Analysing the rate of change in a longitudinal study with missing data, taking into account the number of contact attempts

Mouna Akacha [1*], Jane L. Hutton[1]

[1] *Department of Statistics, University of Warwick, United Kingdom*

November 3, 2009

SUMMARY

In longitudinal and multivariate settings incomplete data, due to missed visits, dropouts or non-return of questionnaires are quite common.

A longitudinal trial in which potentially informative missingness occurs is the Collaborative Ankle Support Trial (CAST). The aim of this study is to estimate the clinical effectiveness of four different methods of mechanical support after severe ankle sprain. The clinical status of multiple subjects was measured at four points in time via a questionnaire and, based on this, a continuous and bounded outcome score was calculated.

Motivated by this study, a model is proposed for continuous longitudinal data with non-ignorable or informative missingness, taking into account the number of attempts made to contact initial non-responders. The model combines a non-linear mixed model for the underlying response model with a logistic regression model for the reminder process.

The outcome model enables us to analyze the rate of improvement including the dependence on explanatory variables. The non-linear mixed model is derived under the assumption that the rate of improvement in a given time interval is proportional to the current score and the still achievable score. Based on this assumption a differential equation is solved in order to obtain the model of interest.

The response model relates the probability of response at each contact attempt and point in time to covariates and to observed and missing outcomes.

Using this model the impact of missingness on the rate of improvement is evaluated for different missingness processes.

1. INTRODUCTION

In clinical trials it is very common for sets of repeated measurements to be incomplete. Missingness usually occurs for reasons outside of the control of the investigators and may be related to the outcome measurement of interest, hence complicating the data analysis. In general there are three potential problems that arise with missing data: loss of efficiency, complication in data handling and analysis, and bias due to differences between the observed and unobserved data, [1].

Data from such trials can be analysed in four ways:

1. Perform the analysis only on those subjects who complete the trial;
2. Analyse only the available data;
3. Use a single or multiple imputation technique to replace the missing observations with plausible values, then analyse the completed data set(s); and
4. Model the repeated data and missingness process jointly, [2].

*Correspondence to: Mouna Akacha, Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom. E-mail:M.Akacha@warwick.ac.uk

The first option yields a *complete case analysis*. In contrast, the second option can be realised through the *direct likelihood approach*, which is the likelihood-based way of using available information only, [3]. Other, mostly nonparametric, methods of using observed data only are available, [4]. *Single* and *multiple imputation* techniques are well known, [1, 4, 5, 6, 7, 8]. Comparisons of different missing data methods with the main focus on repeated measurement studies are given in [2, 3, 9]. We focus on the fourth approach. This option 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 missing data mechanism to be assessed, [2].

According to the taxonomy of Little and Rubin [10], we distinguish between three *missing data mechanisms*, which concern the relation between the *missingness process* and the outcome variable. These are given by the *missing completely at random* (MCAR), the *missing at random* (MAR) and the *missing not at random* (MNAR) mechanisms.
A missingness process is said to be MCAR, when missingness is not related to any measurements, observed or missing in the study. In particular, the incomplete data set can be seen as a random subsample of the complete data set, which would have been observed without missingness. Furthermore, it is not necessary to construct a model for the missingness process. A missingness process which uses less restrictive assumptions is the MAR mechanism. In this case, missingness depends on observed quantities, which include outcomes and explanatory variables, but not on the missing components. If, in addition to MAR, the parameter vectors associated with the measurement and missingness process are disjoint, in the sense that the joint parameter space is the product of the single parameter spaces (*separability* or *distinctness condition*), the missing data mechanism is termed *ignorable*. Likelihood-based or Bayesian inference for the measurement parameter of interest can then be based on the observed data likelihood while ignoring the missing data mechanism, [4]. Finally, if the missingness probability depends on unknown quantities the missingness process is termed MNAR or *informative*. In the case of *non-ignorability* and MNAR, we need to model the measurement and missingness process jointly. Methods, such as *pattern-mixture models*, *shared parameter models* and *selection models* have been proposed for this case. In a pattern-mixture model the joint density of the full data is factorised into the marginal response density and the outcome density, conditional on the missingness pattern. In a shared parameter model, the density of the full data is modelled through the incorporation of random effects, which drive both the outcome and the response process. A selection model factorises the joint density of the outcome and response mechanism into the marginal outcome density and the response density, conditional on the measurements.

Although the assumption of ignorability can be realistic for certain settings, in most applications it is impossible to exclude the possibility of MNAR or non-ignorability. In particular, we cannot test for MAR itself, [11]. Therefore, many researchers recommend performing a *sensitivity analysis* in order to explore the stability of the conclusions across a range of different MAR and MNAR models.
We will focus on studies where a large number of patients drop out throughout the study, and where the reasons for dropout are expected to be related to the outcome of interest. Within a sensitivity analysis, we aim to account for informative missingness through selection models, in line with the fourth analysis option above.

In order to fit a selection model, we need to formulate models for the marginal measurement process and the conditional missingness process. Assuming a monotone missingness pattern, a logistic model for the drop-out process in combination with a multivariate normal linear model for the measurement process was proposed, [12]. The assumption of monotone missingness has been relaxed, [13, 14]. However, in [13] models for repeated binary data are discussed and the main challenge of selection models - the integration over the missing data - reduces to feasible sums. In contrast, in [14] continuous longitudinal data are analyzed. A logistic and probit model for the missingness process and a multivariate normal linear model for the outcome of interest are proposed. As in [12] and [13], the missing data model allows the probability of non-response to depend on current and previous outcomes. However, in order to facilitate the integration and the construction of the likelihood a first-order Markov dependence structure for the measurement vecotor is chosen.

We extend these models in three ways. Firstly, none of the abovementioned approaches includes additional information about the missingness process, which can be very helpful in obtaining a better understanding of the missing data mechanism, [15]. This information usually consists of proxy outcomes [16], follow-up studies on a sample of non-responders [17], collection of the reasons for drop-out or extended retrieval efforts. The additional information we will be using is of the last type. More precisely, we use the number and nature of attempts made to contact initial non-responders. Following ideas in [15, 18] we will use a multinomial model for the reminder process. The paper [18] focuses on studies with a single time point and uses logistic regression to model the response probabilities at each contact attempt. Based on these probabilities, a Horvitz-Thompson type estimator for the sample moments is proposed. The same assumptions are made in [15], but different fitting procedures and estimators are discussed: conditional likelihood method; EM algorithm and a Bayesian approach using MCMC methods. These approaches will be extended for the longitudinal case.
Secondly, instead of a multivariate linear model, which generally models the overall mean, we will be fitting a non-linear mixed model that focuses on the rate of improvement. Furthermore, this model will partly account for the bounded nature of our score data.
Thirdly, we will begin with discussing our approach for the case of monotone missingness patterns, but relax this assumption later on.

The paper is arranged as follows. The CAST study which motivated the presented work is introduced in Section 2. In Section 3 we present the selection model framework where we use the missingness indicator or the number of attempts to account for non-ignorable or informative missingness. Using this model the impact of missingness on the rate of improvement is evaluated for different missingness processes in Section 4. Concluding remarks are given in Section 5.

2. THE COLLABORATE ANKLE SUPPORT TRIAL (CAST)

The aim of the *Collaborative Ankle Support Trial* was to estimate the clinical and cost effectiveness of three different methods of mechanical support after severe ankle sprain compared to a standard treatment, [17, 19, 20].

The data for this trial were obtained from a randomised, multicentre study conducted by the *Warwick Medical School*. Within this trial patients with a severe sprain of the lateral ligament complex of the ankle and aged 16 years or older were randomised in one of the four treatment groups −*Tubigrip* (standard treatment), *Plaster of Paris* (PoP), *Aircast brace* and *Bledsoe boot*. The clinical status of these patients was measured at four points in time (baseline and follow-up: 4 weeks, 12 weeks and 9 months) via the *Foot and Ankle Outcome Score* (FAOS), which is a valid and reliable questionnaire of 42 items and 5 subscales that ascertains functional limitations and the severity of other symptoms after ligament sprains, [21].
This analysis will concentrate on 553 patients and the *symptoms*-subscale which will be referred to as *FAOSS-scale* (FAOS-symptoms subscale).
A continuous score, with 100 indicating no symptoms and 0 indicating extreme symptoms, was calculated for each subscale.

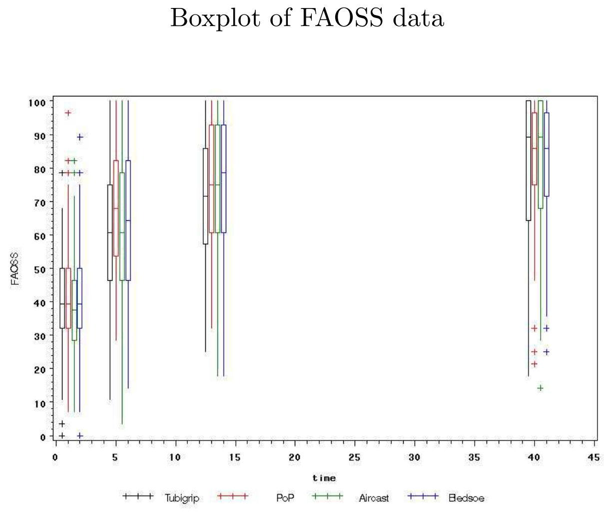

Figure 1. Boxplots of mean measures for the four randomisation groups: Tubigrip (left), PoP-cast (mid-left), Aircast brace (mid-right), Bledsoe boot (right).

For initial exploratory analysis, side-by-side boxplots of the observations for each point in time and each randomisation group are displayed in Figure 1. We use the explanatory variable *randomisation group* rather than the *treatment group* because the analysis will be performed on an intention-to-treat basis, i.e. all participants were analysed in the groups to which they were randomised, regardless of the treatment that they received. Postal questionnaires were used in an attempt to minimise loss to follow-up, and a system of reminder letters and telephone calls was instituted to follow up those who did not return their questionnaire. We distinguish between the following 'reminder categories' $z \in \{0, 1, 2, 3, 4, 5\}$:

- $z = 0$: questionnaire returned - no chasing;
- $z = 1$: questionnaire returned after telephone chase;
- $z = 2$: questionnaire returned after 2nd copy sent with no further telephone chasing;
- $z = 3$: questionnaire returned after 2nd copy sent with further telephone chasing;
- $z = 4$: core outcomes obtained over the telephone;
- $z = 5$: non responder.

The frequency for each category and time point is displayed in Table I. For further details concerning CAST and its design refer to [17, 19, 20].

3. MODEL

In the following subsections, we propose a selection model to adjust for non-ignorable or informative missingness in the case of a longitudinal study with continuous data.

3.1. Notation

Let $y_i = (y_{i,0}; y_{i,4}; y_{i,12}; y_{i,39})^\top$ denote the response vector of subject $i \in \{1, ..., 553\}$, where y_i is a realisation of the random vector Y_i. We assume that Y_i is multivariate normal distributed and denote the joint outcome vector for all subjects by $Y = (Y_1^\top, ..., Y_{553}^\top)^\top$.
Furthermore, let $X_i = (x_{i,0}; x_{i,4}; x_{i,12}; x_{i,39})^\top$ be the matrix of explanatory variables (e.g. time, gender and age (log-transformed)) for subject $i \in \{1, ..., 553\}$. The randomisation group is denoted by $trt_i \in \{1, 2, 3, 4\}$, where $trt_i = 1$ corresponds to Tubigrip, $trt_i = 2$ to PoP-cast, $trt_i = 3$ to Aircast

	# of attempts					
time point	0	1	2	3	4	5
baseline	553 (100%)	0	0	0	0	0
4 weeks	187 (33.8%)	152 (27.5%)	53 (9.6%)	40 (7.2%)	35 (6.3%)	86 (15.6%)
12 weeks	146 (26.4%)	141 (25.5%)	46 (8.3%)	48 (8.7%)	78 (14.1%)	94 (16.7%)
39 weeks	124 (22.4%)	117 (21.3%)	42 (7.6%)	59 (10.7%)	81 (14.7%)	130 (23.5%)
# Total quest. returned	1010	410	141	147	194	310

Table I. Overview of the number of reminders needed to retrieve a questionnaire. In brackets the percentage of the returned questionnaires per attempt category is given for each time.

brace and finally $trt_i = 4$ to Bledsoe boot.
The indicator r_{ij} is a realisation of the random variable R_{ij} which denotes whether y_{ij} was observed, $r_{ij} = 1$, or missing, $r_{ij} = 0$. We summarize the missingness information for subject i through $R_i = (R_{i,0}, R_{i,4}, R_{i,12}, R_{i,39})^\top$ and for all subjects through $R = (R_1^\top, ..., R_{553}^\top)^\top$. Moreover, z_{ij} represents the number of reminders needed and is a realisation of the random variable Z_{ij}.
We aim to analyse the relationship between the response variable Y_i and the explanatory variables X_i for all $i \in \{1, ..., 553\}$, taking into account the missingness process and the number of reminders needed to retrieve a questionnaire.

3.2. *Selection Models*

Suppose the complete data Y follows the parametric model $\mathrm{P}(\theta)$ and R follows the parametric model $\mathrm{P}(\phi)$. We partition the vector Y into the observed, Y_{obs}, and unobserved part, Y_{mis}. If the missingness process is non-ignorable or informative we need to base inference for θ on the joint likelihood of Y_{obs} and the missingness indicator R. A selection model factorises the joint model of the measurement process and the response mechanism into the marginal measurement process and the response process, conditional on the measurements. Thus, the joint likelihood for Y_{obs} and R is given by

$$\begin{aligned} L_{Y_{obs},R}(\theta,\phi) &= \prod_{i=1}^{553} \int f\left(y_{i,obs}, y_{i,mis}, r_i | X_i, \theta, \phi\right) \mathrm{d}y_{i,mis} \qquad (1)\\ &= \prod_{i=1}^{553} \int f\left(y_{i,obs}, y_{i,mis} | X_i, \theta\right) f\left(r_i | X_i, y_{i,obs}, y_{i,mis}, \phi\right) \mathrm{d}y_{i,mis}. \end{aligned}$$

As $z_{ij} \in \{0,1,2,3,4\} \Leftrightarrow r_{ij} = 1$ and $z_{ij} = 5 \Leftrightarrow r_{ij} = 0$ we can extend the selection model by adjusting for non-response through z_{ij} rather than r_{ij}. The motivation for this approach lies in the hypothesis that subjects who reply after several reminders might be more similar to non-responders, than those who reply at the very first attempt. Note that this is not an assumption. Our modelling strategy is flexible enough to explore the plausibility of this hypothesis.
The extension of model (1) is straightforward. Let Z follow the parametric model $\mathrm{P}(\psi)$ then

$$L_{Y_{obs},Z}(\theta,\psi) = \prod_{i=1}^{553} \int f\left(y_{i,obs}, y_{i,mis} | X_i, \theta\right) f\left(z_i | X_i, y_{i,obs}, y_{i,mis}, \psi\right) \mathrm{d}y_{i,mis}. \qquad (2)$$

Fitting the model in equation (1) requires a marginal model for the outcome vector Y_i and a model for the missingness process, conditional on the outcome. Similarly, we need to formulate models for the outcome process and the conditional reminder process when focusing on model (2). We propose a non-linear mixed model for the marginal outcome process with a logistic regression model for the conditional response and the conditional reminder process.

3.3. Outcome Model

The FAOSS score is observed repeatedly over time. Hence, we are confronted with longitudinal data. In this context, classical assumptions - especially independence - are not satisfied. As soon as more than one observation of a characteristic is made for each subject, we have to distinguish carefully between the different sources of variation. Neglecting the correlation between the measurements of each subject is inadvisable as the precision of the results and thereby their significance is usually overestimated, [22].
The most commonly used approaches for analysing continuous longitudinal data are multivariate linear models and *marginal* or *random effect* models. The multivariate linear model formulates a linear model for each component of the observation vector with correlated errors. In marginal models, the regression between the marginal mean of the response vector Y_i and the explanatory matrix X_i is modelled through a linear model; the dependence structure across the entries of Y_i is modelled separately. In contrast, random effect models formulate a mixed model; additional modelling of the intra-individual variation is not necessary.
If we assume normality, these models fit a linear regression, taking into account the intra-individual variation. In general, a linear change with time is assumed. However, due to the bounded nature of the outcome, we expect the score to increase faster at the beginning of the study and to move slowly towards the end. Based on exploratory analysis, see Figure 1, we propose a non-linear mixed model that models the rate of improvement in dependence of explanatory variables and which takes the bounded nature of the score into account.
We propose the following mixed model for the marginal outcome process

$$\begin{aligned} Y_i|U_i &\overset{ind.}{\sim} \mathcal{N}_4\left(\eta_i, \sigma^2 I_4\right); \\ U_i &\overset{iid}{\sim} \mathcal{N}(0, D^2); \\ \eta_{ij} &= g(x_{ij}, \theta_i) \quad \text{for } j \in \{0, 4, 12, 39\}, \end{aligned}$$

where g is the non-linear model function and the parameter vector θ_i varies across subjects. For convenience we omit the i-subscript for θ_i in the derivation of the non-linear model.
The FAOSS score is bounded, therefore we expect the rate of improvement in a given time interval, $g'(x_{ij}, \theta)$, to be proportional to the current score, $g(x_{ij}, \theta)$, and the still achievable score $[\max\{g(x_{ij}, \theta)\} - g(x_{ij}, \theta)]$. Thus, we are interested in solving the differential equation

$$g'(x_{ij}, \theta) = \kappa_{trt_i}\, g(x_{ij}, \theta)\, [\max\{g(x_{ij}, \theta)\} - g(x_{ij}, \theta)]$$

where κ_{trt_i} for $trt_i \in \{1, 2, 3, 4\}$ is the treatment-specific proportion-factor. Reducing the problem to $x_{ij} = t_j \in \{0, 4, 12, 39\}$ yields the solution

$$g(x_{ij}, \theta) = \frac{\beta_1}{e^{-\beta_{2,trt_i} t_j}\left(\frac{\beta_1}{\beta_0} - 1\right) + 1},$$

where

- $\beta_0 = g(0, \theta)$ describes the intercept;
- $\beta_1 = \max\{g(t_j, \theta)\}$ describes the maximum score and
- $\beta_{2,trt_i} = \beta_1 \cdot \kappa_{trt_i}$ describes the acceleration of the non-linear function g.

Incorporating the explanatory variable *age* (log-transformed) is straightforward:

$$g(x_{ij}, \theta) = \frac{\beta_1 + \alpha_1 \cdot age}{\exp(-[\beta_{2,trt_i} + \alpha_2 \cdot age] \cdot t_j)\left(\dfrac{\beta_1 + \alpha_1 \cdot age}{\beta_0 + \alpha_0 \cdot age} - 1\right) + 1}$$

and in order to capture the inter-individual variation, we extend this model to a mixed model:

$$g(x_{ij}, \theta_i) = \frac{\beta_1 + \alpha_1 \cdot age}{\exp(-[\beta_{2,trt_i} + \alpha_2 \cdot age] \cdot t_j) \left(\dfrac{\beta_1 + \alpha_1 \cdot age}{\beta_0 + \alpha_0 \cdot age} - 1 \right) + 1} + U_i. \tag{3}$$

Further explanatory variables can be included in the same fashion. Analyses based on the assumption of ignorability lead us to use model (3) without any age-effect on the intercept. Hence, $\theta_i = (\theta^\top, U_i)^\top$ with $\theta = (\beta_0, \beta_1, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \alpha_1, \alpha_2, \sigma, D)^\top$. This model can easily be reformulated in terms of the multivariate normal model with a compound symmetry covariance structure:

$$Y_i \sim \mathcal{N}_4 (\mu_i, \Sigma), \tag{4}$$

where

$$\mu_{ij} = \frac{\beta_1 + \alpha_1 \cdot age_i}{\exp\{-(\beta_{2,trt_i} + \alpha_2 \cdot age_i)\, t_j\} \left(\dfrac{\beta_1 + \alpha_1 \cdot age_i}{\beta_0} - 1 \right) + 1}$$

and

$$\Sigma = \begin{pmatrix} \sigma^2 + D^2 & D^2 & D^2 & D^2 \\ D^2 & \sigma^2 + D^2 & D^2 & D^2 \\ D^2 & D^2 & \sigma^2 + D^2 & D^2 \\ D^2 & D^2 & D^2 & \sigma^2 + D^2 \end{pmatrix}.$$

3.4. Missingness Process Model

We now consider modelling the missingness process, conditional on the outcome of interest. In the spirit of regression modelling, we propose the following logistic linear model for all $i \in \{1, ..., 553\}$, $j \in \{4, 12, 39\}$:

$$\begin{aligned} R_{ij} = 1 | Y_i, X_{ij} &\sim Bernoulli(\rho_{ij}), \quad \text{where} \\ logit(\rho_{ij}) &= \phi_0 + \phi_1\, age_i + \phi_2\, t_j + \phi_3\, y_{i,j-1} + \phi_4\, y_{ij} \end{aligned} \tag{5}$$

where $t_j \in \{4, 12, 39\}$, y_{j-1} is the previous outcome and y_j the current score. This model corresponds to MNAR if $\phi_3 \neq 0 \neq \phi_4$. The special cases of model (5) corresponding to MAR and MCAR (conditioned on covariates) are obtained from setting $\phi_4 = 0$ or $\phi_3 = 0 = \phi_4$, respectively.
Fitting the following model

$$logit(\rho_{ij}) = \phi_0 + \phi_1\, age_i + \phi_2\, t_j + \phi_3\, y_{i,j-1} + \phi_4\, y_{ij} + \phi_{5,trt_i},$$

which allows dropout rates to differ by treatment unfortunately does not result in a positive definite Hessian matrix, even when assuming $\phi_4 = 0$.

3.5. Reminder Process Model

At first glance the geometric and Poisson model seem realistic to capture the characteristics of the attempt process. However, the lack of monotonic frequencies in the reminder categories discourages use of these models (see Table I). Following ideas in [15, 18] we will therefore focus on a multinomial model for the attempt process.
We develop a model for a single subject. In view of the assumed independence between subjects, it is then easy to build the complete model.
For the time points $j \in \{4, 12, 39\}$ let p_{j0} be the probability of responding at the very first attempt. For $k \in \{1, 2, 3, 4\}$ let p_{jk} denote the probability of responding at the k-th attempt, given that the subject has not responded earlier. According to the study design we know that the probability of

responding at the first attempt at baseline, i.e. p_{00}, is one.
The unconditional probabilities μ_{jk} of replying at attempt k and time point j are then given by:

$$\begin{aligned}\mu_{j0} &= p_{j0};\\ \mu_{j1} &= p_{j1}\,(1-p_{j0});\\ &\vdots\\ \mu_{j4} &= p_{j4}\prod_{k=0}^{3}(1-p_{jk}).\end{aligned}$$

Furthermore, we define $\mu_{j5} = 1 - \sum_{k=0}^{4}\mu_{jk}$ as the probability of not replying at time point $j \in \{4, 12, 39\}$, i.e. $z_j = 5$. Corresponding to these probabilities we redefine the random variable Z_j in terms of an indicator random vector. Let V_j be a six-dimensional random vector, where

$$V_{j\ell} = \begin{cases} 1, & \text{if attempt } Z_j = \ell - 1;\\ 0, & \text{otherwise}\end{cases}$$

for $\ell \in \{1, ..., 6\}$. Thus, for a certain subject all information about Z is now captured through the indicator matrix $V = (V_4, V_{12}, V_{39})^\top$ and the model in (2) can be reformulated as

$$L_{Y_{obs},V}\,(\theta,\psi) \;=\; \prod_{i=1}^{553}\int f\,(y_{i,obs}, y_{i,mis}|X_i,\theta)\; f\,(v_i|X_i, y_{i,obs}, y_{i,mis},\psi)\;\mathrm{d}y_{i,mis}. \qquad (6)$$

We assume

$$V_j \sim \mathit{Multinomial}\,(1, \mu_{j0}, ..., \mu_{j5})\,. \qquad (7)$$

Dependent on the required inference, a generalized linear model for μ_{jk} or p_{jk} can be formulated. The marginal probability μ_{jk} determines the chance of replying at the k-th attempt. In contrast, formulating a model for the conditional probability p_{jk} investigates the effect of covariates on replying at the k-th attempt, given the previous attempts were unsuccessful. Given that the attempt process evolves over time it is sensible to explore the latter case. Nevertheless, we will explore both modelling approaches and compare the inference.

3.5.1. Modelling p_{jk} The generalized linear model we propose for p_{jk} and $j \in \{4, 12, 39\}$, $k \in \{1, 2, 3, 4\}$ is given by

$$\mathrm{logit}(p_{jk}) = \psi_{0k} + \psi_1\, age + \psi_{2,trt}\, t_j + \psi_3\, y_{j-1} + \psi_4\, y_j + \psi_{5,trt} \qquad (8)$$

where y_{j-1} is the previous outcome and y_j the current score. This general model allows for different missingness mechanisms; MAR is implied by $\psi_4 \equiv 0$ and, conditioned on covariates, MCAR is implied by $\psi_3 \equiv 0 \equiv \psi_4$. Fitting this model to the complete cases only, using the `SAS`-procedure `NLMIXED`, showed a high negative correlation for the parameter estimates $\hat{\psi}_{0k}$ and $\hat{\psi}_1$. We therefore excluded ψ_1 from the model. After refitting the model, the corresponding p-values suggested setting $\psi_{00} \equiv 0 \equiv \psi_{02}$. Furthermore, there were no significant differences for the time to treatment interactions and solely $\psi_{5,4}$, i.e. an effect of the treatment Bledsoe boot was shown to be significant. Thus, we replace $\psi_{2,trt}$ by ψ_2 and $\psi_{5,trt}$ by $\psi_5\, \mathbb{1}\,(trt = 4)$. The model of interest is then

$$\mathrm{logit}(p_{jk}) = \psi_{0k} + \psi_2\, t_j + \psi_3\, y_{j-1} + \psi_4\, y_j + \psi_5\, \mathbb{1}\,(trt = 4)\,. \qquad (9)$$

In spite of a high negative correlation between the parameters ψ_3 and ψ_4 we keep both parameters in the model. Also, following [12], we consider a slightly different model:

$$\mathrm{logit}(p_{jk}) = \psi_{0k} + \psi_2\, t_j + \psi_3^*\, [y_{j-1} + y_j] + \psi_4^*\, [y_{j-1} - y_j] + \psi_5\, \mathbb{1}\,(trt = 4) \qquad (10)$$

where ψ_3^* and ψ_4^* are usually less correlated.

3.5.2. Modelling μ_{jk} When modelling the unconditional probabilities of the multinomial distribution given in equation (7), we took the non-responders as the reference category for the dependent variable V_{ij}. Thus we are interested in the odds of being in an attempt category $z_{ij} \in \{0, 1, 2, 3, 4\}$ versus observing $z_{ij} = 5$. Let $\mu_{j,ref}$ denote the probability of not replying at time point j. We propose the following generalized linear model for $j \in \{4, 12, 39\}$ and $k \in \{0, 1, 2, 3, 4\}$:

$$m_{jk} := \log\left(\frac{\mu_{jk}}{\mu_{j,ref}}\right) = \lambda_0 + \lambda_{1,k}\, age + \lambda_2\, t_j + \lambda_3\, y_{j-1} + \lambda_4\, y_j, \tag{11}$$

with

$$\begin{aligned} \lambda_3 = 0 = \lambda_4 &\;\;\Rightarrow\;\; \text{MCAR;} \\ \lambda_4 = 0 &\;\;\Rightarrow\;\; \text{MAR; and} \\ \lambda_4 \neq 0 &\;\;\Rightarrow\;\; \text{MNAR.} \end{aligned}$$

Moreoever, we obtain

$$\mu_{j,ref} = \frac{1}{1 + \exp(m_{j0}) + \exp(m_{j1}) + \exp(m_{j2}) + \exp(m_{j3}) + \exp(m_{j4})}.$$

Allowing the reminder probabilities to differ by randomisation groups results in a non-positive definite Hessian matrix. This might be due to the sparse attempt-by-randomisation group data matrix.

3.6. Full Model for Monotone Missingness

Using these models, we can construct the likelihood in the case of monotone missingness. The derivations will be shown for a selection model that uses the reminder process (via V_{ij}) to account for missingness. The likelihood using the missingness indicator process R_{ij} can be derived by replacing V_{ij} by R_{ij} in all the following equations.
The selection model is given by

$$f(y_i, v_i | X_i, \theta, \psi) = f(y_i | X_i, \theta)\; f(v_i | X_i, y_i, \psi).$$

Let the available data for subject i be denoted by $Y_{i,obs}$, the missing information be summarized in $Y_{i,mis}$ and $V_i = (V_{i,0}^\top, ..., V_{i,39}^\top)^\top$ be the attempt data. The observed data likelihood contribution of a certain subject is then given by:

$$\begin{aligned} f(y_{i,obs}, v_i | X_i, \theta, \psi) &= \int f(y_i | X_i, \theta)\; f(v_i | X_i, y_i, \psi)\, \mathrm{d}y_{i,mis} \\ &= \int f(y_{i,obs}, y_{i,mis} | X_i, \theta)\; f(v_i | X_i, y_{i,obs}, y_{i,mis}, \psi)\, \mathrm{d}y_{i,mis}. \end{aligned} \tag{12}$$

Now assume that dropout for the subject of interest occurs after the second measurement time, i.e. $Y_{i,obs} = (Y_{i,0}, Y_{i,4})^\top$, then

$$\begin{aligned} f(y_{i,obs}, v_i | X_i, \theta, \psi) &= \int f(y_{i,obs}, y_{i,mis} | X_i, \theta)\; f(v_i | X_i, y_{i,obs}, y_{i,mis}, \psi)\, \mathrm{d}y_{i,mis} \\ &= \int\!\!\int f(y_{i,39} | y_{i,12}, y_{i,4}, y_{i,0}, X_i, \theta)\, f(y_{i,12} | y_{i,4}, y_{i,0}, X_i, \theta)\, f(y_{i,4} | y_{i,0}, X_i, \theta)\, f(y_{i0} | X_i, \theta) \\ &\quad \times f(v_{i,39} | v_{i,12}, v_{i,4}, v_{i,0}, X_i, y_i, \psi)\, f(v_{i,12} | v_{i,4}, v_{i,0}, X_i, y_i, \psi) \\ &\quad \times f(v_{i,4} | v_{i,0}, X_i, y_i, \psi)\, \mathrm{d}y_{i,39}\, \mathrm{d}y_{i,12}. \end{aligned}$$

In the case of monotone missingness we observe

$$v_{t_j} = (0, 0, 0, 0, 0, 1) \Longrightarrow v_{t_{j+1}} = (0, 0, 0, 0, 0, 1) \tag{13}$$

for $t_j \in \{4, 12, 39\}$ and $t_{j+1} \in \{12, 39\}$. Therefore,

$$f\{v_{i,39} = (0, 0, 0, 0, 0, 1) | v_{i,12} = (0, 0, 0, 0, 0, 1), v_{i,4}, v_{i,0}, X_i, y_i\} = 1.$$

Rearranging the observed likelihood yields

$$\begin{aligned}
f\left(y_{i,obs}, v_i | X_i, \theta, \psi\right) &= f\left(y_{i,4}|y_{i,0}, X_i, \theta\right) f\left(y_{i,0}|X_i, \theta\right) f\left(v_{i,4}|v_{i,0}, X_i, y_{i,4}, y_{i,0}, \psi\right) \\
&\quad \times \int f\left(y_{i,12}|y_{i,4}, y_{i,0}, X_i, \theta\right) f\left(v_{i,12}|v_{i,4}, v_{i,0}, X_i, y_{i,12}, y_{i,4}, \psi\right) \\
&\quad \times \underbrace{\int f\left(y_{i,39}|y_{i,12}, y_{i,4}, y_{i,0}, X_i, \theta\right) \mathrm{d}y_{i,39}}_{=1} \mathrm{d}y_{i,12}. \\
&= f\left(y_{i,4}|y_{i,0}, X, \theta\right) f\left(y_{i,0}|X_i, \theta\right) f\left(v_{i,4}|v_{i,0}, X_i, y_{i,4}, y_{i,0}, \psi\right) \\
&\quad \times \int f\left(y_{i,12}|y_{i,4}, y_{i,0}, X_i, \theta\right) f\left(v_{i,12}|v_{i,4}, v_{i,0}, X_i, y_{i,12}, y_{i,4}, \psi\right) \mathrm{d}y_{i,12}.
\end{aligned}$$

Furthermore, we assume that the components of the random vector V_i are independent, given the previous or current outcome and covariates, unless a scenario like in equation (13) holds. Thus,

$$\begin{aligned}
f\left(y_{i,obs}, v_i | X_i, \theta, \psi\right) &= f\left(y_{i,4}|y_{i,0}, X_i, \theta\right) f\left(y_{i,0}|X_i, \theta\right) f\left(v_{i,4}|X_1, y_{i,4}, y_{i,0}, \psi\right) \\
&\quad \times \int f\left(y_{i,12}|y_{i,4}, y_{i,0}, X_i, \theta\right) f\left(v_{i,12}|X_i, y_{i,12}, y_{i,4}, \psi\right) \mathrm{d}y_{i,12},
\end{aligned}$$

i.e. the integrals reduce to one-dimensional integrals for $i \in \{1, ..., 553\}$. These integrals can be solved through an adaptive Romberg-type integration technique. This approach produces a quick, rough estimate of the integration result and then refines the estimate until achieving the prescribed accuracy, [23, 24]. It is implemented in `call quad` within the `proc IML` environment in `SAS`. The maximum likelihood estimates for θ and ψ can then be calculated through the Newton-Raphson ridge optimization method (e.g. `call nlpnrr` in `proc IML`). Corresponding macros are available from the authors.

3.7. Full Model for Non-Monotone Missingness

As soon as we relax the assumption of monotone missingness, we are confronted with multidimensional integrals, because condition (13) not longer holds. Attempts to run the `SAS` code which accounts for non-monotone missingness failed, because every iteration step required the calculation of 331 integrals and several hours of computing time.

Therefore, we decided to integrate over the missing data within the Bayesian paradigm using `WinBUGS` which carries out the Gibbs sampling algorithm, [25]. However, due to the complexity of the model and convergence problems of the Monte Carlo Markov Chains produced by `WinBUGS`, we had to simplify the outcome model given in equation (4). Instead, we are considering an outcome model where no age effect is incorporated, i.e. $\alpha_1 = \alpha_2 = 0$.

Furthermore, we focus on the reminder process given in equation (10) and the missingness process MNAR$_r$-2, Section 3.4. Attempts to extend the model in Section 3.5.2 to account for non-monotone missingness failed due to convergence problems of the corresponding Markov chains. We chose vague priors influenced by the likelihood-based estimates obtained under the monotone missingness assumption:

$$\begin{aligned}
\beta_0 &\sim \mathcal{N}(40, 100);\\
\beta_1 &\sim \mathcal{N}(100, 100);\\
\beta_{2,trt_i} &\sim \mathcal{N}(0, 10) \quad \text{for } C \in \{1,2,3,4\};\\
\psi_{0,k} &\sim \mathcal{N}(0, 10) \quad \text{for } k \in \{1,3,4\};\\
\psi_j &\sim \mathcal{N}(0, 100) \quad \text{for } j \in \{1,3\};\\
\psi_j^* &\sim \mathcal{N}(0, 100) \quad \text{for } j \in \{4,5\};\\
\phi_j &\sim \mathcal{N}(0, 10) \quad \text{for } j \in \{0,1,2,3,4\};\\
\frac{1}{\sigma} &\sim \Gamma(0.1, 0.001) \quad \text{and}\\
\frac{1}{D} &\sim \Gamma(0.1, 0.001).
\end{aligned}$$

A burn in of 100,000 and a further 100,000 iterations were performed to make inference. Furthermore, we used the option of over-relaxation which generates several samples at each iteration and then selects one that is negatively correlated to the current state of the chain, [26]. The traces of all simulated variables were examined and different starting values were used in order to verify convergence of the chains. The means of the resulting posterior distributions are used as estimates.

4. RESULTS

In this section we compare the results for the proposed selection models under different assumptions for the missingness mechanism and missingness pattern. In the case of a monotone missingness pattern, we use the outcome model given in equation (4), whereas in the case of non-monotone missingness we remove the age effect. Note that we originally observe a non-monotone missingness pattern in the data set. In order to be able to use our model for monotone missingness, we deleted all those observations that were made after a patient failed to return a previous questionnaire.
The results for the monotone and non-monotone case will be compared with those obtained based on the assumption of ignorability, see Table II and Table VI.

4.1. Results using the Missingness Process Model

When adjusting for missingness through the missingness indicator R_{ij}, we explore the impact of missingness on the rate of improvement for the following missingness processes:

- MCAR$_r$: $\text{logit}(\rho_{ij}) = \phi_0 + \phi_1 age_i + \phi_2 t_j$;
- MAR$_r$: $\text{logit}(\rho_{ij}) = \phi_0 + \phi_1 age_i + \phi_2 t_j + \phi_3 y_{i,j-1}$;
- MNAR$_r$-1: $\text{logit}(\rho_{ij}) = \phi_0 + \phi_1 age_i + \phi_2 t_j + \phi_4 y_{ij}$; and
- MNAR$_r$-2: $\text{logit}(\rho_{ij}) = \phi_0 + \phi_1 age_i + \phi_2 t_j + \phi_3 y_{i,j-1} + \phi_4 y_{ij}$.

Note that here MCAR denotes a mechanism where missingness is allowed to depend on covariates but not on the outcome of interest. Monotone missingness results are given in Table III.
The results of the measurement model are identical under the MCAR and the MAR missing data assumptions. The estimated intercept $\hat{\beta}_0$ is the same under all missing data assumptions.
In contrast, the parameter estimates $\hat{\beta}_1$ and $\hat{\alpha}_1$ are slightly smaller when adjusting for MNAR processes. Our intuition that older participants achieve a lower maximum score than younger participants is confirmed. Furthermore, all models confirm that older participants recover less fast

	Monotone Missingness		
Parameter	Est.	SE	p-val.
β_0	41.04	0.77	< 0.0001
β_1	110.29	6.79	< 0.0001
β_{21}	0.950	0.14	< 0.0001
β_{22}	1.069	0.13	< 0.0001
β_{23}	1.001	0.13	< 0.0001
β_{24}	0.962	0.13	< 0.0001
α_1	-9.292	2.05	< 0.0001
α_2	-0.208	0.04	< 0.0001
σ^2	186.63	7.35	< 0.0001
D^2	144.79	12.32	< 0.0001
$\beta_{21} - \beta_{22}$	-0.119	0.04	0.0054
$\beta_{21} - \beta_{23}$	-0.056	0.04	0.1383
$\beta_{21} - \beta_{24}$	-0.012	0.03	0.7217

Table II. Overview of the parameter estimates, standard errors and p-values for the outcome model (4) based on the assumptions of an ignorable missingness process.

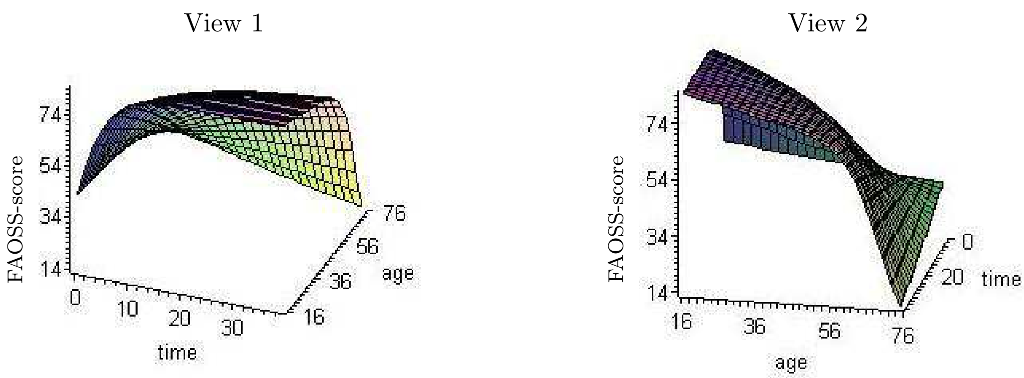

Figure 2. Fitted non-linear mixed model based on the assumption of ignorability and the Tubigrip group. The left view focuses on the development over time, whereas the right view concentrates on the age effect.

than younger patients, see Figure 2.

The treatment effects are marginally smaller for the MNAR processes than for the MCAR and the MAR process. The treatment differences are essentially unchanged. All approaches detect that PoP-cast is significantly better than Tubigrip. Aircast is marginally better and Bledsoe is not measurably different from Tubigrip.

The parameters modelling the asymptote of the non-linear curve, i.e. β_1 and α_1, are different under the assumption of ignorability (Table II) from those obtained in Table III. However, the average age of 30 years leads to approximately the same maximum value. Under ignorability, a significant treatment difference is only observed between PoP-cast and Tubigrip. The inter- and intra-individual variance parameters and their standard deviations are nearly identical under all models.

All models show increasing probabilities of replying with age. The effect for MNAR models is largest and smallest for MCAR model. No time effect is found for any missing data models. Furthermore, the MAR model suggests that patients with high score at the previous occasion are more likely to return their questionnaire (p-value is 0.06). The MNAR$_r$-1 model finds a significant effect of the current

Parameter	MCAR$_r$			MAR$_r$			MNAR$_r$-1			MNAR$_r$-2		
	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.
β_0	41.04	0.77	< 0.0001	41.04	0.77	< 0.0001	41.03	0.78	< 0.0001	41.03	0.78	< 0.0001
β_1	106.89	6.77	< 0.0001	106.89	6.82	< 0.0001	105.57	6.69	< 0.0001	105.70	6.91	< 0.0001
β_{21}	0.968	0.13	< 0.0001	0.968	0.14	< 0.0001	0.939	0.13	< 0.0001	0.942	0.14	< 0.0001
β_{22}	1.087	0.13	< 0.0001	1.087	0.13	< 0.0001	1.059	0.13	< 0.0001	1.061	0.13	< 0.0001
β_{23}	1.024	0.13	< 0.0001	1.024	0.13	< 0.0001	0.994	0.13	< 0.0001	0.997	0.13	< 0.0001
β_{24}	0.979	0.12	< 0.0001	0.979	0.13	< 0.0001	0.952	0.12	< 0.0001	0.955	0.13	< 0.0001
α_1	-8.276	2.03	< 0.0001	-8.276	2.05	< 0.0001	-7.997	2.00	< 0.0001	-8.023	2.06	< 0.0001
α_2	-0.213	0.04	< 0.0001	-0.213	0.04	< 0.0001	-0.207	0.04	< 0.0001	-0.207	0.04	< 0.0001
σ^2	186.27	7.32	< 0.0001	186.27	7.32	< 0.0001	186.53	7.35	< 0.0001	186.47	7.32	< 0.0001
D^2	147.52	12.61	< 0.0001	147.52	12.60	< 0.0001	148.06	12.66	< 0.0001	148.00	12.69	< 0.0001
$\beta_{21} - \beta_{22}$	-0.118	0.04	0.0028	-0.118	0.04	0.0028	-0.120	0.04	0.0023	-0.120	0.04	0.0024
$\beta_{21} - \beta_{23}$	-0.055	0.04	0.0720	-0.055	0.04	0.0721	-0.055	0.04	0.0685	-0.055	0.04	0.0689
$\beta_{21} - \beta_{24}$	-0.011	0.03	0.3771	-0.011	0.03	0.3771	-0.011	0.03	0.3433	-0.013	0.03	0.3469
ϕ_0	-0.707	0.77	0.1784	-1.150	0.85	0.0879	-1.874	1.02	0.0328	-1.827	1.15	0.0560
ϕ_1	0.772	0.23	0.0005	0.824	0.24	0.0004	0.914	0.25	0.0001	0.907	0.27	0.0005
ϕ_2	0.007	0.01	0.1120	0.001	0.01	0.4274	0.002	0.006	0.4037	0.001	0.01	0.4263
ϕ_3				0.006	0.004	0.0613				0.001	0.01	0.4453
ϕ_4							0.011	0.006	0.0315	0.010	0.01	0.1629
-2ℓ	16421.06			16418.64			16417.62			16417.60		

Table III. Parameter estimates, standard errors, p-values and deviances for the outcome model (4), the response model given in equation 5 and different missing data mechanisms.

score on the missingness probabilities: as the score increases, the probability of being a non-responder decreases. This result is counter-intuitive and does not correspond with quantitative findings [27], which suggest that patients who considered themselves to have made fully recovery, did not return their subsequent questionnaire. We will scrutinize this observation in the next subsection. For the MNAR$_r$-2 model, no significant effect of current or previous score is found. This might be partly due to the high correlation of scores at adjacent occasions.

4.2. Results using the Reminder Process Model via p_{jk}

Using the notation in Subsection 3.5.1 we will investigate the following logistic regression models for the conditional reminder process probabilties p_{jk}:

- MCAR$_p$: $\text{logit}(p_{jk}) = \psi_{0k} + \psi_2\, t_j + \psi_{5,trt}\, \mathbb{1}\,(trt = 4)\,;$
- MAR$_p$: $\text{logit}(p_{jk}) = \psi_{0k} + \psi_2\, t_j + \psi_3\, y_{j-1} + \psi_{5,trt}\, \mathbb{1}\,(trt = 4)\,;$
- MNAR$_p$-1: $\text{logit}(p_{jk}) = \psi_{0k} + \psi_2\, t_j + \psi_3\, y_{j-1} + \psi_4\, y_j + \psi_{5,trt}\, \mathbb{1}\,(trt = 4)\,;$
- MNAR$_p$-2: $\text{logit}(p_{jk}) = \psi_{0k} + \psi_2\, t_j + \psi_4\, y_j + \psi_{5,trt}\, \mathbb{1}\,(trt = 4)\,;$ and
- MNAR$_p$-3: $\text{logit}(p_{jk}) = \psi_{0k} + \psi_2\, t_j + \psi_3^*\, [y_{j-1} + y_j] + \psi_4^*\, [y_{j-1} - y_j] + \psi_5\, \mathbb{1}\,(trt = 4)\,.$

where $k \in \{0, 1, 2, 3, 4\}$ and $\psi_{00} \equiv \psi_{02} = \psi_{04} \equiv 0$. The reasons for removing $\psi_{00} = \psi_{02}$ from the reminder process were discussed in Section 3. We set $\psi_{04} = 0$, because fitting some of the above models in the presence of ψ_{04} lead to identifiability problems. The results for the case of monotone missingness are shown in Table IV.

The estimates for the outcome model are practically identical for all reminder processes investigated, and differ from those of the informative missingness processes just by slightly smaller treatment effects. Thus, the conclusions for the outcome model are as discussed in Section 4.1. The treatment effects are slightly larger across all models than under the ignorability assumption (Table II).

For the reminder process z, given the outcome y, we estimate a positive effect of phone calls on the retrieval of questionnaires, although only ψ_{01} is shown to be significant. Furthermore, the probability of replying at a certain attempt decreases as time passes under every reminder mechanism investigated. This effect was not observed in the response models (Section 4.1). Patients allocated to Bledsoe boot are more likely to reply, this effect is only borderline under MCAR.

The MAR results confirm that the reminder process, and therefore the missingness process, depend on the outcome of interest. The probability of returning a questionnaire decreases with the score at the prior occasion: patients with a high score tend to return the questionnaires only after several attempts or not at all. These results are the reverse of those shown in Section 4.1, perhaps due to the incorporated age-effect in Section 4.1. We omitted this effect due to the models' complexity and failed attempts to fit the model with added age-effect. These contradicting results might be explained through the negative dependence of the age of patients and their score. In Section 4.1, we found an increasing probability of replying with increasing age. Thus, the age of patients or their corresponding score might partially adjust for missingness. Nevertheless, the score itself is shown to have a significant effect on the response probabilities, even when an age-effect is included, see Section 4.1, model MNAR$_r$-1; only the conclusion changes.

The probability of replying at a certain time also decreases with the score at that time in case of MNAR$_r$-2, but including both previous and current score is not informative. The alternative parametrization introduced in equation (10) shows the reminder process depends on the mean score, not the difference or improvement. The high correlation between the scores at adjacent occasions means that either score can be used, i.e. an MAR model is adequate. In general, unless the rate of improvement, not the actual health, drives the response process an MAR will be adequate.

Parameter	MCAR$_p$			MAR$_p$			MNAR$_p$-1			MNAR$_p$-2			MNAR$_p$-3		
	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.
β_0	41.04	0.78	< 0.0001	41.04	0.77	< 0.0001	41.04	0.77	< 0.0001	41.04	0.78	< 0.0001	41.04	0.78	< 0.0001
β_1	106.89	6.79	< 0.0001	106.89	6.89	< 0.0001	106.82	6.79	< 0.0001	106.78	6.77	< 0.0001	106.82	7.12	< 0.0001
β_{21}	0.968	0.14	< 0.0001	0.968	0.14	< 0.0001	0.966	0.14	< 0.0001	0.965	0.13	< 0.0001	0.966	0.14	< 0.0001
β_{22}	1.087	0.13	< 0.0001	1.087	0.13	< 0.0001	1.084	0.13	< 0.0001	1.083	0.13	< 0.0001	1.084	0.13	< 0.0001
β_{23}	1.024	0.13	< 0.0001	1.024	0.13	< 0.0001	1.021	0.13	< 0.0001	1.020	0.13	< 0.0001	1.021	0.13	< 0.0001
β_{24}	0.979	0.13	< 0.0001	0.979	0.13	< 0.0001	0.977	0.13	< 0.0001	0.976	0.12	< 0.0001	0.977	0.13	< 0.0001
α_1	-8.276	2.05	< 0.0001	-8.276	2.07	< 0.0001	-8.261	2.04	< 0.0001	-8.252	2.03	< 0.0001	-8.261	2.14	< 0.0001
α_2	-0.213	0.04	< 0.0001	-0.213	0.04	< 0.0001	-0.213	0.04	< 0.0001	-0.213	0.04	< 0.0001	-0.213	0.04	< 0.0001
σ^2	186.27	7.33	< 0.0001	186.27	7.33	< 0.0001	186.25	7.33	< 0.0001	186.26	7.32	< 0.0001	186.25	7.33	< 0.0001
D^2	147.53	12.65	< 0.0001	147.52	12.65	< 0.0001	147.59	12.67	< 0.0001	147.61	12.64	< 0.0001	147.59	12.71	< 0.0001
$\beta_{21} - \beta_{22}$	-0.118	0.04	0.0028	-0.118	0.04	0.0028	-0.118	0.04	0.0028	-0.118	0.04	0.0027	-0.118	0.04	0.0028
$\beta_{21} - \beta_{23}$	-0.055	0.04	0.0720	-0.055	0.04	0.0721	-0.055	0.04	0.0717	-0.055	0.04	0.0716	-0.055	0.04	0.0718
$\beta_{21} - \beta_{24}$	-0.011	0.03	0.3771	-0.011	0.03	0.3772	-0.011	0.03	0.3752	0.011	0.03	0.3743	0.011	0.03	0.3753
ψ_{01}	0.124	0.08	0.0560	0.194	0.08	0.0086	0.205	0.08	0.0062	0.204	0.08	0.0065	0.205	0.08	0.0063
ψ_{03}	0.045	0.12	0.3595	0.107	0.13	0.1971	0.115	0.13	0.1890	0.112	0.13	0.1861	0.115	0.13	0.1814
ψ_2	-0.012	0.002	< 0.0001	-0.006	0.003	0.0383	-0.006	0.003	0.0367	-0.006	0.003	0.0119	-0.006	0.003	0.0367
ψ_3				-0.004	0.001	0.0016	-0.002	0.002	0.1615						
ψ_4							-0.002	0.002	0.2500	-0.003	0.001	0.0012			
ψ_3^*													-0.002	0.001	0.0034
ψ_4^*													-0.0001	0.002	0.4818
ψ_5	0.133	0.08	0.05730	0.212	0.09	0.0083	0.218	0.09	0.0070	0.213	0.09	0.0080	0.218	0.09	0.0070
-2ℓ	19577.16			19568.3			19567.32			19567.78			19567.32		

Table IV. Parameter estimates, standard errors, p-values and deviances for the outcome model (4), the reminder models defined in equations (8) and (10) and the different missing data mechanisms.

4.3. Results using the Reminder Process Model via μ_{jk}

The following missingness processes will be investigated under the model for the unconditional probabilities with the non-responders as reference category:

- MCAR_μ : $\log\left(\frac{\mu_{jk}}{\mu_{j,ref}}\right) = \lambda_0 + \lambda_{1,k}\, age + \lambda_2\, t_j$
- MAR_μ : $\log\left(\frac{\mu_{jk}}{\mu_{j,ref}}\right) = \lambda_0 + \lambda_{1,k}\, age + \lambda_2\, t_j + \lambda_3\, y_{j-1}$;
- MNAR_μ-1: $\log\left(\frac{\mu_{jk}}{\mu_{j,ref}}\right) = \lambda_0 + \lambda_{1,k}\, age + \lambda_2\, t_j + \lambda_4\, y_j$; and
- MNAR_μ-2: $\log\left(\frac{\mu_{jk}}{\mu_{j,ref}}\right) = \lambda_0 + \lambda_{2,k}\, age + \lambda_2\, t_j + \lambda_3\, y_{j-1} + \lambda_4\, y_j$

where $k \in \{0, 1, 2, 3, 4\}$. The results are presented in Table V.
The estimated outcome parameters are consistent with those obtained by modelling the missingness process in Section 4.1. Regarding the reminder process modelled, we observe that the intercept varies substantially across the assumed missingness processes. This is not surprising, as we include more covariates to explain the reminder process. Furthermore, we find a non-significant time effect for all models investigated.
There is an increasing probability of replying with age, which interacts with attempt in all models. This age effect is larger for patients who return their questionnaire after a single telephone reminder or without any chasing, with largest estimates under the MNAR models.
The score has a positive effect on the probability of replying at a certain attempt. As in Section 4.1, we observe a borderline significance of the previous score (i.e. MAR, p-value= 0.06) and a stronger effect with the current score. Once again, an effect of the outcome of interest on the response probabilities could not be verified when incorporating the current and previous score in the reminder process.

4.4. Results for Non-Monotone Missingness

The estimates under the assumption of ignorability, MNAR_p-3 and MNAR_r-2 for the non-monotone case are summarized in Table VI.
The estimates for the mean parameters under ignorability are usually smaller than under the MNAR_p-3 missing data assumption, but larger under MNAR_r-2. The treatment differences under ignorability and MNAR_r-2 are practically identical, but smaller for the MNAR_p-3 process. In all models, both Aircast and PoP-cast are significantly better than Tubigrip. Note that the significant difference between Aircast and Tubigrip is due to the removal of the significant age effect, see Table II.
Similar to the monotone case, Section 4.2, the first telephone call ($z = 1$) significantly increases the response probabilities under MNAR_p-3. Furthermore, time has a negative effect on the response probabilities. In contrast to the monotone case, we observe a borderline non-significant effect of the Bledsoe treatment. The estimates for ψ_3^* and ψ_4^* are both significantly different from zero. The reminder process depends not only on the actual score at the previous or current time point, but also on the improvement between two time points. In contrast to previous models (Section 4.1 and Section 4.3), the probability of replying decreases with the average score, but increases with improvement.
In comparison with the results for monotone missingness patterns and MNAR_r-2, Section 4.1, we observe a smaller intercept and an existing time effect. As time passes the probability of replying decreases. Further, a positive effect of the current score on the missingness process is shown. This effect was not observed in the monotone case.
Note that in all approaches the inter- and intra-individual variations are inflated due to the removed age-effect.
For illustration, we show the probabilities of not replying for different age groups and low/high scores under the MNAR_r-2 and the MNAR_p-3 model, see Table VII.

Parameter	MCAR_μ			MAR_μ			MNAR_μ-1			MNAR_μ-2		
	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.
β_0	41.04	0.78	< 0.0001	41.04	0.78	< 0.0001	41.03	0.78	< 0.0001	41.03	0.78	< 0.0001
β_1	106.89	6.86	< 0.0001	106.89	6.87	< 0.0001	105.57	6.72	< 0.0001	105.70	7.09	< 0.0001
β_{21}	0.968	0.14	< 0.0001	0.968	0.14	< 0.0001	0.939	0.13	< 0.0001	0.942	0.14	< 0.0001
β_{22}	1.087	0.13	< 0.0001	1.087	0.13	< 0.0001	1.059	0.13	< 0.0001	1.061	0.13	< 0.0001
β_{23}	1.024	0.13	< 0.0001	1.024	0.13	< 0.0001	0.994	0.13	< 0.0001	0.997	0.13	< 0.0001
β_{24}	0.979	0.13	< 0.0001	0.979	0.13	< 0.0001	0.951	0.12	< 0.0001	0.955	0.13	< 0.0001
α_1	-8.276	2.06	< 0.0001	-8.275	2.06	< 0.0001	-7.997	2.01	< 0.0001	-8.023	2.12	< 0.0001
α_2	-0.213	0.04	< 0.0001	-0.213	0.04	< 0.0001	-0.207	0.04	< 0.0001	-0.207	0.04	< 0.0001
σ^2	186.27	7.33	< 0.0001	186.27	7.33	< 0.0001	186.53	7.35	< 0.0001	186.47	7.36	< 0.0001
D^2	147.52	12.68	< 0.0001	147.52	12.66	< 0.0001	148.06	12.65	< 0.0001	148.00	12.69	< 0.0001
$\beta_{21} - \beta_{22}$	-0.118	0.04	0.0028	-0.118	0.04	0.0028	-0.120	0.04	0.0023	0.120	0.04	0.0024
$\beta_{21} - \beta_{23}$	-0.055	0.04	0.0720	-0.055	0.04	0.0721	-0.055	0.04	0.0685	-0.055	0.04	0.0688
$\beta_{21} - \beta_{24}$	-0.011	0.03	0.3771	-0.011	0.03	0.3772	-0.013	0.03	0.3433	-0.013	0.03	0.3468
λ_0	-2.166	0.75	0.0021	-2.610	0.85	0.0011	-3.335	0.92	0.001	-3.288	0.96	0.0003
$\lambda_{1,0}$	0.899	0.23	0.0001	0.951	0.25	0.0001	1.041	0.23	< 0.0001	1.035	0.23	< 0.0001
$\lambda_{1,1}$	0.863	0.23	0.0001	0.915	0.25	0.0001	1.005	0.23	< 0.0001	0.999	0.23	< 0.0001
$\lambda_{1,2}$	0.538	0.23	0.0099	0.590	0.25	0.0085	0.680	0.23	0.0016	0.673	0.23	0.0019
$\lambda_{1,3}$	0.539	0.23	0.0097	0.591	0.25	0.0084	0.681	0.23	0.0015	0.675	0.23	0.0019
$\lambda_{1,4}$	0.562	0.23	0.0074	0.614	0.25	0.0065	0.704	0.23	0.0011	0.698	0.23	0.0014
λ_2	0.007	0.01	0.1120	0.001	0.01	0.4273	0.002	0.01	0.4035	0.001	0.01	0.4263
λ_3				0.006	0.004	0.0605				0.001	0.01	0.4437
λ_4							0.011	0.01	0.0277	0.010	0.01	0.1439
-2ℓ	20101.72			20099.30			20098.28			20098.26		

Table V. Parameter estimates, standard errors, p-values and deviances for the outcome model (4), the reminder model given in equation (11) and the different missing data mechanisms.

	Ignorable			MNAR$_p$-3			MNAR$_r$-2	
Parameter	Est.	SE	p-val.	Est.	95% CI	Parameter	Est.	95% CI
β_0	41.16	0.80	< 0.0001	41.15	[39.55,42.73]	β_0	41.70	[40.19,43.22]
β_1	79.29	0.83	< 0.0001	80.35	[78.71,81.98]	β_1	79.44	[77.67,81.25]
β_{21}	0.215	0.03	< 0.0001	0.238	[0.18,0.31]	β_{21}	0.216	[0.16,0.26]
β_{22}	0.379	0.04	< 0.0001	0.394	[0.32,0.47]	β_{22}	0.367	[0.30,0.45]
β_{23}	0.312	0.03	< 0.0001	0.330	[0.27,0.40]	β_{23}	0.296	[0.24,0.36]
β_{24}	0.270	0.03	< 0.0001	0.288	[0.24,0.35]	β_{24}	0.258	[0.21,0.31]
σ^2	192.1	7.37	< 0.0001	196.1	[181.8,211.6]	σ^2	194.8	[180.3,210.5]
D^2	155.8	13.05	< 0.0001	156.2	[131.9,183.4]	D^2	159.1	[134.0,187.0]
$\beta_{21} - \beta_{22}$	-0.164	0.04	0.0002	-0.156	[-0.25,-0.06]	$\beta_{21} - \beta_{22}$	-0.165	[-0.25,-0.08]
$\beta_{21} - \beta_{23}$	-0.098	0.04	0.0169	-0.091	[-0.18,-0.003]	$\beta_{21} - \beta_{23}$	-0.094	[-0.17,-0.02]
$\beta_{21} - \beta_{24}$	-0.056	0.04	0.1299	-0.049	[-0.13,0.03]	$\beta_{21} - \beta_{24}$	-0.057	[-0.12,0.01]
ψ_{01}				0.341	[0.20,0.49]	ϕ_0	-2.143	[-3.60,-0.61]
ψ_{03}				-0.188	[-0.39,0.02]	ϕ_1	0.882	[0.54,1.22]
ψ_2				-0.006	[-0.011,-0.001]	ϕ_2	-0.020	[-0.03,-0.01]
ψ_3^*				-0.006	[-0.007,-0.005]	ϕ_3	-0.005	[-0.01,0.01]
ψ_4^*				0.004	[0.001,0.008]	ϕ_4	0.019	[0.002,0.03]
ψ_5				0.144	[-0.003,0.29]			

Table VI. Overview of the parameter estimates, standard errors and p-values under non-monotone missingness and the assumption of an ignorable missingness mechanism. Further, the means of the posterior distributions and the 95% credible intervals (CI) for the MNAR$_p$-3 and MNAR$_r$-2 missingness processes are shown.

MNAR$_p$-3			
current score	previous score	μ_{j5}	
		$j = 12$	$j = 39$
low	low	0.13	0.21
	high	0.15	0.23
high	low	0.22	0.28
	high	0.23	0.30

MNAR$_r$-2			
age	current score	$1 - \rho_j$	
		$j = 12$	$j = 39$
Young	low	0.20	0.29
	high	0.11	0.18
Middle-Aged	low	0.16	0.24
	high	0.09	0.15
Old	low	0.12	0.20
	high	0.07	0.12

Table VII. Overview of the probabilities of not replying for different age groups and previous / current scores based on the point estimates obtained from fitting the MNAR$_p$-3 and the MNAR$_r$-2 model under non-monotone missingness, see Table VI. Low and high scores denote the first and third quantiles, respectively. The age groups were classified according to the first, second (median) and third quantile.

5. CONCLUSIONS

We have proposed a selection model framework for continuous longitudinal data to adjust for non-ignorable or informative missingness when initial non-responders are reapproached several times. The model presented combines a non-linear mixed model for the underlying outcome model with a logistic regression model for the missingness and the reminder process.

The non-linear mixed model is derived under the assumption that the rate of improvement in a given time interval is proportional to the current score and the still achievable score. For the reminder process, we model the probability of replying at a certain attempt, given not having replied earlier,

dependent on covariates and the outcome score itself.
We investigate the impact of missingness on the rate of improvement for different missingness processes. We distinguish the case of monotone missingness patterns from the case of non-monotone missingness. While we approach the monotone case through likelihood based inference, we estimate the parameters of interest within the Bayesian paradigm for non-monotone missingness patterns.
The conclusions that recovery is slower, and less satisfactory with age, and more rapid with PoP-cast than Tubigrip do not alter materially across models with monotone missingness. The superiority of Aircast brace over Tubigrip is shown to be borderline significant with monotone missingness modelled. Depending on whether the reminder process or the missingness process is explored, and on whether conditional or unconditional reminder probabilities are modelled, we find that the probabilities of replying decrease or increase with the observed outcome at the current or previous occasions. We argue that these results are due to the inconsistent incorporation of an age effect in the missingness process and reminder process models. We conclude that the age of patients or their corresponding score might be used to adjust for missingness as these are negatively associated. Nevertheless, the score itself is shown to have a significant effect on the response probabilities, even when an age-effect is included, see Section 4.1 and Section 4.3. It would be desirable to include an age-effect when modelling the conditional attempt probabilities in Section 4.2. However, the computational complexity prevented us from doing so.
Furthermore, we observe that as time passes the probability of replying decreases for every attempt category with the conditional reminder probabilities modelled. This effect is not verified for monotone missingenss patterns when the unconditional reminder probabilities or the missingness process are modelled. However, a negative effect of time is observed for MNAR$_r$-2 and non-monotone missingness. The results also suggest that phone calls are effective in retrieving questionnaires.
Only about 10% of missingness is non-monotone, so we observe similar results for the non-monotone case. However, the borderline significance of Aircast brace over Tubigrip with monotone missingness modelled becomes significant when we account for non-monotone missingness patterns. Note that these results have to be treated with caution as we are excluding the age-effect in order to reduce the model complexity.
We believe that the selection models presented are valuable for understanding treatment and covariate effects on the outcome and the inclination to reply. More efficient algorithms would facilitate extensions to non-monotone missingness patterns and wider use of these models.

ACKNOWLEDGEMENT

We would like to thank Prof. Sallie Lamb (Warwick Medical School) and the CAST team for the permission to use these data and helpful discussions.

REFERENCES

1. Horton NJ, Lipsitz SR. Multiple imputation in practice: Comparison of software packages for regression models with missing data. *The American Statistician* 2001; **55 (3)**:244–254.
2. Carpenter J, Pocock S, Lamm C. Coping with missing data in clinical trials: A model-based approach applied to asthma trials. *Statistics in Medicine* 2002; **21**:1043–1066.
3. Molenberghs G, Kenward MG. *Missing data in clinical studies.* Wiley, 2007.
4. Little RJA, Rubin DB. *Statistical analysis with missing data.* Wiley Interscience, 2002.
5. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* Wiley, 1987.
6. Rubin DB. Multiple imputation after 18+ years. *Journal of the American Statistical Association* 1996; **91**:473–489.
7. Schafer JL. Multiple imputation: a primer. *Statistical methods in medical research* 1999; **8**:3–15.
8. Schafer JL. *Analysis of incomplete multivariate data.* Chapman & Hall, 1997.
9. Diggle PJ, Farewell D, Henderson R. Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. *Journal of the Royal Statistical Society: Series C* 2007; **56**:1–31.
10. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**:581–592.

11. Molenberghs G, Beunckens C, Sotto C, Kenward MG. Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B* 2008; **70**:371–388.
12. Diggle PJ, Kenward MG. Informative drop-out in longitudinal data analysis. *Applied Statistics* 1994; **43**:49–93.
13. Baker SG. Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics* 1995; **51**:1042–1052.
14. Troxel AB, Harrington DP, Lipsitz SR. Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics* 1998; **47**:425–438.
15. Wood AM, White IR, Hotopf M. Using number of failed contact attempts to adjust for non-ignorable non-response. *Journal of the Royal Statistical Society: Series A* 2006; **169**:525–542.
16. Jackson D, White IR, Morven L. How much can we learn about missing data? an exploration of a clinical trial in psychiatry. To appear in Journal of the Royal Statistical Society: Series A.
17. Cooke MW, Marsh JL, Clark M, Nakash RA, Jarvis RM, Hutton JL, Szczepura A, Wilson S, Lamb SE. Treatment of severe ankle sprain: a pragmatic randomised controlled trial comparing the clinical effectiveness and cost-effectiveness of three types of mechanical ankle support with tubular bandage. The CAST trial. *Health Technology Assessment* 2009; **13: No.13**.
18. Alho JM. Adjusting for nonresponse bias using logistic regression. *Biometrika* 1990; **77**:617–624.
19. Lamb SE, Nakash RA, Withers EJ, Clark M, Marsh JL, Wilson S, Hutton JL, Szczepura A, Dale JR, Cooke C M W. Clinical and cost effectiveness of mechanical support for severe ankle sprains: design of a randomised controlled trial in the emergency department. *BMC Musculoskeletal Disorders* 2005; **6**:1471–2474.
20. Lamb SE, Marsh JL, Hutton JL, Nakash RA, Cooke MW. Mechanical supports for acute, severe ankle sprains: a pragmatic, multi-centre, randomised controlled trial. *Lancet* 2009; **373**:575–581.
21. Roos E, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot & Ankle International* 2001; **22(10)**:788–794.
22. Sherman M, Le Cessie S. A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Communications in Statistics - Simulation and Communication* 1997; **26**:901–925.
23. Sikorsky K, Stenger F. Optimal quadratures in h_p spaces. *Association for Computing Machinery, Transactions on Mathematical Software* 1984; **3**:140–151.
24. *SAS/STAT Users Guide, Version 8. Cary, NC: SAS Institute Inc. ,1999.*
25. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**:325–337.
26. Neal R. *Learning in Graphical Models*, chap. Suppressing random walks in Markov chain Monte Carlo using ordered over-relaxation. Kluwer Academic Publishers, Dordrecht, 1998; 205–230.
27. Nakash RA, Hutton JL, Lamb SE, Gates S, Fisher J. Response and non-response to postal questionnaire follow-up in a clinical trial - a qualitative study of the patient's perspective. *Journal of Evaluation in Clinical Practice* 2008; **14**:226–235.