

I. L. Solis-Trapala (*Lancaster University*)

This stimulating paper begins with a succinct account of existing approaches to the analysis of longitudinal data with drop-out, encouraging the reader to consider carefully the objectives of the study at hand in their own modelling strategies.

Firstly, I would like to reflect on the target of inference which motivates the authors' proposal. Although they state that their target of inference is the mean response, they propose to model the expected increments of the longitudinal process. Indeed, it seems from the context that they are interested in measuring mean contrasts between groups of participants who are assigned to different treatments, rather than mean contrasts within subjects.

This distinction is briefly highlighted in the discussion section, where it is argued that inclusion of residuals in the mean specification rather than previous responses preserves the interpretation of the effects of the exogenous covariates. For example, in the case of the schizophrenia study, measurement of a direct effect of treatment appears to be of primary scientific interest.

Secondly, the way that the measurement error is included in the linear model is not entirely clear to me. Intuitively, I would associate a measurement error, rather than a lagged error, with the response increments.

Thirdly, the authors acknowledge a limitation of their model, namely that it is based on untestable assumptions. This limitation is not specific to their approach, but is well known from other models dealing with missing data. Assuming a martingale for the random effect is, in my opinion, an elegant way of formalizing the key assumption of stability. This reflects that the unobserved increments (due to drop-out) are assumed to follow a process that is similar to that observed in the past.

Jeremy M. G. Taylor (*University of Michigan, Ann Arbor*)

I agree with the authors that for some scientific applications involving longitudinal data it makes sense for the targets of inference to be parameters of a hypothetical drop-out-free world, whereas in other applications this may not make sense. A difficult question is whether we can consider a hypothetical drop-out-free world, when drop-out is due to death. In cancer research a frequently used experiment is one in which tumours grow in laboratory mice, and the response variable is the size of the tumour at 12 months say. Such experiments may include both planned early sacrifices and sacrifices to prevent suffering in animals in which the tumour has grown large. I would be interested in hearing the authors' view on whether it is still reasonable to consider a hypothetical drop-out-free world in this setting when evaluating the mean tumour size at 12 months.

The Diggle, Farewell and Henderson longitudinal model raises the issue of 'what is a statistical model?'. One view, which is implicit in their specification, is that a model should be a plausible approximation to the mechanism that gave rise to the observations. Under this viewpoint, it should be a principle that the observations and drop-out at a certain time cannot depend on the future. But if a model is simply viewed as a way to describe data, using a small number of parameters, then this principle seems less pertinent.

In longitudinal models distinction is made between subject-specific, population-average and transition models. The increments model of the authors has the flavour of a transition model. In continuous time, modelling increments in the response generalizes to that of modelling slopes. This then bears some resemblance to some of our previous work (Taylor *et al.*, 1994). We assumed that the expected slope at time t evolved according to an Ornstein–Uhlenbeck process. This leads to a model for the measured response of the form $Y_{it} = X(t)\beta + a_i + W_i(t) + e_{it}$ where $a_i \sim N(0, \sigma_a^2)$, $e_{it} \sim N(0, \sigma_e^2)$ and $W_i(t)$ is an integrated Ornstein–Uhlenbeck process.

The good efficiency properties of the approach of Diggle and his colleagues compared with fully parametric joint modelling was interesting, but somewhat surprising to me. The very poor efficiency of the inverse probability weighting approach was also striking. Have the authors found similar efficiency comparison results in other applications and in simulations?

D. Zeng and D. Y. Lin (*University of North Carolina, Chapel Hill*)

We congratulate the authors on a clever and intriguing piece of work. The time-specific conditional mean models avoid the ambiguity of counterfactual response after drop-out, which can be an issue in joint modelling. Joint models, however, are useful for prediction and amenable to efficient estimation. We pose two questions.

- (a) Since the model is conditional on the response history, is $\beta(t)$ in equation (10) the most relevant quantity?
- (b) Are there concrete examples that the drop-out process satisfies the assumptions of Section 4.1.2 but violates the assumption of missingness at random?

We offer a simple approach to inference. For $i = 1, \dots, n$ and $t = 1, \dots, \tau$, let $Y_i(t)$ be the response of the i th subject at time t and $X_i(t)$ be the corresponding $p \times 1$ vector of covariates. The estimator $\hat{\beta}(t)$ solves the equation

$$\sum_{i=1}^n X_i(t) \{ \Delta Y_i(t) - X_i(t)' \beta(t) \} = 0, \quad t \in \mathcal{T},$$

where $\Delta Y_i(t) = Y_i(t) - Y_i(t-1)$. Since the estimation function is a sum of independent zero-mean random vectors, standard asymptotic arguments entail that $\{ \hat{\beta}(t) : t = 1, \dots, \tau; t \in \mathcal{T} \}$ is asymptotically (multivariate) normal and the covariance matrix between $\{ \hat{\beta}(t) : t \in \mathcal{T} \}$ and $\{ \hat{\beta}(s) : s \in \mathcal{T} \}$ can be estimated by the sandwich estimator

$$\left\{ \sum_{i=1}^n X_i(t) X_i(t)' \right\}^{-1} \left(\sum_{i=1}^n [X_i(t) \{ \Delta Y_i(t) - X_i(t)' \hat{\beta}(t) \}] [X_i(s) \{ \Delta Y_i(s) - X_i(s)' \hat{\beta}(s) \}]' \right) \left\{ \sum_{i=1}^n X_i(s) X_i(s)' \right\}^{-1}. \tag{30}$$

We then estimate the covariance matrix of $\hat{B}(t)$ on \mathcal{T} by $(I, \dots, I) \hat{\Sigma}_t (I, \dots, I)'$, where $\hat{\Sigma}_t$ is the $pt \times pt$ sandwich covariance matrix estimator for $\{ \hat{\beta}(s) : s = 1, \dots, t \}$ based on expression (30) and I is the $p \times p$ identity matrix. Thus, we can make inference about $B(t)$ by using standard procedures for normal statistics. Since it is a very simple function of data, the sandwich estimator should provide accurate variance estimation in finite samples. It is not necessary to use the bootstrap, although the above arguments imply that the bootstrap is valid.

The authors replied later, in writing, as follows.

We thank all the discussants for their helpful and constructive comments and apologize if we have overlooked any of these in our reply. We have grouped our response under three headings: objectives, general modelling issues, including sensitivity and diagnostic checking, and issues that are specific to our proposed model class and its possible extensions.

Objectives

We agree with Hand that careful consideration of objectives is always important, and not at all specific to longitudinal studies. We suspect that all statisticians would agree, but that not enough statistics degree syllabuses give this topic the attention that it deserves.

Hogan, Cook and Lawless, Molenberghs and Verbeke, and Didelez all comment on the link between our discussion of potential outcomes and the wider topic of causal inference. Hogan asks whether our potential, but unrealized, outcomes are inherent characteristics of the subjects to whom they belong, or purely metaphysical. A partial answer is that this depends on the context. For the data that are analysed in Section 7 of our paper, and borrowing from Didelez’s comments, we can easily conceive of an intervention, albeit an unethical one, that would prevent drop-out. Perhaps a better example is long-term follow-up of dialysis patients, with drop-out corresponding to transplantation. Kidney function in the absence of transplant is definitely a legitimate target for inference. In cases of this kind, our discussion simply makes explicit what is often glossed over—that any analysis treating drop-out as ignorable is, nevertheless, making untestable assumptions about things that, by definition, cannot be observed. In some other contexts, most obviously when drop-out equates to natural death, any inference about a hypothetical drop-out-free population is of dubious practical relevance. Nevertheless, our view is that this need not preclude including a potentially infinite sequence of measurements as part of a joint model for measurements and time of death. In answer to Taylor’s question concerning animal experimentation, planned sacrifices are missing completely at random (MCAR), whereas sacrifices in response to an observed large tumour size are missing at random (MAR). Hence, in conventional terms both kinds of drop-out are ignorable. However, simply to conduct a standard likelihood-based analysis of the non-missing data would be too glib, not because there is anything wrong with modelling a hypothetical drop-out-free process in this setting—on the contrary, this is the natural process that operates in the absence of any intervention by the experimenter—but because the implied target for inference is not necessarily the most sensible interpretation of what precisely is meant by ‘the mean tumour size at 12 months’.

Longford’s Y_c could be construed as a mixture of Y_a and Y_b , with the mixture proportion referring to the rate of compliance in an operational setting; however, we suspect that he is making the stronger point that what happens in a controlled trial setting may or may not be a reliable guide to what happens in clinical practice. This is a fair point, but not specific to the topic of our paper.