PATTERN MIXTURE MODELS FOR MISSING DATA

Mike Kenward

London School of Hygiene and Tropical Medicine

Talk at the University of Turku,

April 10th 2012

CONTENTS

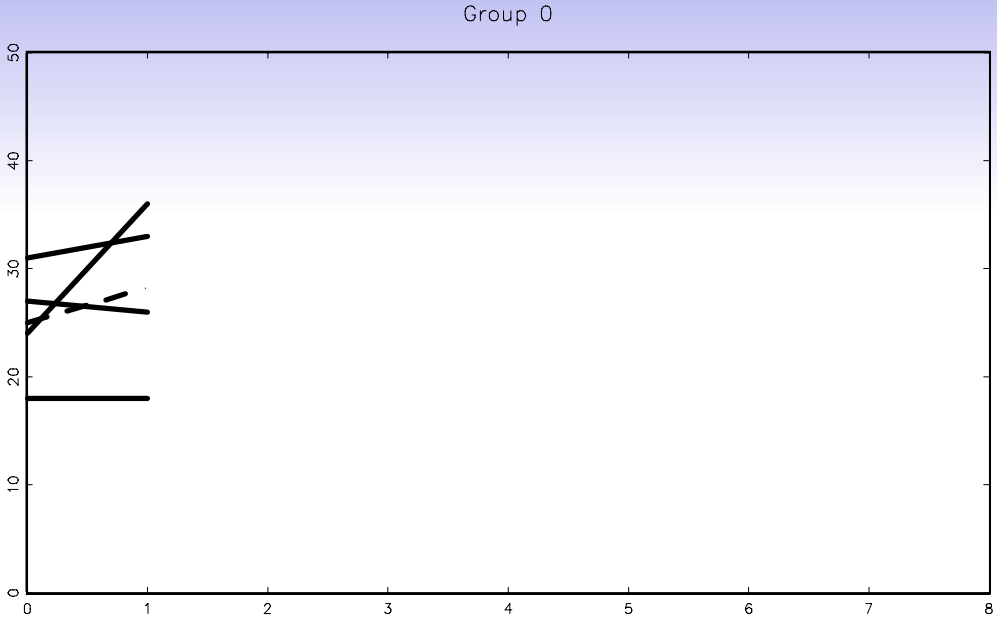
(1) EXAMPLES

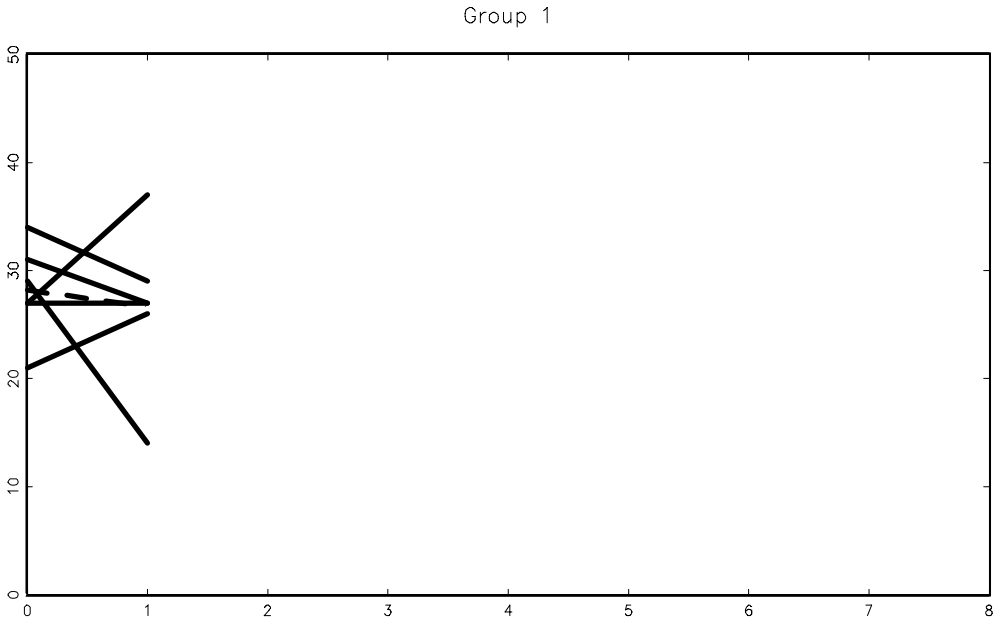
EXAMPLE 1: a randomized trial to compare the efficacy of two treatments for patients suffering from depression following stroke.

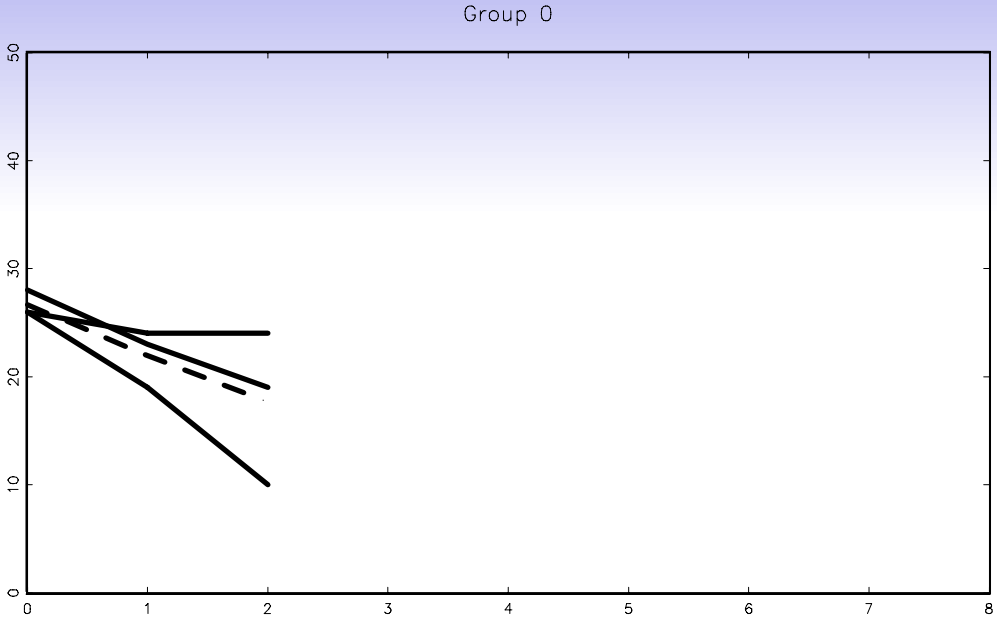
- Group sizes: treatment 0, 114; treatment 1, 111.
- Response: MADRS score.
- Measurements made at baseline and weeks 1, 2, 4, 6 and 8.
- Primary endpoint: difference in group means at week 8.

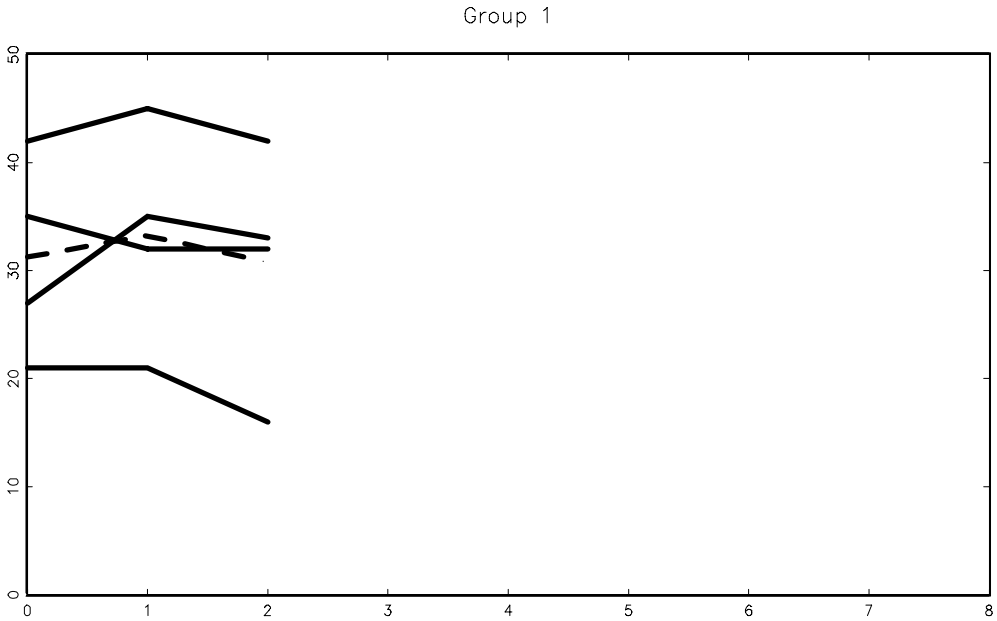
Pattern					Number
Group 0					
×	×	×	×	×	91
×	×	×	×	·	5
×	×	×	·	·	6
×	×	·	·	·	3
×	·	·	·	·	4
·	·	·	·	·	3
·	×	×	×	×	1
×	·	×	×	×	1
×	·	×	·	·	1
×	×	×	·	×	2

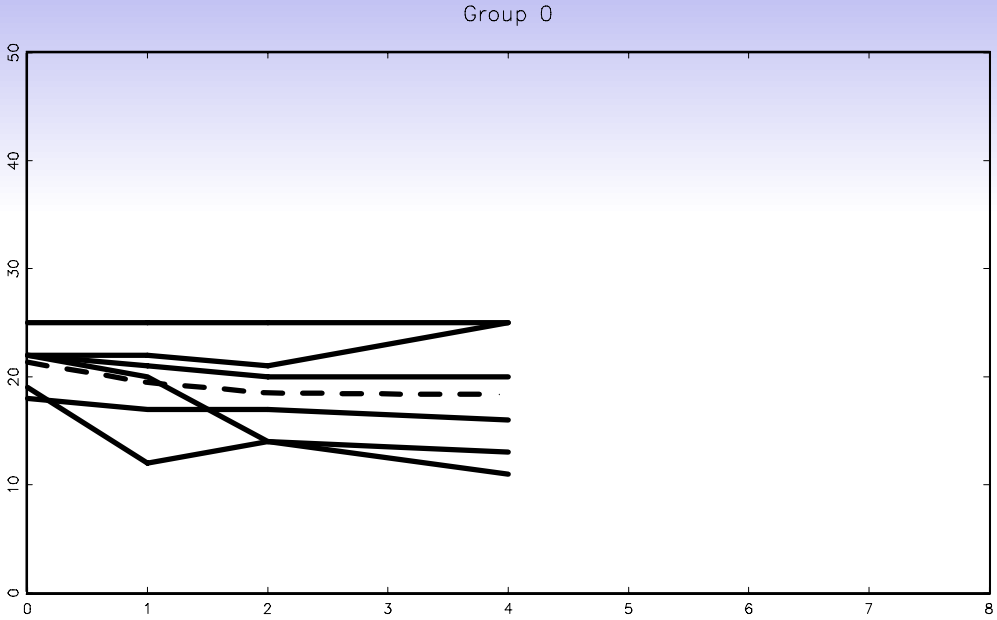
Pattern					Number
Group 1					
×	×	×	×	×	92
×	×	×	×	·	3
×	×	×	·	·	3
×	×	·	·	·	4
×	·	·	·	·	6
·	·	·	·	·	1
×	×	×	·	×	3

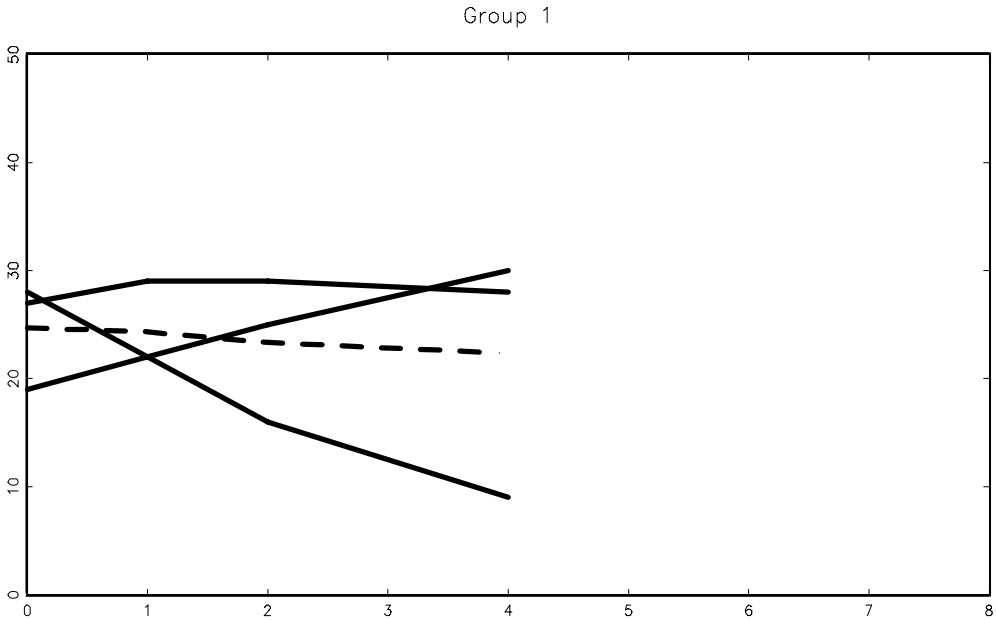

Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

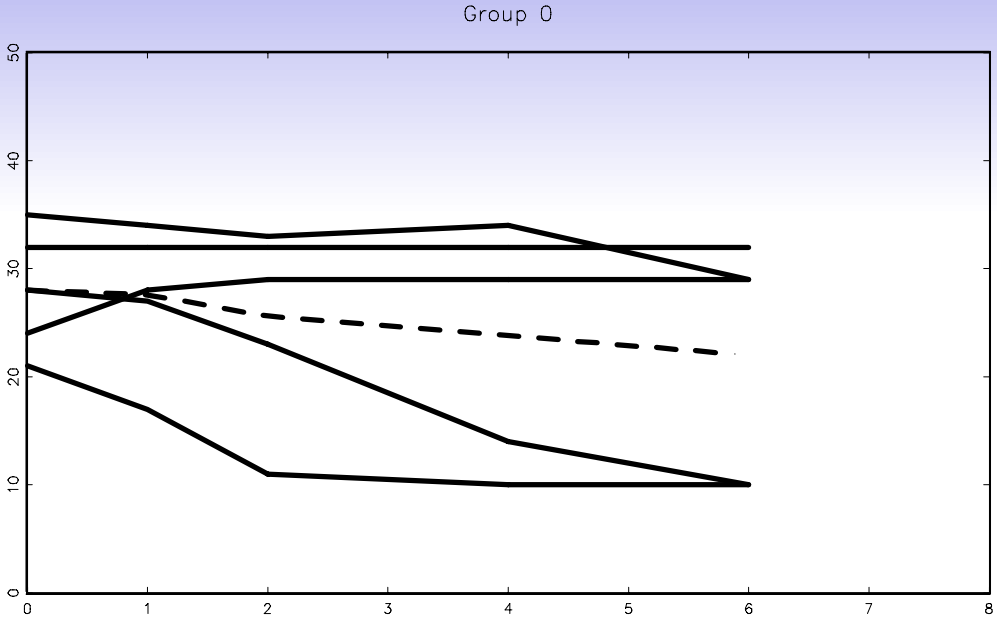

Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

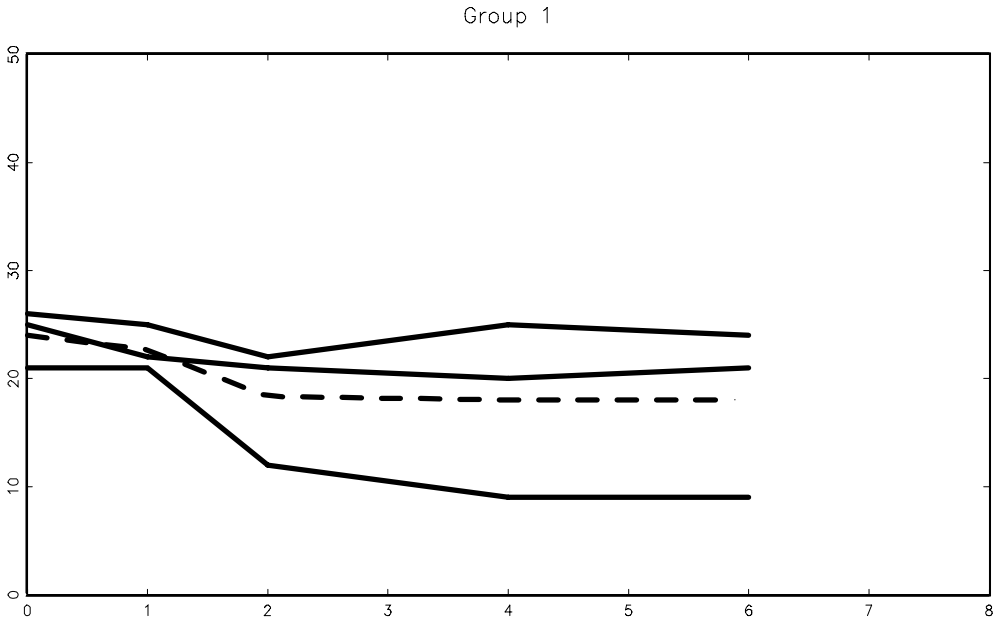

Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

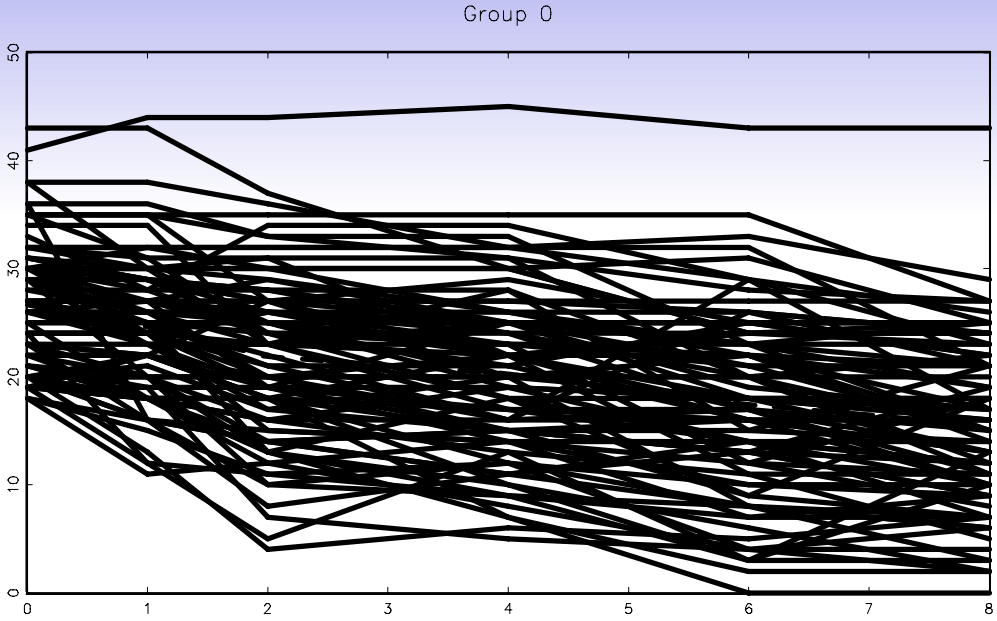

Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

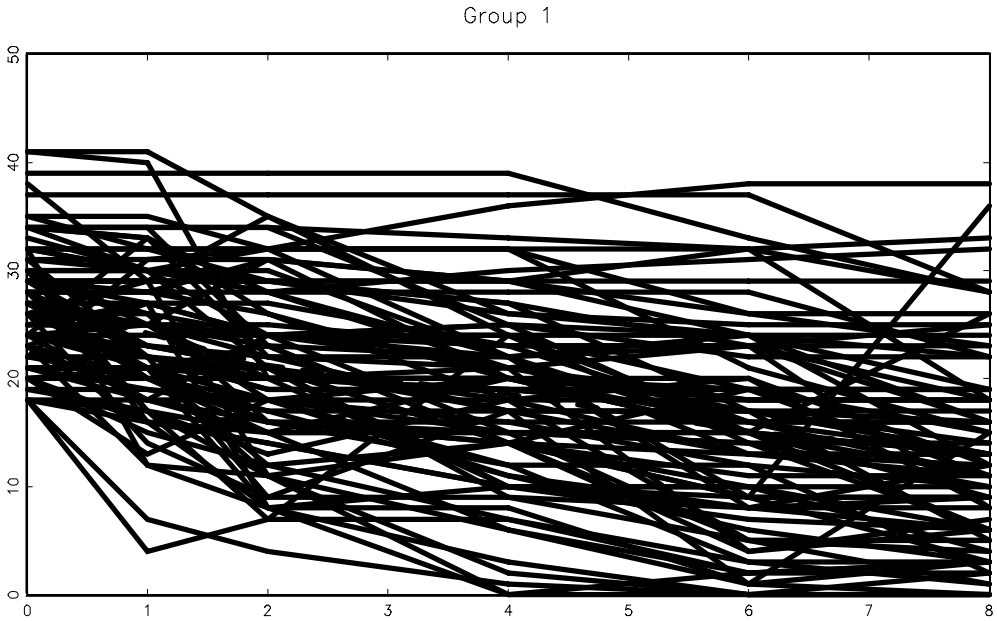

Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

EXAMPLE 2: Asthma study

- Aim: to assess the efficacy and safety of budesonide, a second-generation glucocorticosteroid, on patients with chronic asthma.
- 473 patients with chronic asthma were enrolled in the 12-week randomized, double-blind, multi-centre parallel-group trial, which compared the effect of a daily dose of 200, 400, 800 or 1600 mcg of budesonide with placebo.
- In the following, data will be used from the placebo and lowest active dose (200 mcg) arms.
- After randomization, patients were asked to attend the clinic at 2, 4, 8 and 12 weeks.

- The primary outcome is forced expiratory volume in 1 second (FEV1), recorded at each clinic visit.

- 91 patients were randomized to the placebo arm, of whom 38 completed.

  90 were randomized to the lowest dose active arm, of whom 72 completed.

(2) MODELLING INCOMPLETE DATA

Some notation and definitions

- *All response/outcome data*, whether observed or not:

$$\mathbf{Y} = \{\mathbf{Y}_O, \mathbf{Y}_M\}.$$

  $\mathbf{Y}_O$: observed, $\mathbf{Y}_M$: missing.

- *Covariates:*

$$\mathbf{X} = \{\mathbf{X}_O, \mathbf{X}_M\}.$$

  $\mathbf{X}_O$: observed, $\mathbf{X}_M$: missing.

- Depending on the context these may all refer to one unit, or to an entire dataset.

When there is no need to distinguish between outcome variables and covariates we can use:

$$\begin{aligned} \mathbf{Z} &= \{\mathbf{Y}, \mathbf{X}\} \\ \mathbf{Z}_O &= \{\mathbf{Y}_O, \mathbf{X}_O\} \\ \mathbf{Z}_M &= \{\mathbf{Y}_M, \mathbf{X}_M\} \end{aligned}$$

Missing value indicator:

Corresponding to every element of $\mathbf{Y}$, there is an R:

$$R = \begin{cases} 1 & \text{if observed} \\ 0 & \text{if missing} \end{cases}$$

with $\mathbf{R} = \{R\}$.

Missing Value Mechanism

$$\mathrm{P}(\mathbf{R} \mid \mathbf{Y}, \mathbf{X})$$

We hare here mainly concerned with dropout/withdrawal:

Dropout indicator:

When we confine missing values to dropout (with T repeated measurements), we can define a dropout indicator D, the first time a value is missing and $D = T + 1$ indicates no dropout.

Dropout mechanism:

$$\mathrm{P}(D \mid \mathbf{Y}, \mathbf{X})$$

Goal: make inferences about the substantive model parameters ($\boldsymbol{\theta}$) from

$$f(\mathbf{Y} \mid \mathbf{X}; \boldsymbol{\theta})$$

using $\mathbf{Y}_O$ and $\mathbf{X}_O$ (and $\mathbf{R}$).

How are these connected?

$$f(\mathbf{Y}_O \mid \mathbf{X}_O) = \int \int f(\mathbf{Y}_O, \mathbf{Y}_M \mid \mathbf{X}_O, \mathbf{X}_M) f(\mathbf{X}_M \mid \mathbf{X}_O) \mathbf{dX}_M \mathbf{dY}_M.$$

Estimation:

Suppose we have an unbiased (vector valued) estimating equation for the complete data

$$\mathbf{U}(\mathbf{Y}; \mathbf{X}; \hat{\theta}) = \sum_{i=1}^{n} \mathbf{U}_i(\mathbf{Y}_i; \mathbf{X}_i; \hat{\theta}) = \mathbf{0}.$$

Then with missing responses the following estimating equation remains unbiased (hence consistent for $\boldsymbol{\theta}$)

$$\sum_{i=1}^{n} \{R_i^* \mathbf{U}_i + (1 - R_i^*) \mathrm{E}_g(\mathbf{U}_i)\} = \mathbf{0}.$$

$[R_i^* = 1 \Rightarrow \text{all } R_i = 1 \text{ otherwise } R_i^* = 0]$

where the expectation is taken over the joint distribution of the missing data $\mathbf{Z}_M$ conditional on the observed data: $g(\mathbf{Z}_M \mid \mathbf{Z}_O, \mathbf{R})$.

This will be the score equation in the likelihood setting.

This expectation can be calculated in many ways *e.g.*

- directly or indirectly;
- analytically or numerically;
- in Bayesian a framework with MCMC;
- using hybrid methods (*e.g.* stochastic EM);
- using multiple imputation;
- ...

But, in general it requires an explicit missing value (or dropout) mechanism.

Missing value mechanisms (in a dropout setting).

[Rubin DB (1976)]

Note: I am using these terms in a frequentist way (stronger than Rubin's original definitions).

Definition: Missing Completely at Random (MCAR)

- The probability of dropout does not depend on the outcome variables, whether observed or not.
- The subjects remaining at the end of the trial constitute a random sample of those originally enrolled.
- Hence the data can be analysed as though the withdrawal were pre-planned.
- The completers only analysis is "valid".

Definition: Missing at Random (MAR)

A subject's **history** at dropout: past response, baseline covariates, treatment group (all observed).

- MAR: the probability of a subject dropping out is conditionally independent of future observations, given the history.
- Equivalently:

  MAR: the future statistical behaviour of a subject, conditional on the history, is the same whether the subject drops out or not in the future.

- Under MAR, likelihood based analyses of the outcome only are valid (the actual dropout mechanism can be ignored).
- **Missing Not at Random (MNAR)**: neither MCAR nor MAR.

- Under MAR, **R** can be dropped from the expectation in:

$$\sum_{i=1}^{n} \{R_i^* \mathbf{U}_i + (1 - R_i^*) \mathrm{E}_g(\mathbf{U}_i)\} = \mathbf{0}.$$

- *i.e.* we don't need to model the missing value (dropout) mechanism.
- **But:** the data under analysis **cannot** distinguish between MAR and MNAR.
- Any such distinction based on the observed data is ultimately **assumption driven**.

So, the inferential issues surrounding missing values centre on **assumptions**:

- There can be *no* single "correct" analysis when data are missing, hence no single "correct" answer.
- We need to understand what additional assumptions are associated with any particular analyses.
- We want these assumptions to be as transparent as possible.
- We are then in a position to decide what assumptions are reasonable and sensible in any given setting.

  Contextual/subject matter information is central to this.

- Ideally we should explore the sensitivity of inferences to relevant aspects of the chosen assumptions.

(3) THE PATTERN MIXTURE REPRESENTATION

Modelling frameworks for incomplete data

One goal: formulation of models with non-random missingness with transparent assumptions.

- Selection Models
- Pattern Mixture Models
- Shared Parameter Models

The starting point for all of these is a joint model for

$$f(\mathbf{Y}, \mathbf{R} \mid \mathbf{X}; \xi)$$

Selection Models

These follow naturally from the substantive model:

$$f(\mathbf{Y} \mid \mathbf{X}; \theta_S)$$

and the missing value mechanism:

$$\mathrm{P}(\mathbf{R} \mid \mathbf{Y}, \mathbf{X}; \gamma_S).$$

giving

$$f(\mathbf{Y}, \mathbf{R} \mid \mathbf{X}; \xi) = f(\mathbf{Y}, \mid \mathbf{X}; \theta_S)\mathrm{P}(\mathbf{R} \mid \mathbf{Y}, \mathbf{X}; \gamma_S).$$

These first originated in econometrics in the form of Heckman's (1976) selection Model.

They were formulated for longitudinal data in a biostatistical setting in Diggle and Kenward (1994).

Rubin's classification was originally expressed in terms of selection models.

Pattern Mixture Models

These use the reverse partitioning of the joint distribution:

$$f(\mathbf{Y} \mid \mathbf{X}, \mathbf{R}; \boldsymbol{\theta}_P)$$

and

$$\mathrm{P}(\mathbf{R} \mid \mathbf{X}; \boldsymbol{\gamma}_P).$$

giving

$$f(\mathbf{Y}, \mathbf{R} \mid \mathbf{X}; \boldsymbol{\xi}) = f(\mathbf{Y}, \mid \mathbf{X}, \mathbf{R}; \boldsymbol{\theta}_P)\mathrm{P}(\mathbf{R} \mid \mathbf{X}; \boldsymbol{\gamma}_P).$$

Shared Parameter Models

Here one or more latent variables $\mathbf{U}$ are introduced, and the two components are defined in terms of these

$$f(\mathbf{Y} \mid \mathbf{X}, \mathbf{U}; \boldsymbol{\theta}_U)$$

and

$$\mathrm{P}(\mathbf{R} \mid \mathbf{X}, \mathbf{U}; \boldsymbol{\gamma}_U).$$

and are usually assumed to be conditionally independent given $\mathbf{U}$. Hence the joint model is

$$f(\mathbf{Y}, \mathbf{R} \mid \mathbf{X}; \boldsymbol{\xi}) = \int f(\mathbf{Y}, \mid \mathbf{X}, \mathbf{U}, \boldsymbol{\theta}_U; \boldsymbol{\theta}_P)\mathrm{P}(\mathbf{R} \mid \mathbf{X}, \mathbf{U}; \boldsymbol{\gamma}_U) f(\mathbf{U} \mid \boldsymbol{\tau}) d\mathbf{U}.$$

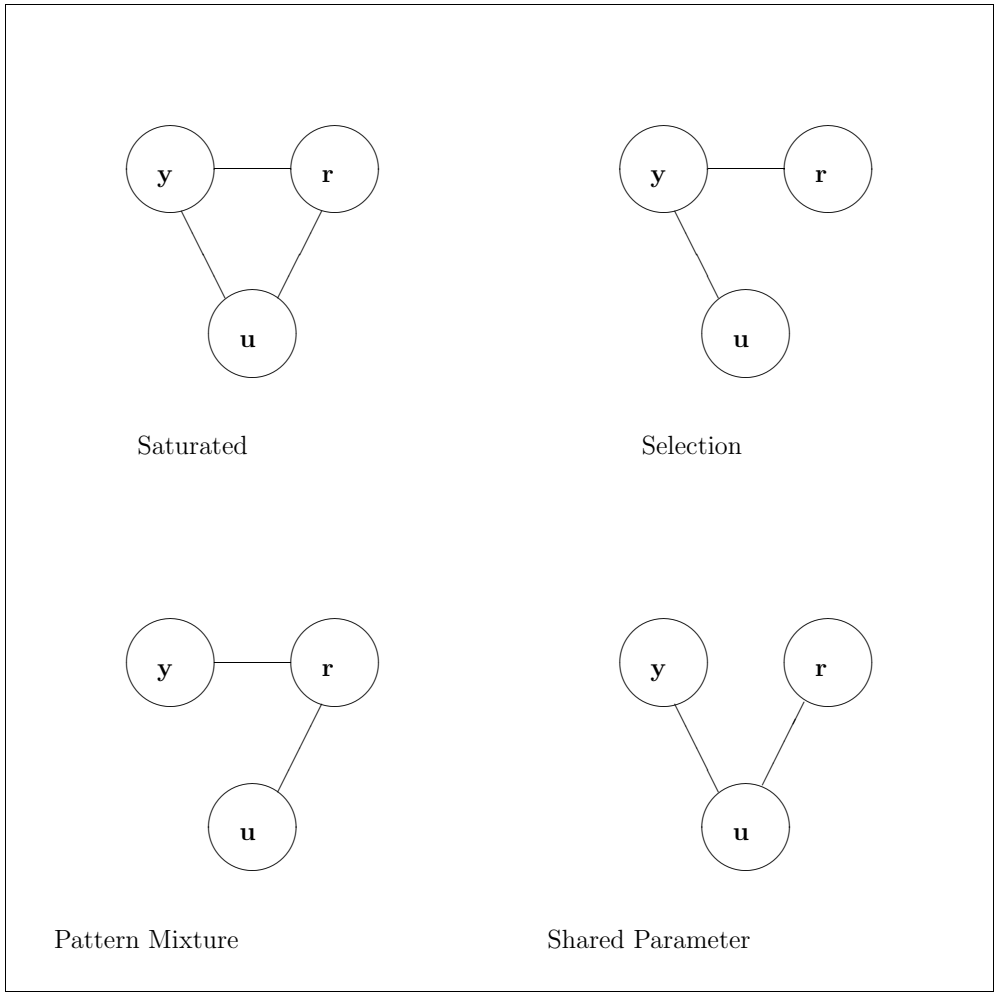
Shared parameter models fit very naturally within the **Structural Equation** framework – and have received a lot of attention in the social science setting.

To me, it is not obvious from these exactly what form is being implied either for the future behaviour of dropouts or for the selection process.

Perhaps useful for some sensitivity analyses?

A recent example in a clinical trial setting: Kenward and Rosenkranz (2011).

Conceptually all three can be represented in a common framework as follows:



Recall, the pattern mixture model takes the generic form (dropping the subscript P on the parameters, and the covariates $\mathbf{X}$ from the dropout mechanism):

$$f(\mathbf{Y}, \mathbf{R} \mid \mathbf{X}; \boldsymbol{\xi}) = f(\mathbf{Y}, \mid \mathbf{X}, \mathbf{R}; \boldsymbol{\theta})\mathrm{P}(\mathbf{R}; \boldsymbol{\gamma}).$$

There is, in principal, one outcome model

$$f(\mathbf{Y} \mid \mathbf{X}, \mathbf{R} = \mathbf{r}; \boldsymbol{\theta})$$

for each possible missing value pattern $\mathbf{r}$

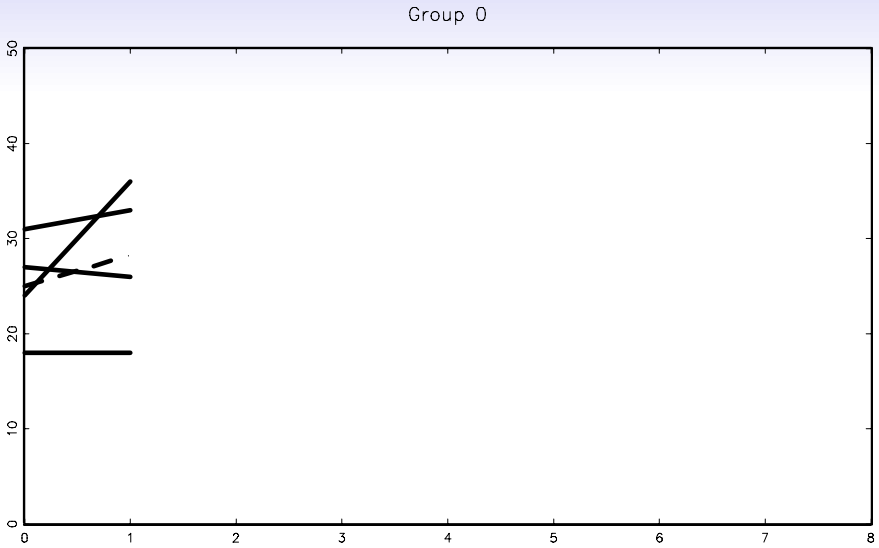
This is too unwieldy for general use.

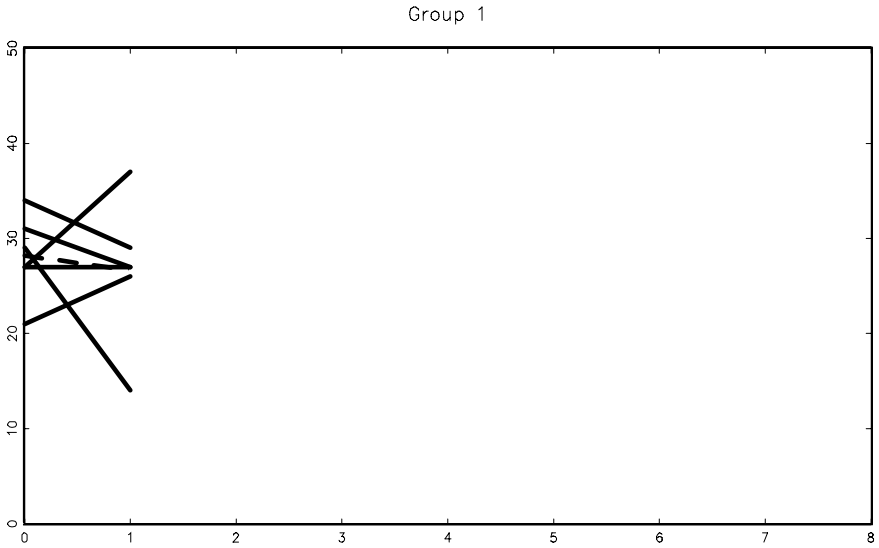
But, in the dropout setting, there are only T such missing value patterns for T times.

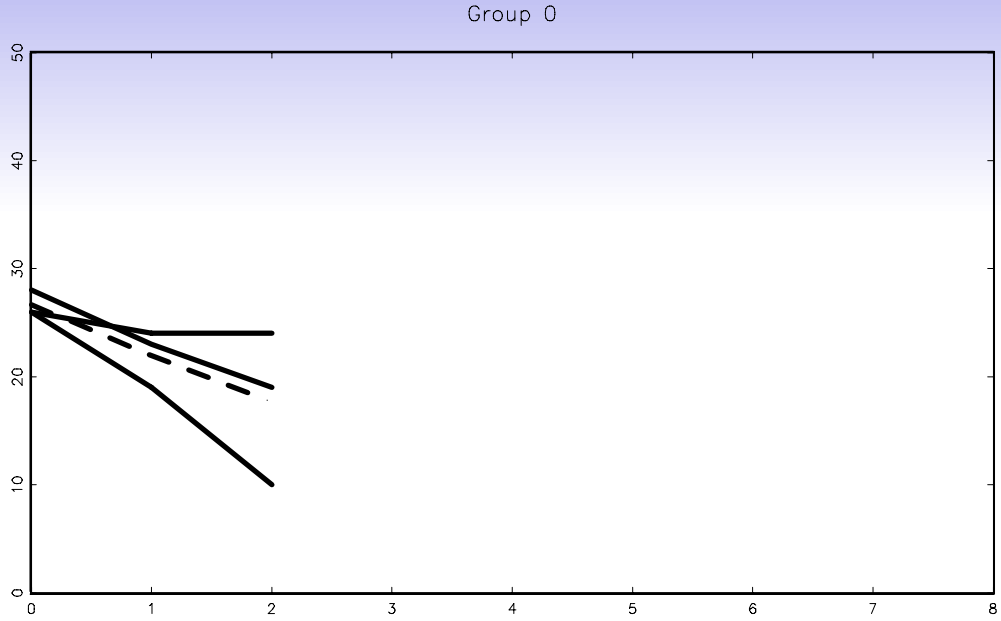
The model can then be rewritten

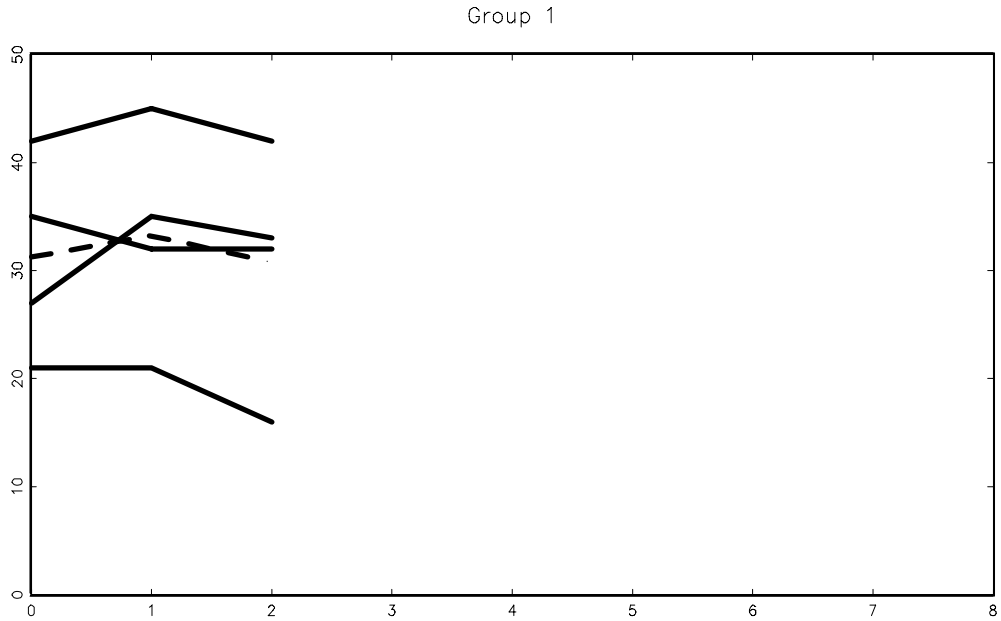
$$f(\mathbf{Y}, D = d \mid \mathbf{X}; \boldsymbol{\xi}) = f_d(\mathbf{Y} \mid \mathbf{X}; \boldsymbol{\theta})\mathrm{P}(D = d; \boldsymbol{\gamma}).$$

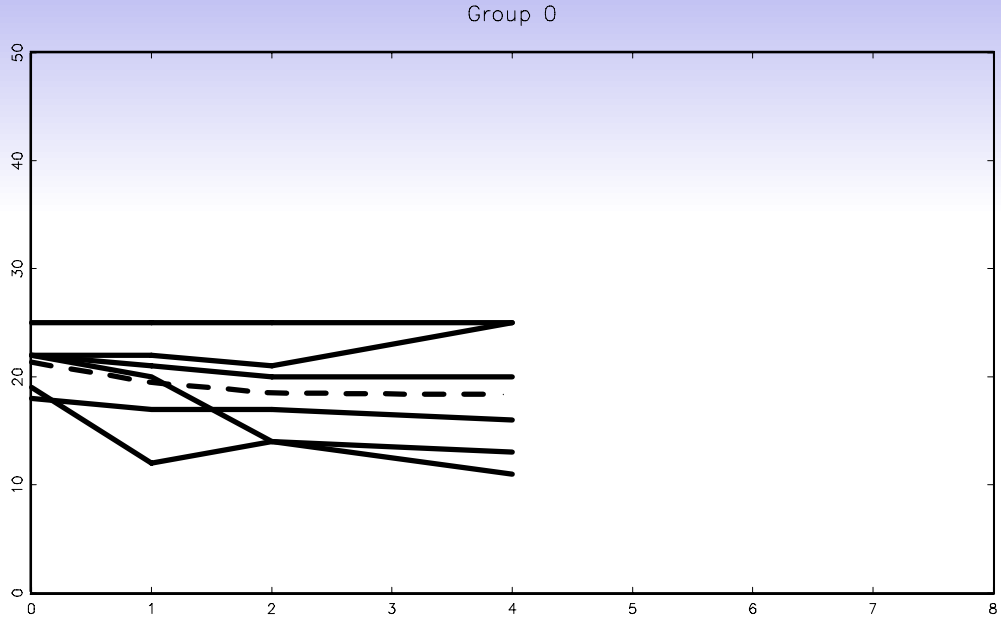
e.g. Example 1, potentially different models are used for each of the dropout sets:

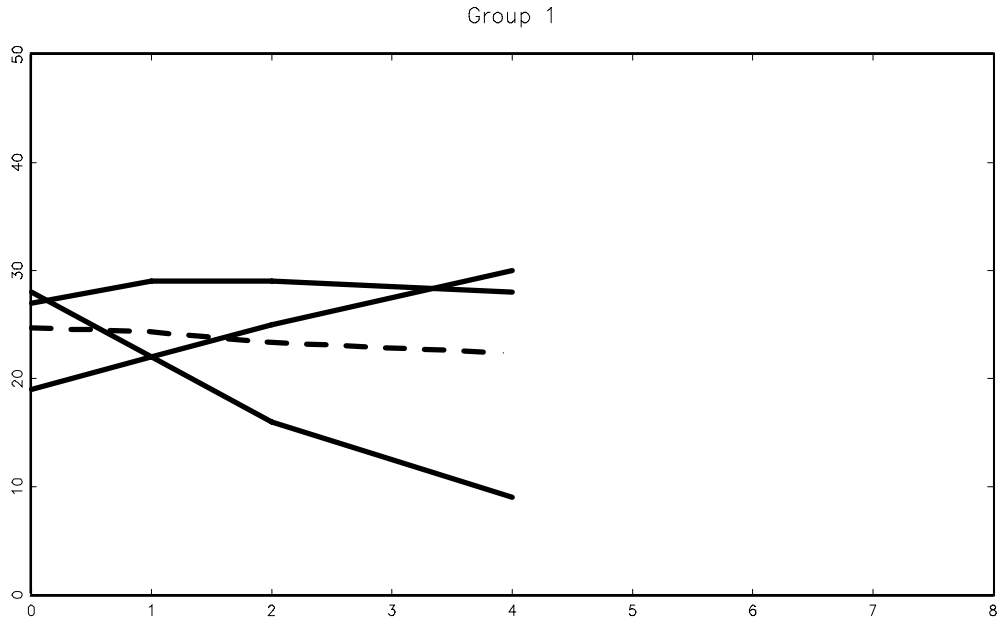


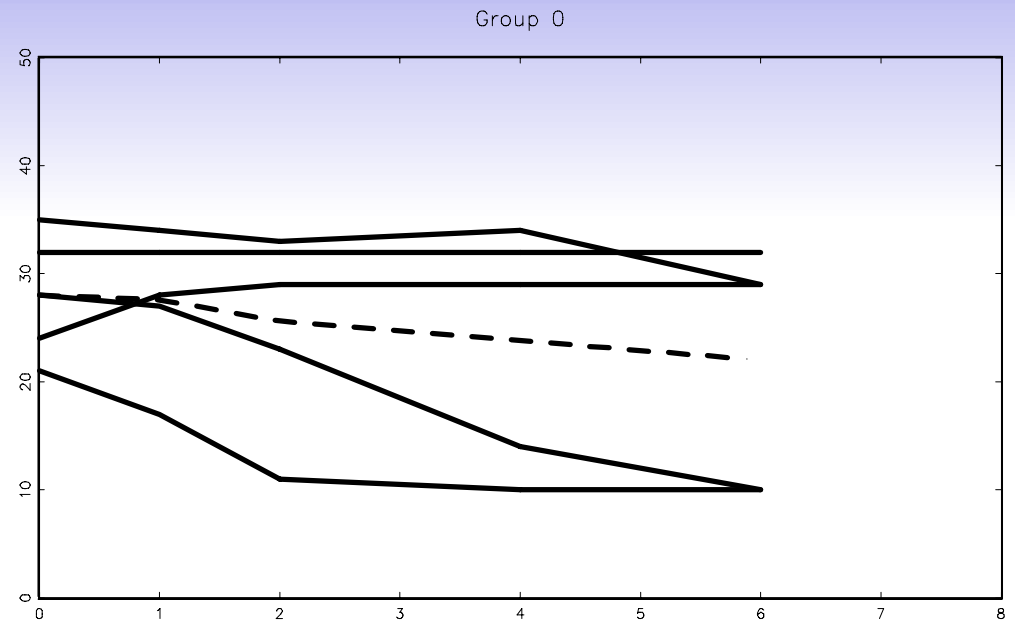


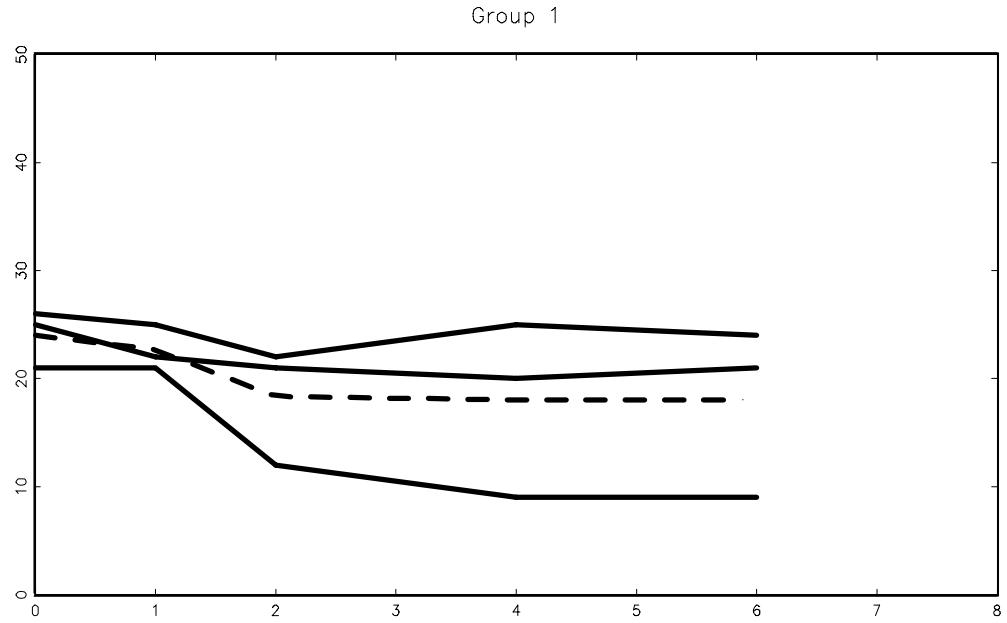

Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

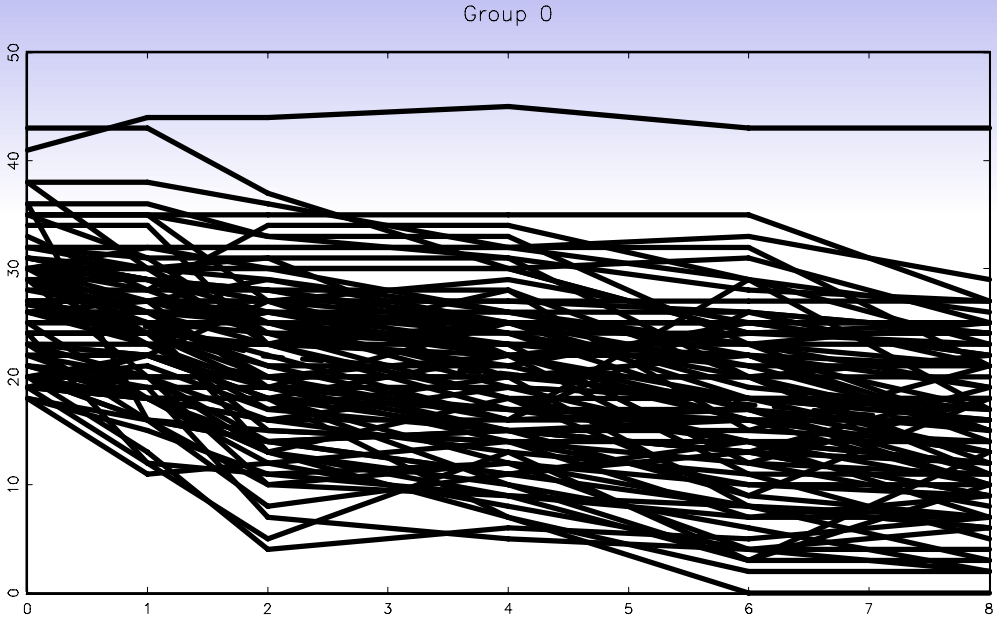

Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

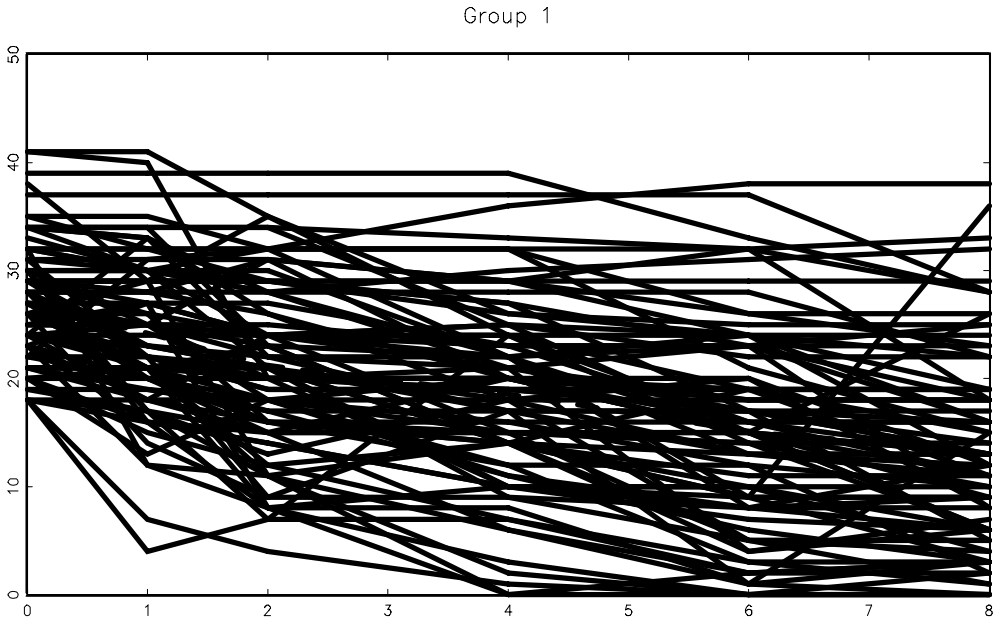

Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 0
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8


Group 1
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8

Setting $\pi_d = \mathrm{P}(D = d; \gamma)$ the substantive model can be written

$$f(\mathbf{Y} \mid \mathbf{X}; \xi) = \sum_{d=2}^{T+1} \pi_d f_d(\mathbf{Y}, \mid \mathbf{X}; \theta).$$

Notes:

- Hence the name *Pattern Mixture Model (PMM)*
- These were first introduced by Little (*e.g.* 1994)
- A thorough development is given in Molenberghs and Kenward (2007, Chapter 16)
- PMM'as are rarely distribution preserving.

So, in pattern-mixture model, there is a (potentially) different model for each dropout group.

Without additional modelling restrictions we cannot estimate such a model for all groups (even supposing there are sufficient numbers in each group).

There are two main classes of approach to this problem.

(1) Use a model than can be estimated within each missing value group.

- *e.g.* use a regression model that allows different parameters for each dropout group.
- This can be done using linear mixed models for example, with terms allowed to differ among dropout groups. [*e.g.* Hogan and Laird, (1997)]
- **But:** it implies very unrealistic polynomial extrapolation.
- Hence it is not as insensitive to modelling assumptions as some have suggested, as shown by *e.g.* Demirtas and Schafer (2003).

(2) "Borrow" information from later-dropout groups.

- What should be borrowed, and why?
- MAR is a special case of this.
- Later on we will see how certain forms of "borrowing" can be useful for constructing sensitivity analyses using multiple imputation.

(4) IDENTIFYING RESTRICTIONS AND NON-FUTURE DEPENDENCE.

A simple example of route (2): three time points, saturated Gaussian model.

$$\mathbf{Y} \mid r \sim \mathrm{N}(\boldsymbol{\mu}^{(d)}; \boldsymbol{\Sigma}^{(d)}), \tag{1}$$

for $d = 1, 2, 3$.
The following *conditional* moments cannot be estimated from the data:

$$\boldsymbol{\mu}^{(1)}_{2,3|1}, \quad \boldsymbol{\Sigma}^{(1)}_{2,3|1}$$

$$\mu^{(2)}_{3|1,2}, \quad \sigma^{(2)}_{3|1,2}$$

How should these be chosen?

Identifying Restrictions:

Consider all the observations from one individual, with dropout at time $t+1$:

$$
\begin{aligned}
&f(y_1, \cdots, y_T, d = t) = \\
&\quad f_t(y_1, \cdots, y_t) f_t(y_{t+1} | y_1, \cdots, y_t) \\
&\quad f_t(y_{t+2}, \cdots, y_T | y_1, \cdots, y_{t+1}) f(d = t).
\end{aligned}
$$

Complete Case restrictions (CCMV)

Constrain the unidentifiable conditional distributions to be equal to those in the **completers** group ($d = T$):

$$f(y_t \mid y_1, \ldots, y_{t-1}, d = j) = f(y_t \mid y_1, \ldots, y_{t-1}, d = T),$$

for all $t \geq 2$, $j < t$.

Neighbouring restrictions (NCMV):

Constrain the unidentifiable conditional distributions to be equal to those in the next group up: $(d = t)$:

$$f(y_t \mid y_1, \ldots, y_{t-1}, d = j) = f(y_t \mid y_1, \ldots, y_{t-1}, d = t),$$

for all $t \geq 2$, $j < t$.

Available Case restrictions (ACMV):

Constrain the unidentifiable conditional distributions to be equal to those in the all the higher groups:

$$f(y_t \mid y_1, \ldots, y_{t-1}, d = j) = f(y_t \mid y_1, \ldots, y_{t-1}).$$

for all $t \geq 2$ and $j < t$.

This equivalent to MAR in the selection model (dropout) context: all terms are identifiable.

[Molenberghs *et al.* (1998)]

These classes (and others) are members of the *interior family*, for which

$$f(y_{t+1}|y_1, \cdots, y_t, d = t) = \sum_{j=t+1}^{T} \alpha_j f(y_{t+1}|y_1, \cdots, y_t, d = j).$$

- ACMV doesn't take us beyond an MAR analysis.
- The set of all MNAR pattern-mixture models is too large to be very useful.
- Arbitrary extensions can lead to strange behaviour in the selection framework:

  e.g. the probability of dropout depends on future values, even after conditioning on the past.

  [Although shared parameter models can imply this as well]

Non-Future Dependence restrictions (NFMV) [Kenward *et al.* (2003)].

Can we link properties of pattern-mixture models with the corresponding selection model?

In a selection model framework, non-future dependence implies:

$$\mathrm{P}(d = t|y_1, \cdots, y_T) = \mathrm{P}(d = t|y_1, \cdots, y_{t+1}).$$

It can be shown that this is equivalent in the pattern-mixture framework to the following NFMV restrictions:

$$f(y_t|y_1, \cdots, y_{t-1}, d = j) = f(y_t|y_1, \cdots, y_{t-1}, r \geq t-1),$$

for all $t \geq 2$ and all $j < t - 1$.

NFMV leaves one conditional distribution per incomplete pattern unidentified:

$$f(y_{t+1}|y_1, \cdots, y_t, d = t).$$

i.e. other words, the distribution of the 'current' unobserved measurement, given the previous ones, is unconstrained.

This implies that the the NFMV class contains members not contained in the *interior family*.

Conversely, the class excludes such mechanisms as CCMV and NCMV (neighbouring-case missing values), showing that there are members of the interior family that are not of non-future missing values type.

Finally, it can be shown that the intersection of the interior and NFMV is given by ACMV.

We can build forms of sensitivity analysis on this, by varying the distributions that are 'free' under the NFMV constraints.

This is effectively what Daniels and Hogan (2008) do in their book: they construct Bayesian sensitivity analyses through NFMV pattern mixture models.

On problem with this "constraint" approach to PMM's is that it can be difficult to link the assumptions being made in NMAR models to the substantive setting.

We now consider a more accessible route to the same problem of sensitivity analysis, but first a brief introduction to multiple imputation will be given.

- We are going to be using pattern-mixture models for sensitivity analysis.
- We will assume that the data generating mechanism follows one model, while the analysis (typically very simple) follows another.
- This is very convenient in MI: the imputation model contains the assumptions about the behaviour of dropouts, the substantive model remains the simple one defined in the protocol (*e.g.* ANCOVA for the final time point).

(5) MULTIPLE IMPUTATION (MI)

- Multiple imputation was developed by Rubin in a survey setting [Rubin (1987)].
- Consequently it has great strengths when dealing with large, complicated sets of data, with missing values on many covariates.
- More recently it has been put forward as a useful technique for clinical trials with missing data: it is important to be clear in this context about its advantages and disadvantages.
- It does have an important role, but in some settings merely reproduces approximately the (more efficient and simple) likelihood analyses.
- We need to distinguish carefully the *computational* and *conceptual* aspects of the method.

Computationally, it uses Bayesian draws from the conditional (posterior) distribution of the missing data given the observed (the **imputation** model) to do the necessary **averaging** (or integration).

Several (M) sets of data are "completed" using these draws.

Rubins's formulae are used to combine estimates and variances from the"completed" datasets (from fitting the **substantive** model), to produce the MI estimates and associated measures of precision.

Bayesian draws (called **proper** by Rubin) are **not** necessary for consistency of the resulting estimators.

They **are** necessary for the validity of Rubin's **variance** formula.

- When only response variables are missing, *e.g.* with longitudinal data, and the imputation model uses only variables in the substantive model, we gain nothing over the corresponding likelihood analysis.
- In fact MI is then slightly inefficient, and more cumbersome.
- We would expect to get very similar answers between the two types of analysis.
- This is often the case in a clinical trial *under MAR*.

However MI can be used in some circumstances when the imputation model is *uncongenial*, that is not strictly compatible with the substantive model.

.

It is important that all the structure in the substantive model is contained in the imputation model, but the latter can have additional structure and variables, for example.

In this latter case we expect MI to behave as it should, but with Rubin's variance formula being a little conservative.

This is how we will use it below for sensitivity analysis, by imputing from a non-random pattern mixture model.

(6) SENSITIVITY ANALYSIS FOR LONGITUDINAL CLINICAL TRIALS.

Carpenter, Roger, Kenward (2011) Relevant, accessible sensitivity analyses using multiple imputation.

Developed from the ideas in Little & Yao (1996) and

Kenward and Carpenter (2009)

We now return to the problem of handling missing data in longitudinal clinical trials.

Aims: de jure and de facto estimands.

Here are some possible scenarios:

- Completely lost to follow-up (CLF), no further information.
- CLF, but reason for withdrawal known.
- Withdrawal from trial, treatment known, no further measurements obtained,
- Withdrawal from trial, treatment known, future measurements obtained,
- ...

A distinction (for this talk only):

Withdrawal: A participant in a clinical trial discontinues prescribed treatment programme (deviates from protocol).

Dropout: A failure to measure the outcome of interest for a trial participant at this and subsequent visits.

Before continuing a vital distinction needs to be made:

- Is withdrawal/dropout defined to be part of the *outcome* or a *nuisance* to be accommodated, (*i.e.* one that prevents us measuring what we want to measure)?
- The former is very common when various form of simple *imputation* are used, with Last Observation Analysed perhaps the best known example.
- Such a procedure *defines away the missing data*, *i.e.* there are no missing data.
- In principle the subsequent analysis will be valid (other aspects being appropriate), but what is the *clinical meaning* of the subsequent conclusions?
- Such an approach runs the risk of making the classic analysis error: when in doubt, provide a valid answer to the wrong question.
- In the following I focus on the former: missingness as a nuisance.

- The starting point for the discussion of how to handle missing data must be *the aims of the analysis* [*Estimand* in the recent FDA commissioned AS report].
- This in turn depends crucially on the aims of the trial: but one trial can have more than one analysis, hence more than one aim.
- To discuss aims we must be clear about the meaning of the terms we use.
- Because common terms are used in different ways, and the same term can mean different things to different people, I begin by introducing two new terms: *de jure* and *de facto*.

Aims of the study/analysis

de jure

Estimate the treatment effect under the best case scenario.

Does the treatment work under the best case scenario?

[Per-Protocol, PP, efficacy]

de facto

Estimate the effect seen in practice if this treatment were applied to the population defined by the trial inclusion criteria.

Is an effect seen in practice if this treatment is applied to the population defined by the trial inclusion criteria?

[Intention To Treat, ITT, effectiveness]

In both cases, the analysis needs to reflect the treatment taken by a patient, that is consistent with the trial (analysis) aims [estimands].

Only in the special settings of

- *no dropout* and a *de facto* hypothesis; and
- *no dropout*, *no withdrawal* and a *de jure* hypothesis;

are there plausible definitive analyses that do not rest on additional untestable assumptions about the statistical behaviour of unobserved measurements.

Otherwise, to construct an analysis that is consistent with the trial aims some assumptions must be made about

treatment use following withdrawal;

and/or

future statistical behaviour of outcomes from dropouts and/or withdrawals.

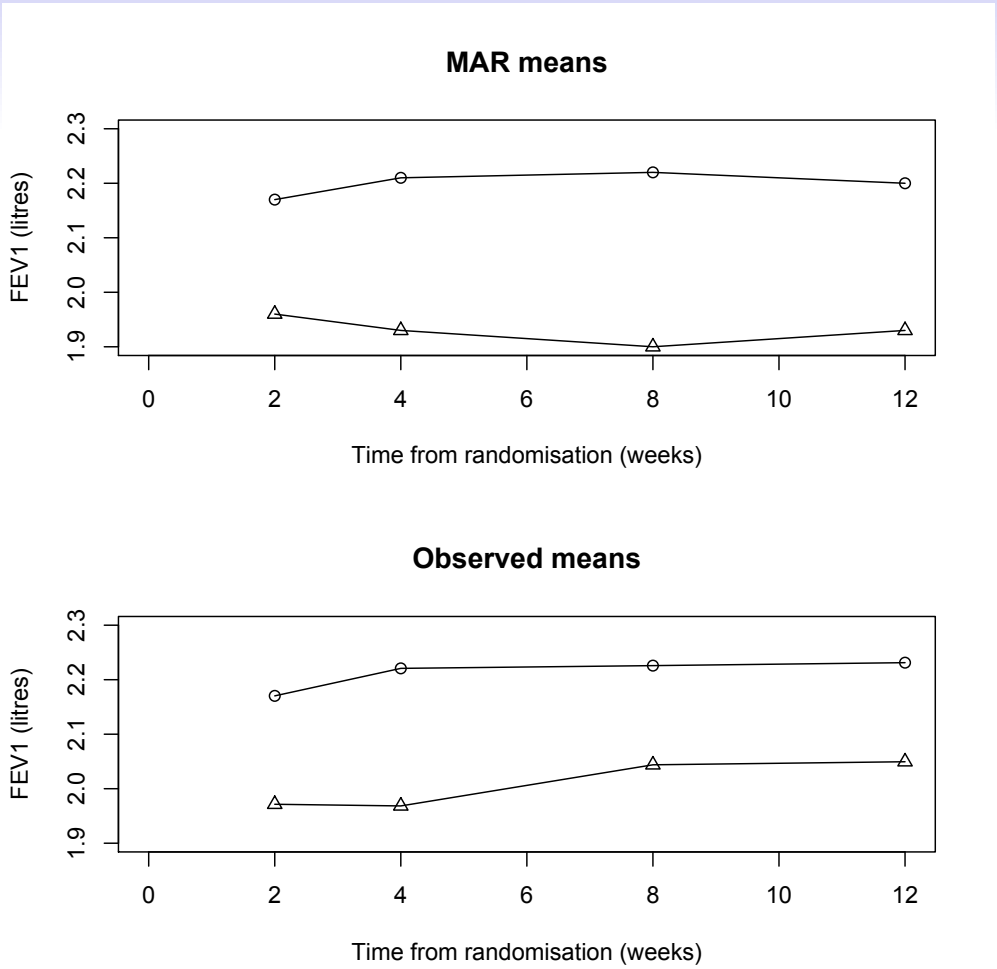
How does Rubin's missing value mechanism fit in with the *de jure* and *de facto* questions?

Recall, MAR dropout implies

the future statistical behaviour of a subject, conditional on the history, is the same whether the subject drops out or not in the future.

That is, both outcome and *future treatment use*, follow the same conditional distributions for all subjects.

- This would lead to the testing of a *de jure* hypothesis if treatment compliance before dropout matches the protocol.
- Or, extending this, if there were withdrawal, all post-withdrawal outcomes were set to missing.
- Or, if future treatment compliance after dropout matched that expected in use in the population, this would lead to the testing of a *de facto* hypothesis.



Mean FEV_1 (litres) from MAR model (top panel) and observed data (bottom panel). Circles — active treatment; triangles — placebo

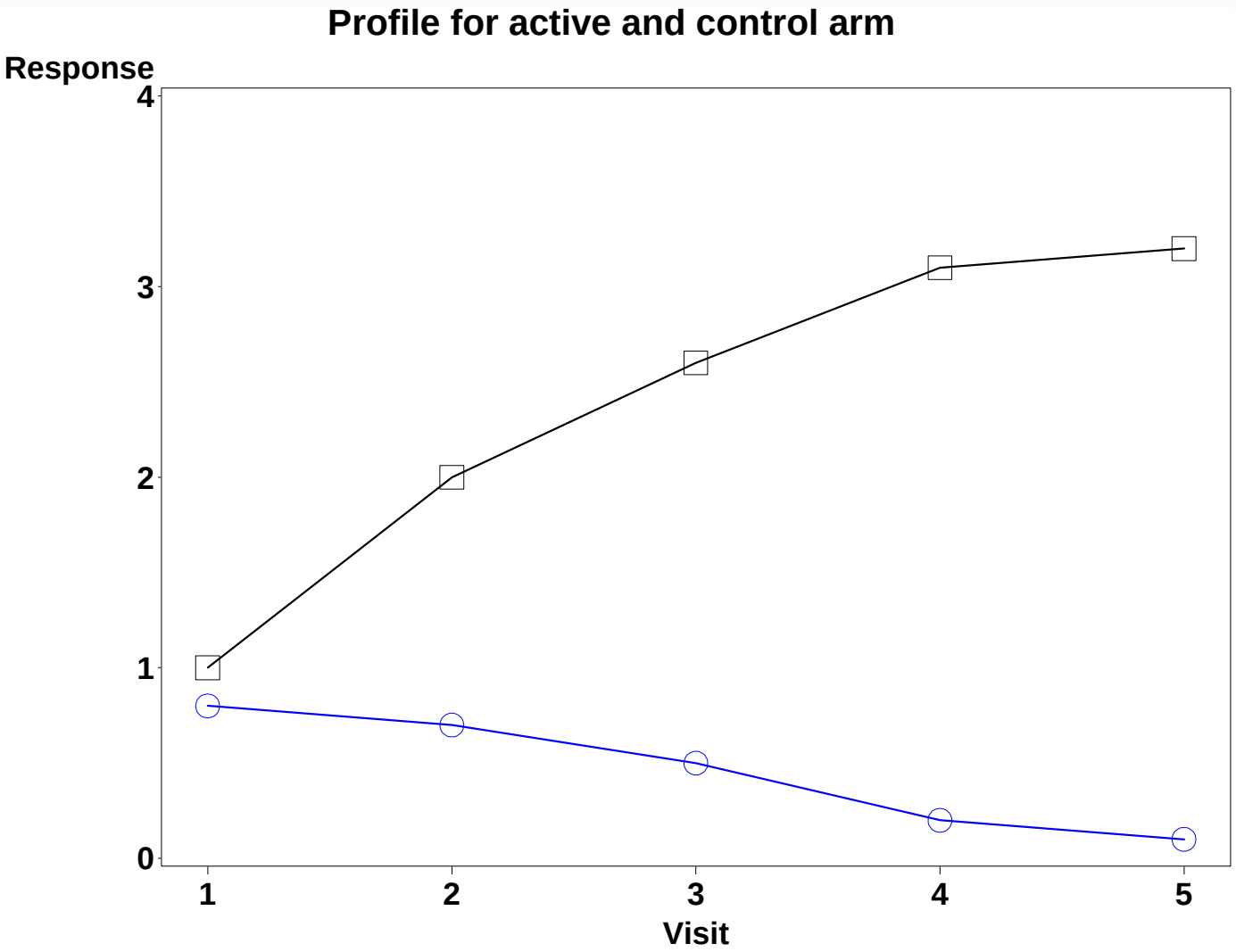
So analyses valid under MAR

- would lead to the testing of a *de jure* hypothesis if treatment compliance before dropout matches the protocol.
- Or, extending this, if there were withdrawal, all post-withdrawal outcomes were set to missing.
- Or, if future treatment compliance after dropout matched that expected in use in the population, this would lead to the testing of a *de facto* hypothesis.

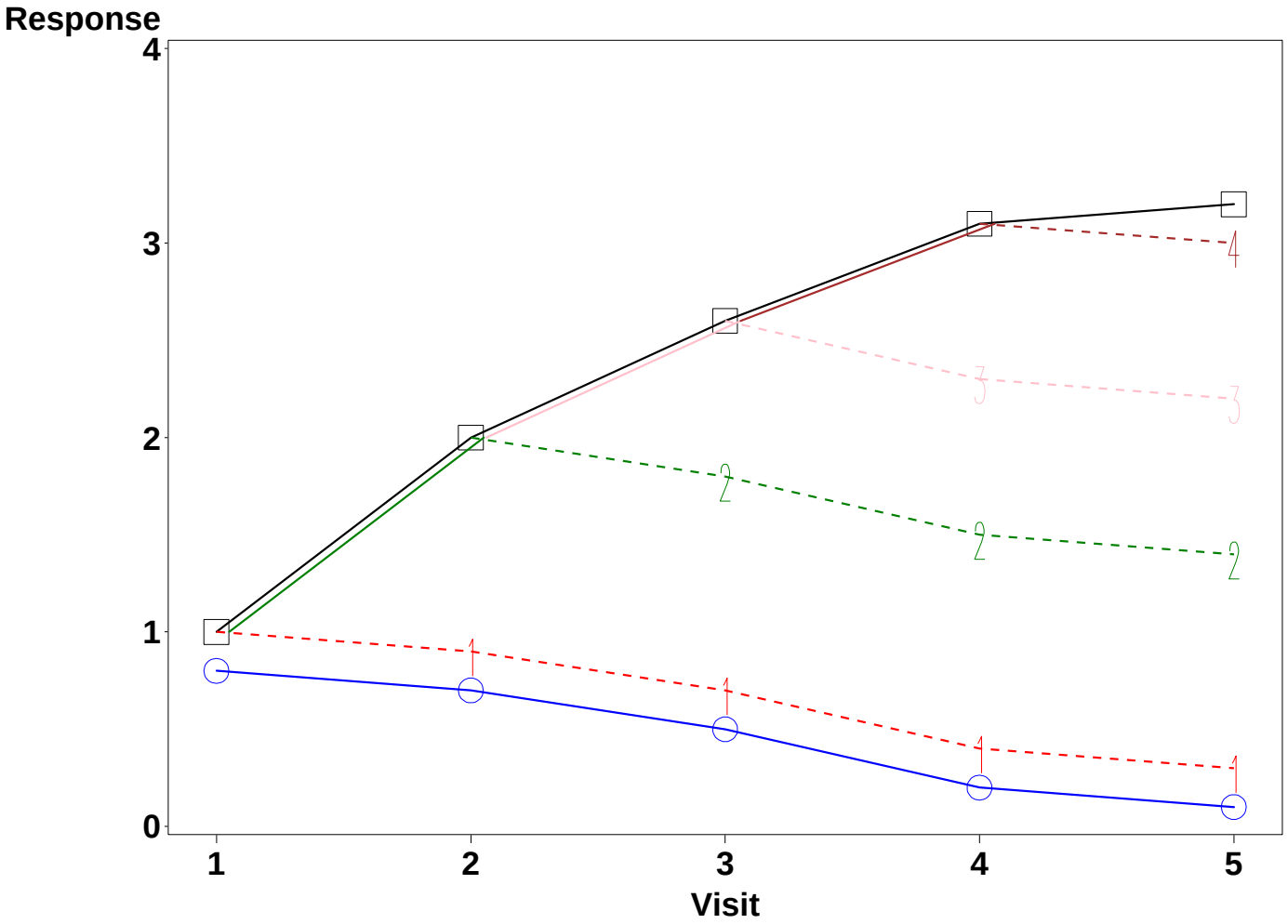
- But this excludes many common settings.
- So, often it is likely that we need to consider NMAR models.
- How should this be done?
- A secondary analyses? As part of sensitivity analyses?

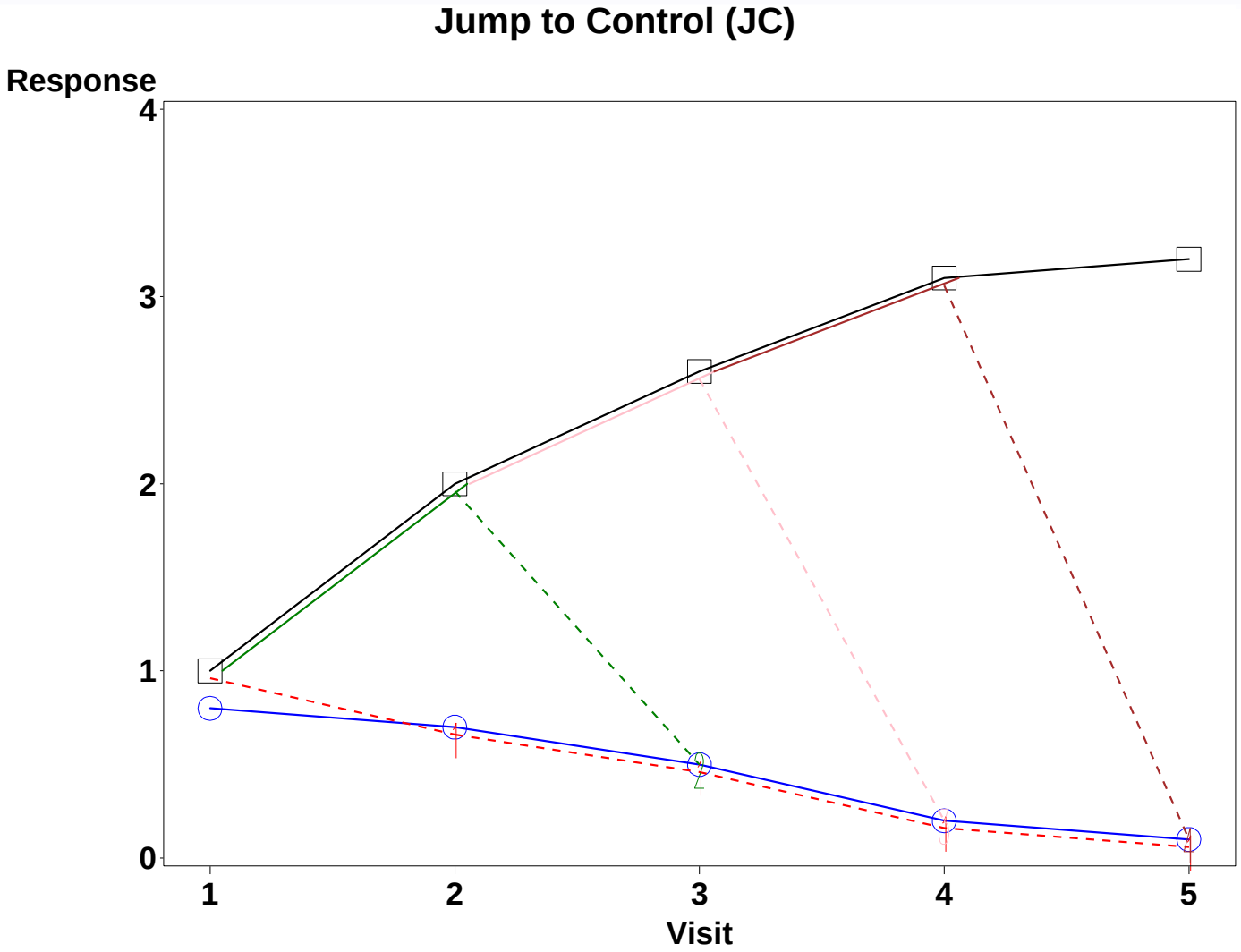
Pattern-mixture models give us a very convenient way of representing alternative NMAR behaviour that is transparent from a substantive viewpoint, and does not require complex (and to some obscure) forms of model.

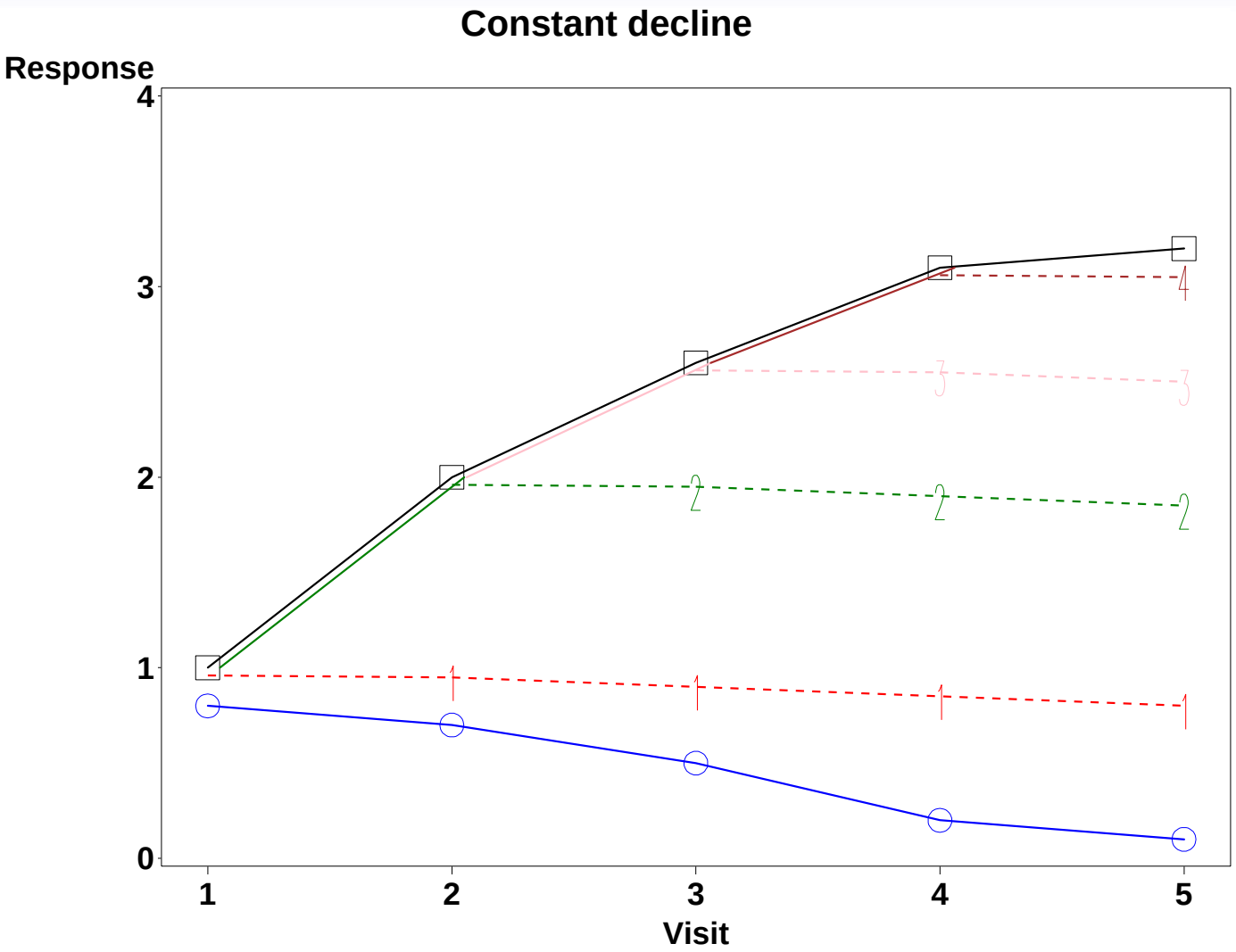
- We see that the MAR assumption allows us to 'borrow' conditional behaviour from those who remain/comply to represent those who do not.
- But what if this is implausible, or more pertinently, *inconsistent with the trial aims*?
- Well-defined departures from the MAR assumption can be formulated in terms of future statistical behaviour of those who drop out/withdraw.
- These alternative analyses can be described as sensitivity analyses: they explore sensitivity to the MAR assumption, in directions determined by the trial aims.


Profile for active and control arm
Response
4
3
2
1
0
1
2
3
4
5
Visit

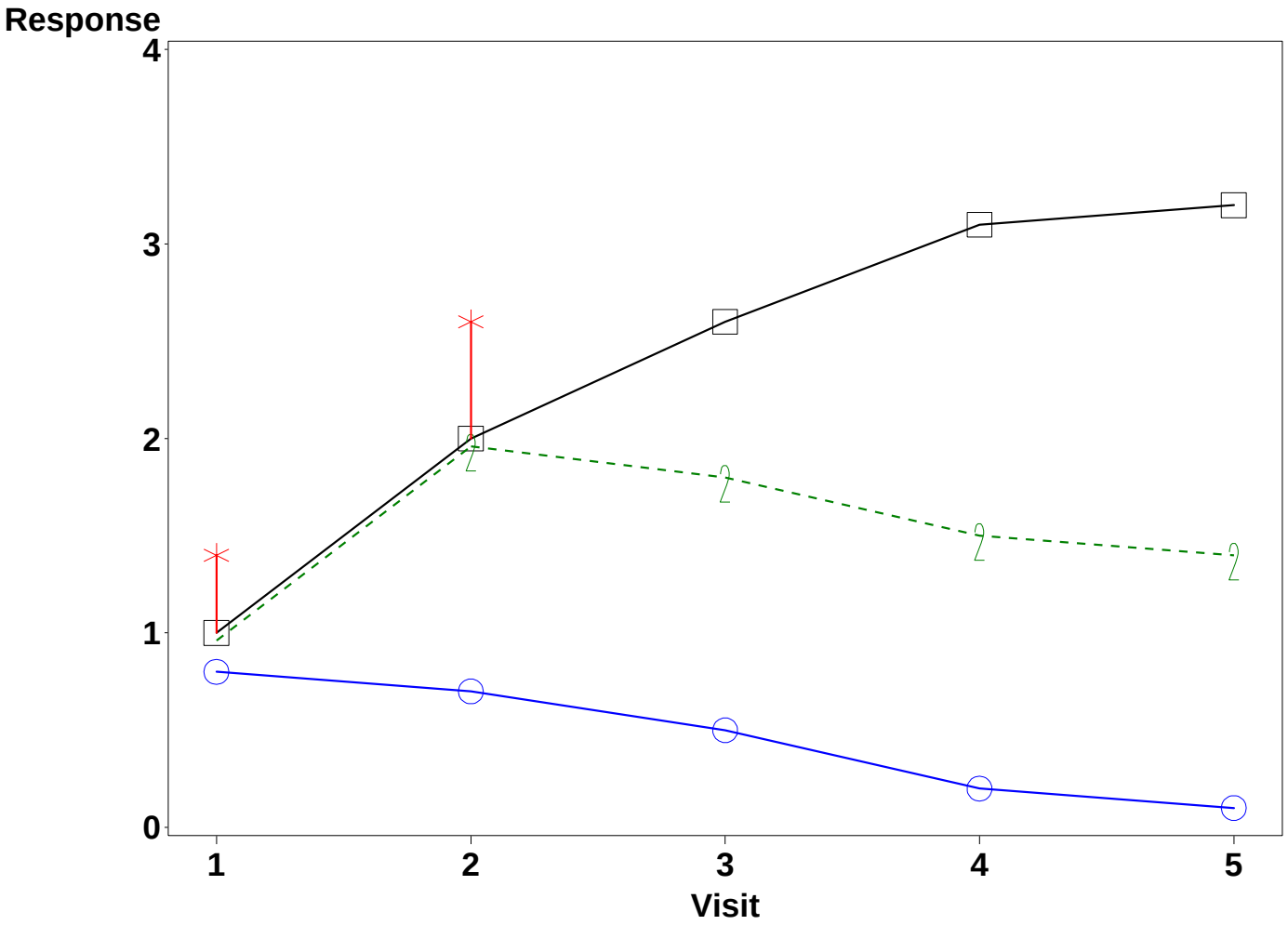
Copy Difference from Control (CDC)



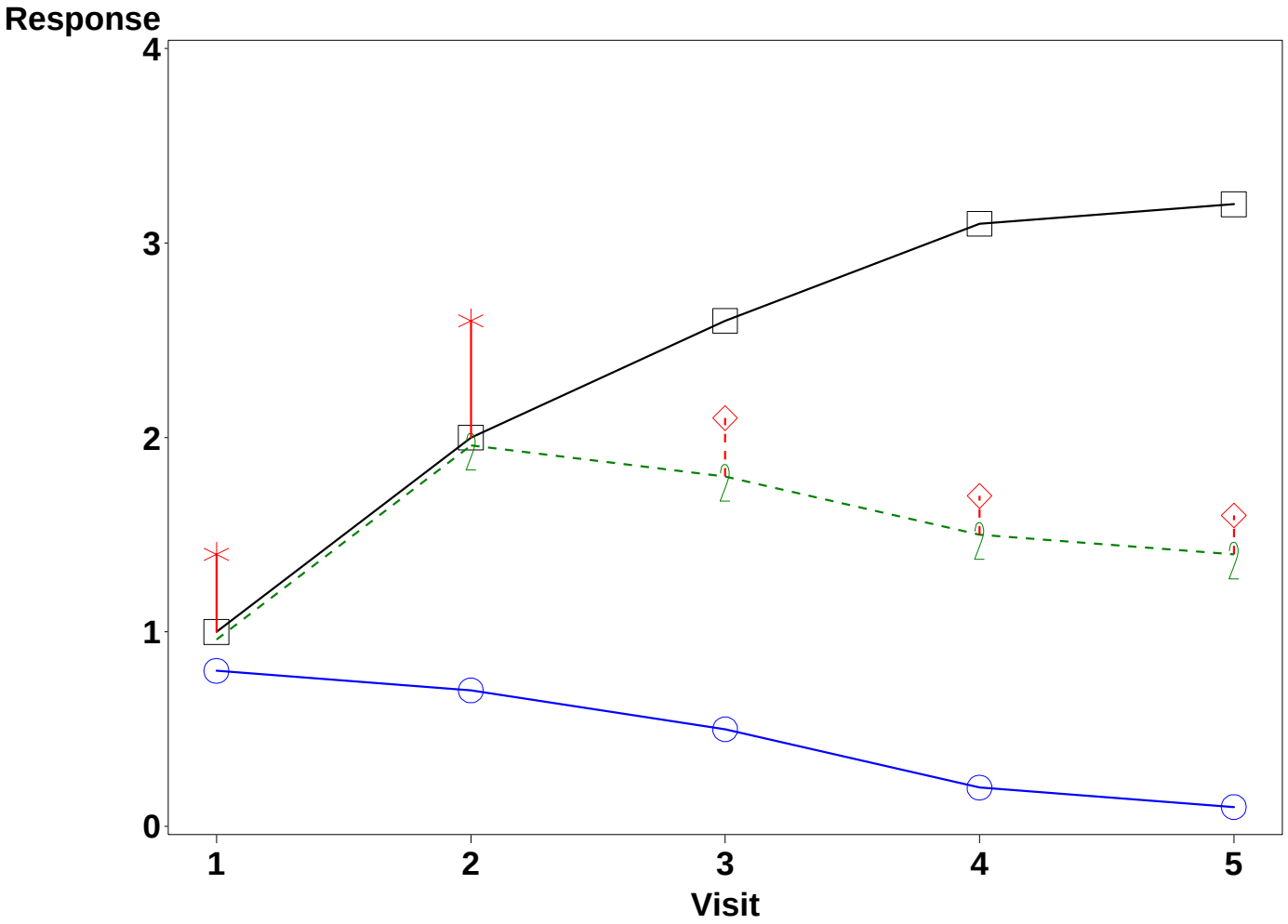

Jump to Control (JC)
Response
4
3
2
1
0
1
2
3
4
5
Visit


Constant decline
Response
4
3
2
1
0
1
2
3
4
5
Visit

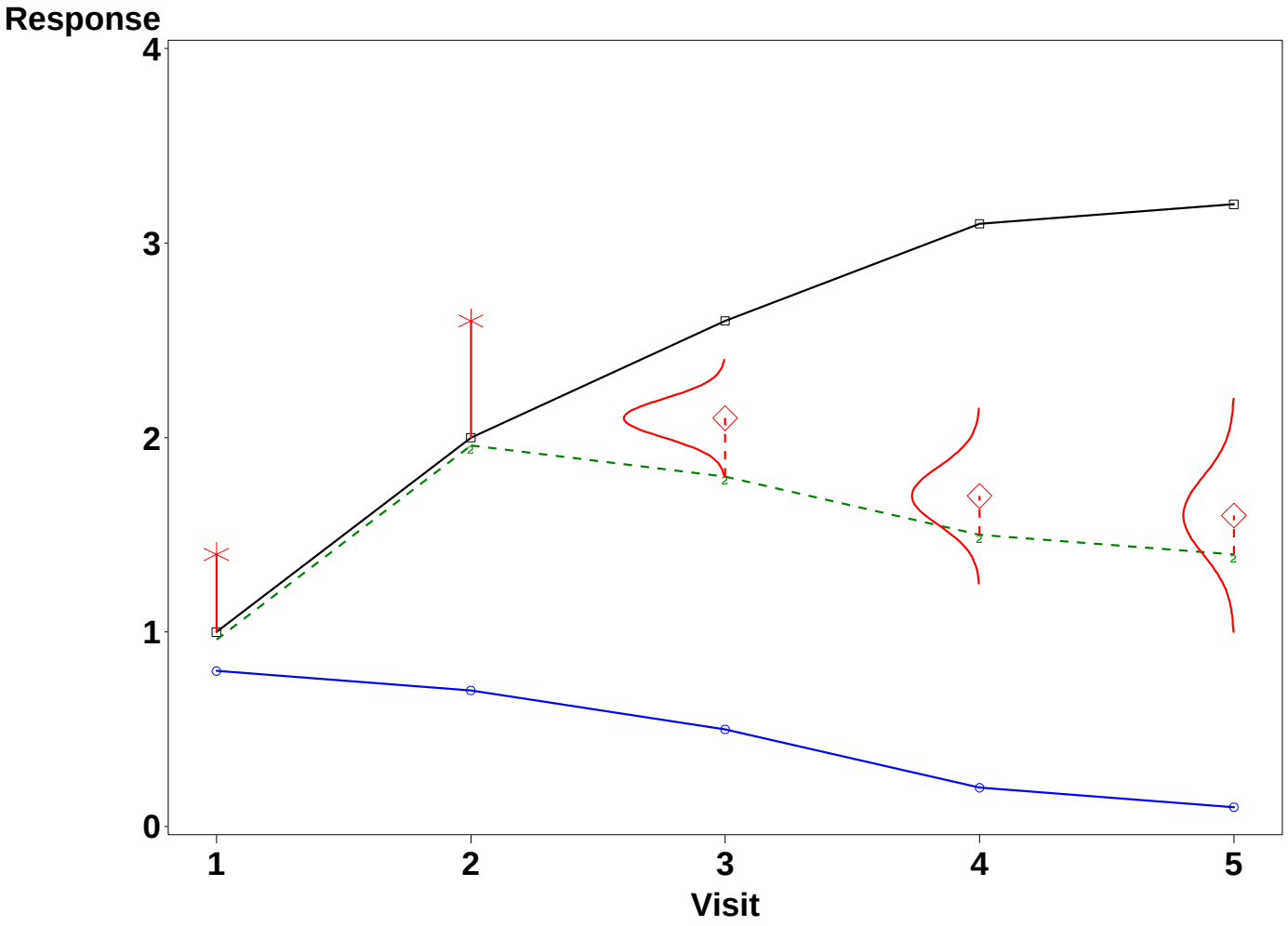
Observed Residuals for an individual subject



Means for conditional distribution for missing data



Samples drawn from conditional distribution for missing data



We are going to use the imputation step of MI to modify the behaviour of the future observations from the dropouts/withrawals (*i.e.* an MNAR imputation model).

The *observed data* will still be represented by an MAR model:

A multivariate linear model is used with

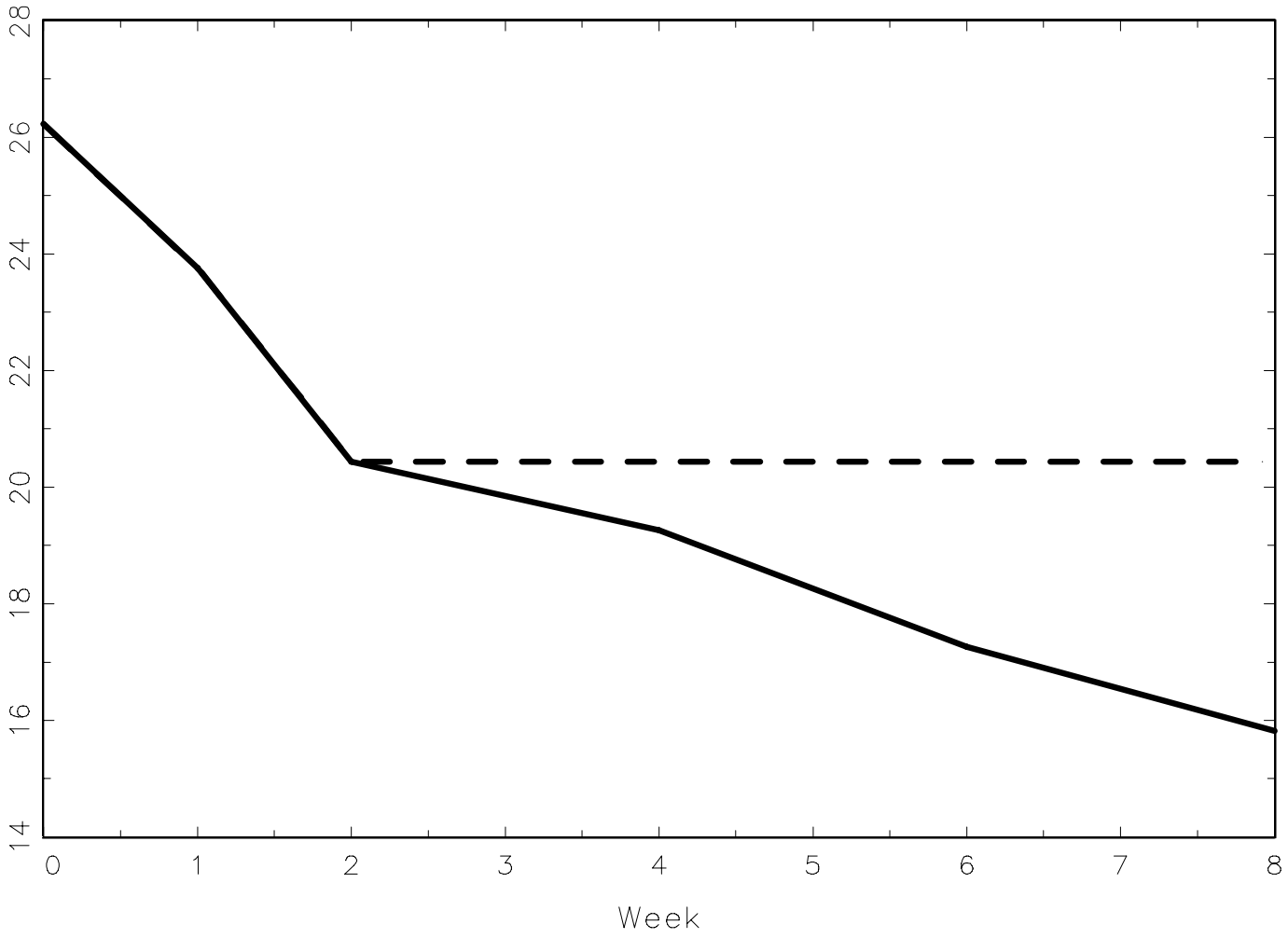
saturated means (treatment-by-time interaction);

and unstructured covariance matrices that can differ between the two groups.

There will be separate imputation models for each group.

- It is important that the parameters of the imputation model are drawn from a Bayesian posterior (at least approximately).
- We can use standard (and relatively simple methods) to do this under the MAR model (*e.g.* SAS PROC MI, PROC MCMC, sequential regression methods.)
- Provided the modified imputation model(s) are expressed in terms of the parameters from the MAR model, we can use the same Bayesian draws.

An example: *Last Mean Carried Forward (LMCF)*
Assume that when a subject drops out the *marginal mean* stays at the same level, *e.g.* dropout at week 4



It is only necessary to modify the mean for this subject for weeks 4,6 and 8.

(The dependence structure is assumed to stay the same.)

There is a slight complication in that the pattern mixture modelling is all done in terms of conditional means from the regression.

We need to go from these, to the marginal means, modify these, and then convert back.

It requires only a couple of matrix calculations:

We have the *conditional* means (intercepts from the regressions):

$$\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_6]^T$$

and the matrix of regressions coefficients

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -\beta_{21} & 0 & 0 & 0 & 0 \\ 1 & -\beta_{31} & -\beta_{32} & 0 & 0 & 0 \\ 1 & -\beta_{41} & -\beta_{42} & -\beta_{43} & 0 & 0 \\ 1 & -\beta_{51} & -\beta_{52} & -\beta_{53} & -\beta_{54} & 0 \\ 1 & -\beta_{61} & -\beta_{62} & -\beta_{63} & -\beta_{64} & -\beta_{65} \end{bmatrix}$$

Then the marginal means are

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{bmatrix} = \mathbf{B}^{-1}\boldsymbol{\alpha}.$$

For a dropout/withdrawal at time 3 (week 4), the modified set of means is, from this,

$$\boldsymbol{\mu}_M = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix}$$

The corresponding conditional means are recovered from

$$\boldsymbol{\alpha}_M = \mathbf{B}\boldsymbol{\mu}_M$$

or more succinctly

$$\boldsymbol{\alpha}_M = \mathbf{B}\mathbf{N}_3\mathbf{B}^{-1}\boldsymbol{\alpha}.$$

where

$$\mathbf{N}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

These modified intercepts are used in the imputation model, together with the original regression coefficients.

Throughout the imputation process these calculations are done using the parameter values drawn from the Bayesian posterior:

the modified intercepts are clearly simple functions of the original linear model parameters in $\boldsymbol{\alpha}$ and $\mathbf{B}$, *i.e.*

$$\tilde{\boldsymbol{\alpha}}_M = \tilde{\mathbf{B}}\mathbf{N}_3\tilde{\mathbf{B}}^{-1}\tilde{\boldsymbol{\alpha}}.$$

To summarize, in general:

- *STEP (1)* Fit a multivariate Gaussian linear model, using likelihood (strictly REML) under the MAR assumption, removing unnecessary constraints: *e.g.*
  - Use a saturated visit-by-treatment interaction.
  - Either use baseline as a response (without a treatment effect), or include it as a baseline that interacts with visit.
  - Allow the covariance matrices to differ between treatment groups.
- *STEP (2)* Using conventional procedures for Bayesian analyses of the multivariate linear Gaussian model, draw, separately for each treatment group, a set of parameters of the MAR model from their posterior distribution.

- *STEP (3)* Non-ignorable (MNAR) models for the future conditional statistical behaviour of dropouts and/or withdrawals are constructed from

  components of the MAR model.

- *STEP (4)* Using the models from (3), constructed from the parameter draws from (2), impute a set of missing data.

- *STEP (5)* Use the analysis method that would have been applied to the full data set (*e.g.* baseline adjusted ANCOVA for the final visit).

- *STEP (6)* Repeat (2)-(5) M times.

- *STEP (7)* Combine using Rubin's rules to get the final inference.

- The imputation and substantive models are *uncongenial.*
- They match exactly on the observed data;
- but the imputation model has structure that is additional to the substantive model for the unobserved.
- Hence we expect the MI analysis to be, at worst, slightly conservative in a long-run sense.

Four example MNAR assumptions we have found useful:

1. Jump to Reference.
2. Last Mean Carried Forward.
3. Copy Differences in Reference
4. Copy Reference.

These four (plus MAR) have been implemented in a SAS macro written by James Roger, which is available from our website: `www.missingdata.org.uk`.

Further Points

- Many other MNAR scenarios are possible, and can be chosen to suit the particular setting.
- The MI framework allows direct examination of the future MNAR behaviour: useful for visualizing the impact of the assumptions, and for assessing their plausibility.
- This approach provides a route for directly incorporating reasons for dropout.

References

Daniels MJ and Hogan JW (2008) *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* Chapman & Hall/CRC.

Demirtas H and Schafer JL (2003) On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, **22**, 1553–2575.

Diggle PJ and Kenward MG (1994) Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–94.

Heckman J (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, **5**, 475–492.

Hogan J and Laird NM (1997) Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, **16**, 259–272.

Kenward MG and Carpenter JC (2009) Multiple imputation. In: Fitzmaurice G *et al.* eds *Longitudinal Data Analysis*, CRC, pp.477-500.

Kenward MG, Molenberghs G and Thijs H. (2003) Pattern-mixture models with proper time dependence. *Biometrika*, **90**, 53–71.

Kenward MG and Rosenkranz G (2011) Joint modelling of outcome, observation time and missingness. *Journal of Biopharmaceutical Statistics*, **21**, 252-262.

Little RA (1994) A class of pattern-mixture models for multivariate incomplete data. *Biometrika*, **81** 471–483.

Little RA and Yao L (1996) Intent-to-treat analysis for longitudinal studies with dropouts. *Biometrics*, **52**, 1324–1333.

Molenberghs G and Kenward MG (2007) *Missing Data in Clinical Studies.* Wiley.

Molenberghs G, Michiels B, Kenward MG and Diggle PJ (1998) Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica*, **52**, 153–161.

Rubin DB (1976) Inference and missing data. *Biometrika* 63, 581–592.

Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys.* Wiley.