

Matthews JNS, Henderson R, Farewell DM, Ho W-K, Rodgers LR. [Dropout in crossover and longitudinal studies: is complete case so bad?](#). *Statistical Methods in Medical Research* 2014, 23(1), 60-73.

Copyright:

© The authors, 2014.

DOI link to article:

<http://dx.doi.org/10.1177/0962280212445838>

Date deposited:

01/07/2015



This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#)

Dropout in Crossover and Longitudinal Studies: Is Complete Case so Bad?

John N. S. Matthews and Robin Henderson, Mathematics & Statistics, Newcastle University UK, **Daniel M. Farewell** Department of Primary Care & Public Health, Cardiff University UK and **Weang-Kee Ho**, Department of Public Health and Primary Care, University of Cambridge, UK, **Lauren R. Rodgers**, Peninsula College of Medicine & Dentistry, Exeter, UK.

We discuss inference for longitudinal clinical trials subject to possibly informative dropout. A selection of available methods is reviewed for the simple case of trials with two timepoints. Using data from two such clinical trials, each with two treatments, we demonstrate that different analysis methods can at times lead to quite different conclusions from the same data. We investigate properties of complete-case estimators for the type of trials considered, with emphasis on interpretation and meaning of parameters. We contrast longitudinal and crossover designs and argue that for crossover studies there are often good reasons to prefer a complete case analysis. More generally, we suggest that there is merit in an approach in which no untestable assumptions are made. Such an approach would combine a dropout analysis, an analysis of complete-case data only, and a careful statement of justified conclusions.

Address for correspondence: Robin Henderson, Mathematics & Statistics, Newcastle University NE1 7RU, UK. Email: Robin.Henderson@ncl.ac.uk

1 Introduction

Over the last 20 or so years an impressive array of methods has been developed for missing data problems in general and dealing with dropout in longitudinal trials in particular¹⁻⁸. Many of these contributions have been deeply thought-provoking and technically highly accomplished. Nevertheless, whatever the technical virtuosity of these methods they cannot make missing data appear. As has been widely recognised⁹, almost all of the methods rely for their validity either on questionable assumptions about the missingness process, or by making untestable assumptions with regard to unobserved outcomes. Consequently it is now generally accepted, and formally recommended⁸, that an analysis of data subject to missingness should include a careful assessment of sensitivity of conclusions to the underlying assumptions.

In this paper we take a different view. We use two simple two-timepoint randomized trials to illustrate some of the techniques now available for dealing with dropout from longitudinal data, and as a basis for discussion of a number of foundational issues. We will argue that dropout from a longitudinal clinical trial should not be treated as a *special case* of a generic missing data problem, but as a *different case*, since the special role of time should be taken into account. We will suggest that a structured analysis that does not rely on untestable assumptions can in certain circumstances be not just a practicable but an appropriate strategy. In particular, for the situations in which a crossover trial is applicable we will propose that a complete case analysis provides a proper basis for inference.

When dealing with generic missing data it is customary to begin by introducing notation $(Y_{\text{obs}}, Y_{\text{miss}})$ for observed and missing components of a response vector Y . Alongside this we have a vector R of response indicators, with elements taking the value one if the corresponding element of Y is observed, zero otherwise. The familiar Rubin classification of missingness mechanisms is Missing Completely at Random (MCAR) if $f(R|Y_{\text{obs}}, Y_{\text{miss}}) = f(R)$, Missing at Random (MAR) if $f(R|Y_{\text{obs}}, Y_{\text{miss}}) = f(R|Y_{\text{obs}})$ and otherwise Missing Not at Random (MNAR). It is recognised that an appropriate analysis based only on Y_{obs} can provide unbiased parameter estimates if the missingness is assumed to be MCAR or MAR. However, the validity of these assumptions cannot be detected without imposing additional untestable structure, such as embedding a MAR model within a broader MNAR one, which cannot be unique⁹.

Not all writers have fully accepted the Rubin missing data taxonomy or its generic application to longitudinal data, for example¹⁰⁻¹². Diggle et al¹¹ assert that the notation $(Y_{\text{obs}}, Y_{\text{miss}})$ is too simplistic when missingness is caused by dropout in a longitudinal study. The argument is that the notation Y_{miss} suggests that the underlying unobserved response is the same whether or not it is observed. This is appropriate when the process that generates the missingness pattern is exogenous to the process that generates the responses. Data may still be MNAR, for example if an experimenter simply chooses not to record very low values, but the extant value attached to the experimental unit does not in itself change depending on whether or not it is observed. However, dropout in a longitudinal trial often corresponds to an event in a subject's life, such as withdrawing from treatment. Meinert¹³ makes the important point that such *treatment dropouts* must be distinguished from *analysis dropouts*, those who simply cease to be observed. Clearly what happens after dropout could be very different from what might have happened in the absence of dropout. In other words if Y_{miss} was not missing it could be something completely different. Considering dropout caused by

death makes this point rather dramatically¹⁴.

Building from Diggle et al. we suggest that the use of counterfactual notation should be taken up more widely. We can think of potential outcomes $Y_t(0)$ and $Y_t(1)$ that correspond, respectively, to the response when external intervention is not, or is, employed to prevent dropout. As Didelez¹⁵ argues in her discussion of Diggle et al., such counterfactuals can be sensibly defined and modelled provided the hypothetical intervention is made explicit; it makes little sense to talk about a response Y_t ‘in the absence of dropout’ without specifying how this absence is to be achieved. Note that we are *indexing* by dropout (prevention) and not *conditioning* on it, a distinction also made in Rubin’s causal model¹⁶. Such notation helps make explicit the implicit assumption in most previous work that the properties of $Y_t(1)$ are usually the targets for inference. Aalen and Gunnes¹² discuss the complexities of attempting to separate a longitudinal response from the process that may cause it to be missing, and describe why different inferential objectives may be appropriate in different circumstances. Buyse¹⁷ takes a holistic view, and invites subject-matter experts to form a hierarchy of contrasts that may include both longitudinal responses or survival outcomes, in order to compare individuals undergoing different treatment regimes.

In this work we consider targets for inference for simple two-period clinical trials. Our notation, assumptions and a short review of a selection of available methods are given in Section 2. In Section 3 we present analyses of two such trials, each designed to compare two treatments and in each of which a number of subjects dropped out before the second measurement was collected. We show that a wide range of treatment effect estimates can be obtained, depending on the method and assumptions adopted. Section 4 includes some technical work on the properties of complete-case estimators for two-treatment two-period designs. In Section 5 we argue that a transparent approach to an analysis of this type of data requires a combination of a dropout analysis, a complete case analysis, and a careful statement of conclusions that takes full account of the context of the analysis.

2 Notation, assumptions and methods

2.1 Methods

Suppose there are n subjects in a study and two observation or treatment periods. Subjects provide a measurement at time 1, denoted Y_{i1} for subject i , but some subjects drop out before measurement time 2. Let $R_i = 1$ indicate continuation to time 2 and $R_i = 0$ otherwise. The possibly counterfactual second response, assuming hypothetically that dropout may be prevented, is $Y_{i2}(1)$. For simplicity we will denote this by Y_{i2} but throughout we bear in mind that our models and assumptions are predicated upon a possible counterfactual. Let $Y_i = (Y_{i1}, Y_{i2})^T$ and assume $E[Y_i] = X_i\beta$ where β is a p -dimensional parameter vector and X_i is a $2 \times p$ design matrix associated with subject i .

A raft of techniques is available for estimation, some of which are summarised here. We do not attempt a full review and our list is certainly incomplete, with pattern mixture modelling, multiple imputation¹⁸ and Bayesian methods notable omissions⁵. We focus in the main on methods where the estimating equations can be written down in fairly simple form.

Complete Case, CC

Estimation is restricted to subjects who provided both measurements and there is no estimation or use of dropout probabilities. A variety of approaches are possible but a random intercept model with Gaussian errors is common. If the marginal $\text{var}(Y_i) = \Sigma$ is known, the maximum likelihood estimator $\hat{\beta}$ solves the score equation

$$\sum_{i=1}^n R_i \{X_i^T \Sigma^{-1} (Y_i - X_i \beta)\} = 0. \quad (1)$$

In practice Σ would be unknown and simultaneous estimation of Σ and β is required.

Last Observation Carried Forward, LOCF

A procedure that has been common in practice, though much criticised^{3,8}, is to carry forward the last observed value and then pretend the data are complete. Let $Y_i^R = (Y_{i1}, Y_{i2})$ if $R_i = 1$ and $Y_i^R = (Y_{i1}, Y_{i1})$ if $R_i = 0$. Then the LOCF score corresponding to (1) is

$$\sum_{i=1}^n \{X_i^T \Sigma^{-1} (Y_i^R - X_i \beta)\} = 0, \quad (2)$$

with the same comments on variance estimation.

Linear Mixed Effect, LME

Maximum likelihood estimation using only observed data yields consistent estimators under MAR, without use of dropout probabilities⁴. Assuming a Gaussian random intercept model, the score equation of current interest is

$$\sum_{i=1}^n \left[R_i \{X_i^T \Sigma^{-1} (Y_i - X_i \beta)\} + \frac{(1 - R_i)}{\sigma_1^2} \{x_{i1} (Y_{i1} - x_{i1}^T \beta)\} \right] = 0, \quad (3)$$

where x_{i1}^T is the first row of X_i and $\sigma_1^2 = \text{var}(Y_{i1})$.

Diggle Kenward, DK

Diggle and Kenward¹⁹ take a fully parametric approach, combining assumptions of multivariate Gaussian response and logistic dropout probabilities. For the remainder of this paper we will assume

$$\Pr(R_i = 1 | Y_{i1}, Y_{i2}) = p_i(\theta, \theta_2) = \text{expit}(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2}), \quad (4)$$

where we use $\text{expit}(u)$ to denote $\exp(u)/(1+\exp(u))$ and $\theta = (\theta_0, \theta_1)$. Taking $\theta_2 = 0$ gives MAR dropout and $\theta_1 = \theta_2 = 0$ gives MCAR.

Diggle and Kenward¹⁹ propose likelihood inference based on a selection factorisation²⁰. In our terms the likelihood is

$$L = \prod_{i=1}^n \{P(R_i = 1 | Y_{i1}, Y_{i2}) f(Y_{i1}, Y_{i2})\}^{R_i} \{P(R_i = 0 | Y_{i1}) f(Y_{i1})\}^{(1-R_i)}. \quad (5)$$

The contribution $P(R_i = 0 | Y_{i1})$ is obtained from the logistic model $P(R_i = 0 | Y_{i1}, Y_{i2})$ by integrating out Y_{i2} with respect to its Gaussian conditional distribution given Y_{i1} . Diggle

and Kenward suggest a probit approximation to logistic in order to obtain a closed form for $P(R_i = 0|Y_{i1})$. The same approximation will prove very useful to us in Section 4.

Linear Increments, LI

Diggle, Farewell and Henderson¹¹ model the *increments* between observations at consecutive measurement times, taking a martingale approach which has links with methods for event history analysis. There are no distributional assumptions for either response or dropout, other than linear models for the mean increments in responses and a martingale assumption for within-subject random effects.

Their method is chiefly designed for processes in which the parameter β in the linear predictor for the mean is assumed to vary with time. It can be extended in several ways²¹, and can be adapted to give estimates of effects that are assumed to be time-constant, in our case as follows. Let $D_i = Y_{i1} - Y_{i2}$ and $d_i = x_{i1} - x_{i2}$; elements of the latter will be zero where covariates are constant between measurement times. Recall that $E[Y_{i1}] = x_{i1}^T \beta$ and note that $E[D_i] = d_i^T \beta$. Ordinary least squares can be used to estimate β , leading to the estimating equation

$$\sum_{i=1}^n \{x_{i1}(Y_{i1} - x_{i1}^T \beta) + R_i d_i (D_i - d_i^T \beta)\} = 0. \quad (6)$$

Inverse probability weighted estimating equations with MAR assumption, IPMAR

In a series of papers²²⁻²⁴, Robins and colleagues proposed inverse probability of observation weighted estimating equations. Again, no distributional assumptions are made about Y_i , other than a linear model for the mean, but a parametric model for dropout probability is required. Although the authors discuss a variety of efficient estimating procedures, for our purposes it is sufficient to concentrate on the simplest version of the proposals. In this case θ_2 in (4) is assumed known and the other dropout parameters $\theta = (\theta_0, \theta_1)$ are estimated using the estimating equations

$$\sum_{i=1}^n \left\{ 1 - \frac{R_i}{p_i(\theta, \theta_2)} \right\} p_i(\theta) \begin{pmatrix} 1 \\ Y_{i1} \end{pmatrix} = 0, \quad (7)$$

where $p_i(\theta) = \text{expit}(\theta_0 + \theta_1 Y_{i1})$. The parameter of interest β is then obtained by solving the weighted estimating equation

$$\sum_{i=1}^n \frac{R_i}{p_i(\theta, \theta_2)} \{X_i^T \Sigma^{-1} (Y_i - X_i \beta)\} = 0. \quad (8)$$

In the sequel we will refer to the combination of (7) and (8) under the MAR assertion that $\theta_2 = 0$ as *IPMAR*.

Inverse probability weighted estimating equations with zero interaction procedure, IPZI, IPZI+ and IPZI-

There are no methods for estimating θ_2 in (4) without further strong assumptions. Instead, in order to apply (8) in the MNAR case it is usually recommended²⁴⁻²⁶ that an expert from the application area be asked to specify either θ_2 itself or a realistic range within which it could feasibly lie. In the latter case the advice is usually to vary θ_2 across the range and

inspect the consequences for estimation of β , or perhaps to elicit a prior distribution for θ_2 or a related quantity and derive an appropriate posterior summary^{27,28}. Application experts with the required abilities are likely to be in high demand, so there is merit in exploring whether, in specific applications, more oblique approaches to sensitivity analysis might pay dividends.

Ho and colleagues²⁹ propose one such alternative, in which one untestable assumption is traded for another. For crossover trials they propose that the value of θ_2 that minimises the estimated treatment by period interaction is sought. The resulting value, say θ_{20} , is then taken as the presumed value of θ_2 in an analysis based on (7) and (8) which now excludes by construction the possibility of interaction. Bootstrap methods can be used to estimate the uncertainty attached to θ_{20} .

Ho et al's suggestion was aimed at crossover trials, where washout periods are specifically included in carefully designed trials precisely to limit the possibility of interaction. It can be applied more generally however, if we are prepared to make a zero interaction assumption. In many cases it may be more straightforward to discuss such an assertion with subject-specialist collaborators than to discuss possible values of the more abstract θ_2 .

We will refer to the use of the inverse probability weighted estimating equations (7) and (8) with $\theta_2 = \theta_{20}$ specified to give zero estimated interaction as *IPZI*. To get some idea of sensitivity we also propose that the parameters be estimated at $\theta_2 = \theta_{20} + 2SE(\theta_{20})$ (*IPZI+*) and at $\theta_2 = \theta_{20} - 2SE(\theta_{20})$ (*IPZI-*).

2.2 Testable assumptions and the role of time

Most of the foregoing approaches are based on variants of selection models, in that they are rooted in the distributions $f(Y_1, Y_2)$ and $f(R|Y_1, Y_2)$. Pattern mixture alternatives^{4,30} would be based on the marginal $f(R)$ and a mixture over R of and $f(Y_1, Y_2|R)$, noting that while $f(Y_1|R=0)$ is identifiable, $f(Y_2 | Y_1, R=0)$ is not.

In Section 1 we stated that dropout in longitudinal trials should be treated as a different type of missing data problem, rather than as a special case of a more generic approach. Both pattern mixture and selection model classes are applicable to generic missing data problems and neither is explicitly based on the fundamental feature of longitudinal studies: events happen and conditions change as time proceeds inexorably. In this section we consider how best to recognise temporal development, and, pre-empting later detailed discussion, suggest that there can be situations where avoiding untestable assumptions can lead to clearer and potentially more pertinent understanding of a longitudinal study.

Our suggestion is to employ only those features of pattern mixture and selection models that do not require untestable assumptions, and in neither case to combine them. An unavoidable feature of this approach is added complexity. Crucially, this complexity lies not in mathematical subtleties, but in the required depth when reporting study findings.

Pattern mixture models offer several advantages in a longitudinal dropout context. Most importantly, definitions of random variables are entirely unambiguous, for longitudinal responses are only modelled conditional on their existence, obviating the need for subtle counterfactual descriptions¹¹. The mixing is done over a clearly identified distribution — the marginal distribution of the dropout indicator R — and familiar tools for longitudi-

nal data analysis may be used to model the conditional distributions $f(Y_1|R = 0)$ and $f(Y_1, Y_2 | R = 1)$. However, making the latter distribution the subject of a model has the disadvantage that it has no mechanistic interpretation. Because the observation of Y_1 necessarily precedes the realisation of R , the model can provide no insight into the causal mechanisms that may underlie the data. The reverse is true, too: any substantive insights into the causal mechanisms linking previous responses to subsequent dropout cannot easily be incorporated into pattern mixture models. Aalen and colleagues^{31,32} have written about the need for greater recognition of the special role of time, and the ability to reason causally, in the analysis of longitudinal data.

Selection models⁴ *do* admit causal (or mechanistic) thinking, by modelling the probability of dropout conditional on previous observed responses, $f(R = 0 | Y_1)$. The corresponding disadvantage comes when attempting to combine these dropout probabilities into a full, joint likelihood: we then require marginal joint distributions of potentially unobserved random variables (e.g. $f(Y_1, Y_2)$) with the accompanying ambiguities of definition.

We can follow a middle path by avoiding the mixing aspect and not attempting to specify a full joint likelihood. Consequently, we are able to restrict ourselves to models requiring no untestable assumptions. Moreover, the features of the analysis strategy we suggest have a natural temporal progression, allowing causal thinking at each stage. We propose

1. **Baseline analysis: model $f(Y_1)$.** The first observation is observed for all subjects, and yields the first opportunity to consider covariate effects on the longitudinal response.
2. **Dropout analysis: model $f(R = 0 | Y_1)$.** The observation of Y_1 occurs strictly before R is realised, and is therefore admissible (since it is previsible) as a covariate in a (discrete-time) event history analysis. Such a model might tell us, for instance, that extreme responders at time 1 were more likely to drop out of the longitudinal study.
3. **Complete case analysis: model $f(Y_2 | R = 1, Y_1)$.** For those who continue to time 2, we can quite legitimately (and without untestable assumptions) describe how their responses relate to their prior observation. Alternatively but essentially equivalently, we might choose to model the changes $f(Y_2 - Y_1 | R = 1, Y_1)$.

The first two points are uncontroversial. In subsequent sections we shall consider in greater detail the third component, including our claim that scientific understanding may in some instances be no less complete when considering only complete cases at the second timepoint.

Our temporal sequence is reminiscent of the careful likelihood construction of Tsiatis and Davidian³³. By deliberately avoiding combining these components, we do not specify a joint likelihood for R, Y_1 and Y_2 . More than this, we *cannot* uniquely specify such a likelihood without further structures or identifying assumptions. Our feeling is that this is a strength, not a weakness.

3 Illustrations

We consider data from two clinical trials, each with two treatments and two measurement times.

3.1 A longitudinal trial with dropout

We first consider results on 422 subjects in an AA/BB longitudinal trial. The data are real but not public. The trial was randomized and observations were taken pre-randomization and at two times post-randomization: in this paper we consider only the post-randomization observations. For confidentiality reasons we have scaled all observed responses by the mean and standard deviation of the responses at time 1 and we will refer simply to treatment and response without describing what they are.

There are no covariates other than treatment type and time. Without treatment by time interaction the parameter vector is $\beta = (\mu, \pi, \tau)$, where μ is a general mean, π measures the effect of time and τ measures the treatment effect. The design matrices X_i are all either

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \end{pmatrix}$$

depending on whether the subject is given treatment A or B respectively. Interest is ostensibly in τ .

All 422 subjects provided a response at time 1. Of 212 subjects receiving Treatment A, only 126 provided a response at time 2 and the other 86 dropped out. We have no information on reason for dropout. Of 210 subjects receiving Treatment B, 193 continued to time 2 and the other 17 did not.

Treatment effect estimates using the techniques described in Section 2 are shown in Figure 1 together with nominal 95% confidence limits. The estimates correspond to the methods summarised in Section 2. The strongest treatment effect is evident under LOCF, with complete case (CC) providing the least evidence of a difference between treatments. The four inverse probability weighted methods are rather similar, and in most cases these techniques have higher standard errors.

It is interesting to explore these data a little further. Figure 2 show a variety of mean responses, alongside each of which is given the number of observations used in the calculation. Standard deviations are not given but all are relatively small. Panel (a) of the figure is essentially that used in the original trial report to summarise the data. It uses LOCF to interpolate missing values at time 2, so that means are formed from the same number of observations at each time point. Evidently there is a large treatment effect. Panel (b) shows the means of only the observed data. Now the treatment effect seems to disappear at time 2. Its presence at time 1 is not surprising because subjects had already been on treatment for some time before this response was collected, it was not a baseline score. Panel (c) is the most informative presentation of the data: it separates the subjects at time 1 into those who subsequently drop out and those who continue to time 2. Subjects who drop out tend to have lower scores at time 1. Amongst the complete cases there is no evidence of a treatment effect.

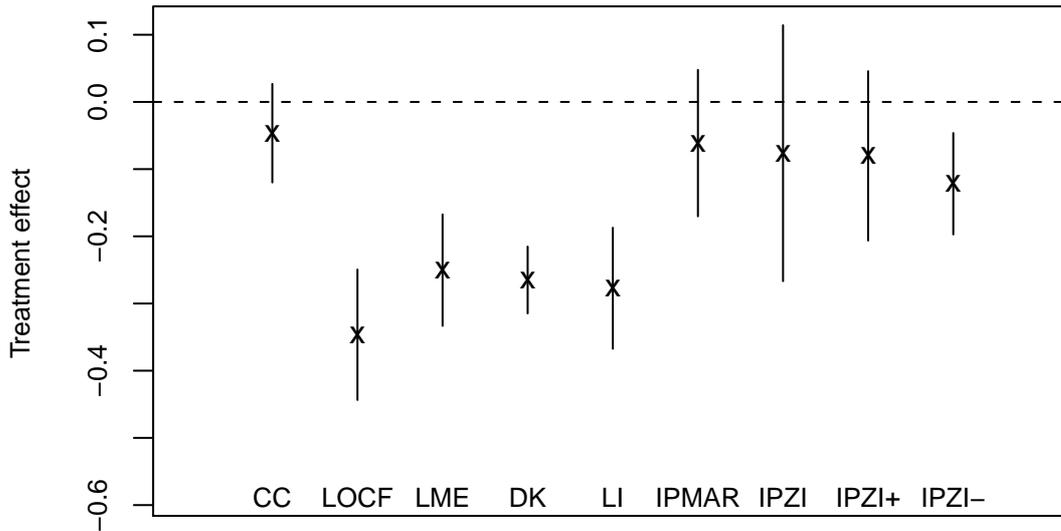


Figure 1: Longitudinal trial: selection of estimated treatment effects and nominal 95% confidence intervals

3.2 A crossover trial with dropout

Our second example is on data from an AB/BA crossover trial. The trial is reported in³⁴ and is used to illustrate the Zero Interaction proposal²⁹. The treatments are analgesics: dihydrocodeine (A) and nabilone (B) and the response was a visual analogue score for pain.

At time 1 we have data on 37 subjects in the AB group and 45 in the BA group. Five of the AB subjects and 10 of the BA subjects dropped out before the time 2 response was recorded. The trial, in common with many crossover trials, required the subjects to receive treatment over a fairly long time interval, in this case six weeks. Thus the subjects who dropped out before the time 2 response are extremely likely to have had some experience of their second treatment.

There are again no covariates other than treatment type and time. Without treatment by time interaction the parameter vector is $\beta = (\mu, \pi, \tau)$ and the design matrices X_i are all either

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

depending on whether the subject is given treatment in order AB or BA respectively.

Figure 3 shows the range of treatment effects obtained using the methods we have described.

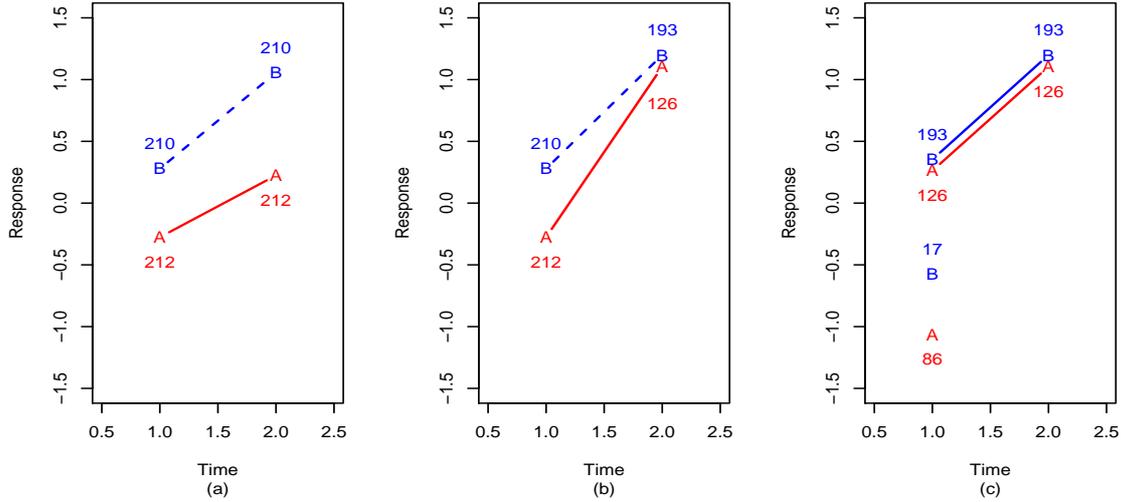


Figure 2: Longitudinal trial: means and sample sizes used for calculation. Panel (a): as reported, using LOCF. Panel (b): observed data. Panel (c): completers and dropouts separated at time 1.

It seems there are relatively little differences between estimates, which is not surprising as the proportion of dropouts is low (15/92 or 16%). However, some of the differences are marked and there is wide variation in estimated standard errors. The estimated treatment effect under complete case for instance is $\hat{\tau} = 2.85$ (SE 1.19), which is little more than half that obtained under the ZI procedure ($\hat{\tau} = 4.862$, SE 2.68).

4 Complete-Case Analyses

4.1 Relationship of CC estimators to MNAR estimators

If the distribution of $Y_{i2}(0)$ is not the same as that of $Y_{i2}(1)$, then an analysis that attempts to compensate for the missing data, for example by assuming data are MNAR, is of no merit. If the data were not missing, then in almost all circumstances they would be expected to have a different distribution. In such instances the temporal strategy of Section 2.2, including a CC analysis, should be preferred. Nevertheless, effect on the required inferences of this choice is of practical importance, so an exploration of the properties of a CC analysis when the data are MNAR is of interest.

One such property is the expectation of the estimators of the treatment effect, $\hat{\tau}_{LONG}$ for the longitudinal study, and $\hat{\tau}_{CO}$ for the crossover design from a CC analysis. For the simple two-period designs the estimators are

$$\hat{\tau}_{LONG} = \frac{1}{4}(\bar{S}_{AA} - \bar{S}_{BB}) \quad \hat{\tau}_{CO} = \frac{1}{4}(\bar{D}_{AB} - \bar{D}_{BA}),$$

where \bar{S}_{AA} denotes the mean of the $S_i = Y_{i1} + Y_{i2}$ for complete cases in sequence AA, with the other terms defined similarly and with $D_i = Y_{i1} - Y_{i2}$. These estimators are unbiased

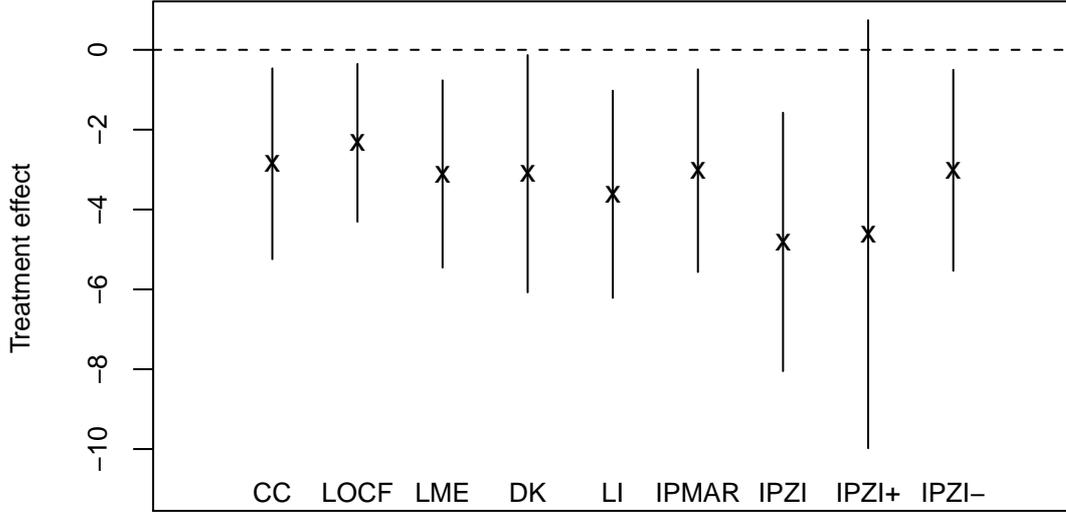


Figure 3: Crossover trial: selection of estimated treatment effects and nominal 95% confidence intervals

in the sample space of complete cases and we denote their expectations by τ_{LONG}^{CC} and τ_{CO}^{CC} , respectively. These parameters will, in general, differ from the treatment effect τ introduced in Section 3.1 that would apply if, e.g., a MNAR analysis were used when the distribution of $Y_{i2}(R_i = 0)$ did coincide with that of $Y_{i2}(R_i = 1)$. Analytic approximations for the relationship between these parameters can be derived if we are prepared to assume that the distribution of $(Y_1, Y_2)^T$ is bivariate Normal with density $f(Y_1, Y_2)$.

The key quantities required are $E(Y_{i1} + Y_{i2} | R_i = 1)$ for each of the sequences in the longitudinal trial, and $E(Y_{i1} - Y_{i2} | R_i = 1)$ for the sequences of the crossover design. Omitting the subscript i , these can be evaluated as

$$E(Y_2 \pm Y_1 | R = 1) = \frac{\int \int (Y_1 \pm Y_2) \text{expit}(\theta_0 + \theta_1 Y_1 + \theta_2 Y_2) f(Y_1, Y_2) dY_1 dY_2}{\int \int \text{expit}(\theta_0 + \theta_1 Y_1 + \theta_2 Y_2) f(Y_1, Y_2) dY_1 dY_2}.$$

If the unconditional mean of Y_{ij} in sequence rs is μ_{rsj} then the above can be written as

$$\mu_{rs1} \pm \mu_{rs2} + \frac{\int \int (z_1 \pm z_2) \text{expit}(\Theta_{rs} + \theta_1 z_1 + \theta_2 z_2) \phi_2(z_1, z_2 | \Omega) dz_1 dz_2}{\int \int \text{expit}(\Theta_{rs} + \theta_1 z_1 + \theta_2 z_2) \phi_2(z_1, z_2 | \Omega) dz_1 dz_2} \quad (9)$$

where $\Theta_{rs} = \theta_0 + \theta_1 \mu_{rs1} + \theta_2 \mu_{rs2}$ and $(Z_1, Z_2)^T$ is a bivariate Normal variable with mean zero, dispersion matrix Ω and density $\phi_2(z_1, z_2 | \Omega)$. We will assume that the variances in Ω are both equal to σ^2 and that the covariance is $\rho\sigma^2$. The integrals can be approximated if we note that $\text{expit}(u) \approx \Phi(cu)$ where $\Phi(\cdot)$ is the standard Normal distribution function

and $c = (16\sqrt{3})/(15\pi)^{35}$. A convenient method of evaluation is to note that $\Phi(c[\Theta_{rs} + \theta_1 z_1 + \theta_2 z_2])\phi_2(z_1, z_2 | \Omega)$ is proportional to the density of an extended skew-Normal (ESN) distribution (see Appendix and³⁶), and the normalising constant is the denominator of (9), after the application of the approximation to $\text{expit}(\cdot)$.

Applying the formula for the expectation of the ESN we obtain

$$\mathbb{E}(Z_1 \pm Z_2) \approx ck_0^{-1}e_{\pm}^T \Omega \theta_{12} \zeta(c\Theta_{rs}/k_0) = \frac{c\sigma^2}{k_0}(1 \pm \rho)(\theta_1 \pm \theta_2)\zeta(c\Theta_{rs}/k_0)$$

where $\theta_{12} = (\theta_1, \theta_2)^T$, $k_0 = \sqrt{1 + c^2\sigma^2(\theta_1^2 + \theta_2^2 + 2\rho\theta_1\theta_2)}$ and $e_{\pm}^T = (1, \pm 1)$. Also $\zeta(u) = \phi(u)/\Phi(u)$, with $\phi(u)$ denoting the density of a standard Normal variable.

From these expressions we can approximate τ_{LONG}^{CC} and τ_{CO}^{CC} , using the parametrisation introduced in Section 3.1, when the data are MNAR with continuation probability $\text{expit}(\theta_0 + \theta_1 Y_1 + \theta_2 Y_2)$. For the longitudinal design we obtain

$$\tau_{LONG}^{CC} \approx \tau + \frac{1}{4} \frac{c\sigma^2}{k_0} (1 + \rho)(\theta_1 + \theta_2) \left(\zeta\left(\frac{c\Theta_{AA}}{k_0}\right) - \zeta\left(\frac{c\Theta_{BB}}{k_0}\right) \right), \quad (10)$$

and for the crossover design

$$\tau_{CO}^{CC} \approx \tau + \frac{1}{4} \frac{c\sigma^2}{k_0} (1 - \rho)(\theta_1 - \theta_2) \left(\zeta\left(\frac{c\Theta_{AB}}{k_0}\right) - \zeta\left(\frac{c\Theta_{BA}}{k_0}\right) \right). \quad (11)$$

For the longitudinal design $\Theta_{AA} = \Theta + \tau(\theta_1 + \theta_2)$ and $\Theta_{BB} = \Theta - \tau(\theta_1 + \theta_2)$, with $\Theta = \theta_0 + \mu(\theta_1 + \theta_2) + \pi(\theta_1 - \theta_2)$, whereas for the crossover design we have $\Theta_{AB} = \Theta + \tau(\theta_1 - \theta_2)$ and $\Theta_{BA} = \Theta - \tau(\theta_1 - \theta_2)$. If the deviations of the Θ_{rs} from Θ are small then (10) and (11) can be written as

$$\tau_{LONG}^{CC} \approx \tau \left(1 + \frac{1}{2} \frac{c^2\sigma^2}{k_0^2} (1 + \rho)(\theta_1 + \theta_2)^2 \zeta'\left(\frac{c\Theta}{k_0}\right) \right) \quad (12)$$

$$\tau_{CO}^{CC} \approx \tau \left(1 + \frac{1}{2} \frac{c^2\sigma^2}{k_0^2} (1 - \rho)(\theta_1 - \theta_2)^2 \zeta'\left(\frac{c\Theta}{k_0}\right) \right) \quad (13)$$

Since $\zeta'(u) = -\zeta(u)(u + \zeta(u)) < 0$ for all u , it follows that both τ_{LONG}^{CC} and τ_{CO}^{CC} are shrunk versions of τ , i.e. they are closer to zero than τ . Consideration of (10) and (11) shows that the biases operate in the same direction even when the deviations are not small.

If the treatment effect τ vanishes then it follows that $\Theta_{rs} = \Theta$ and hence τ_{LONG}^{CC} and τ_{CO}^{CC} also vanish. These expressions also reveal that CC analysis and the MNAR analysis coincide if the data are MCAR ($\theta_1 = \theta_2 = 0$). They do not agree when the data are MAR but not MCAR ($\theta_2 = 0, \theta_1 \neq 0$). However, if data are MNAR they do agree in some special cases. For the longitudinal design this occurs when $\theta_1 = -\theta_2$ and, for the crossover design, when $\theta_1 = \theta_2$. Some insight into this circumstance can be found by rewriting (4) as

$$\Pr(R_i = 1 | Y_{i1}, Y_{i2}) = \text{expit}\left(\theta_0 + \frac{1}{2}(\theta_1 - \theta_2)D_i + \frac{1}{2}(\theta_1 + \theta_2)S_i\right).$$

If $\theta_1 = \theta_2$ then the dropout probability depends on the data only through S_i , that is the chance of dropping out depends on the mean level of the response, whereas if $\theta_1 = -\theta_2$ then the dropout probability depends only on the increments in response, D_i . In the longitudinal design the estimate of the treatment effect depends solely on the mean level of the response.

If $\theta_1 = -\theta_2$, then the probability of dropping out of the study will depend only on the increments in response, and our error model ensures that S_i and D_i are independent, so the coincidence of τ and τ_{LONG}^{CC} is unsurprising. By contrast in the crossover design, the MNAR and the CC estimands coincide if the dropout probability depends only the S_i . This result is reminiscent of Little’s observation that in a pattern-mixture model intended to estimate the mean difference between period 1 and period 2, a CC analysis is to be preferred if the dropout probability depends on S_i ³⁷.

4.2 Meaning of parameters

When interpreting the relationships derived above it is important to remain clear about the meanings of the parameters τ , τ_{LONG}^{CC} and τ_{CO}^{CC} . In Section 3.1 it was stated that τ is the parameter of interest. However, while this would be so for most standard analyses, a case can be made that this is not always the correct target for our inference.

Consider a crossover design to compare two active treatments, in which some subjects are unable to tolerate one or other of the treatments. In this case the response that would have been observed had the subject tolerated the treatment is not meaningful. In these circumstances a MNAR analysis is inappropriate. On the other hand a CC analysis might be justified on the grounds that it compares subjects who can tolerate both treatments. The validity of this analysis stems from the fact that it is precisely for these subjects that the trial comparison is relevant. The decision to specialise the analysis to this subgroup means that the appropriate target for inference is τ_{CO}^{CC} .

Even in circumstances where it makes sense to identify $Y_t(1)$ and $Y_t(0)$, a MNAR approach may not be the right analysis. Consider for instance an analgesic trial in which all subjects can tolerate both treatments, but some drop out because of a side-effect which makes the treatment impractical. For example, subjects on a particular treatment may not be allowed to drive and hence some may refuse the treatment. While one can now sensibly believe that both potential outcomes have the same distribution, the side-effect means that both treatments are a practical proposition only for a subset of subjects and the most pertinent analysis is provided by the complete cases. In this case it is doubtful if the dropout model (4) with the outcomes as the only covariates would adequately describe the missingness process. It could be argued that it is unfair to dismiss a MNAR analysis if the dropout model is mis-specified but, nevertheless, in such an example a CC analysis has the advantages that it is practical, germane and has a readily described rationale.

5 Discussion

In Section 4.2 we argued that that there are circumstances where the complete case parameters τ_{CO}^{CC} or τ_{LONG}^{CC} are the natural parameters of interest. In other situations the underlying parameter τ may indeed be the original target for inference but if the data are blighted by significant dropout it may be unreasonable to attempt to estimate this inestimable parameter given the data to hand. In these circumstances case a CC analysis may be the only

practical alternative, but if τ really is the appropriate target for inference, is this adequate? If the overall level of missingness is low then a CC analysis is unlikely to differ substantially from more sophisticated analyses. Indeed, analyses that link the estimation of treatment parameters to parameters describing dropout may be compromised if there is limited information on the dropout process. On the other hand, as the proportion of the patients in the trial who are retained in a CC falls, the CC analysis will provide useful information about a smaller and rather different population from that originally envisaged. However, as we have argued, this subset of patients may be of particular relevance. In these circumstances a supplementary analysis that seeks to determine the factors affecting dropout becomes necessary.

Consider the longitudinal data of Section 3.1. If the sole focus is on τ then Figure 1 shows that we can choose any or none of a range of estimates. But Figure 2 summarises the true picture and we suggest that an appropriate analysis should proceed as described in Section 2.2. First we analyse Y_1 only and find and report a strong treatment effect to this time of 0.55 units with standard error 0.09. Such an analysis is sensible for these data because the time 1 measurement is not baseline as the subjects had already received treatment before the measurement was obtained. Second we model dropout as a logistic model in either Y_1 only or, with the remarks of Section 4.2 in mind, in Y_1 and treatment (or other covariates). Omitting details, it turns out that for these data there are strong effects on dropout of both Y_1 and treatment group. These should be discussed with the experimenter. Finally, we perform a complete case analysis and state that *for subjects who continue to the second time point* there is no evidence of difference between treatment groups in mean response levels at *either* time point. The conclusions are more complex than initially desired but then so are the data.

A CC analysis of a randomized trials will suffer from the complication that the groups of complete cases will not exhibit the balance conferred by the initial randomization. The effect of this on inferences will need to be considered carefully if a CC analysis is being considered as the primary analysis. In our examples this issue may be less of a concern in the CO study, as here each patient is essentially used as their own control. An imbalance between the sequences could masquerade as a treatment by period interaction but, if the investigator is confident that the design has excluded carryover effects, such interactions are likely to represent differences in overall level of response, and the D_i will be unaffected. The problem of loss of balance in randomized groups is likely to pose a greater challenge for the parallel-group design.

An abiding concern of investigators is the loss of efficiency that results from missing data. This problem may seem to be exacerbated if a CC analysis is used, as data from partially observed cases is ignored. However, the relative efficiencies of alternative analyses should only be a concern if both approaches estimate the same quantity. If data are MNAR then methods such as those exemplified in Section 2 will not, in general, estimate the same quantity. Consequently, the relative performance of these methods cannot be assessed through simple comparisons of standard errors.

While there may be arguments in favour of CC analyses, there could be circumstances where what constitutes a CC analysis is unclear. In our examples matters are straightforward but consider a crossover trial comparing three treatments over three periods. Suppose that it

is required to estimate the contrast between two of the treatments. Does a CC analysis retain only patients providing three observations, or should we use all patients who provide information on the two treatments being compared? The issue becomes even murkier if we consider long sequences of irregularly timed measurements, especially common in observational data rather than designed trials.

Acknowledgment

This paper grew from an invited presentation at the 2011 Erasmus University Spring Symposium: Joint Modeling Techniques. We are grateful to the organisers, Dimitris Rizopoulos and Emmanuel Lesaffre. We are grateful to Dr Bernhard Frank for permission to use the data from the crossover trial. We thank two anonymous referees for their perceptive comments.

Appendix: Properties of the extended skew-Normal distribution

A p -dimensional random variable U has an extended skew-Normal distribution, $ESN(\xi, \Sigma, \alpha, \nu)$, if it has density

$$\frac{\phi_p(u; \xi, \Sigma) \Phi(\alpha^T(u - \xi) + \nu)}{\Phi(\bar{\nu})}, \quad (14)$$

where the parameters ξ and α are p -dimensional vectors, ν is a scalar, Σ is a $p \times p$ dispersion matrix and $\bar{\nu} = \nu(1 + \alpha^T \Sigma \alpha)^{-\frac{1}{2}}$. Here $\Phi(\cdot)$ is the CDF of a univariate standard Normal variable and $\phi_p(\cdot; \xi, \Sigma)$ is the density of a p -dimensional Normal variable with mean ξ and dispersion Σ . It follows that

$$E(U) = \xi + \frac{\Sigma \alpha}{\sqrt{1 + \alpha^T \Sigma \alpha}} \frac{\phi(\bar{\nu})}{\Phi(\bar{\nu})},$$

and $\phi(\cdot)$ is the density of a univariate standard Normal variable.

References

- 1 Laird NM. Missing data in longitudinal studies. *Statistics in Medicine* 1988; **7**:305–315, doi:10.1002/sim.4780070131.
- 2 Little RJA. Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association* 1995; **90**:1112–1121.
- 3 Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004; **5**:445–464, doi:10.1093/biostatistics/kxh001.
- 4 Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*. Wiley: Chichester, 2007, doi:10.1002/9780470510445.

- 5 Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modelling and Sensitivity Analysis*. Chapman and Hall/CRC Press: Boca-Raton, 2008.
- 6 Philipson PM, W-K Ho, Henderson R. Comparative review of methods for handling drop-out in longitudinal studies. *Statistics in Medicine* 2008; **27**:6276–6298, doi:10.1002/sim.3450.
- 7 European Medicines Agency. *Guideline on Missing Data in Confirmatory Clinical Trials: EMA/CPMP/EWP/1776/99 Rev 1*. European Medicines Agency: London, 2010.
- 8 National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press: Washington, DC, 2010.
- 9 Molenberghs G, Beunckens C, Sotito C, Kenward MG. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society, series B* 2008; **70**:371–388, doi:10.1111/j.1467-9868.2007.00640.x.
- 10 Robins J, Gill R. Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine* 1997; **16**:39–56.
- 11 Diggle P, Farewell D, Henderson R. Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal (with discussion). *Applied Statistics* 2007; **56**:499–550, doi:10.1111/j.1467-9876.2007.00590.x.
- 12 Aalen O, Gunnes N. A dynamic approach for reconstructing missing longitudinal data using the linear increments model. *Biostatistics* 2010; **11**:453–472.
- 13 Meinert C. Terminology—a plea for standardization. *Controlled Clinical Trials* 1980; **1**(2):97.
- 14 Dufoil C, Brayne D, Clayton D. Analysis of longitudinal studies with death and drop-out: a case study. *Statistics in Medicine* 2004; **23**:2215–2226, doi:10.1002/sim.1821.
- 15 Didelez V. Discussion of: Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal? (by P. Diggle, D. Farewell and R. Henderson). *Applied Statistics* 2007; **56**:536–537.
- 16 Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974; **66**(5):688.
- 17 Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine* 2010; **29**:3245–3257.
- 18 Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, series A* 2006; **169**:571–584, doi:10.1111/j.1467-985X.2006.00407.x.
- 19 Diggle PJ, Kenward MG. Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* 1994; **43**:49–93, doi:10.2307/2986113.
- 20 Little RJA. Pattern mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; **88**:125–134.

- 21 Farewell DM. Marginal analyses of longitudinal data with an informative pattern of observations. *Biometrika* 2010; **97**:65–78.
- 22 Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression-models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; **90**:106–121, doi:10.2307/2291134.
- 23 Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression-models with missing data. *Journal of the American Statistical Association* 1995; **90**:122–129, doi:10.2307/2291135.
- 24 Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* 1998; **93**:1321–1339, doi:10.2307/2670049.
- 25 Rotnitzky A, Scharfstein DO, Su TL, Robins JM. Methods for conducting sensitivity analysis of trials with potentially non-ignorable competing causes of censoring. *Biometrics* 2001; **57**:103–112.
- 26 Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for non-ignorable drop-out using semiparametric non-response models (with discussion). *Journal of the American Statistical Association* 1999; **94**:1096–1146.
- 27 Scharfstein DO, Daniels M, Robins JM. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* 2003; **4**:495–512.
- 28 White IR, Higgins JP, Wood AM. Allowing for uncertainty due to missing data in meta-analysis—part 1: two-stage methods. *Statistics in Medicine* 2008; **27**:711–727.
- 29 Ho WK, Matthews JNS, Henderson R, Farewell D, Rodgers LR. Dropouts in the AB/BA crossover design. *Statistics in Medicine* 2012; **to appear**, doi:10.1002/sim.4497.
- 30 Little R. Selection and pattern-mixture models. *Longitudinal Data Analysis*, Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds.), Chapman and Hall/CRC Press: Boca Raton, 2008; 409–432.
- 31 Aalen OO. Armitage lecture 2010: Understanding treatment effects: The value of integrating longitudinal data and survival analysis. *Statistics in Medicine* 2012; **to appear**.
- 32 Aalen OO, Røysland K, Gran J, Ledergerber B. Causality, mediation and time; a dynamic viewpoint. *J Royal Statistical Society Series A* 2012; **to appear**.
- 33 Tsiatis A, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; **14**(3):809–834.
- 34 Frank B, Serpell MG, Hughes J, Matthews JNS, Kapur D. Comparison of analgesic effects and patient tolerability of nabilone and dihydrocodeine for chronic neuropathic pain: randomised, crossover, double blind study. *British Medical Journal* 2008; **336**:199–201, doi:10.1136/bmj.39429.619653.80.

- 35 Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions, Volume 2*. 2nd edn., Wiley: Chichester, 1995.
- 36 Arnold BC, Beaver RJ. Hidden truncation models. *Sankhyā A* 2000; **62**:23–35.
- 37 Little RJA. A class of pattern-mixture models for normal incomplete data. *Biometrika* 1994; **81**:471–483.