HANDLING DROPOUT AND WITHDRAWAL IN LONGITUDINAL CLINICAL TRIALS

Mike Kenward
London School of Hygiene and Tropical Medicine

Acknowledgements to

James Carpenter (LSHTM)
Geert Molenberghs (Universities of Hasselt and Leuven)
James Roger (GSK/LSHTM)

Biostatistics Network Symposium:
“Contemporary Statistical Methods in Medical Research”

15th September 2011
CONTENTS

1. Concepts, issues, definitions
2. The position of MAR.
3. Departures from MAR and sensitivity analysis.
4. One approach to sensitivity analysis.
5. Summing up.
(1) CONCEPTS, ISSUES, DEFINITIONS.

1. Aims: *de jure* and *de facto* estimands.

2. Missing data mechanisms.
Aims: *de jure* and *de facto* estimands.

My focus will be on dropout/withdrawal.

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.
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

\textit{de jure}

Estimate the treatment effect under the best case scenario.
Does the treatment work under the best case scenario?
[Per-Protocol, PP, efficacy]

\textit{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.
Missing data mechanisms.

From a frequentist perspective, for dropout:

**Definition: Missing Completely at Random (MCAR)**

- The probability of a subject dropping out is independent of outcomes, seen or unseen, or any other variables in the analysis.
- Any analysis valid for the whole dataset is valid for the observed data.
Definition: Missing at Random (MAR)

- The probability of a subject dropping out is conditionally independent of future (current) observations, given the observed history.

- Under MAR, likelihood based analyses of the outcome only are valid (the actual dropout mechanism can be ignored).

- Non-likelihood methods (e.g. such as GEE/moment based) will need to use the dropout mechanism explicitly to be valid.
Another view of MAR:

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

Implications:

- The future behaviour of dropouts can be modelled using future behaviour of those who remain.
  This implies that treatment behaviour can be borrowed.

- Analyses that capture this behaviour properly, i.e. likelihood based, will be valid, when others such as GEE/moment based, may not be.

- A key issue in the longitudinal setting is correctly representing the regression of the future on the past.
Definition: Missing Not at Random (MNAR)

- The probability of a subject dropping out is conditionally dependent on future (current) observations, given the observed history.

- Equivalently:
  The future statistical behaviour of subjects is not the same for those who drop out and those who don’t, even if their history is identical.

- Valid analyses need to take account of the missing value mechanism, which is usually not known.

- This can be done indirectly.
There has been a lot written about methods of analysis that are valid under MAR.

We know that with likelihood based methods (broadly) the dropout mechanism is ignorable.

There are non-likelihood methods that use inverse probability weighting.

The two can be combined (in a certain sense) using doubly robust methods.

[For example:
We can use comparatively simple likelihood based analyses under MAR, e.g. for a continuous outcome:

the multivariate normal linear model with unstructured covariance matrix and full time-by-treatment interaction.

Sometimes called a Mixed Model Repeated Measurements (MMRM) analysis (unfortunately in my view).

What about other outcomes? *e.g.* binary.

Likelihood can be awkward for *marginal/population averaged* models, *Generalized Estimating Equations* (GEE’s) are commonly used instead.

These are not valid in general under MAR.

Alternatives:

- Use subject specific models for which likelihood analyses are the norm.
  
  But this changes the interpretation of the parameters.

  Scale the parameters appropriately?

- Use *Multiple Imputation* with a sufficiently rich and tractable (*e.g.* loglinear) imputation model which is uncongenial.

- Use *inverse probability weighted estimating equations* with a marginal model, and doubly robust extensions.
How does MAR 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 FEV₁ (litres) from MAR model (top panel) and observed data (bottom panel). Circles — active treatment; triangles — placebo.
(3) DEPARTURES FROM MAR AND SENSITIVITY ANALYSIS

Recall from earlier: how does MAR fit in with the de jure and de facto questions?

Broadly, 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.
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?
Very broadly we can distinguish two important ways of approaching departures from MAR in settings like this.

- **Route 1:** Consider an explicit MNAR dropout mechanisms. *i.e.* modify
  \[
  P(R \mid Y_O, Y_M)
  \]
  This is naturally approached using *selection models*.

- **Route 2:** modify directly the future behaviour of dropouts, *i.e.* modify
  \[
  f(Y_M \mid Y_O, R)
  \]
  This is naturally approached using *pattern-mixture models*. 
In fact there is a third route: so-called shared-parameter models.

These have latent variables that are linked to both outcomes and dropout indicators.

They 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?

The generic sensitivity problem in the dropout setting

1. Define the specific questions, and consequent quantities to be estimates, and hypotheses, that we wish to test;

2. define the nomenclature for departures from protocol;

3. frame the relevant accessible assumptions, and

4. formulate the procedures for estimation and inference.
(4) ONE APPROACH TO SENSITIVITY ANALYSIS


- The common feature is 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.
Continuous outcomes:

*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:

[These four (plus MAR) have been implemented in a SAS macro written by James Roger]

2. Last Mean Carried Forward.
3. Copy Differences in Reference
4. Copy Reference.
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.
A full likelihood approach is also available, the difference is that the fitted model also matches the MNAR scenario.

Together these represent two possible approaches to sensitivity analysis:

1. What is the impact on the “complete data analysis” if the data generating mechanism departs from MAR in the given way?

2. What results are obtained if the substantive analysis itself reflects the given departure from MAR?
(5) SUMMING UP

(1) Analysis of trials with partially observed data requires making untestable assumptions. Therefore:

- design to minimise missing data;
- collect off-treatment data where possible;
- pre-specify key assumptions for patients who withdraw for various reasons;
(2) Apply standard statistical principles:

- make assumptions;
- obtain valid inference under the assumptions;
- check the robustness of the conclusions as the assumptions change;
- Talking about a conservatively biased estimator doesn’t make sense without first specifying the assumptions about the missing data.

(3) Specifying LOCF and then debating whether it is conservative or not in a particular leads to unilluminating discussion.
Methods

1. Think carefully about the question the analysis is addressing.

2. Sensitivity analysis is key, and needs to be accessible and relevant:
   - we would argue that formulating the assumptions through the pattern mixture approach is accessible to most collaborators;
   - the more accessible the assumptions, the more likely the analyses are to be relevant;
   - given the assumptions, valid estimation is straightforward via multiple imputation.

3. Simple inverse probability weighting is too inefficient,
   There are stimulating developments using doubly robust estimators in this area, including doubly robust multiple imputation. More experience is needed before they can be widely recommended in practice.