Author(s): M Akacha, JL Hutton and S Lamb

Article Title: Modelling Treatment, Age- and Gender-Specic Recovery in Acute Injury Studies
Year of publication: 2010
http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2010/paper 10-09

Publisher statement: None

Modelling Treatment, Age- and Gender-Specific Recovery in Acute Injury Studies

Mouna Akacha and Jane L. Hutton
Department of Statistics
University of Warwick, UK

Sallie E. Lamb
Warwick Medical School
University of Warwick, UK

May 17, 2010

Abstract

Background: Acute injury studies often measure physical ability repeatedly over time through scores that have a finite range. This can result in a faster score change at the beginning of the study than towards the end, motivating the investigation of the rate of change. Additionally, the bounds of the score and their dependence on covariates are often of interest.

Methods: We argue that transforming bounded data is not satisfactory in some settings. Motivated by the Collaborative Ankle Support Trial (CAST), which investigated different methods of immobilisation for severe ankle sprains, we developed a model under the assumption that the recovery rate at a specific time is proportional to the current score and the remaining score. This model enables a direct interpretation of the covariate effects. We have re-analyzed the CAST data using these improved methods, and explored novel relationships between age, gender and recovery rate.

Results: We confirm that using below knee cast is advantageous compared with a tubular bandage in relation with the recovery rate. An age and gender effect on the recovery rate and the maximum achievable score is demonstrated, with older female patients recovering less fast (age-effect: -0.21, 95% confidence interval (CI) [-0.28,-0.14]; gender effect: -0.06, CI [-0.12,-0.004]) and achieving a lower maximum score (age-effect: -8.07, CI [-11.68,-4.01]; gender-effect: -5.34, CI [-8.18, -2.50]) than younger male patients.

Conclusions: Our model is able to accurately model repeated measurements on the original scale, while accounting for the bounded nature of a score. We demonstrate that recovery in acute injury trials can differ substantially by age and gender. Older female patients are less likely to recover well from a sprain.

Keywords: Bounded scores, CAST-trial, non-linear mixed models, rate of recovery, repeated measurements

Introduction

In medical research 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[1], the *Neck Disability Index*[2], the *Foot and Ankle Outcome Score*[3] (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 acute ankle sprains of sufficient severity to prevent weight bearing.[4–6]

To show how this model was derived, we first introduce the motivating example in more detail and give descriptive statistics. Subsequently, we introduce our statistical model and apply it to the CAST dataset. Finally, we compare the inference based on our model with those obtained from other approaches. We argue that our model enables a more flexible investigation of covariate effects on rates of recovery and bounds. In addition, our analysis enables the calculation of a variety of auxiliary information, which we believe are of interest to patients suffering from acute soft tissue injuries.

The CAST Trial

The aim of the CAST study was to estimate the clinical and cost effectiveness of three different methods of mechanical support after severe ankle sprain compared to a standard treatment.[4–6]

The data for this trial were obtained from a randomised and multicentre study, which was run in 6 National Health Service trusts (8 hospitals) across the UK. Patients attending the selected emergency departments who had sustained a severe sprain of the lateral ligament complex of the ankle, were unable to weight bear, aged 16 and older, and gave informed consent were randomised in one of the four treatment groups –*Tubigrip* (standard treatment), *Below knee cast* (BKC), *Aircast brace* and *Bledsoe boot*.[4] 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 FAOS questionnaire, which contains 42 items and 5 subscales that ascertains functional limitations and the severity of other symptoms after ligament sprains.[3]

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

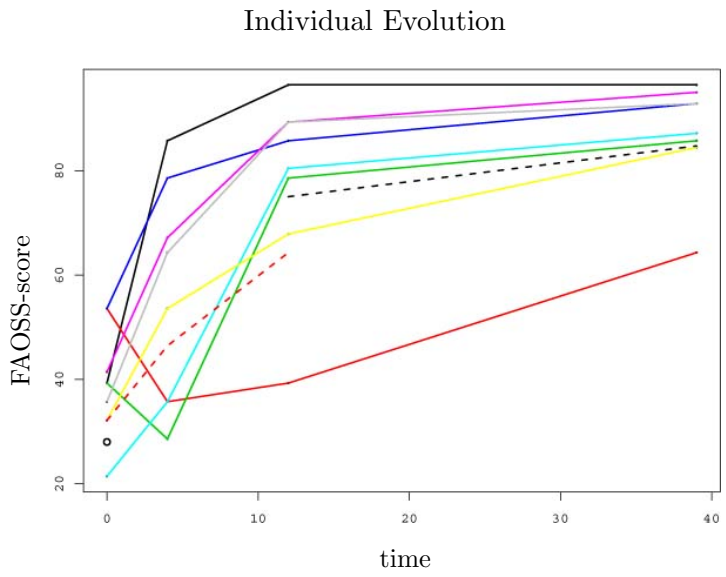

Figure 1: Individual evolution of the FAOSS-score for a random subset of 10 patients. The dashed lines correspond to patients with missing outcomes.

questionnaire called *Ankle Performance Scale*[7] (APS) 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.[4,8–10] 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.[4] 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.[11] 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 was usually an increasing function of time. Also, the score increased much faster at the beginning of the study than towards the end and some patients achieved their maximum score sooner than others. The achieved score at the end of 9 months and the rate at which this score was achieved varied across the subjects. In general, however, the responses exhibited 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 adjust for the explanatory variables gender, age and randomisation group. We use the explanatory variable *randomisation*

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.

group rather than the *treatment group* because the analysis will be performed on an intention-to-treat basis, i.e. all participants are 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 gender and age. There was a slightly larger number of males in the BKC group. Overall there was a greater proportion of men (58%) than women (42%). 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.

Statistical Model

This setting poses several challenges for a statistical model. We need to model the rate of improvement; take into account the bounded nature of the score; model the repeated measurements jointly and handle the missing data. Here, we will focus on all challenges but the last. We will assume an ignorable missingness process and base inference on the observed likelihood.[12,13] For further discussion about the missing data issue for the CAST study we refer to two papers.[14,15]

Let $\mathbf{y}_i = (y_{i,0}; y_{i,4}; y_{i,12}; y_{i,39})^\top$ denote the response vector of subject $i \in \{1, ..., 553\}$, where $\mathbf{y}_i$ is a realisation of the random vector $\mathbf{Y}_i$. As the score is on a continuous scale, we assumed an underlying normal distribution. That is $\mathbf{Y}_i$ is multivariate normal distributed and we denote the joint outcome vector for all subjects by $\mathbf{Y} = (\mathbf{Y}_1^\top, ..., \mathbf{Y}_{553}^\top)^\top$.

Furthermore, let $X_i = (\mathbf{x}_{i,0}; \mathbf{x}_{i,4}; \mathbf{x}_{i,12}; \mathbf{x}_{i,39})^\top$ be the matrix of explanatory variables for subject $i \in \{1, ..., 553\}$. The randomisation group is denoted by $\eta_i \in \{1, 2, 3, 4\}$, where $\eta_i = 1$ corresponds to Tubigrip, $\eta_i = 2$ to BKC, $\eta_i = 3$ to Aircast brace and finally $\eta_i = 4$ to Bledsoe boot. We propose the following model for the outcome process

$$\begin{aligned} \mathbf{Y}_i | U_i &\sim \mathcal{N}_4\left(\mu_i, \sigma^2 I\right); \\ U_i &\sim \mathcal{N}(0, D^2); \\ \mu_{ij} &= g(\mathbf{x}_{ij}, \theta_i) \quad \text{for } j \in \{0, 4, 12, 39\}. \end{aligned} \tag{1}$$

The mean component μ_{ij} of the normal distribution is described by the non-linear model function g. Moreover, I is the four-dimensional identity matrix and the parameter vector $\theta_i = (\theta^\top, U_i)^\top$ varies across subjects. Here, θ denotes the fixed effects and U_i the random subject-specific effect. Assuming that θ_i varies for different subjects

reflects our observation that patients share the same general shape of the response curve. However, the baseline values and the achieved scores at the end of the study vary from patient to patient. In order to account for this inter-individual variation, but also for the intra-individual correlation, regression parameters are assumed to vary for different patients according to an underlying unobservable process U_i. In this way we distinguish between the two sources of variation: the within-individual variation σ^2 and the between-individual variation D^2. This unobservable quantity U_i is patient-specific and measures the deviation from the average evolution, which all patients share. For convenience we omit the i-subscript for θ_i in the derivation of the non-linear model function $g(\mathbf{x}_{ij}, \theta) = \mu_{ij}$.

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 $y_{ij} = 0$ implies a recovery rate of zero. This is motivated by the fact that 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. $g'(\mathbf{x}_{ij}, \theta)$, to be proportional to the current score, $g(\mathbf{x}_{ij}, \theta)$, and the still achievable score $[\max\{g(\mathbf{x}_{ij}, \theta)\} - g(x_{ij}, \theta)]$. Hence, we are interested in solving the differential equation

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

where κ_{η_i} for $\eta_i \in \{1, 2, 3, 4\}$ is the treatment-specific proportion-factor. Note that this relation formulates our assumptions above in mathematical terms, as the derivative of a curve at a certain point measures how a function changes at that point and thus corresponds to the rate of change. Reducing the problem to $x_{ij} = t_j$ for simplicity yields

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

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, previous analyses[4] and exploratory analysis have shown that the scores and the rate of recovery depend on age and gender of the patients. Thus, our aim is to adjust for these covariates. Incorporating the explanatory variables $a_i = age_i - 27$ (age centered around the median) and gender $sex_i \in \{f, m\}$ (f female, m male) 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:[16]

$$\begin{aligned} g(\mathbf{x}_{ij}, \theta_i) &= \frac{\beta_1 + \alpha_1\, a_i + \gamma_1\, \mathbb{1}(sex_i = f)}{\exp\left\{-\left([\beta_{21} + \beta_{2,\eta_i}\, \mathbb{1}(\eta_i \neq 1)] + \alpha_2\, a_i + \gamma_2\, \mathbb{1}(sex_i = f)\right) \cdot t_j\right\} \left(\frac{\beta_1 + \alpha_1 \cdot a_i + \gamma_1\, \mathbb{1}(sex_i = f)}{\beta_0 + \alpha_0\, a_i + \gamma_0\, \mathbb{1}(sex_i = f)} - 1\right) + 1} + U_i \\ &= \tilde{\mu}_{ij} + U_i, \end{aligned} \tag{2}$$

where $\mathbb{1}(\eta_i \neq 1)$ and $\mathbb{1}(sex_i = f)$ are one if $\eta_i \neq 1$ or $sex_i = f$ respectively and zero otherwise. Note that it is possible to include further explanatory variables. The interpretation of all parameters is straightforward:

- $\beta_0 + \alpha_0\, a_i + \gamma_0\, \mathbb{1}(sex_i = f)$ describes the intercept, where α_0 indicates the effect of age on the intercept. For female patients the intercept differs by γ_0 compared

Parameter	Est.	CI	P-val.
β_0	41.11	[39.59,42.63]	-
β_1	82.64	[80.71,84.57]	-
β_{21}	0.29	[0.23,0.36]	-
β_{22}	0.12	[0.04,0.20]	0.0036
β_{23}	0.07	[-0.01,0.14]	0.0798
β_{24}	0.001	[-0.07,0.07]]	0.9780
α_1	-0.24	[-0.37,-0.11]	0.0004
α_2	-0.01	[-0.01,-0.002]	< 0.0001
γ_1	-5.34	[-8.18,-2.50]	0.0002
γ_2	-0.06	[-0.12,-0.004]	0.0353
σ^2	13.63	[13.11,14.14]	-
D^2	12.00	[11.01,13.00]	-

Table 2: Overview of the parameter estimates and confidence intervals (CI) of θ for model (1) based on the assumptions of an ignorable missingness process. The p-values are reported only for the components of θ that might be zero.

to male patients.

- $\beta_1 + \alpha_1\, a_i + \gamma_1\, \mathbb{1}(sex_i = f)$ describes the maximum score (upper bound), α_1 the effect of age on this upper bound and γ_1 the effect of being a female patient. In particular, this model accounts for the bounded nature of the score. As time increases a maximum score, varying according to age and gender, is achieved.
- $[\beta_{21} + \beta_{2,\eta_i}\, \mathbb{1}(\eta_i \neq 1)] + \alpha_2\, a_i + \gamma_2\, \mathbb{1}(sex_i = f)$ indicates the rate of improvement, i.e. how fast the upper bound is achieved. This rate depends on the randomisation group η_i, age and gender. For $\eta_i \in \{2, 3, 4\}$ the parameters β_{2,η_i} denote the contrast to, or increase from, the treatment slope of Tubigrip, i.e. β_{21}.

Note that we could also incorporate a treatment specific effect on the upper bound. However, extending the model reveals that such an effect is non-significant. Due to the randomisation it is not sensible to add an treatment effect on the intercept term.

By adding the subject-specific quantity U_i in equation (2) we assume that the intercept and the upper bound vary across patients, but not the rate of recovery. That is, dependent on the patient-specific quantity the recovery curve is assumed to shift upwards or downwards. However, the shape of the curve remains unchanged. The mean for an average person i, that is with $U_i = 0$, for time point j is given by $\tilde{\mu}_{ij}$, see equation (2).

Analyses based on the assumption of ignorability lead us to use model (2) without any age- and gender-effect on the intercept. Hence, the parameters of the model are $\theta_i = (\theta^\top, U_i)^\top$ with $\theta = (\beta_0, \beta_1, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24},\ \alpha_1, \alpha_2, \gamma_1, \gamma_2, \sigma, D)^\top$.

This model can be reformulated in terms of the multivariate normal model with a compound symmetry covariance structure. Let I be the identity matrix and J a square matrix with all elements unity. Using the notation introduced in equation (2) and $\tilde{\mu}_i = (\tilde{\mu}_{i0}, \tilde{\mu}_{i4}, \tilde{\mu}_{i12}, \tilde{\mu}_{i39})^\top$ we obtain:

$$\mathbf{Y}_i \quad \sim \quad \mathcal{N}_4\left(\tilde{\mu}_i, \Sigma\right) \quad \text{where} \quad \Sigma = \sigma^2\, I + D^2\, J.$$

In particular, the correlation between measurements on an individual at different times is given by $D^2/(\sigma^2 + D^2)$. The maximum likelihood estimate $\hat{\theta}$ based on the joint observed likelihood can be calculated through the *Newton-Raphson* method, which is e.g. implemented in the `SAS`-procedure `NLMIXED`.[17] The parameter estimates for the model defined in equations (1) and (2) are summarized in Table 2.

a) Recovery for $age = 37$ years b) Recovery for $age = 66$ years c) Recovery for Tubigrip

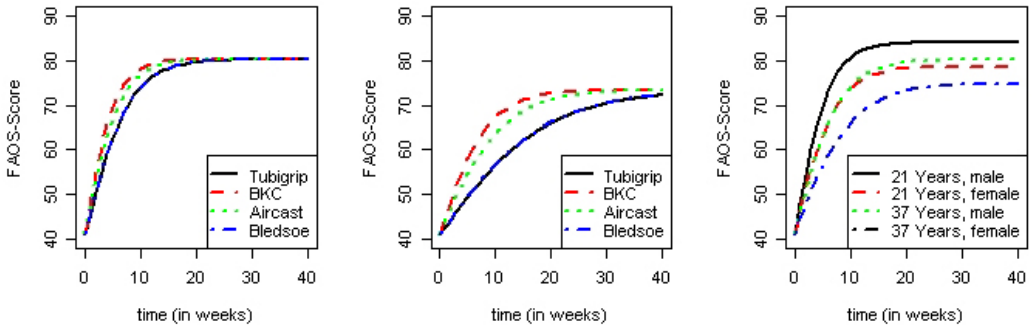

Figure 2: a) Fitted recovery curve versus time for the different randomisation groups and 37 year old male patients. b) Fitted recovery curve versus time for the different randomisation groups and 66 year old male patients. c) Fitted FAOS-score versus time for different genders, age classes and Tubigrip. The age groups were classified according to the first (21 years) and third (37 years) quantiles.

Results

The interpretation of all parameters is straightforward. The intercept for an average patient, i.e. the patient-specific quantity is zero, is given by $\hat{\beta}_0 = 41.11$. The maximum achievable score for an average person is given by $\hat{\beta}_1 = 82.64$, but with a negative age-effect and the upper bound for female patients is in average approximately 5 score points lower than for male patients. Using these point estimates and allowing for the age range of 16 to 72 implies that the upper bound for male patients varied between 73 and 85 score points, whereas for an average female patient the upper bound lay between 68 and 80. Furthermore, we observe a negative age effect on the rate of improvement, i.e. older participants recovered less fast than younger patients. In addition, female patients recovered less fast than male patients as $\hat{\gamma}_2 < 0$. Generally this means that the maximum achievable score for older and female patients was lower than for younger and male patients. In particular, this implies that older and female patients were less likely to recover completely from a severe sprain with the treatment options tested in this trial.

The standard deviations reflecting the within- and between-patient variations are of the same magnitude. Note that the between-subject variation D^2 drives the estimation of the subject-specific effect U_i.[18]

Regarding the treatment comparison, a significant difference in the rates of improvement between Tubigrip and BKC is detected. The rate of recovery for Aircast brace was only marginally higher than Tubigrip. There was no significant difference in recovery rates between Tubigrip and Bledsoe boot. The fitted curves for the different randomisation groups for an average male patients of age 37 or 66 are shown in Figure 2a and Figure 2b respectively. Independent of the randomisation group, patients ended at the same score. However, the rate at which they achieved the upper bound differed substantially, in particular for older patients. Note that the fitted curves for Tubigrip and Bledsoe are indistinguishable due to the insignificant treatment difference. These two plots also show the difference in the upper bounds dependent on age. The dependence of the recovery rate and the upper bound on age and gender is visualized for the randomisation group Tubigrip in Figure 2c.

In order to capture the interplay between the covariate effects of age, gender and randomisation groups with the bounded nature of the score, we present the average

Age	Gender	Treatment	weeks 0-4		weeks 4-12		weeks 12-39	
			Est.	CI	Est.	CI	Est.	CI
16 years	female	Tubigrip	20.3	[15.9,24.8]	16.1	[13.1,19.1]	2.4	[0.1,4.6]
		BKC	26.2	[22.4,30.0]	12.0	[8.6,15.5]	0.6	[-0.03,1.2]
		Aircast	23.8	[20.1,27.5]	13.9	[10.8,17.0]	1.1	[0.1,2.0]
		Bledsoe	20.4	[16.3,24.5]	16.1	[13.2,18.9]	2.3	[0.3,4.4]
	male	Tubigrip	26.2	[21.9,30.4]	16.6	[13.1,20.1]	1.4	[0.1,2.7]
		BKC	32.0	[28.5,35.4]	11.8	[8.5,15.1]	0.3	[0.01,0.7]
		Aircast	29.7	[25.9,33.4]	13.8	[10.4,17.3]	0.6	[0.04,1.2]
		Bledsoe	26.2	[23.0,29.5]	16.5	[13.9,19.2]	1.4	[0.4,2.3]
21 years	female	Tubigrip	18.4	[14.4,22.5]	16.2	[13.9,18.5]	1.3	[0.5,5.6]
		BKC	24.4	[20.8,28.1]	12.4	[9.2,15.6]	0.4	[0.01,1.5]
		Aircast	22.0	[18.5,25.5]	14.2	[11.5,16.9]	0.6	[0.2,2.5]
		Bledsoe	18.5	[14.6,22.3]	16.1	[13.9,18.4]	1.3	[0.6,5.4]
	male	Tubigrip	24.2	[20.2,28.2]	17.0	[14.0,19.9]	0.8	[0.3,3.3]
		BKC	30.2	[26.8,33.5]	12.3	[9.2,15.4]	0.2	[0.03,0.8]
		Aircast	27.8	[24.1,31.5]	14.3	[11.1,17.6]	0.4	[0.1,1.5]
		Bledsoe	24.3	[21.1,27.4]	16.9	[14.6,19.2]	0.6	[0.7,2.9]
27 years	female	Tubigrip	16.2	[12.5,19.8]	16.0	[14.4,17.6]	4.0	[1.1,7.0]
		BKC	22.3	[18.8,25.9]	12.8	[9.9,15.7]	1.0	[0.1,2.0]
		Aircast	19.9	[16.5,23.2]	14.5	[12.2,16.8]	1.9	[0.4,3.3]
		Bledsoe	16.2	[12.5,19.9]	16.0	[14.3,17.6]	4.0	[1.0,7.0]
	male	Tubigrip	21.8	[18.1,25.6]	17.3	[14.9,19.6]	2.4	[0.6,4.2]
		BKC	28.0	[24.7,31.4]	12.9	[9.9,15.9]	0.6	[0.1,1.1]
		Aircast	25.6	[21.8,29.3]	14.9	[11.8,17.9]	1.1	[0.1,2.1]
		Bledsoe	21.9	[18.8,25.0]	17.3	[15.3,19.2]	2.4	[1.0,3.8]
37 years	female	Tubigrip	12.4	[9.5,15.4]	14.9	[13.8,16.1]	6.4	[2.8,9.9]
		BKC	18.9	[15.4,22.4]	13.3	[10.8,15.8]	1.7	[0.2,3.1]
		Aircast	16.3	[13.0,19.6]	14.5	[12.7,16.3]	3.0	[0.8,5.2]
		Bledsoe	12.5	[8.8,16.2]	15.0	[13.8,16.1]	6.3	[1.9,10.8]
	male	Tubigrip	17.9	[14.4,21.4]	17.3	[15.7,18.9]	3.9	[1.4,6.5]
		BKC	24.4	[20.8,28.0]	13.7	[10.8,16.7]	1.0	[0.1,1.9]
		Aircast	21.8	[17.8,25.9]	15.5	[12.7,18.4]	1.8	[0.2,3.4]
		Bledsoe	18.0	[14.6,21.3]	17.3	[15.7,18.9]	3.9	[1.4,6.4]
66 years	female	Tubigrip	2.3	[-1.3,5.9]	4.3	[-1.9,10.5]	10.9	[2.6,19.1]
		BKC	9.2	[4.4,13.9]	11.4	[9.0,13.9]	6.3	[-0.3,12.8]
		Aircast	6.3	[1.3,11.3]	9.7	[5.1,14.3]	10.1	[2.2,18.1]
		Bledsoe	2.4	[-3.3,8.0]	4.4	[-5.2,14.0]	11.0	[-1.3,23.3]
	male	Tubigrip	6.9	[1.9,11.8]	11.2	[5.8,16.5]	13.1	[5.2,21.0]
		BKC	14.1	[8.5,19.6]	14.2	[11.3,17.1]	4.1	[-0.9,9.0]
		Aircast	11.1	[4.7,17.5]	14.0	[11.3,16.8]	7.1	[-1.3,15.5]
		Bledsoe	6.9	[0.7,13.1]	11.2	[4.7,17.7]	13.0	[3.3,22.7]

Table 3: Overview of the average improvements (Est.) between two adjacent time points for different age groups, genders and randomisation groups. The age classes were classified according to the first percentile (16 years), the first (21 years), second (27 years) and third (37 years) quantiles and the 99th percentile (66 years). The confidence intervals are denoted by CI.

		female		male	
Age	Treatment	Weeks	CI	Weeks	CI
16 years	Tubigrip	5.0	[3.6,6.3]	3.6	[2.8,4.3]
	BKC	3.5	[2.8,4.3]	2.7	[2.2,3.1]
	Aircast	4.0	[3.2,4.9]	3.0	[2.4,3.5]
	Bledsoe	4.9	[3.7,6.2]	3.5	[3.0,4.1]
21 years	Tubigrip	5.6	[4.1,7.1]	3.9	[3.1,4.7]
	BKC	3.9	[3.0,4.7]	2.9	[2.4,3.4]
	Aircast	4.5	[3.5,5.4]	3.3	[2.7,3.9]
	Bledsoe	5.6	[4.2,7.0]	3.9	[3.3,4.5]
27 years	Tubigrip	6.7	[4.9,8.4]	4.5	[3.6,5.4]
	BKC	4.4	[3.4,5.4]	3.2	[2.6,3.8]
	Aircast	5.2	[4.1,6.3]	3.7	[2.9,4.4]
	Bledsoe	6.6	[4.9,8.4]	4.5	[3.7,5.2]
37 years	Tubigrip	9.3	[6.8,11.8]	5.8	[4.5,7.1]
	BKC	5.6	[4.2,7.0]	3.9	[3.1,4.7]
	Aircast	6.8	[5.1,8.5]	4.5	[3.4,5.6]
	Bledsoe	9.3	[6.3,12.2]	5.7	[4.5,7.0]
66 years	Tubigrip	73.5	[-40.0,186.9]	18.4	[4.4,32.4]
	BKC	17.1	[4.7,29.5]	8.3	[4.2,12.5]
	Aircast	25.6	[2.2,49.1]	10.9	[3.9,18.0]
	Bledsoe	71.5	[-96.3,239.4]	18.2	[1.4,35]

Table 4: An overview of the expected number of weeks to reach a score of 65 for different genders, age groups and the four randomisation groups.

improvement between two adjacent time points dependent on these covariates, see Table 3. The estimates for the improvements underline the large effect of age and gender on the improvements. Comparing the score gain for 16 year old and 66 year old patients shows that older patients recovered much more slowly than young patients. Furthermore, the effect of the bounded score stands out. In the last time interval, the improvement was generally much smaller than in the previous intervals. However, the effect of the bounded score depended also on the age and gender. On average, older and female patients needed longer to achieve their upper bound, which dependent on age ranges between 68 and 80 score points. In fact, for an average person, i.e. $U_i = 0$, we are able to calculate the expected time to achieve a certain score S based on the fitted curves. For this purpose we need to rearrange equation (2) to solve for time $t(S)$:

$$t(S) = \frac{-1}{[\beta_{21} + \beta_{2,\eta_i}\,\mathbb{1}(\eta_i \neq 1)] + \alpha_2\, a_i + \gamma_2\,\mathbb{1}(sex_i = f)} \times \log\left(\frac{\beta_1 + \alpha_1\, a_i + \gamma_1\,\mathbb{1}(sex_i = f) - S}{\left(\frac{\beta_1 + \alpha_1\, a_i + \gamma_1\,\mathbb{1}(sex_i = f)}{\beta_0} - 1\right)\cdot S}\right)$$

The estimated times to achieve a score of $S = 65$ are shown in Table 4. While a 16 year old male patient under Tubigrip needed approximately 4 weeks to achieve a score of 65, a man at the age of 37 years needed in average 2 weeks longer. The difference for 66 year old patients is even more drastic. However, care should be taken in reporting these results as for 66 year old female patients the maximum achievable score was

nearly 68, which is close to $S = 65$ and thus leads to an imprecise estimation. This is reflected in the wide confidence intervals. In general, however, we believe that this information, together with the ability to quantify the expected upper bound per age band and gender, could be of particular interest to patients.

Discussion

We have proposed a model for continuous, bounded and repeated measurements which enables the investigation of covariate effects on the rate of change and the bounds. The model belongs to the class of non-linear mixed models which have found many biological applications, such as pharmacokinetic analysis, rate of clearance of a drug, studies of growth to adult size and decay.[16,19,20] However, to the best of our 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.

In the CAST study, we observed repeated measurements and thus a sensible model needs to account for the variation among the measurements within a given patient and the variation between different patients. The original analysis[5,6] did not distinguish between these two different sources of variation. Additionally, the repeated measurements were not modelled jointly: the treatment differences were investigated for every time point separately and it was not clear how to combine these estimates into a single inference. This situation was even more complicated as the data were clustered towards the end of the trial. Discrimination between treatments at these time points was practically impossible. Moreover, no satisfactory description of the score evolution over time for different age groups and genders was presented.

Furthermore, we argued 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 over time. 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.

Our model was derived on medical research grounds and reflects 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.

Fitting this model to the CAST data revealed that recovery was more rapid with BKC than with Tubigrip. These results coincide with those of the original analysis, but add to it by allowing the estimation of the time to recovery in each of the groups. The results of this analysis re-enforce the results of the original CAST trial, in so far as it provides further that interventions that immobilise the ankle, such as below knee plaster, are more effective than those which permit early movement. These conclusions are contrary to previous studies.[21] Taking into account the considerable variation in the costs for each treatment, these results might be relevant for the UK National Health Service. Further, we show that older and female patients recovered substantially more slowly than younger and male participants. Also the score at final recovery for older female patients was lower than for young male patients, suggesting that older female patients were less likely to recover completely from an acute soft

tissue injury. We translated these findings into auxiliary information, such as the expected time to achieve a certain score for different patient groups.

Although we believe that our model is superior to standard analysis techniques applied in this field, we note that the model is limited by the covariance structure it assumed for the outcome vector. We worked with a compound symmetry covariance structure, which implied 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. Furthermore, we note that our analysis assumed an ignorable missingness process. A sensitivity analysis scrutinizing this assumptions is available. [14]

Key Messages

- Non-linear models can be preferable to the transformation approach when dealing with bounded data.
- For CAST, we confirm that the use of below knee cast is advantageous compared to a tubular bandage in relation with the rate of recovery.
- We show that older and female patients recover substantially slower than younger male patients and that older patients and female patients are less likely to recover completely from a severe sprain.

References

[1] Mahony F, Barthel D. Functional evaluation: the Barthel index. Maryland State Medical Journal. 1965;14:56–61.

[2] Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. Journal of Manipulative and Physiological Therapeutics. 1991;14:409–415.

[3] 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.

[4] Lamb SE, Nakash RA, Withers EJ, Clark M, Marsh JL, Wilson S, et al. 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.

[5] 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.

[6] Cooke MW, Marsh JL, Clark M, Nakash RA, Jarvis RM, Hutton JL, et al. 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.

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

[8] 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.

[9] 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.

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

[11] 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.

[12] Little RJA, Rubin DB. Statistical analysis with missing data. Wiley Interscience; 2002.

[13] Molenberghs G, Kenward MG. Missing data in clinical studies. Wiley; 2007.

[14] Akacha M, Hutton JL. Analysing the rate of change in a longitudinal study with missing data, taking into account the number of contact attempts;. Available at: http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2010/.

[15] 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.

[16] Davidian M, Giltinan DM. Nonlinear models for repeated measurement data. Chapman & Hall; 1995.

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

[18] Molenberghs G, Verbeke G. Models for discrete longitudinal data. Springer; 2005.

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

[20] Vonesh EF, Chinchilli VM. Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker, Inc.; 1997.

[21] Bridgman SA, Clement D, Downing A, Walley G, Phair I, Maffulli N. Population based epidemiology of ankle sprains attending accident and emergency units in the West Midlands of England, and a survey of UK practice for severe ankle sprains. Emergency Medicine Journal. 2003;20:508–510.