD example: sample elicitation questions

Q1 Which variables do you think will help explain non-response?

A improvement in HAMD score since last visit (HAMD improvement)

Q2 What shape do you expect for the relationship between HAMD improvement and the probability of non-response?



A (d) ‘v’

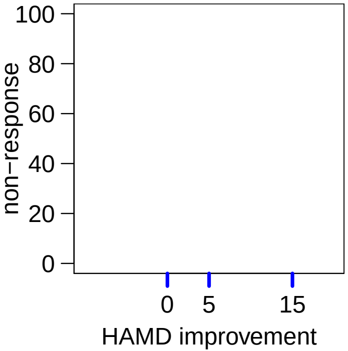
Q3 What value of HAMD improvement will minimise non-response?

A improvement of 5 points

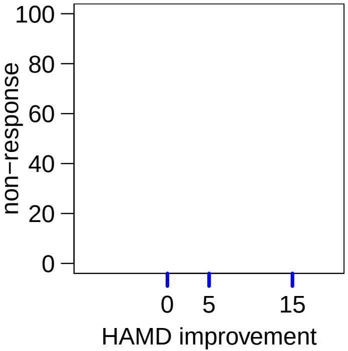
Q4 What other values of HAMD improvement should be used for elicitation?

A no improvement and improvement of 15 points

HAMD example: sample elicitation continued

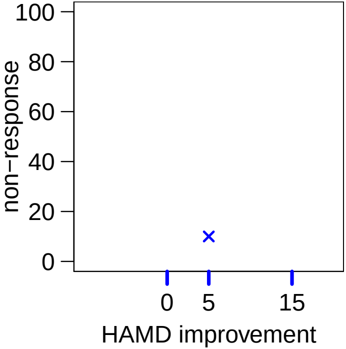


HAMD example: sample elicitation continued



Q5 Out of 100 subjects, how many would you expect not to respond if their HAMD score improves by 5?

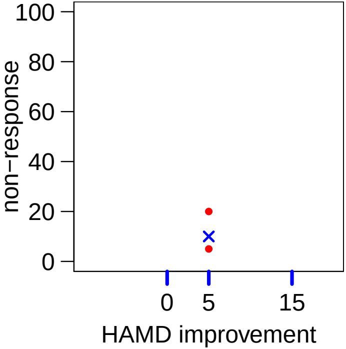
HAMD example: sample elicitation continued



Q5 Out of 100 subjects, how many would you expect not to respond if their HAMD score improves by 5?

A 10

HAMD example: sample elicitation continued

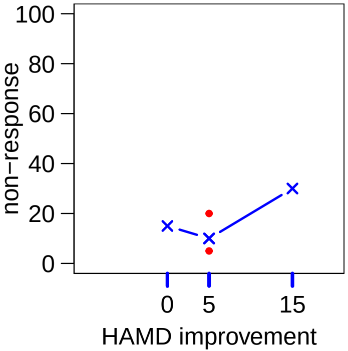


Q5 Out of 100 subjects, how many would you expect not to respond if their HAMD score improves by 5?

A 10

- Similar questions can be used to elicit uncertainty

HAMD example: sample elicitation continued

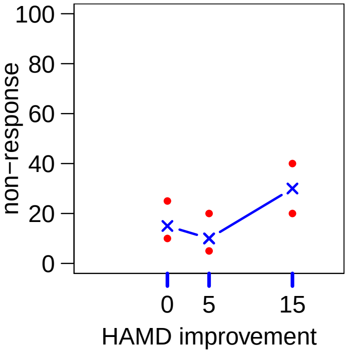


Q5 Out of 100 subjects, how many would you expect not to respond if their HAMD score improves by 5?

A 10

- Similar questions can be used to elicit uncertainty
- Elicit information at other values of HAMD improvement in the same way

HAMD example: sample elicitation continued

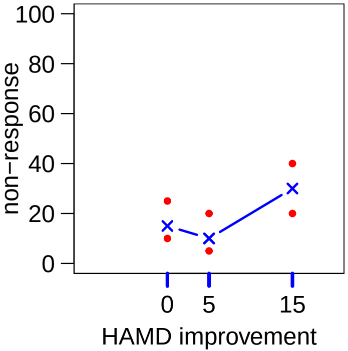


Q5 Out of 100 subjects, how many would you expect not to respond if their HAMD score improves by 5?

A 10

- Similar questions can be used to elicit uncertainty
- Elicit information at other values of HAMD improvement in the same way

HAMD example: sample elicitation continued



Q5 Out of 100 subjects, how many would you expect not to respond if their HAMD score improves by 5?

A 10

- Similar questions can be used to elicit uncertainty
- Elicit information at other values of HAMD improvement in the same way

- This information can be converted into informative prior on δ

$$\text{logit}(p_i) = \theta_0 + f(y_{iw}, y_{i(w-1)}; \delta)$$

- Can be incorporated as part of base case or used to inform sensitivity analysis

HAMD example: potential complications

- Probability of non-response may depend on other factors, e.g.
    - treatment
    - how depressed patient was previous week
- Multiple factors complicate elicitation
    - need to allow for an interaction between factors
    - ask questions of the form:

      Out of 100 subjects, how many would you expect not to respond if their HAMD score improves from 20 to 15?'

    - convert to joint rather than independent priors

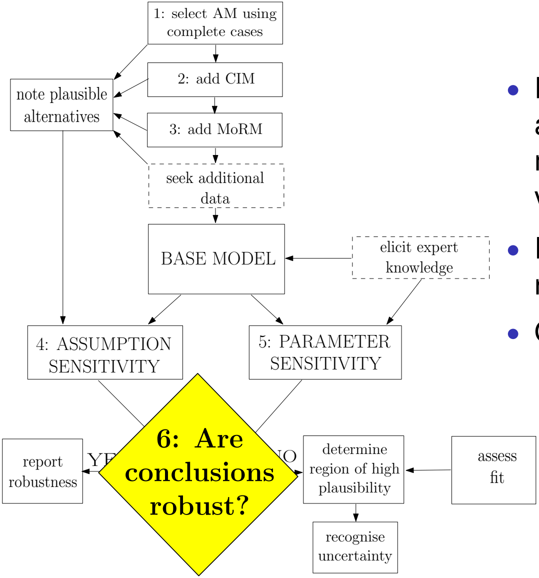
Elicitation: comment

Recall MoRM equation for Base Case

$$\text{logit}(p_{iw}) = \theta_0 + \delta(y_{iw} - \bar{y})$$

- The parameters associated with the response, δ, are identified by the parametric assumptions in
  - the analysis model (AM)
  - the model of response missingness (MoRM)
- This information is limited, so estimation difficulties can be encountered with vague priors
- If detailed elicitation is impractical, some information still helps
- In particular, ask questions that provide guidance on
  - variables to include in the model of response missingness
  - shapes of relationship between variables and probability of non-response
  - signs of parameters

Step 6 - determine robustness of conclusions



- Examine results of sensitivity analyses to establish how much the quantities of interest vary
- If the conclusions are robust, report this
- Otherwise
  - seek more information (optional steps)
  - determine a region of high plausibility
  - recognise uncertainty

Assessing model fit

- A model's fit to observed data can be assessed
- However, its fit to unobserved data cannot be assessed

  $\Rightarrow$ sensitivity analysis is essential
- DIC is routinely used by Bayesian statisticians to compare models, but
  - using DIC in the presence of missing data is not straightforward
  - the DIC automatically generated by WinBUGS is misleading (Mason et al., 2012a)
- Data not used in model estimation may be helpful in assessing model fit
  - compare model predictions against additional data

HAMD example: are conclusions robust? (step 6)

Consider comparison of Treatment 1 (T1) and Treatment 2 (T2)

- Base case - strong evidence T2 is more effective than T1
- Assumption sensitivity suggests strength of effect is uncertain
- In particular, AS3 suggests larger difference between T2 and T1
  - T1: $\delta > 0 \Rightarrow$ patient less likely to dropout if treatment effective
  - T2: $\delta < 0 \Rightarrow$ patient more likely to dropout if treatment effective
  - is this plausible?
  - different side-effects associated with each treatment?
  - are side-effect data available?
- Parameter sensitivity analysis examined AS3 further
  - ordering is only reversed if signs of δ are switched
  - if implausible, conclude treatment ordering robust to plausible sensitivities
  - otherwise report assumptions required to reverse ordering

Adaptions and extensions

- There are situations where it may be necessary to adapt this strategy
- Step 2 can be elaborated to allow MNAR covariates
- Steps 3 and 5 can be omitted if informative missingness in the response is implausible
- Could distinguish between different types of non-response
  - set up a missingness indicator with separate categories for each type of non-response
  - model using multinomial regression
- Bayesian models have the advantage of being fully coherent, but with large datasets or large numbers of covariates with missingness may be computationally challenging to fit

What does the Bayesian approach offer for missing data problems?

Bayesian methods are probably the most powerful and most general methods for dealing with missing data

What does the Bayesian approach offer for missing data problems?

Bayesian methods are probably the most powerful and most general methods for dealing with missing data

- Naturally accommodate missing data without requiring new techniques for inference

What does the Bayesian approach offer for missing data problems?

Bayesian methods are probably the most powerful and most general methods for dealing with missing data

- Naturally accommodate missing data without requiring new techniques for inference
- Bayesian framework is well suited to building complex models by linking smaller sub-models into a coherent joint model for the full data

What does the Bayesian approach offer for missing data problems?

Bayesian methods are probably the most powerful and most general methods for dealing with missing data

- Naturally accommodate missing data without requiring new techniques for inference
- Bayesian framework is well suited to building complex models by linking smaller sub-models into a coherent joint model for the full data
- Bayesian approach lends itself naturally to sensitivity analysis through different choices of prior distributions encoding assumptions about the missing data process

What does the Bayesian approach offer for missing data problems?

Bayesian methods are probably the most powerful and most general methods for dealing with missing data

- Naturally accommodate missing data without requiring new techniques for inference
- Bayesian framework is well suited to building complex models by linking smaller sub-models into a coherent joint model for the full data
- Bayesian approach lends itself naturally to sensitivity analysis through different choices of prior distributions encoding assumptions about the missing data process
- Offers possibility of including informative prior information about missing data process

What does the Bayesian approach offer for missing data problems?

Bayesian methods are probably the most powerful and most general methods for dealing with missing data

- Naturally accommodate missing data without requiring new techniques for inference
- Bayesian framework is well suited to building complex models by linking smaller sub-models into a coherent joint model for the full data
- Bayesian approach lends itself naturally to sensitivity analysis through different choices of prior distributions encoding assumptions about the missing data process
- Offers possibility of including informative prior information about missing data process
- But models can become computationally challenging...

Acknowledgements and References

- Thanks to Sylvia Richardson

- Funding by ESRC: the BIAS project (PI N Best), based at Imperial College, London

`www.bias-project.org.uk/research.htm`

- Daniels, M. J. and Hogan, J. W. (2008).
Missing Data In Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis. Chapman & Hall.

- Diggle, P. and Kenward, M. G. (1994).
Informative Drop-out in Longitudinal Data Analysis (with discussion).
Journal of the Royal Statistical Society, Series C (Applied Statistics), **43**, (1), 49–93.

- Mason, A., Richardson, S., and Best, N. (2012*a*).
Two-pronged strategy for using DIC to compare selection models with non-ignorable missing responses.
Bayesian Analysis, **7**, (1), 109–46.

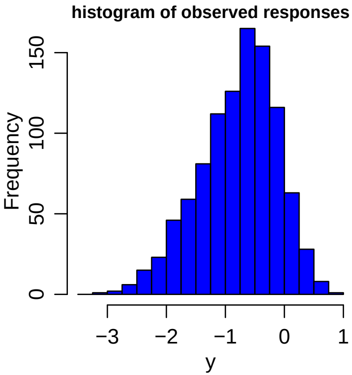
- Mason, A., Richardson, S., Plewis, I., and Best, N. (2012*b*).
Strategy for Modelling Nonrandom Missing Data Mechanisms in Observational Studies Using Bayesian Methods.
Journal of Official Statistics, **28**, (2), 279–302.

How the AM 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 MoM,

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



- If we assume the AM follows a normal distribution, $y_i \sim \mathrm{N}(\mu_i, \sigma^2)$
  - must fill in the right tail $\Rightarrow \delta > 0$
- If we assume the AM follows a skew-normal distribution
  - $\Rightarrow \delta = 0$

Summary of required sub-models for a Bayesian analysis

Type of Variable with Missing Values	Missingness Type	Analysis Model	Covariate Imputation Model	Missing Mechanism Model
response	ignorable	✓	✗	✗
response	non-ignorable	✓	✗	✓
covariate	ignorable	✓	✓	✗
covariate	non-ignorable	✓	✓	✓