



HAL
open science

A hidden Markov model for informative dropout in longitudinal response data with crisis states

Alessandra Spagnoli, Robin Henderson, Richard J. Boys, Jeanine J. Houwing-Duistermaat

► **To cite this version:**

Alessandra Spagnoli, Robin Henderson, Richard J. Boys, Jeanine J. Houwing-Duistermaat. A hidden Markov model for informative dropout in longitudinal response data with crisis states. *Statistics and Probability Letters*, 2011, 10.1016/j.spl.2011.02.005 . hal-00746102

HAL Id: hal-00746102

<https://hal.science/hal-00746102>

Submitted on 27 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

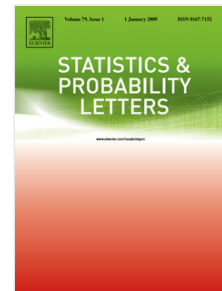
Accepted Manuscript

A hidden Markov model for informative dropout in longitudinal response data with crisis states

Alessandra Spagnoli, Robin Henderson, Richard J. Boys, Jeanine J. Houwing-Duistermaat

PII: S0167-7152(11)00043-5
DOI: [10.1016/j.spl.2011.02.005](https://doi.org/10.1016/j.spl.2011.02.005)
Reference: STAPRO 5903

To appear in: *Statistics and Probability Letters*



Please cite this article as: Spagnoli, A., Henderson, R., Boys, R.J., Houwing-Duistermaat, J.J., A hidden Markov model for informative dropout in longitudinal response data with crisis states. *Statistics and Probability Letters* (2011), doi:10.1016/j.spl.2011.02.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Hidden Markov Model for Informative Dropout in Longitudinal Response Data with Crisis States

Alessandra Spagnoli^{a,*}, Robin Henderson^b, Richard J. Boys^b, Jeanine J. Houwing-Duistermaat^c

^a*Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Italy*

^b*School of Mathematics & Statistics, Newcastle University, UK*

^c*Department of Medical Statistics and Bioinformatics, LUMC, Leiden, The Netherlands*

Abstract

We adopt a hidden state approach for the analysis of longitudinal data subject to dropout. Motivated by two applied studies, we assume subjects can move between three states: stable, crisis, dropout. Dropout is observed but the other two states are not. During a possibly transient crisis state both the longitudinal response distribution and the probability of dropout can differ from the stable state. We adopt a linear mixed effects model with subject-specific trajectories during stable periods and additional random jumps during crises. We place the model in the context of Rubin's taxonomy and develop the associated likelihood. The methods are illustrated using the two motivating examples.

Keywords: Change points, monotonic missing data, mixed models, random effects, state space.

1. Introduction

Longitudinal studies often suffer attrition, in that individuals drop out of the study before the scheduled completion time and thus present incomplete data. A variety of methods have by now been developed to deal with the possibility that dropout is related to responses (Hogan et al., 2004; Molenberghs et al., 2004; Philipson et al., 2008; Tsiatis and Davidian, 2004), though caution in using such methods is always needed (Molenberghs et al., 2004, 2008).

Recently le Cessie et al. (2009) recognised that longitudinal data analysis can be complicated by the fact that during follow-up subjects can change condition or state, an example being remission, relapse and death for cancer patients. Both longitudinal responses and the dropout probability can depend on the current state and this needs to be accounted

*Corresponding address: Alessandra Spagnoli, Dipartimento di Scienze Statistiche, Sapienza Università di Roma. P.le A. Moro, 5, 00185, Roma, Italy.

Email addresses: alessandra.spagnoli@uniroma1.it (Alessandra Spagnoli), Robin.Henderson@ncl.ac.uk (Robin Henderson), Richard.Boys@ncl.ac.uk (Richard J. Boys), J.J.Houwing@lumc.nl (Jeanine J. Houwing-Duistermaat)

for in analysis. The methods developed by le Cessie et al. (2009) are appropriate when the underlying state is observed. If a state is defined by level of a response variable but obscured by measurement error, then the hidden Markov methods of Satten et al. (1996) or Guihenneuc-Jouyaux et al. (2004) can form a basis for analysis. But, as argued by Liestøl and Andersen (2002), there are circumstances where a subject's state is either hidden or vaguely defined. For the liver cirrhosis application considered by Liestøl and Andersen (2002) for example, some subjects experienced apparent "crises", marked by a sudden change in response values. These crises could be transient or could indicate a terminal disease stage.

In this work we build on the ideas of le Cessie et al. (2009) and Liestøl and Andersen (2002) and develop a hidden state modelling approach for longitudinal data subject to dropout. We assume that during follow-up subjects can experience different states, which we will think of as stable state, crisis state and dropout. The first two states are transient and reversible, while the third, dropout, is an absorbing state. The crisis state can be defined as an intermediate phase where significant changes of the response values can be observed and where the probability of dropout is increased. We assume the longitudinal response is associated with underlying state but we assume that states other than dropout are not directly observed, and perhaps not precisely defined. We exclude situations where the state is *defined* by the response, such as AIDS when CD4 T-cell first reaches a given level, or leukaemia relapse when an residual leukaemic cell count is over a defined threshold (De Lorenzo et al., 2005).

In Section 2 we provide brief details of two applications which motivated our work. Our model is introduced and estimation outlined in Section 3, where we also argue the merits of proper treatment of time ordering when considering missingness mechanisms for longitudinal data. Section 4 includes summaries of our analyses of the motivating data sets, and some brief comments in Section 5 conclude the paper.

2. Motivating examples

2.1. Schizophrenia data

We consider data from a trial into treatment of schizophrenia, previously described by Henderson et al. (2000) and Diggle et al. (2007). There are three treatment groups (standard, placebo and experimental) and the response of interest is the Positive and Negative Symptom Scale (PANSS), which is high for subjects with poor condition. Six assessments were scheduled, at weeks 0,1,2,4,6 and 8, but of the 518 subjects under consideration only 269 completed the trial. Of the remainder, 183 subjects dropped out of the trial because of "inadequate response", which was considered to be potentially informative, and 66 dropped out before the sixth observation for a variety of other reasons, which we treat as censoring in our analysis, following previous authors. Further information on the dropout reasons is provided by Diggle (1998). In their analysis of these data, Henderson et al. (2000) suggested that informative dropout (inadequate response) was associated with poor condition (high PANSS) or deteriorating condition, which they modelled as a sustained subject-specific trend. As well as a sustained trend, there is evidence for many subjects of an abrupt

change in response just before informative dropout. Figure 1 illustrates. The figure shows mean first difference in responses at assessments 2-6 (baseline being assessment one), for the group of subjects about to experience informative dropout and for the subjects who returned for the next assessment. For instance, 40 subjects had informative dropout between assessments two and three. For these subjects the mean *increase* in PANSS score between assessments one and two was over 10 units. This is to be compared with a mean *decrease* of almost 10 units over the same period for the 451 subjects who were present at assessment two and returned for assessment three. A similar difference is seen for the later assessments. Standard errors are around 2-3 for the informative dropout cohorts and 0.6-0.7 for the larger continuing cohorts.

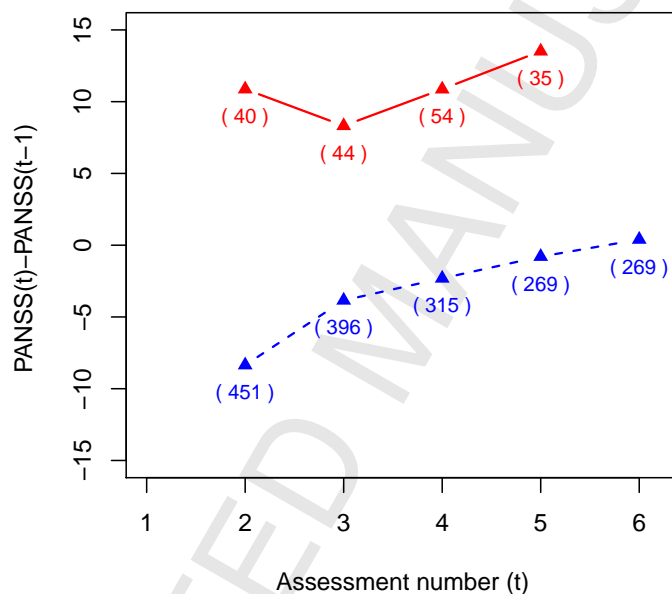


Figure 1: Schizophrenia data: mean differences in responses between assessments $t - 1$ and t . The solid line is for subjects who have informative dropout between assessments t and $t + 1$ and the broken line for subjects who return for assessment $t + 1$. Baseline assessment is coded $t = 1$. Numbers in brackets give the cohort sizes.

2.2. Leiden 85+ data

Our second example is on data from the Leiden 85+ Study, carried out by the Leiden University Medical Centre (der Wiel et al., 2002; van Vliet et al., 2010). The purpose of the study is the identification of biomarkers and genetic factors for healthy aging. Between September 1997 and September 1999, 705 inhabitants of Leiden, The Netherlands,

reached the age of 85 years, and in the month after their 85th birthday, they were asked to participate in the study. We consider here the 541 subjects who agreed to participate and provided complete covariate information. Covariates were gender, educational status (two levels: primary or higher) and Apolipoprotein E (APOE) genotypes, which is considered a strong genetic risk factor for dementia as well as human longevity (Christensen et al., 2006). The Mini Mental Status Examination (MMSE) instrument for cognitive ability was determined annually until death/dropout or age 90. Of the 541 participants, 240 died during the study and are considered as possibly informative dropouts, and 35 withdrew or were withdrawn for non-informative causes.

Figure 2 corresponds, for the Leiden data, to Figure 1 for the schizophrenia data. The pattern is similar, though reversed because for MMSE low values indicate poor condition. Standard errors are around 0.5-0.7 for the informative dropout (death) cohorts and 0.12-0.15 for the continuing cohorts. Note that a selection effect complicates interpretation of the plot. Even if there is linear (or no) underlying trend, we may not expect the observed mean differences to be constant given that different risk sets are being compared.

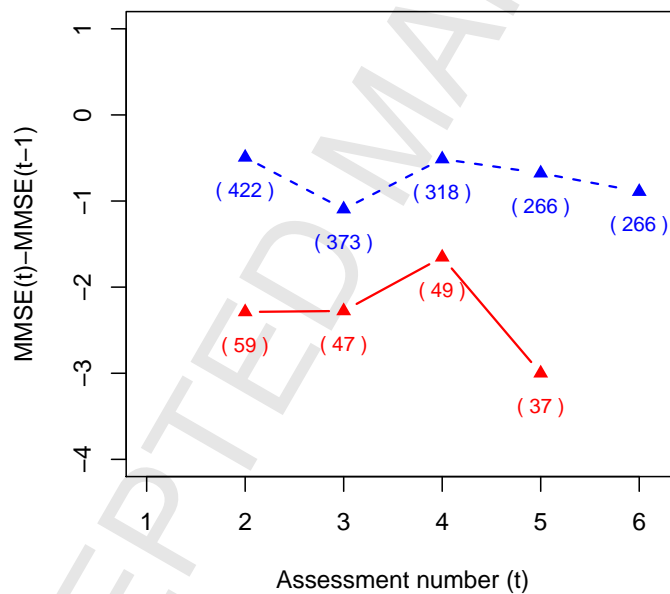


Figure 2: Leiden 85+ data: mean differences in responses between assessments $t - 1$ and t . The solid line is for subjects who have informative dropout between assessments t and $t + 1$ and the broken line for subjects who return for assessment $t + 1$. Baseline assessment is coded $t = 1$. Numbers in brackets give the cohort sizes.

3. Theoretical Development

3.1. Model and assumptions

We will begin with the general situation. We assume a balanced design with common scheduled assessment times for all n subjects recruited into the study. We label the assessment times $t = 1, 2, \dots, T$ but do not require equal spacing in calendar time. For the moment we will consider a single generic subject and do not use a subscript to differentiate between individuals. We identify two stochastic processes in t : a partly unobservable finite state first-order Markov chain, S_t , and an observable Gaussian response process, Y_t , the latter recorded only if the subject is under observation. We will use m ($\leq T$) for the number of recorded observations. Associated with each subject is a p -vector of covariates \mathbf{x} . For notational simplicity we will assume these are time-constant, though extension to time-varying covariates is automatic. We define $\bar{\mathbf{S}}_t = (S_1, S_2, \dots, S_t)^T$ and $\bar{\mathbf{Y}}_t = (Y_1, Y_2, \dots, Y_t)^T$ as the partial state and response vectors up to assessment t . The distribution of response Y_t can depend on the state path $\bar{\mathbf{S}}_t$ up to and including t .

We assume a partially hidden Markov model in which subjects can move over time between a finite number of disease states, S_t , with space state $S = \{1, 2, \dots, N\}$, some of which are not observed. We implicitly assume that all patients start in state 1 at time $t = 1$. The Markov chain parameters are then the transition probabilities $\phi_{jk}(t) = \Pr(S_t = k | S_{t-1} = j)$, defined as the probabilities of being in state k at time t conditional on being in state j at time $t - 1$. We do not require the chain to be homogeneous: transition probabilities may not be constant over time. In principle, transitions may also depend on covariates but for simplicity we suppress this in notation.

Turning to our specific model, we assume there are three possible states, with 1=stable, 2=crisis and 3=dropout, so that $N=3$. The dropout state is observed rather than hidden, and is absorbing since we allow monotonic dropout only. The transition matrix is

$$\mathbf{Q}(t) = \begin{pmatrix} (1 - \eta)(1 - \phi_{13}(t)) & \eta(1 - \phi_{13}(t)) & \phi_{13}(t) \\ \phi_{21}(t) & 1 - \phi_{21}(t) - \phi_{23}(t) & \phi_{23}(t) \\ 0 & 0 & 1 \end{pmatrix}.$$

The probability of dropping out from stable state is thus $\phi_{13}(t)$. We will take a logistic model

$$\phi_{13}(t) = \frac{\exp(\alpha_{0t} + \alpha_{1t}Y_{t-1})}{1 + \exp(\alpha_{0t} + \alpha_{1t}Y_{t-1})}, \quad (1)$$

in which the most recent observed response can influence dropout. This can be extended to incorporate the prior response path $\bar{\mathbf{Y}}_{t-1}$ but we specifically *exclude* any dependence on Y_t or any other as yet unobserved responses. This point will be discussed in a later section.

From $\mathbf{Q}(t)$, the conditional probability of moving from stable state to crisis given no dropout is η . Again this can be time-varying or covariate-dependent but in the applications to come we will assume it to be constant, so that the chance of moving to crisis state cannot be anticipated from observation history. The transition probabilities from the crisis state, $\phi_{21}(t)$ and $\phi_{23}(t)$, might be allowed to depend on covariates or observed responses

as required.

Turning to the longitudinal responses, let $\mathbf{Y}_m = \bar{\mathbf{Y}}_m = (Y_1, \dots, Y_m)^T$ denote the full vector of observed longitudinal measurements for the subject, where m is the number of completed assessments. Conditional on state history $\bar{\mathbf{S}}_t$ we assume a linear mixed model

$$\mathbf{Y}_m = \mathbf{X}_m \beta + \mathbf{Z}_m \mathbf{b} + \mathbf{c}(\bar{\mathbf{S}}_m) d + \epsilon_m. \quad (2)$$

Here \mathbf{X}_m and β are an appropriate design matrix and parameter vector respectively for fixed effects, including time trends. The vector \mathbf{b} represents unobserved individual specific sources of heterogeneity common to each measure within the same individual, and \mathbf{Z}_m is an associated design matrix. We will assume a standard Laird-Ware slope and intercept model in which

$$\mathbf{Z}_m = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}$$

and $\mathbf{b} = (b_0, b_1)^T$ is zero mean bivariate Gaussian with variance matrix

$$\mathbf{B} = \begin{pmatrix} \sigma_{b_0}^2 & \rho \sigma_{b_0} \sigma_{b_1} \\ \rho \sigma_{b_0} \sigma_{b_1} & \sigma_{b_1}^2 \end{pmatrix}.$$

The term $\mathbf{c}(\bar{\mathbf{S}}_t) d$ describes the effect of being in crisis state. We assume that on entry to crisis state there is a random change d in mean, which persists thereafter so that repeated entries into crisis have a cumulative effect. To model, we assume $d \sim N(\mu_d, \sigma_d^2)$ independently of \mathbf{b} , and that element t of the m -vector $\mathbf{c}(\bar{\mathbf{S}}_t)$ counts the number of entries to crisis up to assessment t , ($t = 1, 2, \dots, m$), namely $\mathbf{c}(\bar{\mathbf{S}}_m)_t = \sum_{j \leq t} I(S_j = 2)$. The model is completed by the error term ϵ_m , which consists of independent zero-mean Gaussian random variables with common variance σ_ϵ^2 .

3.2. Filtrations and longitudinal missing data mechanisms

In dealing with missing data it is common to use $(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}})$ to denote the observed and missing components of the intended response vector. The Rubin missing data taxonomy (Rubin, 1976) can then be summarised as the following.

- *Missing completely at random* (MCAR): conditional on covariates \mathbf{x} , the observation probability is independent of $(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}})$.
- *Missing at random* (MAR): conditional on covariates \mathbf{x} and observed responses \mathbf{Y}_{obs} , the observation probability is independent of \mathbf{Y}_{miss} .
- *Missing not at random* (MNAR): conditional on covariates \mathbf{x} and observed responses \mathbf{Y}_{obs} , the observation probability is not independent of \mathbf{Y}_{miss} .

The terminology was developed for general missing data problems, and implicitly assumes that \mathbf{Y}_{miss} is extant whether or not it is observed. In dealing with longitudinal data subject to dropout however, the crucial role of time ordering is not explicitly recognised, and nor is there recognition that dropout is not simply a missing data indicator, but an event in subjects' lives which can materially determine even the existence of further responses (Diggle et al., 2007). An extreme example is when dropout is because of death, as in the Leiden 85+ study, and discussed in detail by Dufoil et al. (2004). Interpretation of MNAR models incorporating \mathbf{Y}_{miss} in dropout mechanisms is at best challenging.

Our preference, following (Diggle et al., 2007), is to recognise the role of time and *exclude* any dependence on future observations in a model which attempts to describe the *system* of data generation as opposed to empirical relationships between observables. We think instead of two forms of filtration or history. Let \mathcal{F}_t denote the filtration generated by observables up to and including assessment t . In our case these are $\{\mathbf{x}, \bar{\mathbf{Y}}_t\}$. Next, let \mathcal{G}_t be the filtration generated by non-observables to time t : for us these are $\{\mathbf{b}, \bar{\mathbf{S}}_t, d\}$. We suggest that any model that purports to be realistic for the underlying system of data generation should be MAR with respect to the combined filtration $\mathcal{H}_t = \{\mathcal{F}_t, \mathcal{G}_t\}$. In other words, given *everything* in the past there is no dependence of dropout on the future. However, the implicit marginal model generated by \mathcal{F}_t alone may be MNAR, as illustrated in Figure 3. It can *seem* that there is dependence on the future given the observed past, but this is because of shared dependence on unobserved past. The full system model described in the previous subsection is of this type: MAR given \mathcal{H}_t because dropout depends only on previous observations and states, but MNAR given only \mathcal{F}_t because future observations provide information about the missing past states, which are in turn informative for dropout. Thus we have a MNAR data-model constructed from an underlying MAR system-model, and the crucial role of time is explicitly recognised.

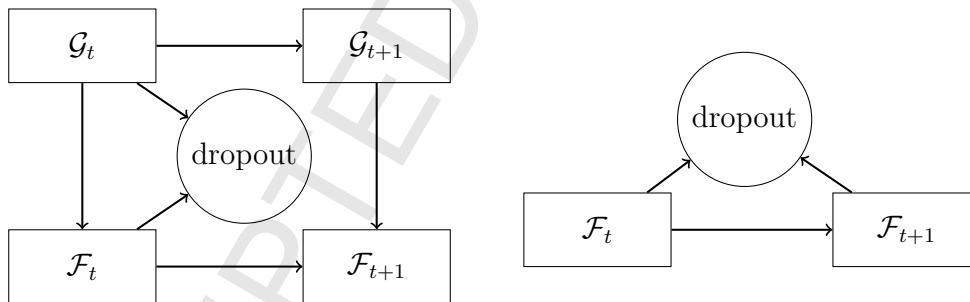


Figure 3: MAR underlying system model (left) and marginal MNAR empirical model (right)

3.3. Estimation and Inference

We assume between-subject independence and introduce subscript i to differentiate between individuals. Conditional upon the state paths $\bar{\mathbf{S}}_{im}$, the responses \mathbf{Y}_{im} are conditionally independent Gaussian vectors. Since we have a linear mixed model it is straightforward

to integrate out the random effects \mathbf{b}_i and d_i to obtain a multivariate Gaussian distribution with mean $\mathbf{X}_i\beta + \mathbf{c}(\bar{\mathbf{S}}_{im})\mu_d$ and variance matrix $\mathbf{V}_i = \mathbf{Z}_i\mathbf{B}\mathbf{Z}_i^T + \sigma_d^2\mathbf{J}(\bar{\mathbf{S}}_{im}) + \sigma_\epsilon^2\mathbf{I}_m$, where \mathbf{I}_m is the m -dimensional identity matrix and $\mathbf{J}(\bar{\mathbf{S}}_{im}) = \mathbf{c}(\bar{\mathbf{S}}_{im})\mathbf{c}^T(\bar{\mathbf{S}}_{im})$. We will write $f(\mathbf{Y}_{im}|\bar{\mathbf{S}}_{im};\theta)$ for the corresponding density, where θ is the combined vector of unknown parameters.

Let \mathcal{S}_{im} denote the set of state paths which are consistent with the dropout time for subject i , and recall that states 1,2 and 3 denote stable, crisis and dropout, that dropout is absorbing, and that we assume all subjects start in stable state. Then \mathcal{S}_{im} will have 2^{m-1} elements for subjects who drop out for informative reasons after m assessments. For example if $m = 2$ then $\mathcal{S}_{im} = \{(1, 1, 3), (1, 2, 3)\}$. If dropout is for non-informative reasons (censoring) then in principle there are $2^T - 1$ state paths which are consistent with the observed data. However, under our independent censoring assumption the later states are ignorable and for these subjects we consider only states up to the number m of observed assessments, such as $\mathcal{S}_{i2} = \{(1, 1), (1, 2)\}$ if $m = 2$.

The marginal likelihood of the observed data can now be derived from the Markov state space property and missingness assumptions as

$$\begin{aligned} L(\theta; \mathbf{Y}_{im}) &= \prod_{i=1}^n \sum_{\bar{\mathbf{s}}_{im} \in \mathcal{S}_{im}} \left[f(Y_{i1}; \theta) \prod_{t=2}^m \{f(Y_{it}|\mathbf{Y}_{i,t-1}, \bar{\mathbf{s}}_{it}; \theta) \Pr(s_{it}|\mathbf{Y}_{i,t-1}, s_{i,t-1}; \theta)\} \right] \\ &= \prod_{i=1}^n \sum_{\bar{\mathbf{s}}_{im} \in \mathcal{S}_{im}} \left[f(\mathbf{Y}_{im}|\bar{\mathbf{s}}_{im}; \theta) \prod_{t=2}^m \Pr(s_{it}|\mathbf{Y}_{i,t-1}, s_{i,t-1}; \theta) \right]. \end{aligned}$$

There are closed form expressions for all terms in the square brackets, meaning that the likelihood can be evaluated very quickly provided T is not too large. A numerical search to find the maximum likelihood estimate is feasible as is a Markov chain Monte Carlo approach if Bayesian inference is preferred.

4. Applications

4.1. Schizophrenia data

For the fixed effects component we assume separate quadratic time trends within each of the three treatment groups. We take the logistic model (1) for the probability $\phi_{13}(t)$ of a direct transition from stable state to dropout, but we assume the coefficient of previous response is time-constant, ie $\alpha_{1t} = \alpha_1$. We assume transitions out of crisis state are time-homogeneous: $\phi_{21}(t) = \phi_{21}$ and $\phi_{23}(t) = \phi_{23}$.

We performed three analyses of these data: standard maximum likelihood, Bayes using MCMC and flat priors, and Bayes/MCMC with informative priors. The maximum likelihood estimates were almost identical to the posterior means when we used MCMC with flat priors, and all posterior distributions appeared close to Gaussian. Table 1 summarises results for the parameters which distinguish our model from a standard Laird-Ware model with MAR dropout. These are the conditional probability η of moving from stable to crisis

state given not dropping out; the probabilities ϕ_{21} and ϕ_{23} of leaving crisis state for stable or by dropout respectively, and the parameters μ_d and σ_d^2 for the random change d on entry to crisis state.

Our *a priori* assumption was that entry into crisis state would be relatively rare, but that once in crisis state dropout would be likely. Hence the informative prior chosen for η was a beta distribution with parameters 2 and 8, and for (ϕ_{21}, ϕ_{23}) a Dirichlet distribution with parameters (2,12,1). We selected more vague Gaussian and inverse gamma priors for μ_d and σ_d^2 respectively. Prior means and standard deviations are shown in Table 1 alongside the associated posterior summaries from an MCMC run of 50,000 iterations, following 10,000 iterations of burn-in. Movement into crisis state was relatively rare and there seemed to be a lower dropout probability from that state than expected *a priori*. The variance of the random change d on entry to crisis was high, leading to the low precision mean and standard deviation shown in Table 1. The mean and standard deviation for σ_d directly were 24.03 and 2.00 respectively from the posterior using informative priors.

Broadly similar results were obtained when flat priors were used throughout, and, as stated, in that case posterior means were very close to the maximum likelihood estimates. The overall maximised partial log likelihood for the responses Y was -10502.41. This is partial as the full log likelihood also includes a component for the dropout events. The partial form can be compared with -10587.7 from the corresponding single-state model (constraining $\eta = \mu_d = \sigma_d = 0$). Note that we cannot compare the models using a standard likelihood ratio test, since the problem is non-regular (Davies, 1977, 1987). Nonetheless there is very strong evidence that of the two models under consideration, the state-space model should be preferred to the standard. Of course we should make the proviso that as always untestable dropout assumptions have had to be made. We cannot claim that our model is correct in any sense: we simply claim that it fits the observed data better than the standard MAR model without crisis state.

Parameter	Informative Prior		Posterior Summaries			
	Mean	SD	Informative prior Mean	Informative prior SD	Flat prior Mean	Flat prior SD
η	0.2000	0.1206	0.1001	0.0176	0.1391	0.0221
ϕ_{21}	0.1333	0.0085	0.3560	0.0594	0.6462	0.0633
ϕ_{23}	0.8000	0.1000	0.4577	0.0613	0.2724	0.0540
μ_d	0	25	1.1617	0.9581	4.4549	2.3675
$1/\sigma_d^2$	500	500	0.0017	0.0003	0.0020	0.0003

Table 1: Schizophrenia data: selection of results from Bayesian analyses with flat or informative priors.

Figure 4 shows as solid lines how the observed mean responses declined during the course of the study, with each point in the plot based on all available data. The decline

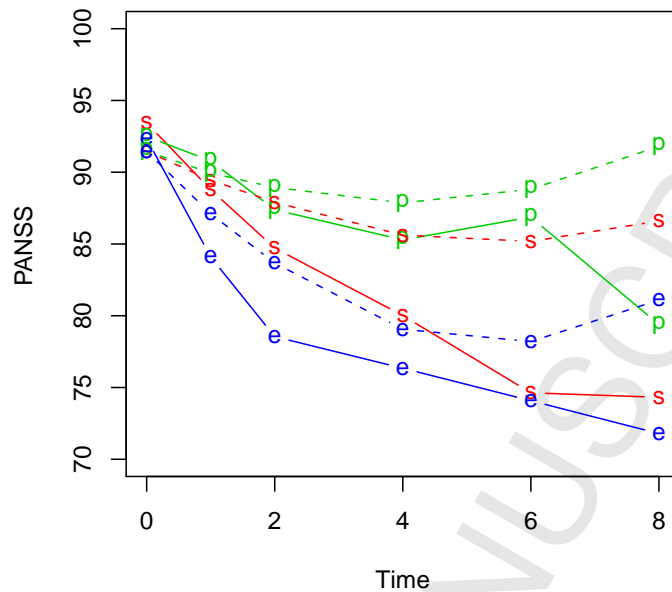


Figure 4: Schizophrenia data. Observed (solid lines) and estimated dropout-free (broken lines) mean response profiles for placebo (p), standard (s) and experimental (e) treatment groups

may be real, as subjects improve over time, or could reflect a selection effect as subjects in poor condition and hence with high responses tended to drop out, leaving a lower mean for continuing subjects. Given our model, we can estimate profiles in a hypothetical dropout-free situation (Diggle et al., 2007). These are the dotted lines in Figure 4, which indicate a substantial improvement would still occur in the experimental treatment group, a slight improvement in the standard treatment group, and, reassuringly, no change over time in the placebo group, assuming no change in other conditions or treatments. Diggle et al. (2007) discuss analysis objectives when longitudinal trials are subject to dropout: in this case the targeting of a conceptual hypothetical dropout-free population is appropriate.

Figure 5 provides further illustration, concentrating on subjects with at least four observations. The left column shows the response of the ten subjects most likely to be in crisis state at the fourth observation time, which is week four. For clarity the traces have been split over two plots, with the five most-likely subjects in the top plot. The right plot shows traces for ten people chosen randomly from the remainder of the subjects still at risk, again split into two groups of five, this time in no particular order. In the plot we use a broken line from time t to $t + 1$ whenever the probability of dropping out for informative reasons, given observed history \mathcal{F}_t , is higher than an arbitrary threshold of

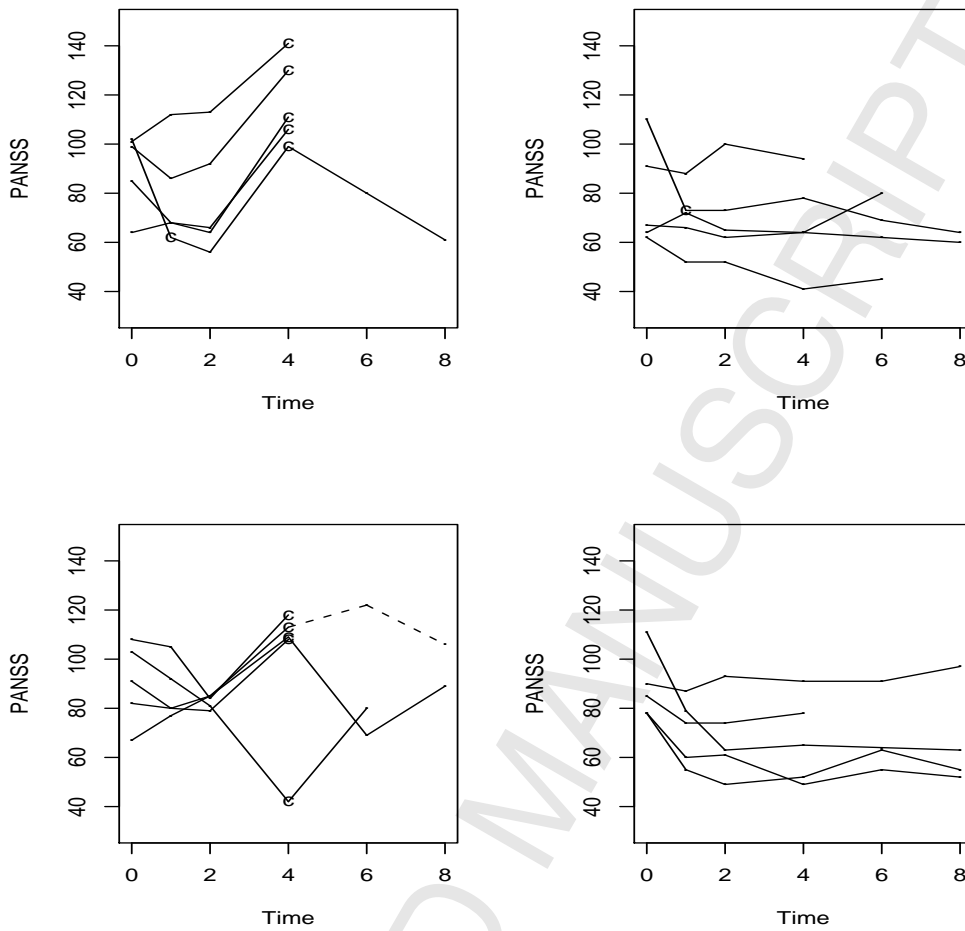


Figure 5: Example profiles of schizophrenia data. The two left plots show profiles of the ten people most likely to be in crisis state at week four, and the right plots show the profiles of ten people randomly selected from the remainder of the risk set at week four. See text for further information.

0.3. The symbol “c” indicates that the conditional probability of being in crisis state at time t given observation history \mathcal{F}_t to t is above another arbitrary threshold of 0.3. In the left column the actual probability of being in crisis state at time four exceeds 0.7 for all ten subjects. In the right column all corresponding probabilities are below 0.1. Nine of the ten subjects with high probability of being in crisis state at time four show an increase in response leading to that time, and one has a decrease. This is the pattern which motivated our analysis. Six of the ten had informative dropout before the next observation time. The traces in the right column of people not likely to be in crisis are more stable. One of these subjects also had informative dropout before the next observation time.

4.2. Leiden 85+ data

Crisis-state modelling was also applied to the Leiden 85+ data. The aim was to assess the effects of the crisis state and the covariates on the response variable and to compare results from fitting our crisis state model with those from a more standard Laird-Ware MAR approach. Table 2 shows the parameter estimates and the corresponding standard errors for MAR and MNAR (crisis state) analyses. For this analysis we used likelihood methods only, since this is the method used for the Laird-Ware model in the standard R software we used. Standard errors were obtained from 100 bootstrap resamples. We considered the following covariates: gender (female=0), educational status classified into two levels (primary=0 or higher levels), and apolipoprotein E (APO-E) clustered into four genotypes (33=0,22-23,24,34-44). We see that male subjects with at most primary school level were generally more cognitively impaired than subjects with higher educational level, and age and the APOE (34-44) genotypes negatively influenced cognitive function. The transition probability estimates from the crisis state to dropout and from the crisis state to stable state are similar. The transition probability from the stable state to dropout decreases with increasing values of MMSE at time $t - 1$. No differential intercept terms (time varying) are significant. The age effect is much stronger under the MAR model than under the crisis state model and subjects that enter the crisis state tend to have a decrease in level of MMSE. That is, the crisis state effect absorbs part of the age effect. The same holds in less extent to carriers of the APOE 2 alleles: in the MAR model APOE2 show a non significant protective effect on MMSE while in the new model the estimate is almost zero. The maximized partial log likelihood under the MNAR crisis state model was -6741.7, whereas under MAR it was -6963.58, providing strong evidence that our model should be preferred.

As a final comment for this section we point out that a plot equivalent to Figure 4 could be produced for the Leiden data, to compare observed and dropout-free profiles. We have not considered such a plot as the concept of dropout-free responses when dropout corresponds to death makes no real sense.

5. Discussion

We have proposed an approach to modelling longitudinal data subject to dropout which might be useful when there are indications that subjects can have high risk or crisis periods during which the response variable can change dramatically and the probability of dropout be affected. Our model is MAR given complete data filtrations but MNAR given only observed data filtrations. We do not claim that our approach will always be appropriate, but we do consider it potentially useful. In both the schizophrenia and Leiden applications our model provided a better fit to the observed data than a standard MAR approach based on combination of a Laird-Ware longitudinal component and a logistic dropout component. This is not to say that all MAR models would be inferior: to the contrary, as pointed out by Molenberghs et al. (2008), for every MNAR model there is a MAR equivalent which provides an equally good fit to the observed data. Although such models would not usually include standard parametric components, it would be interesting to examine what features

	Model			
	MNAR (Crisis)		MAR	
	Coeff.	St.Err.	Coeff.	St.Err.
α_{01}	-0.41	0.55	-0.50	0.80
α_{02}	-0.29	0.63	-0.19	0.81
α_{03}	-0.59	0.73	-0.32	0.78
α_{04}	-0.29	0.77	-0.13	0.78
α_{05}	-0.53	0.89	-0.28	0.77
α_1	-0.08	0.02	-0.08	0.03
η	0.11	0.05	-	-
ϕ_{21}	0.38	0.13	-	-
ϕ_{23}	0.34	0.17	-	-
constant	23.24	0.81	23.11	0.73
age	-0.38	0.06	-0.92	0.04
gender	0.98	0.27	1.27	0.17
educational	3.53	0.26	3.59	0.17
apoe _{22,23}	-0.01	0.25	0.10	0.19
apoe ₂₄	-0.85	0.60	-0.41	0.64
apoe _{34,44}	-3.07	0.28	-2.77	0.21
μ_d	-4.04	1.87	-	-
σ_d^2	3.93	2.24	-	-
σ_ϵ^2	1.51	0.21	1.95	0.07
σ_{b0}^2	5.38	0.45	5.37	0.42
σ_{b1}^2	0.39	0.36	1.08	0.16
ρ_{b0b1}	0.33	0.20	0.33	0.16
Partial log-likelihood	-6741.7		-6963.58	

Table 2: Leiden 85+ data: maximum likelihood estimates and standard errors of the transition probability and explanatory variable effects for the MAR and MNAR models

are required of a MAR model which matches our crisis state model in terms of observed data fit. This is an area for future research. Many extensions are possible and might be explored in further work. For example, in our work we assume that entry into crisis state is accompanied by a step change d in response, with d being a time-fixed Gaussian random effect. We might allow the step changes to vary over time, or we might constrain d to lie on only one of the positive or negative half lines rather than both. In that way we might associate an entry into crisis with a deterioration in health: a positive step for the schizophrenia data or a negative step for the Leiden 85+ data.

Acknowledgements

We gratefully acknowledge Dr Ton de Craen and Dr Rudi Westendorp of the Leiden University Medical Centre, for kindly providing the analyzed data. We thank the guest editor and an anonymous reviewer for helpful comments on an earlier version of the manuscript.

References

- le Cessie, S., de Vries, E.G.E., Buijs, C., Post, W.J., 2009. Analyzing longitudinal data with patients in different disease states during follow-up and death as final state. *Statistics in Medicine* 28, 3829-3843.
- Christensen K., Johnson T.E., Vaupel J.W., 2006. The quest for genetic determinants of human longevity: challenges and insights. *Nature Reviews Genetics* 7, 436-448.
- Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247-254.
- Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33-43.
- De Lorenzo, P., Henderson, R. and Valsecchi, M.G., 2005. Analysis of interval-censored longitudinal data with application to onco-haematology. *Statistics in Medicine* 24, 3945-3957.
- Diggle, P.J., 1998. Dealing with missing values in longitudinal studies. *Recent Advances in the Statistical Analysis of Medical Data* editors Everitt and Dunn, London: Arnold, 203-228.
- Diggle, P.J., Farewell, D. and Henderson, R., 2007. Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal (with discussion). *Applied Statistics* 56, 499-550.
- Dufoil, C., Brayne, D. and Clayton, D., 2004. Analysis of longitudinal studies with death and drop-out: a case study. *Statistics in Medicine* 23, 2215-2226.
- Guihenneuc-Jouyaux, C., Richardson, S. and Longini, I.M., 2000. Modeling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline. *Biometrics* 56, 733-741.
- Henderson, R., Diggle, P.J. and Dobson, A., 2000. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1, 465-480.
- Hogan, J.W., Roy, J. and Korkontzelou, C., 2004. Tutorial in biostatistics - handling drop-out in longitudinal studies. *Statistics in Medicine* 23, 1455-1497.

- Liestøl, K. and Andersen, P.K., 2002. Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. *Statistics in Medicine* 21, 3701-3714.
- Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M.G., Mallinckrodt, C. and Carroll, R., 2004. Analysing incomplete longitudinal clinical trial data. *Biostatistics* 5, 445-464.
- Molenberghs, G., Beunckens, C., Sotito, C. and Kenward, M. G., 2008. Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B* 70, 371-388.
- Philipson, P.M., Ho, W.K. and Henderson, R., 2008. Comparative review of methods for handling drop-out in longitudinal studies. *Statistics in Medicine* 27, 6276-6298.
- Rubin, D.B., 1976. Inference and missing data. *Biometrika* 63, 581-592.
- Satten, G.A. and Longini, I.M., 1996. Markov chains with measurement error: estimating the true course of a marker of the progression of human immunodeficiency virus disease (with discussion). *Applied Statistics* 45(3): 275-309.
- Tsiatis, A.A. and Davidian, M.A., 2004. Joint modelling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 14, 809-834.
- van Vliet, P., Westendorp, R.G., van Heemst, D., de Craen, A.J. and Oleksik, A.M., 2010. Cognitive decline precedes late-life longitudinal changes in vascular risk factors. *J Neurol Neurosurg Psychiatry* 81(9), 1028-1032.
- der Wiel, A.B., van Exel, E., de Craen, A.J., Gussekloo, J., Lagaay, A.M., Knook, D.L. and Westendorp, R.G.J., 2002. A high response is not essential to prevent selection bias: Results from the Leiden 85-plus study. *Journal of Clinical Epidemiology* 55, 1119-1125.

Figure 1: Schizophrenia data: mean differences in responses between assessments $t - 1$ and t . The solid line is for subjects who have informative dropout between assessments t and $t + 1$ and the broken line for subjects who return for assessment $t + 1$. Baseline assessment is coded $t = 1$. Numbers in brackets give the cohort sizes

Figure 2: Leiden 85+ data: mean differences in responses between assessments $t - 1$ and t . The solid line is for subjects who have informative dropout between assessments t and $t + 1$ and the broken line for subjects who return for assessment $t + 1$. Baseline assessment is coded $t = 1$. Numbers in brackets give the cohort sizes

Figure 3: MAR underlying system model (left) and marginal MNAR empirical model (right)

Figure 4: Schizophrenia data. Observed (solid lines) and estimated dropout-free (broken lines) mean response profiles for placebo (p), standard (s) and experimental (e) treatment groups

Figure 5: Example profiles of schizophrenia data. The two left plots show profiles of the ten people most likely to be in crisis state at week four, and the right plot shows the profiles of ten people randomly selected from the remainder of risk set at week four. See text for further information