Modelling the rate of change in a longitudinal study with missing data, adjusting for contact attempts

Mouna Akacha and Jane L. Hutton

May 17, 2010

Abstract

The Collaborative Ankle Support Trial (CAST) is a longitudinal trial in which interest lies in the rate of improvement, the effectiveness of reminders and potentially informative missingness. A model is proposed for continuous longitudinal data with non-ignorable or informative missingness, taking into account the nature of attempts made to contact initial non-responders. The model combines a non-linear mixed model for the outcome model with a logistic regression model for the reminder process. A sensitivity analysis is used to contrast this model with the traditional selection model, where we adjust for missingness by modelling the missingness process.

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].

The first option yields a *complete case analysis*. 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]. Different missing data methods with the main focus on repeated measurement studies are investigated 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].

We distinguish between three *missing data mechanisms*, which concern the relation between the *missingness process* and the outcome variable [10]. 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 the case of MAR 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

CRiSM Paper No. 10-08, www.warwick.ac.uk/go/crism

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 dropout 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, 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 vector is chosen.

We extend these models in two 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 dropout 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 the ideas in [15, 18] we will use a multinomial model for the reminder process. In [18] the focus lies on studies with a single time point and a logistic regression model is used to analyze 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, we will be fitting a non-linear mixed model that focuses on modelling the rate of a response curve. This model is motivated by medical research, where it is very common to measure physical or mental ability repeatedly over time through questionnaires or scales. Based on the answers, summary measures such as scores can be derived for every point in time. In many applications, these scores will have finite range, where one bound indicates 'no symptoms' and the other bound 'extreme symptoms'. Examples are the *Barthel* index [19], the *Neck Disability Index* [20], the *Foot and Ankle Outcome Score* [21] (FAOS) and visual analogue scales. In studies where we expect most patients to recover, we often observe that later measurements are clustered towards one end of the range. In this case, different patients might have the same initial and the same final scores. However, the rate at which they achieve the final score might differ substantially dependent on explanatory variables, for example, treatment or age. The bounds themselves can also be of scientific interest, e.g. a maximum achievable score can differ substantially for different ages and genders.

For a continuous and bounded score, the classical approach is to transform the data such that fitting a linear regression model seems reasonable. For some scores, however, a non-linear dependence of the transformed outcome score on covariates persists due to the bounded nature of the score. In addition, models based on transformations cannot investigate the dependence of bounds on covariates as the bounds need to be specified prior to the transformation. Using transformations can also complicate the interpretation of covariate effects on the original score.

In this paper, we present a model for the outcome score on the original scale as a function of covariates. The model is constructed for scores where the rate of recovery changes over time and was motivated by the *Collaborative Ankle Support Trial* (CAST), which is the first large randomized controlled trial comparing four types of mechanical support for ankle sprains of sufficient severity to prevent weight bearing [17, 22, 23].

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

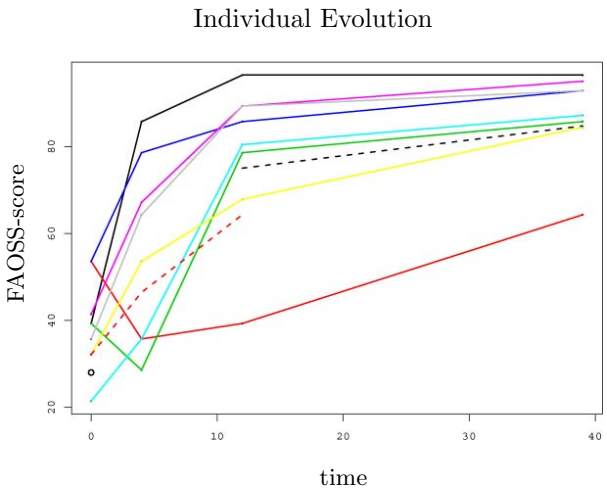

Figure 1: Individual evolution of the FAOSS-score for a random subset of 10 patients. The dashed lines correspond to patients with missing outcomes.

and nature 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, 22, 23].

The data for this trial were obtained from a randomised, multicentre study, which was run in 6 National Health Service trusts across the UK. 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), *Below knee cast* (BKC), *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]. A continuous score, with 100 indicating no symptoms and 0 indicating extreme symptoms, was calculated for each subscale. The total sample size was $N = 584$. Due to the fact that some patients did not receive the FAOS questionnaire but another questionnaire called *Ankle Performance Scale* (APS) [24] during the baseline assessment, the data of 553 persons instead of 584 persons will be investigated in this report. Moreover, this analysis will concentrate on the *symptoms*-subscale score which will be referred to as *FAOSS-score* (FAOS-symptoms score).

As with many studies which measure recovery from acute injuries, the natural time course of recovery of ankle sprains is likely to stabilise within a certain period (here: 3 to 9 months) and it is expected that the difference between the treatments will narrow in the longer term because the majority of people will recover [22, 25, 26, 27]. An important aim of treatment is to accelerate the rate of recovery. Understanding the impact of explanatory covariates on the rate of recovery is important for guiding patients and clinicians expectations.

The original analysis included randomisation group and adjusted for gender, age and baseline score [22]. The recovery was analysed at every time point separately, thus neglecting the correlation between the four measurements of each subject. This can reduce the precision of the analysis and thereby the significance of the results can be overestimated [28]. Additionally, the comparison of the different treatments was reduced to per time point conclusions and did not enable an overall statement about the rate of recovery.

For initial exploratory re-analysis, the individual evolution of the FAOSS-score for a small subset of patients was plotted against time, see Figure 1. We connected the four scores per subject to demonstrate the evolution over time. The dashed lines correspond to patients with missing observations. From this plot we see the score is usually an increasing function of time. Also, the score increases much faster at the beginning of the study than towards the end and some patients achieve their maximum score sooner than others. The achieved score at the end of 9 months and the rate at which this score is achieved varies across the subjects. In general, however, the responses exhibit

Time		Randomisation Groups ($N = 553$)			
		Tubigrip	BKC	Aircast	Bledsoe Boot
Baseline	Mean	40.3	41.8	38.8	41.1
	SD	14.1	16.4	15.1	16.9
4 weeks	Mean	60.7	67.4	62.8	61.6
	SD	19.5	19.0	20.5	20.7
12 weeks	Mean	70.0	76.0	73.8	75.1
	SD	20.5	18.4	20.6	20.4
39 weeks	Mean	80.4	82.8	81.0	81.2
	SD	20.4	17.0	20.3	19.0

Table 1: Summary Statistics (SD: standard deviation) for the FAOSS-score and the different randomisation groups and time points.

similarly shaped curves.

Our aim is to model the recovery rate and the bounds by modelling the responses at the four time points jointly. In this context, we want to adjust for the explanatory variables age and randomisation group. 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 will be analysed in the groups to which they were randomised, regardless of the treatment that they received. The randomisation groups were generally well matched in terms of age. The mean age of participants was 30 years (SD 10.8, median 27, range $16 - 72$). Summary statistics for the FAOSS-score and the different randomisation groups and time points are given in Table 1.
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 2. Note that we observe a non-monotone missingness pattern in this study. For the FAOSS-scale 10% of the patients exhibit a non-monotone pattern. Although our methodology is able to account for non-monotone missingness patterns, we will focus on monotone missingness. In particular, we deleted all those observations that were made after a patient failed to return a previous questionnaire. Further discussion for the non-monotone case can be found in the discussion, Section 5.

3 MODEL

In the following subsections, we propose a selection model for continuous longitudinal data to adjust for non-ignorable or informative missingness when initial non-responders are reapproached several times.

3.1 Notation

Let $\mathbf{y}_i = (y_{i,1}, ..., y_{i,M})^\top$ denote the M-dimensional response vector of subject $i \in \{1, ..., N\}$, where $\mathbf{y}_i$ is a realisation of the random vector $\mathbf{Y}_i$. We assume that $\mathbf{Y}_i$ is multivariate normal distributed and denote the joint outcome vector for all subjects by $\mathbf{Y} = (\mathbf{Y}_1^\top, ..., \mathbf{Y}_N^\top)^\top$.

Furthermore, let $X_i = (\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,M})^\top$ be the matrix of explanatory variables (e.g. time, gender and age) for subject $i \in \{1, ..., N\}$. The randomisation group is denoted by η_i, $i \in \{1, ..., L\}$, and the observation times by t_j, $j \in \{1, ..., M\}$.

The indicator $r_{i,j}$ is a realisation of the random variable $R_{i,j}$ which denotes whether $y_{i,j}$ was observed, $r_{i,j} = 1$, or missing, $r_{i,j} = 0$. We summarize the missingness information for subject i through $\mathbf{R}_i = (R_{i,1}, ..., R_{i,M})^\top$ and

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

for all subjects through $\mathbf{R} = (\mathbf{R}_1^\top, ..., \mathbf{R}_N^\top)^\top$. Moreover, $z_{i,j} \in \{0, 1, ..., K\}$ represents the reminder category and is a realisation of the random variable $Z_{i,j}$. In particular, we distinguish $K + 1$ reminder categories. In the CAST study we observe $N = 553$, $M = 4$, $L = 4$ and $K = 5$. Note that the observations times for CAST are given by $t_j \in \{0, 4, 12, 39\}$ for $j \in \{1, 2, 3, 4\}$.
We aim to analyse the relationship between the response variable $\mathbf{Y}_i$ and the explanatory variables X_i for all $i \in \{1, ..., N\}$, taking into account the missingness process and the reminders needed to retrieve a questionnaire.

3.2 Selection Models

Suppose the complete data $\mathbf{Y}$ follows the parametric model $\mathrm{P}(\theta)$ and $\mathbf{R}$ follows the parametric model $\mathrm{P}(\phi)$. We partition the vector $\mathbf{Y}$ into the observed, $\mathbf{Y}_{obs}$, and unobserved part, $\mathbf{Y}_{mis}$. If the missingness process is non-ignorable or informative we need to base inference for θ on the joint likelihood of $\mathbf{Y}_{obs}$ and the missingness process $\mathbf{R}$. 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 $\mathbf{Y}_{obs}$ and $\mathbf{R}$ is given by

$$L_{\mathbf{Y}_{obs},\mathbf{R}}(\theta, \phi) = \prod_{i=1}^{N} \int f(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}|X_i, \theta)\, f(\mathbf{r}_i|X_i, \mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}, \phi)\, \mathrm{d}\mathbf{y}_{i,mis}. \quad (1)$$

As $z_{i,j} \in \{0, 1, 2, 3, K-1\} \Leftrightarrow r_{i,j} = 1$ and $z_{i,j} = K \Leftrightarrow r_{i,j} = 0$ we can extend the selection model by adjusting for non-response through $z_{i,j}$ rather than $r_{i,j}$. 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. In particular, we can see $r_{i,j}$ as a special case of $z_{i,j}$. The extension of the likelihood in equation (1) to adjust for the reminder process is straightforward. Let $\mathbf{Z}$ follow the parametric model $\mathrm{P}(\psi)$, then we simply need to replace $\mathbf{r}_i$ by $\mathbf{z}_i$ and ϕ by ψ in equation (1). Fitting these selection models requires a marginal model for the outcome vector Y_i and models for the conditional response process $R_i|Y_i$ and the conditional reminder process $Z_i|Y_i$. We propose a non-linear mixed model for the marginal outcome process and various plausible regression models for the conditional response and the conditional reminder process. In this way we want to assess the influence of misspecification and sensitivity on our conclusions.

3.3 Outcome Model

The outcome model we propose in this section is motivated by the CAST study, where we observe continuous and bounded longitudinal data. We argue that the traditional approach of transforming the data, for example by using the log or logit transformation does not always resolve the problem of a non-linear relationship of the response and time. For example, we investigated several transformations for the CAST study, but a non-linear relation with time persisted due to the bounded nature of the outcome. Also the inclusion of higher order time effects did not lead to a satisfying fit. Importantly, using transformations we were no longer able to investigate covariate effects on the bounds. For these reasons, we refrain from using multivariate linear models and *marginal* or *random effect* models. Based on exploratory analysis, see Section 2, we propose a non-linear mixed model for the original score data that models the rate of recovery 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

$$\mathbf{Y}_i|U_i \overset{ind.}{\sim} \mathcal{N}_M\left(\mu_i, \sigma^2 I\right); \qquad U_i \overset{iid}{\sim} \mathcal{N}(0, D^2); \quad \text{and} \qquad \mu_{i,j} = g(\mathbf{x}_{i,j}, \theta_i) \quad \text{for } j \in \{1, ..., M\},$$

where g is the non-linear model function, I the M-dimensional identity matrix 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 increasing over time and bounded, thus motivating our proposal that the recovery rate should change over time. We expect a very low rate of recovery when patients suffer from extreme symptoms, in particular $y_{i,j} = 0$ implies a recovery rate of zero, e.g. for the FAOSS-score worst symptoms indicate a very swollen and stiff ankle, which delays the start of recovery. Additionally, we know that the recovery rate is zero when the upper bound of the score is achieved. This means that the rate of recovery at a certain time point depends on the distance of the current score to the lower and the upper bound. In mathematical terms, we expect the rate of improvement in a given time interval, i.e. $g'(\mathbf{x}_{i,j}, \theta)$, to be proportional to the current score, $g(\mathbf{x}_{i,j}, \theta)$, and the still achievable score $[\max\{g(\mathbf{x}_{i,j}, \theta)\} - g(x_{i,j}, \theta)]$. Hence, we are interested in solving the differential equation

$$g'(\mathbf{x}_{i,j}, \theta) = \kappa_{\eta_i}\, g(\mathbf{x}_{i,j}, \theta)\, [\max\{g(\mathbf{x}_{i,j}, \theta)\} - g(\mathbf{x}_{i,j}, \theta)],$$

where κ_{η_i} for $\eta_i \in \{1, ..., L\}$ is the treatment-specific proportion-factor.

Reducing the problem to $x_{i,j} = t_j$ yields the solution

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

In this model β_0 denotes the intercept, β_1 the upper bound, i.e. maximum achievable score, and $\beta_{2,\eta_i} = \kappa_{\eta_i} \cdot \beta_1$ the treatment specific recovery rate of the outcome curve. However, usually the scores and the rate of recovery depend on explanatory variables. Incorporating the covariates $x_{i,j}$ is straightforward; and in order to capture the inter-individual variation, we extend this model to a *non-linear mixed model* by adding the subject-specific quantity U_i [29]:

$$g(\mathbf{x}_{i,j}, \theta_i) = \frac{\beta_1 + \alpha_1^\top x_{i,j}}{\exp\{-([\beta_{21} + \beta_{2,\eta_i}\, \mathbb{1}(\eta_i \neq 1)] + \alpha_2^\top x_{i,j}) \cdot t_j\}\left(\frac{\beta_1 + \alpha_1^\top x_{i,j}}{\beta_0 + \alpha_0^\top x_{i,j}} - 1\right) + 1} + U_i,$$

where $\mathbb{1}(\eta_i \neq 1)$ is one if $\eta_i \neq 1$ and zero otherwise, i.e. without loss of generality we are assuming that interest lies in the treatment contrasts compared to the standard treatment $\eta = 1$.

The interpretation of all parameters is straightforward:

- $\beta_0 + \alpha_0^\top x_{i,j}$ describes the intercept, where α_0 indicates the effects of the covariates on the intercept.
- $\beta_1 + \alpha_1^\top x_{i,j}$ describes the maximum score (upper bound) and α_1 the covariate effects on this upper bound. In particular, this model accounts for the bounded nature of the score. As time increases a limiting score, varying according to $x_{i,j}$ is achieved.
- $[\beta_{21} + \beta_{2,\eta_i}\, \mathbb{1}(\eta_i \neq 1)] + \alpha_2^\top x_{i,j}$ indicates the rate of improvement, i.e. how fast the upper bound is achieved. This rate depends on the randomisation group η_i and the covariates $x_{i,j}$ of the patients. For $\eta_i \in \{2, ..., L\}$ the parameters β_{2,η_i} denote the contrast to, or increase from, the treatment slope of $\eta = 1$, i.e. β_{21}.

In particular, the parameter of interest is given by $\theta_i = (\theta^\top, U_i)^\top$ with $\theta = (\beta_0, \beta_1, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \alpha_0^\top, \alpha_1^\top, \alpha_2^\top, \sigma, D)^\top$. This model can easily be reformulated in terms of the multivariate normal model with a compound symmetry covariance structure:

$$\mathbf{Y}_i \sim \mathcal{N}_M\left(\tilde{\mu}_i, \Sigma\right), \text{ where } \tilde{\mu}_{ij} = \frac{\beta_1 + \alpha_1^\top x_{i,j}}{\exp\{-([\beta_{21} + \beta_{2,\eta_i}\, \mathbb{1}(\eta_i \neq 1)] + \alpha_2^\top x_{i,j}) \cdot t_j\}\left(\frac{\beta_1 + \alpha_1^\top x_{i,j}}{\beta_0 + \alpha_0^\top x_{i,j}} - 1\right) + 1}. \quad (2)$$

Let I be the identity matrix and J a square matrix with all elements unity, then $\Sigma = \sigma^2\, I + D^2\, J$.

3.4 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 2. 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 $j \in \{2, ..., M\}$ let $p_{j,0}$ be the probability of responding at the very first attempt. For $k \in \{1, ..., K-1\}$ let $p_{j,k}$ denote the probability of responding to 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. $p_{1,0}$, is one.

The unconditional probabilities $\mu_{j,k}$ of replying to attempt k at time point j are then given by:

$$\mu_{j,0} = p_{j,0}; \quad \mu_{j,1} = p_{j,1}\,(1 - p_{j,0}); \quad \dots \quad \mu_{j,K-1} = p_{j,K-1} \prod_{k=0}^{K-2} (1 - p_{j,k}).$$

Furthermore, we define $\mu_{j,K} = 1 - \sum_{k=0}^{K-1} \mu_{j,k}$ as the probability of not replying at time point $j \in \{2, ..., M\}$, i.e. $z_j = K$. Corresponding to these probabilities we redefine the random variable Z_j in terms of an indicator random vector. Let V_j be a $(K+1)$-dimensional random vector, where

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

for $\ell \in \{1, ..., K+1\}$. Thus, for a certain subject all information about Z is now captured through the indicator matrix $V = (\mathbf{V}_2, ..., \mathbf{V}_M)^\top$ and the likelihood for (Y_{obs}, V) can be derived by replacing $R_{i,j}$ with $V_{i,j}$ and ϕ with ψ in equation (1). We can write

$$\mathbf{V}_j \sim \textit{Multinomial}\,(1, \mu_{j,0}, ..., \mu_{j,K})\,. \tag{3}$$

Dependent on the required inference, a generalized linear model for $\mu_{j,k}$ or $p_{j,k}$ can be formulated. The marginal probability $\mu_{j,k}$ determines the chance of replying to the k-th attempt. In contrast, formulating a model for the conditional probability $p_{j,k}$ investigates the effect of covariates on replying to 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.

The generalized linear model we propose for p_{jk} and $j \in \{2, ..., M\}$, $k \in \{1, ..., K-1\}$ is given by

$$\text{logit}(p_{j,k}) = \psi_{0k} + \psi_1^\top \breve{x_j} + \psi_2\, t_j + \psi_3\, y_{j-1} + \psi_4\, y_j \tag{4}$$

where $\breve{x_j}$ denotes covariates we wish to include in the reminder process model and t_j the observation times. Further, y_{j-1} is the previous outcome and y_j the current score. This model assumes that the covariate effect on $p_{j,k}$ is the same for all attempts k. However, the intercept varies across the different reminder categories. Moreover, this model assumes that the model for V_j does not depend on later observations V_ℓ, where $\ell > j$. Theoretically speaking, we could also include future outcomes, however we believe that for most settings, including the CAST data set, this is rather unnecessary.

Under monotone missingness, this general model allows for different missingness mechanisms; MNAR is implied by $\psi_4 \neq 0$, MAR by $\psi_4 \equiv 0$ and, conditioned on covariates, MCAR is implied by $\psi_3 \equiv 0 \equiv \psi_4$.

With bounded scores and highly correlated scores at adjacent occasions, however, we will usually observe a high correlation between ψ_3 and ψ_4. This problem will always persist when including previous and current scores linearly. Take for example two perfectly positively correlated scores at adjacent time points; a distinction between MAR and MNAR is not possible. Therefore, we will also consider a different parametrization [12]:

$$\text{logit}(p_{j,k}) = \psi_{0k} + \psi_1^\top \breve{x_j} + \psi_2\, t_j + \psi_3^* \,[y_{j-1} + y_j] + \psi_4^* \,[y_j - y_{j-1}] \tag{5}$$

where ψ_3^* and ψ_4^* are usually less correlated. Again, this model allows for different missingness processes: MCAR corresponds to $\psi_3^* \equiv 0 \equiv \psi_4^*$, MAR to $\psi_3^* \equiv -\psi_4^*$ and MNAR to $\psi_3^* \neq -\psi_4^*$. In order to account for the dependence across the attempts at the different observation times of a certain patient, we need to extend this model, see Section 3.6.

3.5 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 $i \in \{1, ..., N\}$, $j \in \{2, ..., M\}$:

$$R_{i,j} = 1|\mathbf{Y}_i, \mathbf{X}_{i,j} \quad \sim \quad Bernoulli(\rho_{i,j}), \text{ with } \quad logit(\rho_{i,j}) = \phi_0 + \phi_1^\top \tilde{x}_{i,j} + \phi_2\, t_j + \phi_3\, y_{i,j-1} + \phi_4\, y_{i,j} \tag{6}$$

where $\tilde{x}_{i,j}$ denotes covariates we wish to include in the missingness process model and t_j the observation times. Further, $y_{i,j-1}$ is the previous outcome and $y_{i,j}$ the current score. Note that we do not specify a model for $R_{i,1}$ as all scores are observed at baseline.

The presented model corresponds to MNAR if $\phi_4 \neq 0$ and to MCAR (conditioned on covariates) if $\phi_3 \equiv 0 \equiv \phi_4$. As we focus on monotone missingness, a MAR missingness process is obtained by setting $\phi_4 \equiv 0$. Note, that the last relation holds for monotone missingness only. In the case of non-monotone missingness it is a less than trivial manner to construct sensible MAR models [11]. We will also consider a model similar to that given in equation (5), see Section 4.2.

3.6 Joint Model

For monotone missingness, we can now construct the joint likelihood of (Y_{obs}, R) and (Y_{obs}, V) respectively. The derivations will be shown for a selection model that uses the reminder process (via $\mathbf{V}_{i,j}$) to account for missingness. The likelihood using the missingness indicator process $R_{i,j}$ can be derived by replacing $\mathbf{V}_{i,j}$ by $R_{i,j}$ in all the following equations.

The observed data likelihood contribution of a certain subject is given by:

$$f\left(\mathbf{y}_{i,obs}, v_i|X_i, \theta, \psi\right) \quad = \quad \int f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}|X_i, \theta\right)\, f\left(v_i|X_i, \mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}, \psi\right) \mathrm{d}\mathbf{y}_{i,mis}.$$

For simplicity, we assume that $M = 4$ and that dropout for the subject of interest occurs after the second measurement time, i.e. $\mathbf{y}_{i,obs} = (y_{i,1}, y_{i,2})^\top$, then

$$\begin{aligned} f\left(\mathbf{y}_{i,obs}, v_i|X_i, \theta, \psi\right) \quad = \quad & \int\int f\left(y_{i,4}|y_{i,3}, y_{i,2}, y_{i,1}, X_i, \theta\right) f\left(y_{i,3}|y_{i,2}, y_{i,1}, X_i, \theta\right) f\left(y_{i,2}|y_{i,1}, X_i, \theta\right) f\left(y_{i,1}|X_i, \theta\right) \\ & \times f\left(\mathbf{v}_{i,4}|\mathbf{v}_{i,3}, \mathbf{v}_{i,2}, \mathbf{v}_{i,1}, X_i, \mathbf{y}_i, \psi\right)\, f\left(\mathbf{v}_{i,3}|\mathbf{v}_{i,2}, \mathbf{v}_{i,1}, X_i, \mathbf{y}_i, \psi\right)\, f\left(\mathbf{v}_{i,2}|\mathbf{v}_{i,1}, X_i, \mathbf{y}_i, \psi\right) \mathrm{d}y_{i,4}\, \mathrm{d}y_{i,3}. \end{aligned}$$

In the case of monotone missingness we observe

$$\mathbf{v}_{i,j} = (0, 0, 0, 0, 0, 1) \Longrightarrow \mathbf{v}_{i,j+1} = (0, 0, 0, 0, 0, 1) \tag{7}$$

for $j \in \{2, 3, 4\}$ and $j + 1 \in \{3, 4\}$. Therefore,

$$f\left\{\mathbf{v}_{i,4} = (0, 0, 0, 0, 0, 1)|\mathbf{v}_{i,3} = (0, 0, 0, 0, 0, 1), \mathbf{v}_{i,2}, \mathbf{v}_{i,1}, X_i, \mathbf{y}_i\right\} = 1.$$

Rearranging the observed likelihood yields

$$\begin{aligned} f\left(\mathbf{y}_{i,obs}, v_i|X_i, \theta, \psi\right) \quad = \quad & f\left(y_{i,2}|y_{i,1}, X_i, \theta\right) f\left(y_{i,1}|X_i, \theta\right) f\left(\mathbf{v}_{i,2}|\mathbf{v}_{i,1}, X_i, y_{i,2}, y_{i,1}, \psi\right) \\ & \times \int f\left(y_{i,3}|y_{i,2}, y_{i,1}, X_i, \theta\right) f\left(\mathbf{v}_{i,3}|\mathbf{v}_{i,2}, \mathbf{v}_{i,1}, X_i, y_{i,3}, y_{i,2}, \psi\right) \\ & \times \underbrace{\int f\left(y_{i,4}|y_{i,3}, y_{i,2}, y_{i,1}, X_i, \theta\right) \mathrm{d}y_{i,4}}_{=1}\, \mathrm{d}y_{i,3}, \end{aligned} \tag{8}$$

i.e. the integrals reduce to one-dimensional integrals for $i \in \{1, ..., N\}$. We note that the likelihood terminates after the time of the first missing observation due to the relation shown in equation (7). In particular, the implication (7) does not hold for non-monotone missingness; and we would be confronted with multidimensional integrals as soon as we relax the assumption of monotone missingness.

The likelihood contribution shown in (8) stresses that we ought to consider the dependence structure across the reminders at the different time points for a given subject. When modelling the reminder process such a dependence structure can be included in various ways, e.g. by formulating a random-effect model for $p_{j,k}$. However,

	Monotone Missingness		
Parameter	Est.	SE	p-val.
β_0	41.01	0.78	-
β_1	80.05	0.85	-
β_{21}	0.27	0.03	-
β_{22}	-0.12	0.04	0.0028
β_{23}	-0.05	0.04	0.0869
β_{24}	-0.004	0.04	0.4573
α_1	-0.30	0.07	< 0.0001
α_2	-0.006	0.001	< 0.0001
σ^2	186.1	7.33	-
D^2	148.17	12.67	-

Table 3: Overview of the parameter estimates and standard errors for the outcome model (2) based on the assumptions of an ignorable missingness process. The p-values are reported only for the components of θ that might be zero.

this would require the computation of further integrals which seriously complicates the evaluation of the likelihood. Alternatively, we can account for the dependence by formulating a model for $V_{i,j}$ conditional on $V_{i,j-1}, ..., V_{i,1}$. For simplification, we decide to model $V_{i,j}$ conditional on $V_{i,j-1}$, that is we extend the models given in equation (4) and equation (5) by adding the term $\psi_{5,v_{i,j-1}}$, which indicates which attempt category $z_{i,j-1} \in \{0, ..., K-1\}$ was observed at the previous point in time. We note that modelling the missingness process R_i under the assumption of monotone missingness does not enable the incorporation of a dependence structure.

The integral in the likelihood (8) 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 [30, 31]. The maximum likelihood estimates for θ and ψ (or ϕ) can then be calculated through the Newton-Raphson ridge optimization method. Corresponding macros using the software `SAS` are available from the authors.

4 CAST and Selection Models

In this section, we want to apply the selection models proposed in Section 3 to the CAST data set. Our aim is to explore the impact of missingness on the rate of improvement through a sensitivity analysis. Furthermore, we want to investigate the effectiveness of reminders to re-approach initial non-responders. In this context, we adjust for missingness by modelling the reminder process, see Section 3.4. We contrast this model with the traditional selection model, where we adjust for missingness by modelling the missingness process, see Section 3.5.

For CAST we focus on $x_{i,j} = \breve{x}_{i,j} = \tilde{x}_{i,j} = age_i - 27$, i.e. age centered around the median age. Different assumptions for the missingness mechanisms will be made and the results will be compared with those obtained based on the assumption of ignorability, see Table 3.

4.1 Results using the Reminder Process Model via $p_{j,k}$

Using the notation used in Subsection 3.4, we will investigate the following logistic regression models for the conditional reminder process probabilties $p_{j,k}$:

- MCAR$_p$: $\text{logit}(p_{j,k}) = \psi_{0k} + \psi_1\, a_i + \psi_2\, t_j + \psi_{5,v_{i,j-1}}$;
- MAR$_p$: $\text{logit}(p_{j,k}) = \psi_{0k} + \psi_1\, a_i + \psi_2\, t_j + \psi_3\, y_{j-1} + \psi_{5,v_{i,j-1}}$;
- MNAR$_p$-1: $\text{logit}(p_{j,k}) = \psi_{0k} + \psi_1\, a_i + \psi_2\, t_j + \psi_3\, y_{j-1} + \psi_4\, y_j + \psi_{5,v_{i,j-1}}$;
- MNAR$_p$-2: $\text{logit}(p_{j,k}) = \psi_{0k} + \psi_1\, a_i + \psi_2\, t_j + \psi_4\, y_j + \psi_{5,v_{i,j-1}}$; and
- MNAR$_p$-3: $\text{logit}(p_{j,k}) = \psi_{0k} + \psi_1\, a_i + \psi_2\, t_j + \psi_3^*\, [y_{j-1} + y_j] + \psi_4^*\, [y_{j-1} - y_j] + \psi_{5,v_{i,j-1}}$.

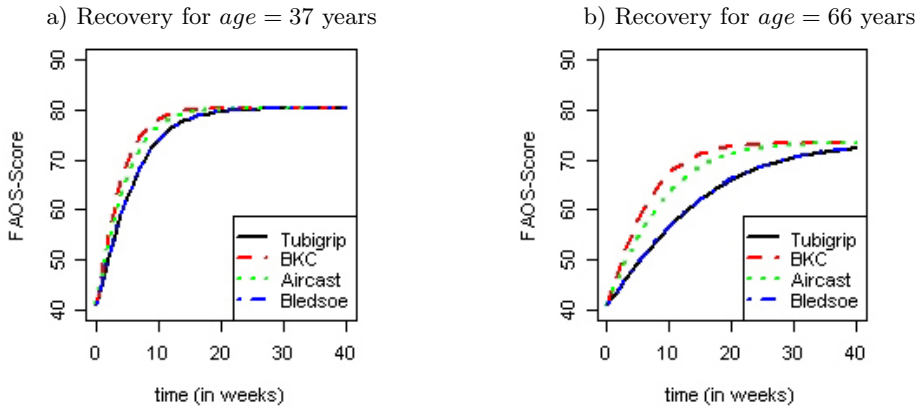

Figure 2: a) Fitted recovery curve versus time for the different randomisation groups and 37 year old male patients. b) Fitted recovery curve versus time for the different randomisation groups and 66 year old male patients.

where $k \in \{0,1,2,3,4\}$ and $j \in \{2,3,4\}$, i.e. $t_j \in \{4,12,39\}$. Note that here MCAR denotes a mechanism where missingness is allowed to depend on covariates but not on the outcome of interest. Initial analysis shows that the inclusion of $\psi_{5,v_{i,j-1}}$ is not necessary, as these parameters are equal, in the sense that their pairwise contrasts are not significant. Furthermore, the age-effect on the intercept, i.e. α_0, was shown to be not significant. The results for the remaining model and the case of monotone missingness are shown in Table 4.

The estimates for the outcome model parameters, i.e. θ, are practically identical under all reminder processes investigated, including the estimates under the assumption of an ignorable missingness process. The parameter estimates of $\hat{\beta}_1$ are just slightly smaller under the MNAR$_p$-1 and MNAR$_p$-2 model; and larger for MNAR$_p$-3.

Our intuition that older participants achieve a lower maximum score than younger participants is confirmed, as $\hat{\alpha}_1 < 0$. Furthermore, all models confirm that older participants recover less fast than younger patients, see Figure 2. Note that due to the non-significant treatment difference between Tubigrip and Bledsoe, the recovery curves for these treatments are practically undistinguishable.

The treatment effect of Tubigrip, the treatment differences and the associated p-values are essentially equal for all models. All approaches detect that BKC is significantly better than Tubigrip. Aircast is marginally better and Bledsoe is not measurably different from Tubigrip.

For the reminder vector $\mathbf{z}$, given the outcome $\mathbf{y}$, the results vary substantially under the different missingness processes. Especially, the MNAR$_p$-3 process leads to different conclusions compared to the other models.

Under MCAR$_p$ we observe a positive effect of sending a second questionnaire on the return of the copy. The other reminder categories do not significantly affect the response probabilities. We observe a positive age-effect, i.e. the probability of replying at a certain attempt increases with age. Furthermore, the probability of replying at a certain attempt decreases as time passes. The age- and time-effects persist under MAR$_p$, MNAR$_p$-1 and MNAR$_p$-2; the effect sizes are practically identical.

The MAR$_p$ results confirm that the reminder process, and therefore the missingness process, depends on the outcome of interest. The probability of returning a questionnaire decreases with the score at the prior occasion: patients with high scores at the previous observation times tend to return the questionnaires only after several attempts or not at all. This result is in line with quantitative findings presented in [32], which suggest that patients who considered themselves to have made fully recovery, did not return their subsequent questionnaire. Furthermore, we observe a positive effect of phone calls on the retrieval of questionnaires, see $\hat{\psi}_{01}$ and $\hat{\psi}_{03}$. In contrast, sending a second questionnaires decreases the response probability and the attempt of obtaining core outcomes over the phone appears to be not effective. The same conclusions are carried forward for the models under the MNAR$_p$-1 and the MNAR$_p$-2 assumption. However, the effect sizes and the associated p-values vary across the models.

For the MNAR$_p$-1 model, no significant effect of current or previous score on the response probabilities is found. As mentioned in Section 3.6, this is likely to be due to the high correlation of scores at adjacent occasions. The empirical correlation based on the observed data are given by: $\text{Corr}(\mathbf{Y}_0, \mathbf{Y}_4) = 0.34$, $\text{Corr}(\mathbf{Y}_4, \mathbf{Y}_{12}) = 0.65$ and $\text{Corr}(\mathbf{Y}_{12}, \mathbf{Y}_{39}) = 0.68$.

Removing the previous outcome, that is fitting the model under a MNAR$_p$-2 process, reveals a negative effect of the current score on the probability of replying. The effect size is comparable with the effect of the previous score, $\hat{\psi}_3$, in the MAR$_p$ model. Thus it seems that towards the bounds the effects of y_{ij} and $y_{i,j-1}$ become difficult to

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	41.01	0.78	-	41.01	0.78	-	41.01	0.78	-	41.01	0.78	-	41.01	0.78	-
β_1	80.05	0.85	-	80.05	0.85	-	79.98	0.85	-	79.95	0.85	-	80.36	0.85	-
β_{21}	0.28	0.03	-	0.28	0.03	-	0.27	0.03	-	0.27	0.03	-	0.28	0.03	-
β_{22}	-0.12	0.04	0.0028	-0.12	0.04	0.0028	-0.12	0.04	0.0027	-0.12	0.04	0.0027	-0.12	0.04	0.0032
β_{23}	-0.05	0.04	0.0869	-0.05	0.04	0.0869	-0.05	0.04	0.0857	-0.05	0.04	0.0853	-0.05	0.04	0.0904
β_{24}	-0.004	0.04	0.4573	-0.004	0.04	0.4573	-0.005	0.04	0.4496	-0.005	0.04	0.4466	-0.002	0.04	0.4802
α_1	-0.30	0.07	<0.0001	-0.30	0.07	<0.0001	-0.30	0.07	<0.0001	-0.30	0.07	<0.0001	-0.31	0.07	<0.0001
α_2	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001
σ^2	186.41	7.33	-	186.41	7.33	-	186.40	7.33	-	186.39	7.33	-	186.39	7.33	-
D^2	148.17	12.67	-	148.17	12.68	-	148.22	12.69	-	148.26	12.68	-	148.68	12.75	-
ψ_{00}	-0.14	0.07	0.0225	0.02	0.11	0.4395	0.14	0.15	0.1740	0.12	0.14	0.1957	0.39	0.16	0.0072
ψ_{01}	0.09	0.08	0.1562	0.24	0.12	0.0231	0.37	0.16	0.0102	0.35	0.15	0.0115	1.19	0.18	<0.0001
ψ_{02}	-0.61	0.12	<0.0001	-0.45	0.15	0.0010	-0.33	0.18	0.0311	-0.35	0.17	0.0235	0.46	0.19	0.0091
ψ_{03}	0.07	0.13	0.3036	0.23	0.16	0.0733	0.35	0.19	0.0304	0.34	0.18	0.0342	1.23	0.21	<0.0001
ψ_{04}	13.02	20.10	0.2587	14.87	29.13	0.3050	13.32	26.59	0.3084	22.43	74.45	0.3817	15.49	54.04	0.3872
ψ_1	0.008	0.004	0.0108	0.007	0.004	0.0211	0.006	0.004	0.0357	0.007	0.004	0.0341	0.002	0.004	0.3278
ψ_2	-0.009	0.003	0.0003	-0.006	0.003	0.0236	-0.006	0.003	0.0255	-0.007	0.003	0.0058	0.011	0.003	0.0003
ψ_3	-	-	-	-0.003	0.0019	0.0363	-0.002	0.002	0.2087	-	-	-	-	-	-
ψ_4	-	-	-	-	-	-	-0.003	0.002	0.1067	-0.004	0.002	0.0203	-	-	-
ψ_3^*													-0.013	0.001	<0.0001
ψ_4^*													0.023	0.002	<0.0001
-2ℓ	19357.72			19354.48			19352.92			19353.58			18933.3		

Table 4: Parameter estimates, standard errors and deviances for the outcome model (2), the reminder process and the different missing data mechanisms, see Section 4.1. The p-values are reported only for the components of (θ, ψ) that might be zero.

distinguish.

The alternative parametrization, i.e. MNAR$_p$-3, shows that the reminder process depends on the mean score and the improvement of the score between two adjacent time points. The probability of replying decreases with the mean but increases with the improvement. In contrast to the previous models, the age effect is shown to be not significant and the time effect is positive. Furthermore, for MNAR$_p$-3 all attempts, except obtaining core outcomes, are shown to have a positive effect on the retrieval of questionnaires. However, the effect sizes of the telephone reminders are the largest. Thus, the conclusions for this missingness process are generally quite different to the previously discussed models; and the deviance suggests that the MNAR$_p$-3 leads to the best fit.

4.2 Results using the Missingness Process Model

When adjusting for monotone missingness through the missingness indicator $R_{i,j}$, we explore the impact of dropout on the rate of improvement for the following missing data mechanisms:

- MCAR$_r$: $\text{logit}(\rho_{i,j}) = \phi_0 + \phi_1 a_i + \phi_2 t_j$;
- MAR$_r$: $\text{logit}(\rho_{i,j}) = \phi_0 + \phi_1 a_i + \phi_2 t_j + \phi_3 y_{i,j-1}$;
- MNAR$_r$-1: $\text{logit}(\rho_{i,j}) = \phi_0 + \phi_1 a_i + \phi_2 t_j + \phi_4 y_{i,j}$;
- MNAR$_r$-2: $\text{logit}(\rho_{i,j}) = \phi_0 + \phi_1 a_i + \phi_2 t_j + \phi_3 y_{i,j-1} + \phi_4 y_{i,j}$.; and
- MNAR$_r$-3: $\text{logit}(\rho_{i,j}) = \phi_0 + \phi_1 a_i + \phi_2 t_j + \phi_3^* (y_{i,j-1} + y_{i,j}) + \phi_4^* (y_{i,j} - y_{i,j-1})$.; and

The results under monotone missingness are given in Table 5.

The estimated outcome parameters are consistent with those obtained by modelling the reminder process in Section 4.1. Regarding the missingness processes modelled, we observe that the intercept varies substantially across the assumed models. This is not surprising, as we include more covariates to explain the missingness process. The probability of replying increases with age for all investigated models. Furthermore, we observe a non-significant time effect.

The MAR$_r$ model suggests that the probability of replying at a certain time increases with the score at that time (p-value= 0.06). Note that this result is contrary to the results based on modelling the reminder process. Including both previous and current scores using MNAR$_r$-1 is not informative. The MNAR$_r$-2 model finds a significant effect of the current score on the missingness probabilities: as the score increases, the probability of replying increases. The effect is stronger with the current score for MNAR$_r$-2 than with the previous score for MAR$_r$. Under the assumption of MNAR$_r$-3, we obtain that missingness depends positively on the average score but not on the improvement of the scores at adjacent observations times. All the results, except the age-effect, disagree with the findings in Section 4.1.

For illustration, we show the probabilities of not replying for different age groups and low/high scores under the MNAR$_p$-3 and the MNAR$_r$-3 model, see Table 6. Note that under MNAR$_p$-3 some of the probabilities of not replying are very high. This is because the third quantile of the scores at the previous time point is larger than the first quantile at the following time point, at least this holds for the scores at observation times $t_j \in \{4, 12, 39\}$. Therefore, the improvement between the adjacent scores is negative for these occasions and leads to a reverse effect of $\hat{\psi}_4^*$. As the score evolution of most patients is monotone increasing, this reverse effect would occur only very rarely, i.e. the probabilities of not replying are generally much lower.

5 CONCLUSIONS

We have proposed a selection model for continuous longitudinal data to adjust for non-ignorable or informative missingness when initial non-responders are reapproached several times. In addition we have contrasted this model with the traditional selection model framework, where we adjust for missing data by modelling the missingness process.

The models presented combine a non-linear mixed model for the underlying outcome model with a logistic regression model for the missingness and the reminder process, respectively.

The class of non-linear mixed models has found many biological applications, such as pharmacokinetic analysis, rate of clearance of a drug, studies of growth to adult size and decay [29, 33, 34]. However, to the best of our

Parameter	$MCAR_r$			MAR_r			$MNAR_r$-1			$MNAR_r$-2			$MNAR_r$-3		
	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.	Est.	SE	p-val.
β_0	41.01	0.78	-	41.01	0.78	-	41.00	0.78	-	41.00	0.78	-	41.00	0.78	-
β_1	80.05	0.85	-	80.05	0.85	-	79.65	0.93	-	79.63	0.88	-	79.65	0.93	-
β_{21}	0.28	0.03	-	0.28	0.03	-	0.27	0.03	-	0.27	0.03	-	0.27	0.03	-
β_{22}	-0.12	0.04	0.0028	-0.12	0.04	0.0028	-0.12	0.04	0.0023	-0.12	0.04	0.0022	-0.12	0.04	0.0023
β_{23}	-0.05	0.04	0.0869	-0.05	0.04	0.0869	-0.06	0.04	0.0805	-0.06	0.04	0.0801	-0.06	0.04	0.0805
β_{24}	-0.004	0.04	0.4573	-0.004	0.04	0.4573	-0.008	0.04	0.4144	-0.008	0.04	0.4119	-0.008	0.04	0.4144
α_1	-0.30	0.07	<0.0001	-0.30	0.07	<0.0001	-0.29	0.07	<0.0001	-0.29	0.07	<0.0001	-0.29	0.07	<0.0001
α_2	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001	-0.006	0.001	<0.0001
σ^2	186.41	7.34	-	186.41	7.34	<0.0001	186.68	7.38	-	186.71	7.38	-	186.68	7.38	-
D^2	148.17	12.70	-	148.17	12.73	<0.0001	148.66	12.72	<0.0001	148.70	12.78	-	148.66	12.76	-
ϕ_0	1.79	0.12	<0.0001	1.52	0.21	<0.0001	1.08	0.43	0.0063	1.06	0.38	0.0028	1.08	0.44	0.0071
ϕ_1	0.03	0.01	0.0004	0.03	0.01	0.0002	0.03	0.01	0.0001	0.03	0.01	0.0001	0.03	0.01	0.0001
ϕ_2	0.007	0.006	0.1141	0.001	0.007	0.4352	0.001	0.007	0.4353	0.001	0.006	0.4220	0.001	0.01	0.4353
ϕ_3	-	-	-	0.007	0.004	0.0594	0.001	0.007	0.4710	-	-	-	-	-	-
ϕ_4	-	-	-	-	-	-	0.011	0.01	0.1341	0.012	0.01	0.0255	-	-	
ϕ_3^*	-	-	-	-	-	-	-	-	-	-	-	-	0.006	0.003	0.0303
ϕ_4^*	-	-	-	-	-	-	-	-	-	-	-	-	0.005	0.01	0.2565
-2ℓ	16422.36			16419.90			16418.69			16418.70			16418.69		

Table 5: Parameter estimates, standard errors and deviances for the outcome model (2), the response process and the different missing data mechanisms defined in Section 4.2. The p-values are reported only for the components of (θ, phi) that might be zero.

MNAR$_p$-3			
current score	previous score	μ_{j5}	
		$t_j = 12$	$t_j = 39$
low	low	0.07	0.07
	high	0.52	0.43
high	low	0.03	0.03
	high	0.49	0.34

MNAR$_r$-2					
current score	previous score	$1 - \rho_j$			
		$t_j = 12$		$t_j = 39$	
		young	old	young	old
low	low	0.17	0.11	0.15	0.10
	high	0.14	0.09	0.12	0.08
high	low	0.15	0.09	0.13	0.09
	high	0.11	0.07	0.11	0.07

Table 6: 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$_r$-3 models. Low and high scores denote the first and third quantiles, respectively. The age groups were also classified according to the first (21 years) and third (37 years) quantile.

knowledge they are not yet used in the health-care context. We argue that in our specific case these models are preferable to standard techniques. Our 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. These assumptions and the model are based on medical research grounds and reflect the knowledge of experts in the specific research area. It enables a very flexible incorporation of exploratory variables and is easy to interpret, which is a valuable advantage to the alternative of data transformation. In addition, it accounts for the two different sources of variation and enables us to model the upper bound, that is, final recovery, which might be of particular interest to patients.

Although we believe that our outcome model is superior to standard analysis techniques applied in this field, we note that the model is limited by the covariance structure it assumes for the outcome vector. We work with a compound symmetry covariance structure, which implies equal correlation of any two different measurements on the same subject regardless of the length of the time interval between these measurements. However, the design of the CAST study had unequally spaced time points and with repeated measurements we expect more correlation when the measurements are closer in time than when they are further apart. Additionally, with bounded data, correlations increase as measurements reach the bounds regardless of the distance of the measurements in time, thus complicating the situation even more. Further work concerning modelling covariance structures for bounded continuous data is in progress.

For the reminder process, we model the probability of replying at a certain attempt, given not having replied earlier through a multinomial model. We explore the dependence of the response probabilities at a certain attempt on covariates and the outcome of interest by logistic regression. Such insight is important to understand the effectiveness of reminder systems and to improve the design of future studies. The missingness process is modelled through logistic regression.

Generally, we investigate the impact of missingness on the rate of improvement for different model families and missingness processes within a sensitivity analysis. We focus on the case of monotone missingness patterns. The simplicity of the described model fitting heavily relies on this assumption. As soon as we relax this assumption, we are confronted with multi-dimensional integrals. Attempts to run the extended `SAS` code which accounts for non-monotone missingness failed due to slowness. The calculation of the likelihood in every iteration step requires the computation of several hundred integrals and every iteration step ran for several hours. Therefore, we moved to the Bayesian paradigm to fit models based on non-monotone missingness using `WinBUGS`. However, the complicated outcome model and correlated parameters make the model fitting very difficult. This work is currently in progress.

For CAST, the conclusions that recovery is slower, and less satisfactory with age, and more rapid with BKC than Tubigrip do not alter materially across all models investigated. The superiority of Aircast brace over Tubigrip is shown to be borderline significant with monotone missingness modelled. These conclusions were not evident from linear models.

Depending on whether the reminder process or the missingness process is explored, we find that the probabilities of replying decrease or increase with the observed outcome at the current or previous occasions. Due to the high correlation between the scores at adjacent time points, problems arise when including current and previous scores jointly. The MNAR$_p$-3 model suggests that the improvement and the average score, rather than the actual scores, affect the missingness process and this model leads to the best fit. When modelling the missingness process, solely an effect of the average score was found.

In general, we observe disagreeing covariate effects dependent on whether we model the reminder or the missingness process. However, also within the two model families conclusions depend on the assumed missingness mechanism. For example, for all reminder process models, except MNAR$_p$-3, we observe a positive effect of age on the response probabilities. The MNAR$_p$-3 suggests that age does not have a significant effect.

We argue that modelling the reminder process uses richer information about the missingness process than the missingness indicator; e.g. we are able to incorporate the dependence across the reminders at the different observation times for a given patient. Therefore, we are inclined to weight the results based on the reminder process more than those based on the missingness indicator. However, and most importantly, both model families suggest that missingness depends on the outcome of interest.

Regarding the reminder process, we observe that phone calls are most effective in retrieving questionnaires.

Overall, the outcome parameters of interest are estimated very robustly across all models investigated. However, we believe that care has to be taken with such conclusions. The identification of all models presented is driven by the parametric assumptions on the marginal outcome model and the conditional missingness and the conditional reminder process, respectively. It is not clear to which extent these conclusions would differ under other assumptions; e.g. assuming an underlying skewed normal distribution, use of the probit link-function, incorporation of explanatory variables such as gender or occupation in the missingness and reminder process models. Furthermore, we include the previous and current score linearly in the reminder and missingness process. Other functional relationships might lead to differing conclusions. Nevertheless, we believe that the model families explored are valuable for understanding treatment and covariate effects on the outcome as well as 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] Mahony F, Barthel D. Functional evaluation: the barthel index. *Maryland State Medical Journal* 1965; **14**:56–61.

[20] Vernon H, Mior S. The neck disability index: a study of reliability and validity. *Journal of Manipulative and Physiological Therapeutics* 1991; **14**:409–415.

[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] 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.

[23] 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.

[24] Karlsson J, Peterson L. Evaluation of ankle joint function: the use of a scoring scale. *The Foot* 1991; **1**:15–19.

[25] Avci S, Sayli U. Comparison of the results of short-term rigid and semi-rigid cast immobilization for the treatment of grade 3 inversion injuries of the ankle. *Injury* 1998; **29**:581–584.

[26] Linde F, Hvass I, Juergensen U, Madsen F. Early mobilizing treatment in lateral ankle sprains. course and risk factors for chronic painful or function-limiting ankle. *Scandinavian Journal of Rehabilitation Medicine* 1986; **18**:17–21.

[27] Schapp G, de Keizer G, Marti K. Inversion trauma of the ankle. *Archives of Orthopaedic and Trauma Surgery* 1989; **108**:273–275.

[28] 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.

[29] Davidian M, Giltinan D. *Nonlinear models for repeated measurement data*. Chapman & Hall, 1995.

[30] Sikorsky K, Stenger F. Optimal quadratures in h_p spaces. *Association for Computing Machinery, Transactions on Mathematical Software* 1984; **3**:140–151.

[31] *SAS/STAT Users Guide, Version 8. Cary, NC: SAS Institute Inc.* 1999.

[32] 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.

[33] Davidian M, Giltinan D. Some general estimation methods for non-linear mixed models. *Journal of Biopharmaceutical Statistics* 1993; **3**:23–55.

[34] Vonesh E, Chinchilli V. *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker, Inc., 1997.