Modelling Disease Activity in Juvenile Dermatomyositis: A Bayesian Approach

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Modelling Disease Activity in Juvenile Dermatomyositis: A Bayesian Approach

E.H. Pieter van Dijkhuizen[1, 2], Claire T. Deakin[3], Lucy R. Wedderburn[3, 4] and Maria De Iorio[5]

Abstract

Juvenile dermatomyositis (JDM) is the most common form of the juvenile idiopathic inflammatory myopathies (IIM) characterised by muscle and skin inflammation, leading to symmetric proximal muscle weakness and cutaneous symptoms. It has a fluctuating course and varying prognosis. In a Bayesian framework, we develop a joint model for four longitudinal outcomes, which accounts for within individual variability as well as inter-individual variability. Correlations among the outcome variables are introduced through a subject-specific random effect. Moreover, we exploit an approach similar to a hurdle model to account for excess of a specific outcome in the response. Clinical markers and symptoms are used as covariates in a regression set-up. Data from an ongoing observational cohort study are available, providing information on 340 subjects, who contributed 2725 clinical visits. The model shows good performance and yields efficient estimations of model parameters, as well as accurate predictions of the disease activity parameters, corresponding well to observed clinical patterns over time. The posterior distribution of the by-subject random intercepts shows a substantial correlation between two of the outcome variables. A subset of clinical markers and symptoms are identified as associated with disease activity. These findings have the potential to influence clinical practice as they can be used to stratify patients according to their prognosis and guide treatment decisions, as well as contribute to on-going research about the most relevant outcome markers for patients affected by JDM.

Introduction

The juvenile idiopathic inflammatory myopathies (IIM) are a group of heterogeneous chronic diseases whose prime manifestation is skeletal muscle inflammation. Juvenile dermatomyositis (JDM) is the most common of the juvenile IIM, with an incidence of about 1.9-4.1 per million children per year[1]. Besides the skeletal muscle inflammation, JDM is characterised by skin inflammation and manifests itself clinically as symmetric proximal muscle weakness and cutaneous symptoms such as papules over the knuckles (Gottron's papules), skin rashes, nail fold erythema, ulceration and calcinosis, (deposition of subcutaneous or intramuscular calcium). JDM is a systemic disease, potentially involving other organ systems as well. If not adequately treated, the inflammation may cause muscular damage, leading to permanent muscle weakness and life-long disabilities, and skin atrophy[2;3]. The prognosis of JDM is variable with about 24-40% of patients experiencing a monocyclic course, and the majority of patients (50-60%) showing a chronic disease pattern. The mortality rate of JDM is about 2-3%[3].

This clinical picture highlights that JDM is a disease with a significant medical, social and economic burden[1;4]. Some studies suggest that early aggressive treatment improves the outcome for children with JDM, thereby alleviating this burden[5]. On the other hand, the treatment itself is associated with serious side effects, such as growth failure, pubertal delay and osteoporosis. As a consequence, it is good practice to administer as few drugs as possible, especially in paediatrics, in order to limit the negative effects on the normal physiology of the developing child. As such, the clinical challenge is to stratify patients according to their prognosis and adapt their treatment plans in concordance with their expected prognosis.

To answer the clinical question of interest, we need to identify clinical markers associated with disease activity and capable of predicting its evolution. Insight has been gained by the study of the various antibody patterns in the juvenile IIM[2;3;6] mainly to distinguish between JDM and the other subtypes of the juvenile IIM. Two previous studies have evaluated the potential of clinical variables to predict disease activity[1;7] but these studies have the limitation of measuring disease activity as

[1] Paediatric Rheumatology, IRCCS G. Gaslini, Genoa, Italy
[2] Paediatric Rheumatology, University Medical Centre Utrecht Wilhelmina Children's Hospital, Utrecht, The Netherlands
[3] University College London GOS Institute of Child Health, London, United Kingdom
[4] Great Ormond Street Hospital for Children, London, United Kingdom
[5] Department of Statistical Science, University College London, United Kingdom

Corresponding author:
E.H. Pieter van Dijkhuizen, University Medical Centre Utrecht, Wilhelmina Children's hospital, department of paediatric rheumatology, Lundlaan 6, 3584 EA, Utrecht, The Netherlands
Email: E.H.P.Dijkhuizen@umcutrecht.nl

a binary outcome (active vs. non-active). Moreover, whereas Stringer et al.[7] analyse the time to remission, Ravelli et al.[1] evaluate disease activity cross-sectionally at approximately 7.7 years after disease onset.

In contrast, the aim of the current study is to identify variables associated with disease activity of JDM by modelling clinically relevant, continuous response variables over time using clinical markers and symptoms of the disease as covariates. Such a model could be used to inform physicians about features associated with disease activity, as well as guide clinicians in the management of the disease.

Data

Data collection

Data are available from a national longitudinal cohort study, in which children with definite or probable JDM according to Bohan and Peter criteria[8;9] are enrolled across the UK. The study has been described previously[4]. Enrolment started in 2000 and is ongoing. Patients are enrolled at diagnosis, or as soon as they come to the attention of one of the participating centres, and are followed up every 3 months for 2 years and subsequently at least annually. Data collected concern disease manifestations, such as muscular involvement, skin disease and other organ involvement. Routine laboratory blood tests are also performed.

At each visit four outcome parameters[10] are measured as proxy of disease activity:

- *Serum creatine kinase level* (CK), a muscle enzyme which is released into the blood stream in increased amounts in case of muscular inflammation. The serum level of this enzyme is usually elevated in active JDM, but tends to be relatively low[3]. Its theoretical range is in $(0,\infty)$ and values exceeding 150 U/L are considered pathological[10]. Many laboratories report a censored value of <20 U/L if the level is below the laboratory's detection limit.
- *Childhood myositis assessment scale* (CMAS) which assesses muscular strength and endurance[11]. The CMAS consists of a series of little tasks, which the child is asked to perform, such as raising his/her head for at least 2 minutes while laying supine, raising to his/her feet from a sitting position and performing sit ups, with and without counterbalance. A predefined number of points is assigned based on the performance of the child and the total CMAS score is the sum over all tasks, ranging from 0 to 53, with a lower score indicating more active disease. Scores of at least 48 points are considered normal[10].
- *Manual muscle testing of 8 muscle groups* (MMT8), which consists in the assessment of muscle strength of 8 predefined muscle groups. Each muscle group is then assigned by the examiner a score ranging from 0 (equal to a complete loss of muscle contraction) to 10 (equal to full muscle strength). The total MMT8 is the sum over all muscle groups, thus ranging from 0 to 80, with a lower score indicating higher disease activity. A value of at least 78 points is considered normal[10].

- *Physician's global assessment of disease activity* (PGA), which is the physician's rating of disease activity, taking into account muscular disease, skin involvement and all other organ involvement. The PGA ranges from 0.0, indicating no disease activity, to 10.0, indicating maximal disease activity. The PGA should be at most 0.2 points to be considered normal[10].

Ethical approval of the study was obtained and all participants and their parents (as appropriate) provided informed consent. The study was conducted according to good clinical practice guidelines and the declaration of Helsinki.

Data cleaning

At the time of analysis, 4122 visits of 469 patients are available. Patients not fulfilling the inclusion criteria of having probable or definite JDM according to Bohan and Peter criteria are excluded, leaving 413 patients contributing 3881 visits. Missing data are imputed in the Bayesian model, except for missing data on history variables, because these are based on patient's recollection. This leads to the exclusion of 73 patients. Therefore, the final model includes 340 patients, contributing 2725 visits. Baseline characteristics of included and excluded patients are compared in Table 1. **As expected, due to the exclusion of visits with missing data, patients having contributed fewer visits are more likely to be excluded altogether from the study. This, in turn, is more likely to occur in patients enrolled later in the course of their disease, therefore presenting less disease activity at enrolment (Table 1).**

Table 1. Baseline variables. Values are the median, [first quartile, third quartile], except where indicated otherwise. Abbreviations: CK, creatine kinase; cm, centimeter; CMAS, childhood myositis assessment scale; MMT8, manual muscle testing of 8 muscle groups; PGA, physician's global assessment of disease activity; U/L, units per litre; y, year.

Parameter	Included $N = 340$	Excluded $N = 73$
Female, n (%)	236 (69.4)	54 (74.0)
Age at diagnosis, y	7.4 [4.5, 10.5]	7.3 [4.1, 11.1]
Disease duration at diagnosis, y	0.3 [0.2, 0.6]	0.3 [0.2, 1.0]
Time after diagnosis at enrolment, y	0.2 [0.1, 1.1]	2.3 [0.4, 5.4]
Duration follow up, y	4.1 [1.6, 7.1]	1.2 [0.1, 2.6]
Disease activity at enrolment:		
CK, U/L	103 [64, 440]	98 [45, 256]
CMAS, points	41 [21, 50]	46 [37, 52]
MMT8, points	65 [45, 80]	80 [64, 80]
PGA, cm	3 [1.3, 6.0]	2.3 [0.5, 4.0]

Of the 340 included subjects, five fully observed participants (i.e. no missing values in the dependent and independent variables for all visits) are randomly selected for the out of sample predictions and are therefore excluded from the training set.

The original dataset contains 83 covariates that can be used as predictor variables, including the time elapsed since

diagnosis at the visit, disease signs and symptoms and treatment variables. **The number of covariates is reduced by excluding those hardly containing any information. To this end, a pre-selection is made by fitting linear mixed effects models, using the package lme4 [12] available in the R software [13]. More in details, for each covariate we fit four independent linear mixed models, one for each outcome, including also time from diagnosis and a random intercept for the time from diagnosis. We obtain a p-value for the covariate by employing a Wald t test and retain the smallest one among the four models. We then rank the covariates according to their most significant p-value and in such a way we select the 50% most significant predictors.** This procedure leaves 47 covariates in the model. All treatment variables and the time elapsed since diagnosis are included in the model, regardless of their performance in the linear mixed effects model, because of clinical interest.

Model

Model specification

As discussed previously, the data consist of $N = 340$ patients, each of them contributing multiple visits. For each patient i at visit j, a vector of four outcome variables is measured, $Y_{ij} = (Y_{ij}^{(\text{CK})}, Y_{ij}^{(\text{CMAS})}, Y_{ij}^{(\text{MMT8})}, Y_{ij}^{(\text{PGA})})$, $i = 1, 2, \ldots, N, j = 1, 2, \ldots, m_i$, where m_i is the number of observations for the i-th patient. Additionally, for each subject i at visit j, a vector of covariates X_{ij} is available and t_{ij} denotes the time elapsed since diagnosis. To account for the longitudinal nature of the study and for dependency between visits for the same patients, we introduce a subject-specific random effect for each of the response variables. We then link the response variables by assuming a joint model for the random effects.

The continuous outcome variables are transformed to ensure normality: we take a log-transformation for the CK levels, whereas we choose a square-root transformation for CMAS, MMT8 and PGA, as it is a well-known variance stabilising transformation that can deal with the presence of zeros in the data. For ease of notation, in what follows $\log(\text{CK})$, $\sqrt{\text{CMAS}}$, $\sqrt{\text{MMT8}}$ and $\sqrt{\text{PGA}}$ are referred to as CK, CMAS, MMT8 and PGA respectively. The empirical distribution of the transformed outcome measures is shown in Figure 1. These variables are modelled using a Bayesian approach.

Let us first consider the outcome variable CK, which we assume normally distributed:

$$Y_{ij}^{(\text{CK})} \mid \mu_{ij}^{(\text{CK})}, \sigma_{(\text{CK})}^2 \sim N(\mu_{ij}^{(\text{CK})}, \sigma_{(\text{CK})}^2) \, ,$$

where

$$\mu_{ij}^{(\text{CK})} = \alpha^{(\text{CK})} + \eta_i^{(\text{CK})} + (\gamma^{(\text{CK})} + \theta_i^{(\text{CK})}) \times t_{ij} + X_{ij}\beta^{(\text{CK})}$$

Here, $\alpha^{(\text{CK})}$ is a mean effect common to all subjects, $\beta^{(\text{CK})}$ is the vector of coefficients for the covariates X_{ij}, $\gamma^{(\text{CK})}$ is the common effect for the time elapsed since diagnosis t_{ij}, $\eta_i^{(\text{CK})}$ is a subject-specific random intercept, while $\theta_i^{(\text{CK})}$ is a
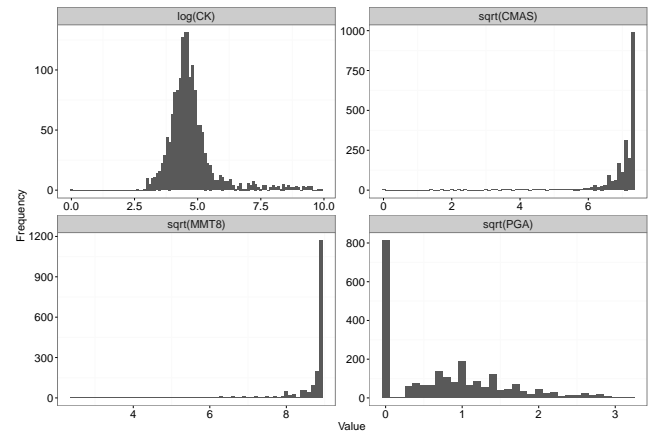

Figure 1. Empirical distribution of the four transformed outcomes.

patient-specific random coefficient for the time elapsed since diagnosis.

As mentioned above, some of the CK values ($n = 14, 0.5\%$ of the observations) are below the laboratory's detection level and reported as <20 or <21 (depending on the laboratory). These instances are modelled as truncated values, by specifying this observation as right-truncated at 20 or 21.

Most patients reach disease remission after a certain period of time and a large number of non-pathological values for CMAS, MMT8 and PGA are observed. From Figure 1 it is clear that these variables are not normally distributed, but they show an excess of non-pathological values. For these clinical outcomes we assume a mixture of a point mass at the most frequent value ($\sqrt{53}$ for CMAS, $\sqrt{80}$ for MMT8 and 0 for PGA) and a truncated Normal distribution for the other observations. **This strategy is similar to the hurdle model [14] as we recognise the possibility that the mechanism determining non-pathological values may differ from those that influence the distribution of pathological observations.** We specify a probit regression to link the probability of observing a non-pathological value for subject i at visit j to **the time elapsed since diagnosis**. Moreover, we assume a linear regression model for the continuous part of the mixture.

More in details, let $k \in \{\text{CMAS}, \text{MMT8}, \text{PGA}\}$ denote one of the clinical outcomes of interest. Then,

$$Y_{ij}^{(k)} \mid w_{ij}^{(k)}, \mu_{ij}^{(k)}, \sigma_{(k)}^2 \sim w_{ij}^{(k)}\delta_{Y_\star^{(k)}} + (1 - w_{ij}^{(k)}) \times TN(\mu_{ij}^{(k)}, \sigma_{(k)}^2) \, , \quad (1)$$

where δ is the Kronecker delta and the non-pathological values are:

$$Y_\star^{(k)} = \begin{cases} \sqrt{53} & \text{if } k = \text{CMAS}, \\ \sqrt{80} & \text{if } k = \text{MMT8}, \\ 0 & \text{if } k = \text{PGA}. \end{cases}$$

Here $TN(\mu, \sigma^2)$ denotes a truncated Normal distribution with mean μ and variance σ^2. The truncation limit changes depending on the outcome variable, but the idea is to truncate the normal distribution according to the range of possible values the response can assume (since they represent clinical scores) and the non-pathological

value. As such, on the transformed scale, the distribution of CMAS is truncated on the interval $(0, \sqrt{53})$, the distribution of MMT8 on $(0, \sqrt{80})$ and the distribution of PGA on the interval $(0, \sqrt{10})$.

We assume a probit regression for the weights of the mixture:

$$w_{ij}^{(k)} = \Phi(z_{ij}^{(k)}) \, ,$$

where Φ is the cumulative distribution function of the standard Normal distribution and

$$z_{ij}^{(k)} = \alpha^{(k)} + \eta_i^{(k)} + \gamma^{(k)} \times t_{ij} \qquad (2)$$

As in the case of creatine kinase, $\alpha^{(k)}$ and $\gamma^{(k)}$ are the common intercept and regression coefficient, respectively, common to all subjects, while η_i is the patient specific random effect. We complete the model by specifying a linear regression model for the continuous component of the mixture (see equation 1):

$$\mu_{ij}^{(k)} = \psi^{(k)} + \lambda^{(k)} \times t_{ij} + X_{ij}\beta^{(k)},$$

Note that we do not include an individual random effect for $\mu_{ij}^{(k)}$.

The four disease outcomes are measured in the same patient at the same time and capture different aspects of the same biological entity, i.e. disease activity. As such, they are correlated, especially CMAS and MMT8. Therefore it is important to account for these correlations in the model. Following Li et al.[15], we specify a multivariate Normal distribution on subject-specific random effects to take the inherent correlation of the longitudinal outcomes into account. This approach leads to a more efficient estimation of the model parameters. Hence, the four by-subject random intercepts for the outcome variables are modelled as follows:

$$\eta_i = \begin{pmatrix} \eta_i^{(\text{CK})} \\ \eta_i^{(\text{CMAS})} \\ \eta_i^{(\text{MMT8})} \\ \eta_i^{(\text{PGA})} \end{pmatrix} \sim \text{N}(\mathbf{0}, \Omega) \qquad (3)$$

where $\mathbf{0}$ is a four-dimensional vector of zeros and Ω denotes the precision matrix.

Prior Specification

The model is completed by specifying uninformative prior distributions on the remaining parameters. We use independent Normal$(0, 1000)$ priors for the intercepts $\alpha^{(k)}$ and $\psi^{(k)}$, as well as for the regression coefficients $\gamma^{(k)}$ and $\lambda^{(k)}$, with $k \in \{\text{CK, CMAS, MMT8, PGA}\}$. We place a Normal$(0, \sigma_\theta^2)$ prior on the patient-specific random slope θ_i^{CK} for the CK model, as well as a Normal$(0, \sigma_\beta^2)$ prior on the regression coefficients $\beta^{(k)}$, with common variance σ_β^2 shared by all $\beta^{(k)}$. We elicit independent Gamma distributions with hyper-parameters $(0.001, 0.001)$ for the observation precision $\tau^{(k)} = \frac{1}{\sigma_{(k)}^2}$, $k \in \{\text{CK, CMAS, MMT8, PGA}\}$, $\tau_\theta = \frac{1}{\sigma_\theta^2}$ and $\tau_\beta = \frac{1}{\sigma_\beta^2}$.

We assume a Multivariate Normal prior as random effects distribution. We choose for computational reasons

a conjugate Wishart prior for the precision matrix Ω in equation (3) with degrees of freedom $\nu = 6$ and centering matrix $S = I \times \nu \times 10$, where I is the identity matrix of appropriate dimension. As such the prior mean is given by $\mathcal{E}(\Omega) = 0.1I$[15].

Bayesian variable selection

To identify the most influential variables on disease progression and to obtain a sparser and more interpretable model, **we perform Bayesian variable selection of the 47 covariates entered into the model (except for time from diagnosis),** following the approach proposed by Kuo & Mallick[16]. See O'Hara et al.[17] for a review of Bayesian variable selection methods.

In short, for each covariate we introduce an indicator variable $\delta_p, p = 1, 2, \ldots, P$ taking values in $\{0, 1\}$. Here P denotes the total number of available covariates. When $\delta_p = 1$, we include the p-th predictor in the regression model. When $\delta_p = 0$, we omit the p-th predictor. We then add the indicator functions to the regression term by defining new coefficients

$$\widetilde{\beta}_p = \delta_p\beta_p \, ,$$

We complete the model by eliciting independent prior distributions on the δ_p:

$$\begin{aligned} \delta_p &\sim \text{Bernoulli}(\pi_p) \, , \qquad p = 1, 2, \ldots, P \\ \pi_p &\sim \text{Beta}(0.1, 0.1) \end{aligned}$$

The Beta hyperprior induces sparsity as it places most of its mass on 0 or 1. The same indicator variable is used for all instances of each covariate in the model, so that the covariate would be included or not simultaneously in the models describing the four longitudinal outcomes, i.e. the mean of the CK model, and the linear parts of the CMAS, MMT8 and PGA models. The prior on the coefficients β_p is still a Normal distribution with mean 0 and common precision τ_β, to which we assign a Gamma distribution with parameters $(0.001, 0.001)$, as described previously.

Results

Posterior inference is performed in the R software version 3.2.2[13] and JAGS, using the package rjags[18] as interface. We run the chain for 30000 iterations with a burn-in of 5000 and 2000 adaptation iterations. We thin the chain every 10 iterations.

The goodness of fit of the model is assessed visually in six randomly selected patients, each having contributed at least 10 visits. Their observed values of CK, CMAS, MMT8 and PGA are plotted over time, together with the model fit and 95% credible intervals (CI). The plot for PGA is shown in Figure 2, plots for CK, CMAS and MMT8 were similar. The plots show a good fit of the model to the data. In particular, fluctuating patterns over time are predicted well by the model. **Some observed values in Figure 2 are outside the posterior 95% CI, however, this happens only for 1.8% of the observations.**

Moreover, to better understand the predictive ability of the model, we consider the posterior predictive probability
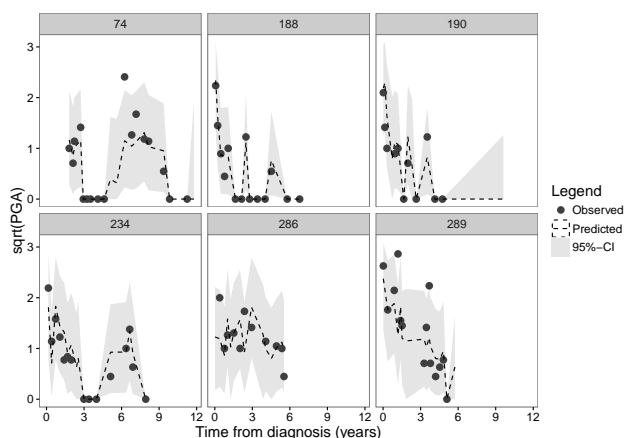
Figure 2. Model fit for PGA. Observed values of sqrt(PGA) over time of six randomly selected subjects are represented by dark grey dots, the fitted trajectory by the dashed line and the 95% credible interval around the fit is given by the shaded area.
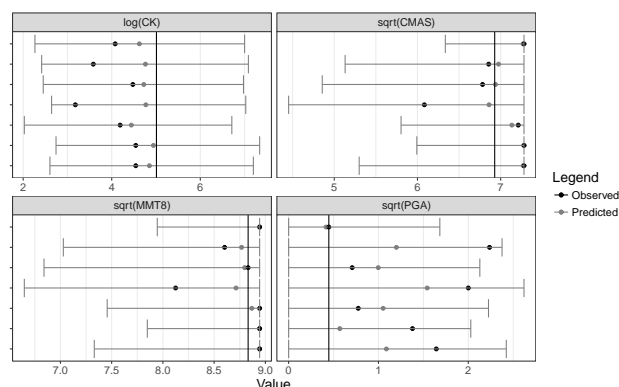

Figure 3. Out-of-sample predictions. Observed values are represented by black dots, while predicted values with corresponding 95% credible intervals are represented by grey dots with error bars for all four outcomes in 7 visits of 5 randomly selected patients. The vertical line indicates the pathological value as defined in literature.

of an observation falling **beyond the cut off point for pathological values for each of the four longitudinal outcomes, as defined in literature (i.e. >150 for CK, <48 for CMAS, <78 for MMT8 and >0.2 for PGA; note that these cut off points are defined on the original measurement scale[10]). These predictions concern pathological values and are, therefore, relevant from a clinical point of view.**

In order to assess the predictive performance of the proposed model we employ the Brier statistic[19]. This statistic assesses the quality of predictions when the response variable is binary:

$$\text{Brier} = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2,$$

where o_i is a binary observation, f_i is its predicted probability and N is the number of observations in the sample.

In our case the response variable is continuous, so in order to apply this statistic we discretize the response variable. We are interested in predicting whether each clinical outcome is above or below the above-mentioned thresholds. Then, the predicted f_{ij} is the predictive probability of obtaining values larger or smaller than the specified threshold for each patient i in the sample and visit j. Hence, the Brier statistic becomes:

$$\text{Brier} = \frac{1}{\sum_{i=1}^{N} m_i} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \left(f_{ij} - Y_{ij}^d\right)^2$$

where Y_{ij}^d is the i-th response discretized with respect to the critical threshold (equal to 1 if Y_{ij} assumes pathological values and 0 otherwise), N is the number of patients and m_i is the number of observations for individual i. The sum is taken over all the patients in the sample and the observation times. The Brier score assumes values between 0 and 1, with lower scores indicating a better performance[20]. We obtain Brier scores equal to 0.12, 0.12, 0.08 and 0.05 for CK, CMAS, MMT8 and PGA respectively. Following Steyerberg *et*

***al.*[20], we calculate rescaled Brier scores, which are normalised with respect to the maximum obtainable Brier score for a non-informative model based on the observed frequency of the outcomes. The scaled Brier score is very similar to Pearson's R^2 statistic (high values indicate a good predictive performance). In our analysis we obtain 49%, 42%, 63% and 80% for CK, CMAS, MMT8 and PGA, respectively, indicating moderate to good performance of the model.**

We also perform out of sample predictions on the four outcomes for five randomly selected, fully observed patients collectively contributing 7 observations (Figure 3). Once again, the results show good accuracy, with the model's predictions being close to the observed values and almost always on the same side of the cut-off point discriminating pathological values from non-pathological values, as established in the literature (150 U/L for CK, 48 points for CMAS, 78 points for MMT8 and 0.2 points for PGA)[10]. In only 2 cases involving CMAS the observed value is below the cut-off point, indicating a pathological value, whereas the predicted value is above. However, on a closer inspection we notice that in both these cases the observed values are close to the cut-off point (46 and 47 respectively). The relatively wide credible intervals reflect the uncertainty in the predictions, the estimation of the parameters, as well as in the imputation of missing values.

The posterior estimates of the regression coefficients of the time elapsed since diagnosis and the 46 covariates are reported in Figure 4. **The variability associated with the estimates is reasonable as shown by the credible intervals.** As expected, CMAS, MMT8 and PGA tend to normalise over time, whereas the CK level does not significantly change from the time from diagnosis.

Various clinical signs and symptoms are associated with the four outcomes, most notably a range of manifestations of skin disease, such as periorbital rash, rash on the trunk, rash over large joints, nail fold changes and facial swelling. These parameters are not only associated with PGA, but also with the strictly muscular outcomes CK and CMAS,
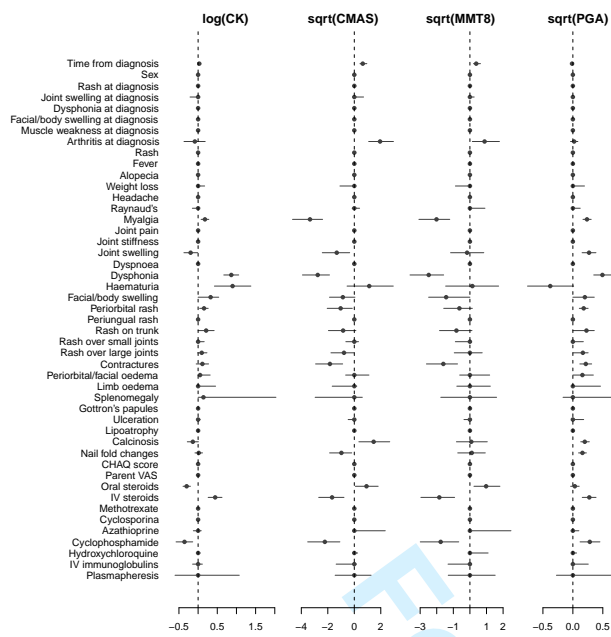
Figure 4. Medians (dots) and 95% credible intervals (horizontal lines) for the fixed effects for the time elapsed since diagnosis and all the covariates. We plot the results for each of the four outcomes.

indicating that patients who are doing worse with respect to muscular disease activity, tend to have higher cutaneous disease activity as well. As expected, patients having myalgia (muscle pain) or dysphonia (altered voice pitch due to inflammation of the muscles of the vocal cords) have much worse disease activity. Of signs and symptoms present at diagnosis, only arthritis is associated with improved disease activity with regard to the CMAS and MMT8 at the follow up visits, probably reflecting a subset of patients with an overlap syndrome between JDM and juvenile idiopathic arthritis (JIA).

The methodology developed by Li et al.[15] for modelling multiple correlated outcomes is adopted, because of the prior belief that the four outcomes for JDM are correlated. The posterior distribution of the correlation between the by-subject random intercepts for the four outcomes supports this prior belief. The estimates are shown in Table 2, evidencing a substantial correlation between $\eta_i^{(\mathrm{CMAS})}$ and $\eta_i^{(\mathrm{MMT8})}$ (0.54, 95% CI 0.42-0.65), highlighting the appropriateness of the adopted methodology.

Table 2. Estimated correlations with 95% credible intervals of subject random intercepts for the four outcomes CK, CMAS, MMT8 and PGA. Abbreviations: CI, credible interval; CK, creatine kinase; CMAS, childhood myositis assesment scale; MMT8, manual muscle testing of 8 muscle groups; PGA, physician's global assessment.

Comparison	$\hat{\rho}$	95% CI
CK *vs.* CMAS	0.07	(-0.07, 0.23)
CK *vs.* MMT8	-0.001	(-0.16, 0.15)
CK *vs.* PGA	0.0007	(-0.15, 0.16)
CMAS *vs.* MMT8	0.54	(0.42, 0.65)
CMAS *vs.* PGA	0.17	(0.01, 0.32)
MMT8 *vs.* PGA	0.23	(0.08, 0.38)

In Appendix A we present four extensions of the proposed modelling strategy and compare their performance to the one of the model presented in the previous Section. Appendix B contains the `JAGS` code used to fit the original model.

Discussion

The aim of this work is to jointly model four correlated longitudinal outcomes that measure disease activity in JDM, including clinical markers and symptoms as covariates. To the best of our knowledge, this is the first study to model disease activity in JDM, taking account of (i) longitudinal disease activity; (ii) continuous outcome measures; (ii) four different response variables, measured simultaneously, accounting for correlations among them; and (iv) the particular characteristics of the distribution of the outcome variables due to the fact that many patients attain disease remission at a certain time point.

To achieve this, we propose a statistical approach based on the work of Li et al.[15] and specify a joint random effects distribution for the four outcomes (see equation (3)). Moreover, since most enrolled patients present disease remission at a certain point in time, normal (i.e. non-pathological) values are observed for CMAS, MMT8 and PGA for a large number of follow up visits. Therefore we observe an excess of non-pathological values in the empirical distribution of these variables. To account for this peak in the empirical distribution, we adopt an approach similar to a hurdle model for CMAS, MMT8 and PGA[14].

Our results show that the proposed model leads to accurate prediction of the outcome variables, corresponding well to observed patterns in the data over time. Moreover, even though the credible intervals of the predictions are often large, **the variability associated with the estimates for the regression coefficients is reasonable, as shown by relatively narrow credible intervals.** Indeed, Li et al.[15] demonstrated that their methodology leads to a more efficient estimation of the model parameters compared to a strategy that considers each outcome individually. Moreover the posterior distribution of Ω shows evidence of correlation between the by-subject random intercepts, especially between those for CMAS and MMT8 ($\hat{\rho} = 0.54$, 95% CI 0.42-0.65, Table 2), indicating that those subjects with a higher CMAS tend to have a higher MMT8 as well. This corresponds to existing clinical knowledge.

Our analysis shows that disease activity tends to decrease over time, especially with respect to CMAS, MMT8 and PGA, whereas CK levels in general do not change significantly over time (the change in CK level is estimated to be $e^{0.031} = 1.03$ U/L per year follow up; Figure 4). This is consistent with the clinically observed fact that CK levels tend to be relatively low in patients with JDM[3]. In the described cohort, the median CK level at baseline, shortly after diagnosis, is 103 U/L (Table 1), which is considered a normal value[10]. Nonetheless, patients with myalgia, dysphonia or haematuria, all signs of more severe (systemic) disease, present more elevated CK levels (Figure 4).

Furthermore, we find that different cutaneous manifestations of the disease, most notably periorbital rash, rash on

the trunk, rash over large joints, nail fold changes and facial swelling, are associated with disease activity, especially with PGA, which is the only clinical response taking cutaneous disease activity into account directly. However, some of these skin manifestations, such as periorbital rash and nail fold changes, are associated with the strictly muscular disease activity responses as well. These findings are in line with previous studies, even though based on a different methodology[1;7].

While Ravelli et al.[1] found associations between disease activity and sex, age at onset and dysphonia/dysphagy at onset, this study does not replicate their results (although we find an association between dysphonia at the visit and all four outcome variables). These differences may be explained by differences in ethnic composition of the samples (a mix of European and Latin-American patients for Ravelli et al.[1] while this study recruits only UK patients) and methodology. Specifically, Ravelli et al.[1] analysed each outcome individually using logistic regression at a median of 7.7 years after disease onset, therefore discretizing the response and not accounting for the development over time, while we propose joint modelling of the longitudinal outcome on their original scale. Furthermore, this analysis includes time-varying clinical markers and symptoms determined at each study visit, whereas Ravelli et al.[1] mainly considered baseline variables. It can be hypothesised that the association between current signs and symptoms and disease activity parameters overwhelms the associations between baseline parameters and disease activity at the study visits. Indeed, most baseline variables are not shown to be influential when we perform Bayesian variable selection (Figure 4). This is probably due to the fact that we integrate in the model information from different sources and across time, allowing us to better explain variability in the responses.

Since we did not impute history variables, we have excluded visits with missing values for these variables. Owing to this procedure, 73 patients are excluded altogether from the study. From a probabilistic point of view, patients contributing less visits are less likely to have any visit with completely observed history variables. This, in turn, is more likely to happen in patients who are enrolled in the study not immediately at the time of diagnosis, but later during their course of disease. Consequently, they have lower disease activity at enrolment. These observations explain the differences between included and excluded patients (Table 1). These patients represent a biased subset of the study, often contributing just a few visits, and we believe their exclusion is not problematic. Missing values in non-history variables are imputed in the Bayesian model, by assuming missingness at random and as a result we are able to retain a large number of patients and visits in the model.

In conclusion, the proposed methodology based on joint modelling of four correlated longitudinal measures of disease activity in JDM and exploiting a hurdle model approach, leads to a well-fitting model able to explain observed patterns over time, resulting in clinically meaningful inferences. Various clinical markers and symptoms of JDM are shown to be associated with

disease activity. The clinical implications of these findings will be discussed in a clinical paper.

Appendix A: Additional models

We fit four additional models to the data and their performance is assessed and compared to that of the original model using the deviance information criterion (DIC)[21]. Results are reported in Table 3. Note that the original model makes use of the `JAGS dinterval()` distribution to deal with censored CK values, which has to be removed in order to calculate the DIC. Posterior inference is performed in the R software[13] and `JAGS`, using the package `rjags`[18]. In order to calculate DICs, two chains are required. Each chain is run for 15000 iterations with a burn-in of 5000 and 2000 adaptation iterations. As before, we thin the chain every 10 iterations.

The DIC of the model presented in the paper is 4.1×10^6 (Table 3).

Table 3. Deviance information criterion (DIC) for four extensions of the model. *TVE*: model with time dependent error terms. *CI*: model with imputation based on time from diagnosis and sex. *OM*: original model. *CW*: model that specifies a regression term for the weights of hurdle models for CMAS, MMT8 and PGA. *RIS*: model that includes a random intercept and slope for the coefficient of time from diagnosis for the mean and the weights of all four response variables.

Model	Mean deviance	Penalty	DIC
TVE	2.1×10^4	4.0×10^6	4.0×10^6
CI	2.0×10^4	4.0×10^6	4.0×10^6
OM	2.2×10^4	4.0×10^6	4.1×10^6
CW	2.2×10^4	7.2×10^6	7.2×10^6
RIS	2.2×10^4	13.5×10^6	13.5×10^6

The first additional model we consider contains a time-dependent error term for all four outcomes in the model. Thus, instead of eliciting independent Gamma distributions with hyper-parameters $(0.001, 0.001)$ for the observation precision $\tau^{(k)} = \frac{1}{\sigma_{(k)}^2}$, $k \in \{CK, CMAS, MMT8, PGA\}$, $\sigma_{(k)}^2$ is assumed to follow:

$$\log(\sigma_{(k),ij}^2) = \alpha_\sigma^{(k)} + \beta_\sigma^{(k)} \times t_{ij}$$

where t_{ij} is the time elapsed from diagnosis for individual i at visit j. Independent Normal(0, 1000) priors are specified for the coefficients $\alpha_\sigma^{(k)}$ and $\beta_\sigma^{(k)}$. The DIC of this model is 4.0×10^6 (Table 3). Even though this is slightly lower than the DIC of the original model, we judge this improvement too little to justify the increased complexity of the model with time-dependent error terms. Moreover, visual inspection of the out-of-sample predictions reveal that they do not improve in comparison with the original model (same accuracy and precision).

We also consider a model in which missing values are imputed conditionally on the sex of the patient and the time elapsed since diagnosis. This model, too, yields a slightly improvement in DIC (4.0×10^6, Table 3), but once again the reduction in DIC is judged too small to justify the increased complexity.

Furthermore, we investigate the effect of including covariate information in the model for the weights $w_{ij}^{(k)}, k \in$ {CMAS, MMT8, PGA}. In this case, equation 2 becomes:

$$z_{ij}^{(k)} = \alpha^{(k)} + \eta_i^{(k)} + \gamma^{(k)} \times t_{ij} + X_{ij}\beta_w^{(k)},$$

where we added the term $X_{ij}\beta_w^{(k)}$. The $\beta_w^{(k)}$ are subject to the same Bayesian variable selection procedure as described for the original model. Thus, we define new coefficients $\widetilde{\beta_p} = \delta_p\beta_p$ and specify independent Normal$(0, \frac{1}{\tau_\beta})$ priors for β_p. Here, δ_p and τ_β are defined as previously. The same δ_p is used for all instances of the covariate in the model, i.e. the linear part of CK and the linear parts and weights of CMAS, MMT8 and PGA. The DIC of this model is 7.2×10^6, which is higher than the original model (Table 3).

Finally, we extend the original model by introducing random intercepts and random slopes for the effect of time since diagnosis in all components of the model, i.e. the linear term of the CK response as well as the linear terms and weights of CMAS, MMT8 and PGA models. Due to the significantly larger number of parameters, the DIC is much higher, 13.5×10^6 (Table 3).

Appendix B: `JAGS` code

This appendix includes the `JAGS` code of the model presented in this paper. For the sake of brevity, imputations are removed from the code presented here.

```
data {

mu.ranint[1] <- 0
mu.ranint[2] <- 0
mu.ranint[3] <- 0
mu.ranint[4] <- 0

}

model {

# Imputations omitted

# Random intercepts and random slope
# for time from diagnosis
for (j in 1:J) {
  ranint[j, 1:4] ~ dmnorm(mu.ranint,
                  Omega)
  ranslope.ck[j] ~ dnorm(0,
                  tau.ranslope.ck)
}

# Model for each observational unit
for (i in 1:N) {

  # X[, 1] contains time from diagnosis
  # X[, 2:nvar] all other predictors

  # CK
  mu.ck[i] <- alpha.ck +
              ranint[subj[i], 1] +
              (gamma.ck +
```

```
              ranslope.ck[subj[i]]) *
              X[i, 1] +
              inprod(X[i, 2:nvar],
              beta.ck)
y.ck[i] ~ dnorm(mu.ck[i], tau.e.ck)
censored[i] ~ dinterval(y.ck[i],
cut[i])


# CMAS
probit(p.cmas[i]) <- zeta.cmas +
              ranint[subj[i], 2] +
              theta.cmas * X[i, 1]
ind.cmas[i] ~ dbern(p.cmas[i])

mu.cmas.temp[i] <- alpha.cmas +
              gamma.cmas * X[i, 1] +
              inprod(X[i, 2:nvar],
              beta.cmas)
mu.cmas[i] <- ifelse(ind.cmas[i] == 0,
              sqrt(53),
              mu.cmas.temp[i])

y.cmas[i] ~ dnorm(mu.cmas[i],
        tau.e.cmas[ind.cmas[i] + 1])
        T(0, sqrt(53))


# MMT8
probit(p.mmt[i]) <- zeta.mmt +
              ranint[subj[i], 3] +
              theta.mmt * X[i, 1]
ind.mmt[i] ~ dbern(p.mmt[i])

mu.mmt.temp[i] <- alpha.mmt +
              gamma.mmt * X[i, 1] +
              inprod(X[i, 2:nvar],
              beta.mmt)
mu.mmt[i] <- ifelse(ind.mmt[i] == 0,
              sqrt(80),
              mu.mmt.temp[i])

y.mmt[i] ~ dnorm(mu.mmt[i],
        tau.e.mmt[ind.mmt[i] + 1])
        T(0, sqrt(80))

# PGA
probit(p.pga[i]) <- zeta.pga +
              ranint[subj[i], 4] +
              theta.pga * X[i, 1]
ind.pga[i] ~ dbern(p.pga[i])

mu.pga.temp[i] <- alpha.pga +
              gamma.pga * X[i, 1] +
              inprod(X[i, 2:nvar],
              beta.pga)
mu.pga[i] <- ifelse(ind.pga[i] == 0,
              0,
              mu.pga.temp[i])

y.pga[i] ~ dnorm(mu.pga[i],
        tau.e.pga[ind.pga[i] + 1])
        T(0, sqrt(10))
```

```
}

# Prior for Omega
# I is 4 x 4 identity matrix
Omega ~ dwish(R, 6)
R <- I * 60

# Priors for fixed intercepts
# and slopes
alpha.ck ~ dnorm(0, 0.001)
gamma.ck ~ dnorm(0, 0.001)
alpha.cmas ~ dnorm(0, 0.001)
gamma.cmas ~ dnorm(0, 0.001)
zeta.cmas ~ dnorm(0, 0.001)
theta.cmas ~ dnorm(0, 0.001)
alpha.mmt ~ dnorm(0, 0.001)
gamma.mmt ~ dnorm(0, 0.001)
zeta.mmt ~ dnorm(0, 0.001)
theta.mmt ~ dnorm(0, 0.001)
alpha.pga ~ dnorm(0, 0.001)
gamma.pga ~ dnorm(0, 0.001)
zeta.pga ~ dnorm(0, 0.001)
theta.pga ~ dnorm(0, 0.001)

for (i in 1:(nvar - 1)) {

  # Variable selection. Use same
  # indicator for all four outcomes
  # (i.e. same variables are "in"
  # and "out")

  ind[i] ~ dbern(pind[i])
  pind[i] ~ dbeta(0.1, 0.1)

  betaT.ck[i] ~ dnorm(0, taub)
  betaT.cmas[i] ~ dnorm(0, taub)
  betaT.mmt[i] ~ dnorm(0, taub)
  betaT.pga[i] ~ dnorm(0, taub)

  beta.ck[i] <- ind[i] * betaT.ck[i]
  beta.cmas[i] <- ind[i] * betaT.cmas[i]
  beta.mmt[i] <- ind[i] * betaT.mmt[i]
  beta.pga[i] <- ind[i] * betaT.pga[i]

}

taub ~ dgamma(0.001, 0.001)

# Prior for tau of random slope
tau.ranslope.ck ~ dgamma(0.001, 0.001)

# Prior for error terms
# High precision in case the observation
# is estimated to be in the point mass.
# Non-informative prior otherwise.
tau.e.ck ~ dgamma(0.001, 0.001)
tau.e.cmas[1] <- 10000000
tau.e.cmas[2] ~ dgamma(0.001, 0.001)
tau.e.mmt[1] <- 10000000
tau.e.mmt[2] ~ dgamma(0.001, 0.001)
tau.e.pga[1] <- 10000000
tau.e.pga[2] ~ dgamma(0.001, 0.001)

}
```

References

1. Ravelli A, Trail L, Ferrari C, et al. Long-term outcome and prognostic factors of juvenile idiopathic arthritis: A multinational, multicenter study of 490 patients. *Arthritis Care Res (Hoboken)* 2010; 62: 63–72.

2. Nistala K and Wedderburn LR. Update in juvenile myositis. *Curr Opin Rheumatol* 2013; 25: 742–746.

3. Rider LG, Katz JD and Jones OY. Developments in the classification and treatment of the juvenile idiopathic inflammatory myopathies. *Rheum Dis Clin N Am* 2013; 39: 877–904.

4. Martin N, Krol P, Smith S, et al. A national registry for juvenile dermatomyositis and other paediatric idiopathic inflammatory myopathies: 10 years' experience; the Juvenile Dermatomyositis National (UK and Ireland) Cohort Biomarker Study and Repository for Idiopathic Inflammatory Myopathies. *Rheumatology (Oxford)* 2011; 50: 137–145.

5. Kim S, M El-Hallak M, Dedeoglu F, et al. Complete and sustained remission of juvenile dermatomyositis resulting from aggressive treatment. *Arthritis Rheum* 2009; 60: 1825–1830.

6. Shah M, Mamyrova G, Targoff IN, et al. The clinical phenotypes of the juvenile idiopathic inflammatory myopathies. *Medicine* 2013; 92: 25–41.

7. Stringer E, Singh-Grewal D and Feldman BM. Predicting the course of juvenile dermatomyositis: Significance of early clinical and laboratory features. *Arthritis Rheum* 2008; 58: 3585–3592.

8. Bohan A and Peter JB. Polymyositis and dermatomyositis (first of two parts). *N Engl J Med* 1975; 292: 344–347.

9. Bohan A and Peter JB. Polymyositis and dermatomyositis (second of two parts). *N Engl J Med* 1975; 292: 403–407.

10. Lazarevic D, Pistorio A, Palmisani E, et al. The PRINTO criteria for clinically inactive disease in juvenile dermatomyositis. *Ann Rheum Dis* 2013; 72: 686–693.

11. Lovell D, Lindsley CB, Rennebohm RM, et al. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies. II. The Childhood Myositis Assessment Scale (CMAS): a quantitative tool for the evaluation of muscle function. The Juvenile Dermatomyositis Disease Activity Collaborative Study Group. *Arthritis Rheum* 1999; 42: 2213–2219.

12. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. *J Stat Soft* 2015; 67: 1–48.

13. R Core team. R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria, https://www.R-project.org/ (2015, accessed 2 November 2016).

14. Mullahy J. Specification and testing of some modified count data models. *Journal of econometrics* 1986;33(3): 341–65.

15. Li Q, Pan J and Belcher J. Bayesian inference for joint modelling of longitudinal continuous, binary and ordinal events. *Stat Methods Med Res.* Epub ahead of print 19 November 2014. DOI: 10.1177/0962280214526199.

16. Kuo L and Mallick B. Variable selection for regression models. *Sankhya Ser. B* 1998; 60: 65–81.

17. O'Hara RB and Sillanpää MJ. A review of Bayesian variable selection methods: what, how and which. *Bayesian Anal* 2009; 4: 85–117.

18. Plummer M. rjags: Bayesian graphical models using MCMC. https://CRAN.R-project.org/package=rjags (2016, accessed 2 November 2016).

19. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly weather review* 1950; 1: 1–3.

20. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology* 2010; 21: 128–138.

21. Spiegelhalter DJ, Best N, Carlin B, et al. Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol* 2002; 64: 583–639.