Coping with missing data in clinical trials: A model-based approach applied to asthma trials

James Carpenter\(^1\)^\(^*\), Stuart Pocock\(^1\) and Carl Johan Lamm\(^2\)

\(^1\)Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, U.K.
\(^2\)AstraZeneca, S-221 87, Lund, Sweden

SUMMARY

In most clinical trials, some patients do not complete their intended follow-up according to protocol, for a variety of reasons, and are often described as having ‘dropped out’ before the conclusion of the trial. Their subsequent measurements are missing, and this makes the analysis of the trial’s repeated measures data more difficult. In this paper we briefly review the reasons for patient drop-out, and their implications for some commonly used methods of analysis. We then propose a class of models for modelling both the response to treatment and the drop-out process. Such models are readily fitted in a Bayesian framework using non-informative priors with the software \textsc{Bugs}. The results from such models are then compared with the results of standard methods for dealing with missing data in clinical trials, such as last observation carried forward. We further propose the use of a time transformation to linearize an asymptotic pattern of repeated measures over time and therefore simplify the modelling. All these ideas are illustrated using data from a five-arm asthma clinical trial. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: clinical trial; repeated measures; drop-out

1. INTRODUCTION

By their nature, clinical trials involve the follow up of patients over a period of time so that their response to particular treatments can be described. It is inevitable that some patients will drop-out before completing the trial, either by discontinuing their prescribed course of treatment (non-compliance) or by ceasing to be evaluated (drop-out) or both. Here we consider drop-out. Since the subsequent measurements on patients who drop out are missing, the data analysis is no longer straightforward.

\(^*\)Correspondence to: James Carpenter, Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, U.K.
\(^\dagger\)E-mail: james.carpenter@lshtm.ac.uk

Contract/grant sponsor: AstraZeneca

Copyright © 2002 John Wiley & Sons, Ltd.
In general, data from such trials can be analysed in four ways:

1. Discard data from all patients who did not complete the trial and analyse the remaining data.
2. Analyse only the observed data.
3. Use a single or multiple imputation to replace the missing observations with plausible values, then analyse the ‘completed’ data set.
4. Build a longitudinal model for the data which includes a model for the drop-out process.

Option 4 is usually the most complex computationally, but it is also the most useful, as it elucidates the often unexpectedly subtle assumptions behind the others, and allows the sensitivity of the conclusions to assumptions about the dependence between drop-out and response to be assessed.

In this paper we briefly review the reasons for patients discontinuing a trial and the implications for the more common methods of analysis. We then propose using a class of models similar to those discussed by Diggle and Kenward [2] to tackle option 4. We fit these models using the BUGS [1] software using non-informative priors, since this approach is both easier to program and, once programmed, quicker to estimate. Moreover, it is flexible; a variety of models can be fitted without the effort required for familiar approaches which use the EM algorithm. Throughout, we use data from a five-arm asthma clinical trial to illustrate our ideas. The outline of the paper is as follows. Section 2 describes the asthma trial data used to illustrate our approach. Section 3 reviews how drop-outs arise and describes the implications for analysis. Simple graphical and tabular methods that can help identify the nature of the drop-out mechanism are described in Section 4 and this is extended in Section 5 with logistic regression models. Sections 6 and 7 describe our class of models. They also detail our proposal for transforming the time variable, when the response to treatment appears to have an asymptote, so that the response is approximately linear with time. Finally, Section 8 presents the results of applying such models to the asthma trial and Section 9 draws some conclusions.

2. DATA

Throughout, we illustrate our ideas with data from a five-arm asthma clinical trial to assess the efficacy and safety of budesonide, a second-generation glucocorticosteroid, on patients with chronic asthma. A total of 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.

The two outcome variables of particular interest were forced expiratory volume in 1 second (FEV$_1$) and peak expiratory flow (PEF). FEV$_1$ represents the maximum volume of air, in litres, an individual can exhale in one second. Usually, it can only be recorded at clinic visits. PEF represents the maximum rate an individual can exhale air, in litres per second. Unlike FEV$_1$, it can be readily recorded by an individual on a twice daily basis at home, using a peak flow meter.

FEV$_1$ was recorded at clinic visits at baseline, 2, 4, 8 and 12 weeks and PEF was recorded by individuals morning and evening, together with asthma severity score. The number of puffs of bronchodilator taken was recorded once daily. Treatment efficacy was assessed by monitoring change in FEV$_1$ and PEF from baseline.
In summary, the trial found a statistically significant dose–response effect for the mean change from baseline over the study for both morning PEF, evening PEF and FEV1 at the 5 per cent level. Budesonide treated subjects also showed reduced asthma symptoms and bronchodilator use compared with placebo, while there were no clinically significant differences in treatment related adverse experiences between the treatment groups. Further details about the conduct of the trial, its conclusions and the variables collected can be found elsewhere [3]. Here, we focus on the response FEV1.

For an individual, the importance of a change in FEV1 depends on his or her expected FEV1. The latter can be estimated from an individual’s height, age and gender using an established formula [4]. Thus actual FEV1 as a percentage of that predicted is a useful dimensionless measure of asthma severity. In this paper we therefore take as the response, change in per cent predicted FEV1.

3. DROP-OUT IN CLINICAL TRIALS

3.1. Notation

Suppose \( i \in 1, \ldots, N \) subjects are randomized into a trial and that each subject has their response measured at times \( t_{ij}, j \in 1, \ldots, J \), on and after randomization. Set \( t_{i1} = 0 \). Let the change in the \( i \)th individual’s (univariate) primary response from baseline at time \( t_{ij} \) after randomization be denoted by \( y_{ij} \), so that \( y_{i1} = 0 \). Let the column vector \( y_i = (y_{i1}, \ldots, y_{ij})^T \) denote the complete set of primary responses for that individual. Further, let the column-vector \( x_i \) denote baseline covariates measured on the \( i \)th individual. Then the data are represented by \((y_i, x_i)\).

For example, in the asthma trial, \( y_i \) denotes change in per cent predicted FEV1, \( x_i \) denotes baseline readings such as age, sex, blood pressure and the degree of asthma reversibility.

Now let \( M_{ij} \) be an indicator variable being 1 if measurement \( j \) on individual \( i \) is missing, and zero otherwise. For clarity, in this discussion we assume that once an individual drops out, they do not return, so that the drop-out vector \( M_i = (M_{i1}, M_{i2}, \ldots, M_{ij}) \) consists of a set of 0’s followed by 1’s. Of course, an individual who completes would have a complete set of 0’s.

The complete information on a patient can be viewed as a combination of observed and unobserved, or missing, components which we identify with the superscripts obs and miss, respectively. Thus the complete data on an individual, \((y_i, x_i)\), can be rewritten as \((y_{i}^{\text{obs}}, y_{i}^{\text{miss}}, M_i, x_i)\), where the combined length of the vector \((y_{i}^{\text{obs}}, y_{i}^{\text{miss}})\) is \( J \).

3.2. Probability model for drop-out

We now describe a framework for modelling the data, which enables us to elucidate the three general categories of missing data proposed by Little and Rubin [5] and Diggle and Kenward [2]. We then make the link with actual scenarios that arise in clinical trials.

Divide the parameter space into two subsets \( \theta \) and \( \phi \). Let \( \theta \) be the subset of parameters relating the outcome \( y \) to the covariates \( x \), and \( \phi \) be the subset of parameters relating drop-out \( M \) to the same covariates \( x \). Note that these two subsets may have some parameters in common.
Let \( P(y, M; \theta, \phi, x) \) represent the joint likelihood (probability) of the data, given the baseline covariates, at particular values of the parameter vectors. Then, using the notation above

\[
P(y, M; \theta, \phi, x) = P(y_{\text{obs}}^i, M; \theta, \phi, x)
\]

(1)

Since this likelihood depends on missing data, it cannot be maximized in the usual way. A way around this problem is to calculate from the joint distribution (1), the distribution of \( (y_{\text{obs}}^i, M) \) alone by integrating out \( y_{\text{miss}}^i \) to give

\[
P(y_{\text{obs}}^i, M; \theta, \phi, x) = \int P(y_{\text{obs}}^i, y_{\text{miss}}^i, M; \theta, \phi, x) dy_{\text{miss}}^i
\]

(2)

Since this integral is usually difficult, it makes sense to avoid it if possible. It turns out that if the data are missing for a particular class of reasons, the integral (2) can be avoided.

To elucidate these reasons, we need to consider the right hand side of (2). For clarity, we consider the likelihood for the \( i \)th individual alone. Since individuals are considered independent, the joint likelihood, as represented by (2), is then the product of the likelihoods for the individuals.

We can write the \( i \)th component of the integrand, a joint density, as either

\[
P(y_{\text{obs}}^i, y_{\text{miss}}^i, M_i; \theta, \phi, x_i) = P(M_i | y_{\text{obs}}^i, y_{\text{miss}}^i, \phi, x_i)P(y_{\text{obs}}^i, y_{\text{miss}}^i, \theta, x_i)
\]

(3)

or as

\[
P(y_{\text{obs}}^i, y_{\text{miss}}^i, M; \theta^*, \phi^*, x) = P(y_{\text{obs}}^i, y_{\text{miss}}^i | M_i, \theta^*, x_i)P(M_i; \phi^*, x_i)
\]

(4)

Equation (3) views the probability of drop-out as dependent on the response variable, \( y \); in particular, in a trials context, patient drop-out would depend on treatment response. In other words, patients are selected for drop-out by their response – hence models based on (3) are known as selection models.

Conversely, equation (4) has the probability distribution of the response depending on drop-out status, so that different patterns of response can be proposed for patients who drop out and continue. For this reason models based on (4) are known as pattern mixture models.

Although both models have their advantages [6–8], we believe that a selection model is the most natural for modelling drop-out in the asthma trial, since drop-out is often plausibly explained by a steady decline in a patient’s condition to a level at which they do not wish to participate in the trial any more. Henceforth, our discussion focuses on this.

We are thus concerned with (3). The key is the term \( P(M_i | y_{\text{obs}}^i, y_{\text{miss}}^i, \phi, x_i) \). The precise form of this defines the drop-out mechanism, and hence whether the integral in (2) can be avoided. In the spirit of Little and Rubin [5] and Diggle and Kenward [2] there are three drop-out mechanisms: drop-out completely at random; drop-out at random, and non-random drop-out. These, and their implications in a clinical trials context, are now described.

### 3.2.1. Drop-out completely at random (DCAR)

In this case, drop-out \( M_i \) does not depend on either the unobserved or observed responses, \( (y_{\text{obs}}^i, y_{\text{miss}}^i) \). Mathematically

\[
P(M_i | y_{\text{obs}}^i, y_{\text{miss}}^i, \phi, x_i) = P(M_i; \phi, x_i)
\]

(5)

which implies that, conditional on a complete set of relevant baseline covariates \( x_i \)

\[
P(y_{\text{obs}}^i, M_i; \theta, \phi, x_i) = P(M_i; \phi, x_i)P(y_{\text{obs}}^i, \theta, x_i)
\]

(6)
We begin with two observations on (6). First, the same covariates \( x_i \) in the joint density on the left hand side appear in the right hand side of each of the two terms in the product \( P(M_i; \phi, x_i)P(y_i^{\text{obs}}; \theta, x_i) \). Secondly, joint likelihood on the left hand side has factored into two terms, one involving the observed data and \( \theta \), and the other involving the drop-out indicator vector and \( \phi \).

The implication of the first observation is that the same set of baseline covariates \( x_i \) must be conditioned on in the marginal models for the response and drop-out, \( P(y_i^{\text{obs}}; \theta, x_i) \) and \( P(M_i; \phi, x_i) \), respectively. Hence if a baseline covariate is an important predictor of drop-out, it must be included in the model for the response, and vice versa. This fact differentiates the missing completely at random model for the response \( y_i \), from a model that would be fitted to the observed response alone.

The implication of the second observation depends on whether (i) the parameter sets \( \theta \) and \( \phi \) are distinct or (ii) they overlap. In case (i), there is no information about any parameters in \( \theta \) in \( P(M_i; \phi, x_i) \). Hence, inference for \( \theta \) can be drawn from the observed data alone, not neglecting to adjust for any covariates \( x_i \) that are predictive of drop-out. In case (ii), Diggle and Kenward [2] show that ignoring the information in \( P(M_i; \phi, x_i) \) about those parameters in the set \( \theta \) that also belong to \( \phi \), leads to consistent, but not necessarily statistically fully efficient, estimates of such parameters. To obtain fully efficient estimates, the parameter values that maximize both terms on the right hand side of (6) jointly must be found. This problem is uncommon in practice, however, since the same parameters are rarely used to relate the outcome \( y_i \) and drop-out \( M_i \) to treatment related covariates of interest.

In a clinical trial context, the assumptions of DCAR correspond to an individual dropping out for any reason, possibly noted at baseline, which, conditional on the baseline covariates in the model, is not associated with their post-randomization response. This could range from moving away through to the presence of other medical conditions. It includes the possibility that the patient enrolled late in the trial and could not complete the treatment course before the trial wound up.

Interim drop-out, whereby a patient misses a clinic visit or a few days of diary card data, usually occurs for transitory reasons such as an ‘off day’ perhaps caused by an unrelated illness, and is therefore almost always equivalent to drop-out completely at random. To see this formally, we would have to recast the model above to allow for interim drop-out, which is beyond our current scope. An example from the asthma study is that, because clinic visits were scheduled early in the morning, patients’ morning diary card data is almost always missing on the day of their clinic visit.

In conclusion, with completely random drop-out, whether interim or not, it is not necessary to construct a model for the drop-out mechanism or the missing observation (subject to the caveat on efficiency outlined above if \( \theta \) and \( \phi \) are not distinct). However, it is necessary to include in the model for the response, covariates that are predictive of drop-out.

### 3.2.2. Drop-out at random (DAR).

Here, drop-out can depend on post-randomization response. Mathematically

\[
P(M_i | y_i^{\text{obs}}, y_i^{\text{miss}}, \phi, x_i) = P(M_i | y_i^{\text{obs}}, \phi, x_i)
\]

Applying the same argument which gives (6), we see that

\[
P(y_i^{\text{obs}}, M_i; \theta, \phi, x_i) = P(M_i | y_i^{\text{obs}}, \phi, x_i)P(y_i^{\text{obs}}; \theta, x_i)
\]
Thus, exactly as in the DCAR case, correct inference (subject to the caveat on efficiency if \( \theta \) and \( \phi \) are not distinct outlined in the DCAR discussion) can be obtained by modelling the observed data alone, including covariates that predict drop-out in the model for \( y \). If past observations predict drop-out, arguably the current response should be adjusted for these too. A random effects model achieves this because the random effects ensure that 

\[
P(y_{i,j} | y_{i,j-1}, \theta, x_i) \neq P(y_j; \theta, x_i).
\]

In passing, note that if the expected, rather than observed, information is used for inference, some care needs to be taken with the choice of appropriate sampling distribution \([6, 9]\).

The clinical trial context in which this missing data mechanism might arise is when drop-out is determined by the response of the patient as the trial progresses, but, critically, not on any unobserved information. In other words, if a patient drops out for a known reason and that reason can be used as a predictor in the observed data model, that patient’s drop-out can be treated as ‘at random’. Thus, in practice, we have DAR when we know the reason why each individual dropped out, and this reason is associated with trial-related post-randomization events. For example:

1. it may be that the treatment is working well in terms of improving the principal outcome variable, but generating unacceptable recorded side-effects which eventually cross a prespecified level of acceptability, so that the patient is withdrawn;
2. the primary outcome variable might cross some prespecified threshold, causing the patient to be withdrawn.

In fact, the reason for drop-out can vary widely between patients; all that is required is that each patient’s drop-out conditional on the observed information does not depend on any unobserved information.

In practice, a judgment has to be made as to the strength of association needed between drop-out and a covariate for that covariate to be included in the observed data model. Usually, it is preferable to include covariates of borderline significance and risk type I error.

3.2.3. Non-Random drop-out (NRD). In this situation, the reason for a patient’s drop-out depends on \( y^{\text{miss}}_i \), so we cannot simplify

\[
P(M_i | y^{\text{obs}}_i, y^{\text{miss}}_i, \theta, \phi, x_i)
\]

to take it through the integral sign in the manner exemplified by (6). Thus, in order to find the maximum likelihood estimates, we have to carry out the integration on the right hand side of (2). Usually, this is not possible analytically, so it has to be done numerically at many different values of the parameters to identify the maximum. This is often difficult; an alternative is to use an EM algorithm or a Markov chain Monte Carlo approach.

If non-random drop-out is present, ignoring it means choosing not to model the relationship between the unobserved response and drop-out. This means ignoring the fact that drop-out is indicative of an unexpected deviation, usually decline, in response and ignoring the knock-on effect of this on the estimate of parameters of interest such as treatment efficacy.

It is thus important not to ignore non-random drop-out if it is present. The approach adopted by Diggle and Kenward [2] is to incorporate into the model parameters mediating the observed response on drop-out. Their statistical significance can then be tested in the usual way. However, the statistical significance of such parameters can depend critically on the precise model fitted. For example, a test that is significant when the error structure is chosen to be Gaussian.
can disappear if a \( t \)-distribution is chosen. Further, it is not possible to distinguish definitively between these two distributions on the basis of the data, because they are missing [10]. Thus formal tests of non-random drop-out should be avoided.

However, as noted by Diggle and Kenward [2], the principal motivation for investigating non-random drop-out is to explore the effect it has on the scientific conclusions. Therefore, we advocate a ‘sensitivity’ analysis, where the conclusions of a drop-out at random model are compared with the conclusions of non-random drop-out models.

The non-random drop-out models need to be selected to build an envelope of conclusions, bounded by the results of the drop-out at random model and the ‘worst case’ non-random drop-out model. Clearly, selection of the latter and the form of the non-random model both depend on scientific judgement. The way in which the non-random mechanism operates should be simple yet scientifically plausible; there is little point in showing the conclusions are very sensitive to scientifically implausible mechanisms.

3.3. Implications for design

The above discussion underlines the importance of discovering why patients have dropped out of a trial, for if all the predictors of patient drop-out were known and could be used in the model (in practice a daunting prospect!), then by definition a drop-out at random model would be sufficient.

Therefore, it is sensible to try and identify the likely causes of drop-out so that variables that measure them can be recorded. Relevant are any patient characteristics at baseline which could be related to drop out. Further, one could consider designing a study whereby patient withdrawal is triggered by a response variable deteriorating beyond a prespecified bound, although such a dichotomy is often difficult to achieve in practice.

4. TABULAR AND GRAPHICAL METHODS FOR ASSESSING DROP-OUT MECHANISMS

The foregoing discussion has emphasized the need to distinguish the various drop-out mechanisms. It is useful to (i) distinguish between drop-out completely at random and drop-out at random and (ii) identify covariates that predict drop-out and hence should be included in any model. Failure to identify and include the latter could bias the DAR model and make the differences between DAR and NRD models unduly sensitive to the postulated non-random drop-out mechanism.

Taking (i) first, if data are missing completely at random, then drop-out itself is not response related. Thus, conditional on baseline covariates, there should be no difference in the distribution of the observation \( Y_{ij} \) between those patients who drop out in the time interval \((j, j + 1)\) and those who do not.

The extent to which this is true can be evaluated by comparing the means and standard errors by future drop-out status, either analytically [11] or in tabular or graphical form. Figure 1 shows, for each treatment group in the asthma study, the mean (± two standard errors) change in per cent predicted FEV\(_1\) from baseline at each clinic visit, adjusted for baseline per cent predicted FEV\(_1\). The graph shows two categories of patient: those who do and do not attend their next scheduled clinic. The number of patients in each category at each time is shown above the \( x \)-axis in each panel.
Figure 1. Mean change from baseline of per cent predicted FEV₁, adjusted for baseline per cent predicted FEV₁ (± 2 standard errors) by patients attending or not at their next scheduled clinic. Panels are for placebo: 92 patients initially; 200 mcg: 91 patients initially; 400 mcg: 93 patients initially; 800 and 1600 mcg: 197 patients initially. At week 2 in the 400 mcg arm, only one patient is not present at week 4 and consequently the error bar extends off the graph. The patterns for the 800 mcg and 1600 mcg arms are very similar so they have been combined.
There is both a drop-out trend with time (patients tend to drop out sooner rather than later) and differential drop-out rate between arms, with patients receiving the higher treatment doses less likely to drop-out. Further, patients who drop-out are showing a consistently lower per cent predicted FEV₁ than those who remain. Thus drop-out is clearly not completely at random.

Similar plots for baseline variables can be used to informally address (ii). If the distribution of baseline variables for groups of patients who do and do not drop-out are distinct, then we have a good predictor of drop-out which should be included in all drop-out models. We illustrate with baseline per cent predicted FEV₁. Figure 2 shows the mean baseline per cent predicted FEV₁, as per cent of predicted normal (± 2 standard errors), by treatment group and final drop-out status. See text for further details. The patterns for the 800 mcg and 1600 mcg arms are very similar so they have been combined.

Figure 2. Baseline FEV₁, as per cent of predicted normal (± 2 standard errors), by treatment group and final drop-out status. See text for further details. The patterns for the 800 mcg and 1600 mcg arms are very similar so they have been combined.
predicted FEV₁ (a measure of asthma severity) by treatment group and final drop-out status. In the placebo group, patients with a low baseline per cent predicted FEV₁, as a percentage of that predicted, are less likely to complete the trial, but this pattern is less marked for the active treatments. There may be a subgroup of severely ill patients who do not respond to the treatment, however active, and drop out. Thus, baseline per cent predicted FEV₁ is an important variable to include in models for the asthma data.

5. USE OF LOGISTIC REGRESSION TO IDENTIFY PREDICTORS OF DROP-OUT

In this section we use logistic regression to identify predictors of drop-out. As discussed in Section 3, it is important to include these in DAR, DCAR and NRD models. We report the model for drop out by week 4 since more patients drop out in weeks 0–4 than 4–8 or 8–12. We discuss these cases briefly at the end of this section. We construct a binary response that is 1 if a patient drops out by week 4 and 0 otherwise. We begin by looking for post-randomization variables that could influence drop-out and then consider whether any baseline variables give any further improvement.

We first considered last observed per cent predicted FEV₁. As expected, this is more predictive of drop-out than either last FEV₁ or change in per cent predicted FEV₁ from baseline. Next, we investigated the association between the following variables and drop-out (for those covariates derived from diary card data the mean of the recorded measurements was used): last forced vital capacity (FVC); last forced expiratory flow (FEF); mean morning and evening PEF over the two weeks before the last FEV₁ reading as percentages of predicted normal [4]; mean day time and mean night time symptom scores over the two weeks before the last FEV₁ reading; mean number of puffs of bronchodilator over the last two weeks before the last FEV₁ reading and allocated treatment. Only allocated treatment significantly predicted drop-out once last observed per cent predicted FEV₁ was in the model. The 400–1600 mcg treatment arms (active treatments 2–4) were combined, without significant loss of prediction, but an additional quadratic term for last per cent predicted FEV₁ was also significant. The model is

\[
\text{logit}\{\text{Probability of drop-out by week 4}\} = \alpha + \beta_1 \text{I(treatment 1)} + \beta_2 \text{I(treatment 2–4)} + \beta_3 (\text{last observed per cent predicted FEV}_1) + \beta_4 (\text{last observed per cent predicted FEV}_1)^2
\]

where \(\text{I(\cdot)}\) is an indicator random variable. The estimated coefficients and standard errors for model (8) are given in Table I.

We then investigated whether any of the baseline variables added significantly to the model at the 5 per cent level, and found that none did. Model (8) therefore gives our preferred set of predictors for patient drop-out by four weeks. The model differentiates well between the two groups, giving a receiver operating characteristic of 0.85 with an exact 95 per cent binomial confidence interval of (0.82, 0.88) [12]. (Note 1.0 is perfect prediction and 0.5 is no better than chance.) Additional regressions of drop-out between weeks 4 and 8 on measurements at baseline and week 4, and drop-out between weeks 8–12 on measurements at baseline and
week 8 establish that last observed per cent predicted FEV1 is by far the strongest predictor of drop-out, while its square and treatment decline in importance with time. We next fit a drop-out at random model to the data and then compare it with a non-random drop-out model in Section 7.

6. DROP-OUT AT RANDOM (DAR) MODEL

In this section we fit a DAR model to the observed change in per cent predicted FEV1 from baseline. As these are approximately multivariate normal, there are a wide choice of options, from structural equations to random effects and non-linear models [13], which can be fitted in a variety of software packages, such as SAS (proc MIXED) [14], MLwiN [15] and BUGS [1]. Here, we use BUGS (which stands for Bayesian analysis Using Gibbs Sampling), since within the BUGS framework, the sensitivity of the model to non-random drop-out can be readily assessed (see Section 7). We use very vague priors, so that the parameter estimates are, to a good approximation, maximum likelihood estimates.

First, it is necessary to choose a model for an individual’s FEV1 over the course of the trial. Figure 3(a) shows how mean per cent predicted FEV1 varies from baseline over the course of the trial by treatment group. There appears to be substantial change over the first four weeks of the trial, followed by a relatively steady, possibly asymptotic, pattern. Clearly, a straight line is going to fit the data rather poorly. One option would be to fit an autoregressive model to the response [16], but this is not generally straightforward; neither is fitting an exponential model

\[ Y_{ij} = \alpha(1 - e^{-\beta t}) \quad \alpha, \beta > 0 \]

Hence, we therefore decided on the following approach.

6.1. Transforming the time axis to obtain a linear response

We propose transforming the time axis in longitudinal data to make the response more nearly linear and allow a simpler class of models. This approach is quite generally applicable, and is useful when the response tends to an asymptote. It is especially useful when the initial readings are much closer together in time than the later ones. The reason is that on the untransformed scale the later observations will usually have greater influence on the fitted slope than the early ones. Thus, a time transformation which pulls the later readings closer to the earlier ones therefore has the effect of increasing the influence of the earlier ones on the fitted slope.
We illustrate with the asthma data. Reciprocal, square root and log transformations, motivated by the discussion above, were considered. From graphs of the response against time for various transformations, it was clear that the change in FEV\textsubscript{1} was most nearly linear with a reciprocal transformation

\[
\text{time} \mapsto -\frac{1}{\text{time} + \tau} + \eta
\]  

(9)

This has an additional attraction. The responses in this trial change quite quickly early on and then stabilize; hence the increasing time between observations as the trial progresses. Thus it makes sense to estimate this stable level, or asymptote. This can be readily done with a reciprocal transformation. From (9) the eventual treatment effect, as time $\to \infty$, can be estimated as $\eta$ times the fitted slope.

To identify plausible $\tau$ and $\eta$ we treated each observation on each individual as independent and fitted a different straight line through the origin to each treatment group simultaneously.
We then obtained a likelihood ratio confidence region for plausible values of \( \eta \) and \( \tau \) (which are estimated consistently, although the region is too tight because of the extra information injected by assuming each observation is independent).

This is shown in Figure 4. The maximum likelihood estimates of \((\tau, \eta)\) are \((0.98, 1.01)\), respectively. However, the likelihood surface is relatively flat, and takes the form of a crescent around the line \( \eta = 1/\tau \), so there are a wide range of possibilities.

We use the transformation \( \eta = \tau = 1 \) below, but we discuss the results obtained with other values in Section 9. The result of using this transformation is shown in Figure 3(b). A random selection of individual patient profiles on the transformed scale are shown in Figure 5. Although the model does not successfully describe all patient profiles, it provides a reasonable approximation overall.

### 6.2. Drop-out at random model

We now describe our drop-out at random model. The response is change in per cent predicted FEV\(_1\) from baseline. The primary baseline covariate of interest is treatment. In addition, as discussed in Section 3, we should adjust for any baseline covariates that are predictive of drop-out. Model (8) shows that a patient’s most recent per cent predicted FEV\(_1\) is strongly predictive of drop-out. This can be regarded as comprising two components: (i) baseline per cent predicted FEV\(_1\) and (ii) change in per cent predicted FEV\(_1\) from baseline. This explains why, after treatment, baseline per cent predicted FEV\(_1\) is the only baseline covariate predictive of drop-out; it is thus included in the model below.

Recall \( i \) indexes patient and \( j \) clinic visit. Let \( Y_{ij} \) be the \( i \)th patient’s change in per cent predicted FEV\(_1\) from baseline at clinic visit \( j \), so that \( Y_{i,1} = 0 \). Then the model is

\[
Y_{ij} \sim N(\mu_{ij}, \sigma^2) \\
\mu_{ij} = \beta_{\text{slope},i} t_{ij} \\
\beta_{\text{slope},i} \sim N(\mu_{\text{slope},i}, \sigma_{\text{slope}}^2) \\
\mu_{\text{slope},i} = \sum_{l=1}^{6} \beta_l x_{l,i} 
\]

\[\text{(10)}\]
Figure 5. Selection of individual patient profiles against transformed time (see (9)), from: placebo group (first row); 200 mcg group (second row); 400 mcg group (third row); 800 mcg group (fourth row).

where \( t_{ij} \) is transformed time since randomization, \( \beta_{\text{slope},i} \) is the \( i \)th patient’s slope, normally distributed about the \( i \)th patient’s mean slope, which is a linear function of six covariates. In order, these are a constant (\( x_{1,i} = 1 \) for all \( i \)), mean-centred baseline per cent predicted FEV\(_1\) and indicators for the four active treatments.

Since we are using BUGS to obtain approximate maximum likelihood estimates (see discussion at the start of this section) we specify the following uninformative priors for the parameters. Each \( \beta \) parameter has a N(0,100,000) prior. Rather than put a prior on the variances \( \sigma^2 \) and \( \sigma_{\text{slope}}^2 \), we put one on their reciprocal, \( \tau = 1/\sigma^2 \) and \( \tau_{\text{slope}} = 1/\sigma_{\text{slope}}^2 \), as suggested in the BUGS users’ guide [1]. We choose these to be flat gamma distributions, \( f(x; r, \mu) = \mu^r x^{r-1} \exp(-\mu x)/\Gamma(r) \), with \( \mu = r = 0.001 \).
6.3. Convergence checks

The program **BUGS** uses the Gibbs sampler to draw a correlated sample from the joint posterior distribution of all the parameters. An excellent introduction to Gibbs sampling is given by Gilks *et al.* [17]. Starting from initial values specified by the user, in the initial phase, dubbed ‘burn in’, the sampler moves towards the true posterior distribution. Thereafter, in the ‘stationary’ phase, it yields correlated samples from the true joint posterior.

Thus results from the Gibbs sampler need to be examined to check that they are likely to be from the true posterior distribution rather than the ‘burn in’ phase. The correlation between the samples also needs to be examined, since if it is high, the sampler is not moving quickly around the posterior distribution, so it will have to be run for longer to explore the posterior distribution adequately.

We used the software **CODA** [18] to do this, and found that after a ‘burn in’ of 5000 samples from the joint distribution, a further sample of 10,000 points passed the stationarity test of Heidelberger and Welch [19]. The autocorrelations in the chain also appeared acceptably low (maximum absolute autocorrelation at lag 50 0.05), and visual inspection of the traces showed no evidence of non-stationarity.

![Figure 6](image)

*Figure 6. (a) Post-randomization placebo means, and (b) 200 mcg treatment arm means, for some common procedures for handling missing data and the models described in the text. Examination of the standard errors shows that in the placebo group, the means from the non-random drop-out model with \( \delta = -0.14 \) are different to those from the ‘completers’ and ‘all observed data’ approaches with borderline 5 per cent significance at week 12. For the 200 mcg group, no differences are near significance at the 5 per cent level.*
Table II. Asymptotic change in per cent predicted FEV$_1$ from baseline value for a patient who has an average baseline per cent predicted FEV$_1$. Results obtained using drop-out at random and non-random drop-out models; 95 per cent highest posterior density (HPD) intervals are shown in parentheses.

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Drop-out mechanism</th>
<th>At random</th>
<th>Non-random ($\delta = -0.07$)</th>
<th>Non-random ($\delta = -0.14$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>At random</td>
<td>-5.9</td>
<td>-10.4</td>
<td>-12.6</td>
</tr>
<tr>
<td></td>
<td>Non-random ($\delta = -0.07$)</td>
<td>(-9.1, -2.8)</td>
<td>(-13.9, -7.0)</td>
<td>(-16.0, -9.2)</td>
</tr>
<tr>
<td>200 mcg BID</td>
<td>At random</td>
<td>6.0</td>
<td>4.6</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Non-random ($\delta = -0.07$)</td>
<td>(3.1, 8.9)</td>
<td>(1.4, 7.7)</td>
<td>(0.6, 7.0)</td>
</tr>
<tr>
<td>400 mcg BID</td>
<td>At random</td>
<td>10.0</td>
<td>9.2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Non-random ($\delta = -0.07$)</td>
<td>(7.1, 12.9)</td>
<td>(6.1, 12.3)</td>
<td>(5.7, 12.0)</td>
</tr>
<tr>
<td>800 mcg BID</td>
<td>At random</td>
<td>8.7</td>
<td>7.4</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td>Non-random ($\delta = -0.07$)</td>
<td>(5.9, 11.4)</td>
<td>(4.4, 10.3)</td>
<td>(3.9, 9.9)</td>
</tr>
<tr>
<td>1200 mcg BID</td>
<td>At random</td>
<td>10.0</td>
<td>9.2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Non-random ($\delta = -0.07$)</td>
<td>(7.3, 12.9)</td>
<td>(6.2, 12.2)</td>
<td>(5.8, 11.9)</td>
</tr>
</tbody>
</table>

6.4. Results

Fitting model (10) gave as parameter estimates the first eight rows of column 1 in Table III. Further, for missing $Y_{ij}$ BUGS drew samples from the posterior distribution $N(\mu_i, \sigma^2)$. This enables a direct comparison with previous simple analyses. The results are shown in Figure 6, for the placebo and 200 mcg BID groups only. Since there is so little missing data in the other treatment arms, the means for these treatments are virtually unchanged by the model. The left hand column of Table II also gives the asymptotic treatment estimates and their highest posterior density intervals. The reciprocal time transformation adopted in Section 6.1 ($\eta = \tau = 1$) means these are estimated by the mean slope.

We defer discussion of the results until Section 8; first we describe the non-random drop-out model.

7. NON-RANDOM DROP-OUT MODEL

Recall from Section 3 that the non-random drop-out mechanism has the drop-out depending on the unseen observation, so that we cannot bring the probability of drop-out through the integral as exemplified by equation (6). A frequentist approach would thus entail finding the maximum of the likelihood which entails performing the integral (2) numerically. However, this is relatively non-trivial and non-routine for a model such as (10). By contrast, the Bayesian framework regards all the missing observations as parameters [20], as previously discussed, if vague priors are used, then posterior means will be a good approximation to maximum likelihood estimates. The Gibbs sampler gives a sample from the full joint distribution of the parameters and hence the marginal distributions. Gibbs sampling for this class of model can be done in a fairly routine way using the software BUGS and a program such as the one in the Appendix.

The non-random drop-out model has the same structure for the response as the dropout at random model. The extra complexity comes from terms relating the presence or absence of a reading from individual $i$ at time $j$ to that individual’s value at that time $j$ and earlier outcomes plus associated baseline covariates (see model (8)). Thus the non-random drop-out
model below assesses the sensitivity of the conclusions obtained with model (10) by relating each individual’s presence or absence to their (sometimes unobserved) response.

Let $M_{ij}$ be 1 for a missing observation and 0 otherwise. The model for the response is identical to (10); that for drop-out is as follows:

$$M_{ij} \sim \text{Bernoulli}(\pi_{ij})$$
\[
\text{logit } \pi_{ij} = \sum_{l=1}^{4} \gamma_l z_{l;j} + \sum_{l=1}^{5} \gamma_l x_{l+1;i} + \delta Y_{ij}
\]

(11)

Here, the covariates \( x_{li} \) are the same as those in (10) (omitting the constant \( x_1; i \)) and the additional covariates \( z_{1;j}, \ldots, z_{4;j} \) are indicator variables for clinic visits at weeks 2, 4, 8 and 12.

While the response model is exactly the same as (10), the second half of the model describes the non-random drop-out. Model (8) suggests a patient’s drop-out by week 4 depends on both treatment and last observed per cent predicted FEV1. More generally, it is also likely to vary with clinic visit. Noting that a patient’s current per cent predicted FEV1 is the sum of (i) baseline per cent predicted FEV1 (\( x_{2;i} \)) and (ii) change from baseline in per cent predicted FEV1 (\( Y_{ij} \)), this leads to the model for logit \( \pi_{ij} \) above. (While (18) suggests including a quadratic term in \( Y_{ij} \) in the model for logit \( \pi_{ij} \), for the relatively inexact purposes of sensitivity analysis, we judge a linear term sufficient.) Model (11) thus relates the logit of the probability of patient \( i \)’s presence at clinic visit \( j \) in a linear way to clinic visit, treatment, baseline per cent predicted FEV1 (\( x_{2;i} \)) and change from baseline in per cent predicted FEV1 (\( Y_{ij} \)). The change in per cent predicted FEV1 from baseline, \( Y_{ij} \), is related to logit \( \pi_{ij} \) via the non-random drop-out parameter \( \delta \). Note, if \( \delta = 0 \), we can do separate maximizations of the response and drop-out parts of (11) yielding estimates of the \( \beta \)'s which will coincide with those obtained from (10).

Next, we specify vague prior distributions for all the parameters. For those parameters in the response model, we use the same priors as in model (10). Additionally, each \( \gamma \) and \( \gamma \) parameter has a \( \text{N}(0,100,000) \) prior. In the case of \( \delta \), while, in theory, this parameter should be estimable (see, for example, Diggle and Kenward [2]), Spiegelhalter (personal communication) found that in many examples, the likelihood for this parameter was very flat. Accordingly, we prefer to compare the drop-out at random and non-random drop-out models for several specific plausible values of \( \delta \), whose selection is described in the next paragraph.

Fitting a variant of model (8), which regresses drop-out by week 4 on treatment, baseline per cent predicted FEV1 and last observed change from baseline per cent predicted FEV1, showed that the odds ratio of remaining in the trial after week 4 given a percentage point drop in last observed ‘change in per cent predicted FEV1 from baseline’ is 0.93, a 7 per cent drop. A plausible scenario is that the unseen per cent predicted FEV1 may be the same or more strongly associated with drop-out than this. Accordingly we assessed the sensitivity of the conclusions to two values of \( \delta \): \(-0.07 \) and \(-0.14 \), equivalent to an odds ratio of 0.93 and 0.87, respectively.

The models were fitted in BUGS; after a ‘burn in’ of 5000 iterations of the Gibbs sampler, a sample of 10,000 was obtained from the posterior. The convergence tests described in Section 6.3 were again employed; all parameters passed them.

8. RESULTS AND THEIR INTERPRETATION

Figure 6 shows the means of the observed (and imputed where appropriate) data under various scenarios, for the placebo and 200 mcg treatment arms, respectively. We focus on these
two treatment arms, since relatively few patients dropped out in the other treatment arms. Several points are of interest. First, the means for the completers are systematically higher than those for all observed data, which suggests drop-out is not completely at random. Secondly, the drop-out at random model results are initially close to the all observed data results, but systematically slightly below these later. However, drop-out at random results are above those for last observation carried forward. This suggests that, if drop-out were truly at random, last observation carried forward gives unduly pessimistic results in any treatment group. Thirdly, in the placebo arm, the non-random drop-out models give the worst outcome, while in the 200 mcg arm, last observation carried forward gives the worst outcome. This illustrates the difference between last observation carried forward, where the ‘best guess’ for missing data from an individual is their last observation, and a model-based approach, in which the ‘best guess’ for missing data from an individual is influenced by their previous observations, treatment group and the fact they dropped out.

The estimated asymptotic per cent changes for all five treatment groups are presented in Table II. These confirm that the results for the placebo arm are very sensitive to the relationship between decline in per cent predicted FEV₁ and drop-out. This is because 63 per cent of patients drop-out in the placebo arm, compared to no more than 25 per cent in any treatment arm.

Turning to the parameter estimates in Table III, for the drop-out at random model the average slope is negative for the placebo arm ($\beta_1$) and positive for all the other treatments ($\beta_1+\beta_3$ for 200 mcg; $\beta_1+\beta_4$ for 400 mcg; $\beta_1+\beta_5$ for 800 mcg; $\beta_1+\beta_6$ for 1600 mcg). There is a clear difference between the placebo arm, treatment group 1 and treatment groups 2–4. The coefficient for baseline per cent predicted FEV₁ ($\beta_2$) is negative, suggesting that patients whose initial per cent predicted FEV₁ is high are likely to have smaller increases, as one would expect from regression to the mean.

Turning to the non-random drop-out models, as $\delta$ decreases from zero (meaning drop-out is more strongly associated with low unseen per cent predicted FEV₁), the baseline mean slope, $\beta_1$, which describes the placebo group, declines. Indeed, the 95 per cent highest posterior density intervals for $\beta_1$ when $\delta = 0$ and $\delta = -0.14$ do not overlap. Conversely, since there is less drop-out in the other treatment arms, the fitted slopes to these arms are not so greatly affected by drop-out; in order to keep the average slope for these groups similar as $\beta_1$ declines, $\beta_3, \ldots, \beta_6$ therefore increase. The coefficient, $\beta_5$, for baseline per cent predicted FEV₁ remains virtually unchanged, suggesting little change in the effect on the slope of regression to the mean.

The logit baseline probabilities of drop-out at weeks 2, 4, 8 and 12, ($x_1, \ldots, x_4$) decrease as $\delta$ decreases. This is because the probability of drop-out in the placebo arm is now increasingly dependent on a decline in per cent predicted FEV₁ from baseline, rather than week. Conversely, the treatment parameters for drop-out, $\gamma_2, \ldots, \gamma_5$, increase. To see why, consider the probability of drop-out for a patient with average baseline per cent predicted FEV₁ in the 1600 mcg treatment arm at week 12. This is $x_4 + \gamma_5 + \delta \times \text{change in per cent predicted FEV₁ from baseline.}$ As $\delta$ decreases, both $x_4$ and the last term decrease; however the overall probability of drop-out remains approximately constant. Hence $\gamma_5$ needs to increase. Lastly, the coefficient of baseline per cent predicted FEV₁, $\gamma_1$, decreases slightly from zero with $\delta$, presumably because those patients who had a high baseline per cent predicted FEV₁, which subsequently declines in a regression to the mean effect, otherwise have too great a probability of drop-out when $\delta < 0$. 

To conclude this section, we briefly discuss the results obtained with other plausible values of \( \tau \) and \( \eta \) for the time transformation (9). We re-ran the analyses reported in Table III with \( \tau \) closer to the boundary of the region shown in Figure 4, specifically \( \tau = 1/\eta = 0.25 \) and \( \tau = 1/\eta = 2 \). Since these transformations change the fitted slopes, the slope parameters \( (\beta) \) and the standard deviation \( \sigma_{\text{slope}} \) change markedly. However the remainder of the parameters remain well within the 95 per cent highest posterior density intervals of the corresponding estimates in Table III. Turning to the asymptotic treatment effect, for the transformation \( \tau = 1/\eta = 0.25 \), all the asymptotic effects are reduced in absolute value; those for the placebo arm by 20–21 per cent compared with the corresponding estimates in Table II, and those for the active treatments by 4–13 per cent. However, all remain in the asymptotic intervals of the corresponding estimates in Table III. For the transformation \( \tau = 1/\eta = 2 \), an equivalent but opposite effect is observed. That is to say, all the asymptotic effects are increased in absolute value, those for the placebo arm by 25–29 per cent compared with the corresponding estimates in Table II, and those for the active treatments by 3–16 per cent. Again, all remain in the asymptotic intervals of the corresponding estimates in Table III.

9. DISCUSSION AND CONCLUSIONS

If a substantial number of patients do not complete a trial it is important to understand how this affects the conclusions. We have described a model-based approach for assessing how sensitive the conclusions of a longitudinal study are to missing data.

We began with a review of the different kinds of drop-out mechanism and how these affect the modelling. Two aspects of this are worth highlighting. The first is the difference between a model for all the observed data and a dropout at random model. The latter, because it is derived from a model for the response and drop-out, has a different conditional framework. The implication of this is that covariates that are predictive of drop-out should be adjusted for in the response model. The second aspect is that the most preferable way to assess the sensitivity of the conclusions to non-random drop-out is to modify the dropout at random model to allow for various plausible non-random drop-out scenarios and see how the conclusions vary.

Section 8 gives the results of adopting this approach for the asthma trial. We see that, under plausible non-random drop-out assumptions, estimated changes in response over time for the active treatments are altered little, while those for the placebo alter considerably. Of particular interest is the comparison between data imputed using last observation carried forward and that imputed from our models; the latter allow the imputations to depend not only on an individual’s previous observations, but on their treatment group and the fact they dropped out. In particular our informative drop-out models suggest last observation carried forward underestimates the effect of the 200 mcg treatment but overestimates the effect of placebo treatment. This, coupled with the greater insight into the data given by a modelling approach, argues strongly for its use where possible.

In this paper, we have only considered continuous outcomes, but our approach should extend readily to discrete and categorical data, by appropriately modifying the response model (see the \textsc{BUGS} manual for examples [1]).

One important assumption that may not be generally appropriate is that of the selection model for drop-out. This assumes that drop-out is strongly associated with a poor response to treatment and that the same model structure (for example, a simple slope in the asthma example) is appropriate for those who do and do not drop-out. This seems plausible for many
problems, but if a different model structure for the completers and drop-outs were required, in other words if the model for the response depended on drop-out, then a pattern mixture model might be more be appropriate.

Computationally, our model can be estimated in either a frequentist or a Bayesian framework; in the former we need to use some form of EM algorithm; in the latter we need to adopt uninformative priors, so that the results will approximate maximum likelihood. Then, the Gibbs sampler can be used to obtain a sample from the posterior. Similar models to the one we used can be readily fitted with the Gibbs sampler using the program BUGS. To facilitate this, we give our program code in the Appendix.

In conclusion, our modelling approach is appropriate for, and readily adaptable to, data from a range of longitudinal studies. Further, a model-based approach to missing data is the most appropriate way to assess the sensitivity of the conclusions to various drop-out mechanisms, and preferable to automatic procedures such as ‘last observation carried forward’.

APPENDIX: BUGS CODE

Here we give the BUGS code for fitting the non-random drop-out model. Code for the random drop-out model is simply obtained by deleting the portion of code between the ‘+++’ lines or setting delta equal to zero.

model asthma;

#Informative dropout model
#Code for non-random dropout between ++++++++++++++++++++++’s

const
  N=473, # number of subjects
  T=5, # number of time points
  Ncov=5, # number of fixed covariates in data file (rev, weight,t2-t5)
  L = 2; # No of missingness levels.

var
  mu[N,T], # true mean response
  cov[N,Ncov], # true response
  pfev[N], # covariates
  Y[N,T], # true response
  id1[N],id2[N], # identifiers
  beta[N], # slope
  tau.beta,sigma.beta,# precisions and standard deviations slope
  beta.slope[Ncov+1], # regression coefficients for slope
  tau,sigma, # precision and standard deviation of FEV
  Y.plac[5], # average observed at each wave for each treat
  Y.treat1[5], # average observed at each wave for each treat
Y.treat2[5],  # average observed at each wave for each treat
Y.treat3[5],  # average observed at each wave for each treat
Y.treat4[5],  # average observed at each wave for each treat
mu.beta[N],  # expected gradient for individual i

time[T],  # variable of transformed times of measurements

miss[N,T],  # missing data indicator matrix.
p[N,T,L],phi[N,T,L],  # probabilities for missing data
delta,alpha.2[T],  # missing data parameters

gamma.2[Ncov],  # contrasts with placebo for mean dropout rate;
teffect[T];  # estimate of asymptotic treatment effect.

#Data files:

data id1,Y in "fev.dat",
id2, cov in "predout.dat",
miss in "missfev.dat";

inits in "initial.dat";

{
# set elements of time for transformation with tau=eta=1
  time[1] <- 0.0;
  time[2] <- 0.67;
  time[3] <- 0.8;
  time[4] <- 0.89;
  time[5] <- 0.92;

# calculate summary statistics
  for (j in 1:T) {
    Y.plac[j]  <- mean(Y[1:92,j]);
    Y.treat1[j]<-mean(Y[93:183,j]);
    Y.treat2[j]<-mean(Y[184:276,j]);
    Y.treat3[j]<-mean(Y[277:375,j]);
    Y.treat4[j]<-mean(Y[376:473,j]);
  }

# estimates of asymptotic treatment effect:
  teffect[1]<-1*beta.slope[1];

  # mean centre covariate
for (i in 1:N) {
    pfev[i] <- cov[i,1]-mean(cov[,1]);
}

# set up observation likelihood
for (i in 1:N) {
    for (j in 1:T) {
        Y[i,j] ~ dnorm(mu[i,j],tau);
        mu[i,j] <- beta[i]*time[j];
    }
}

# joint distribution of slopes
for (i in 1:N) {
    beta[i] ~ dnorm(mu.beta[i],tau.beta);
                 + beta.slope[3]*cov[i,2] + beta.slope[4]*cov[i,3]
                 + beta.slope[5]*cov[i,4] + beta.slope[6]*cov[i,5];
}

# prior distributions
for (m in 1:(Ncov+1)) {
    beta.slope[m] ~ dnorm(0,0.00001);
}

tau ~ dgamma (1.0E-3,1.0E-3); # vague prior on precision
sigma <- sqrt (1/tau); # standard deviation

tau.beta ~ dgamma(1.0E-3,1.0E-3); # vague prior on slope precision
sigma.beta <- sqrt(1/tau.beta); # standard deviation of slope

# missing data model
for (j in 2:T) { # only bring in for second observation onwards
    for (i in 1:N) {
        miss[i,j] ~ dcat( p[i,j,] );
        for (1 in 1:L){
            p[i,j,1]<-phi[i,j,1] / sum(phi[i,j,]);
        }
        log(phi[i,j,1]) <- 0;
        log(phi[i,j,2]) <- alpha.2[j] + delta*Y[i,j] +
            gamma.2[1]*pfev[i] + gamma.2[2]*cov[i,2] +
            gamma.2[3]*cov[i,3] + gamma.2[4]*cov[i,4] +
            gamma.2[5]*cov[i,5]; #dropouts
\begin{verbatim}
}
alpha.2[j] ~ dnorm(0,.00001);
}
delta <- -0.07;
for (k in 1:Ncov) {
  gamma.2[k] ~ dnorm(0,0.00001);
}

#++++++++++++++++++++++++++++++++++++++++++++++++++++++

\end{verbatim}

ACKNOWLEDGEMENTS

We thank AstraZeneca for providing the data, a number of stimulating meetings and financial support for this research. Thanks are also due to Chris Frost and David Spiegelhalter for helpful advice and a BUGS tutorial. We are grateful to the two referees for their constructive comments.

REFERENCES
