

A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial

Graeme L. Hickey

University of Liverpool, UK

Pete Philipson

Northumbria University, Newcastle-upon-Tyne, UK

and Andrea Jorgensen and Ruwanthi Kolumunnage-Dona

University of Liverpool, UK

[Received August 2016. Final revision November 2017]

Summary. Joint modelling of longitudinal data and competing risks has grown over the past decade. Despite the recent methodological developments, there are still limited options for fitting these models in standard statistical software programs, which prohibits their adoption by applied biostatisticians. We summarize four published models, each of which has software available for model estimation. Each model features a different hazard function, latent association structure between the submodels, estimation approach and software implementation. Of the four models considered here, the model specifications and association structures are substantially different, thus complicating model-to-model comparison. The models are applied to the 'Standard and new anti-epileptic drugs' trial of anti-epileptic drugs to investigate the effect of drug titration on the treatment effects of lamotrigine and carbamazepine on the mode of treatment failure. Notwithstanding the vastly different association structures, we show that the inference from each model is consistent, namely, that there is a beneficial effect of lamotrigine on unacceptable adverse events over carbamazepine and a non-significant effect on the hazard of inadequate seizure control. The association between anti-epileptic drug titration and treatment failure was significant in most models. To allow for the routine adoption of joint modelling of competing risks and longitudinal data in the analysis of clinical data sets, further work is required on the development of model diagnostics to aid model choice.

Keywords: Competing risks; Epilepsy; Joint modelling; Longitudinal analysis; Software

1. Introduction

In many clinical studies, it is common that both longitudinal measurement data and time-to-event data are collected during follow-up (Ibrahim *et al.*, 2010; Asar *et al.*, 2015). When focus lies on the time-to-event process, so-called naive approaches such as Cox regression with a time varying covariate or application of the two-stage model have historically been applied (Cox, 1972; Tsiatis *et al.*, 1995; Kalbfleisch and Prentice, 2002). However, numerous studies have demonstrated that it is preferable to model both outcomes jointly, particularly when the censoring mechanism is informative or the time varying covariate is measured with error, as

Address for correspondence: Ruwanthi Kolumunnage-Dona, Department of Biostatistics, University of Liverpool, Waterhouse Building, 1–5 Brownlow Street, Liverpool, L69 3GL, UK.
E-mail: ruwanthi.kolumunnage-dona@liverpool.ac.uk

© 2018 The Authors Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/18/1811105
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

efficiency is increased and bias decreased (Ibrahim *et al.*, 2010; Chen *et al.*, 2011). Joint models of longitudinal and time-to-event data can also be exploited to improve inference about longitudinal profiles with informative drop-out (Henderson *et al.*, 2000). This has been an active research area for the past two decades. The literature on this topic is extensive, with comprehensive reviews given by Hogan and Laird (1997), Tsiatis and Davidian (2004) and Gould *et al.* (2015). Reviews specific to joint latent class models and multivariate longitudinal data are given by Proust-Lima *et al.* (2012) and Hickey *et al.* (2016) respectively.

Much of the research has been focused on data with a single event time and a single mode of failure, combined with an assumption of independent censoring of event times (Tsiatis and Davidian, 2004). However, in some situations interest lies with more than one possible cause of event or where the censoring is informative. Such data are termed competing risks data (Bakoyannis and Touloumi, 2012). For example, the time to treatment failure (here defined as the withdrawal of a randomized drug or addition of another) has been recommended by the International League Against Epilepsy to be one of the primary end points for clinical trials of anti-epileptic drugs (AEDs) (Commission on Antiepileptic Drugs, 1998). Patients may decide to switch to an alternative AED, or to begin an additional AED, because of inadequate seizure control (ISC). Alternatively, patients may withdraw from a treatment because of an unacceptable adverse effect (UAE). Overall analysis of treatment failure may miss differential effects of AEDs on the reasons for withdrawal, which may differ in terms of their relative importance for patients (Williamson *et al.*, 2007b).

Research into joint modelling of longitudinal data and competing risks time-to-event data has grown over the past few years (Elashoff *et al.*, 2007, 2008; Williamson *et al.*, 2008; Li *et al.*, 2009, 2010, 2012; Hu *et al.*, 2009, 2012; Huang *et al.*, 2010, 2011; Yu and Ghosh, 2010; Deslandes and Chevret, 2010; Rizopoulos, 2012; Gueorguieva *et al.*, 2012; Andrinopoulou *et al.*, 2014, 2017; Ko, 2014; Proust-Lima, 2016, 2017; Blanche *et al.*, 2015). In almost all cases, the standard linear mixed model for the longitudinal outcome submodel was used. Nonetheless, several extensions have been proposed, including robust standard errors (Li *et al.*, 2009; Huang *et al.*, 2010), modelling the random effects heterogeneously (Huang *et al.*, 2010, 2011), modelling of ordinal longitudinal data (Li *et al.*, 2010, 2012; Andrinopoulou *et al.*, 2014, 2017), modelling of multivariate longitudinal data (Li *et al.*, 2012; Andrinopoulou *et al.*, 2014, 2017) and non-parametric modelling of the random effects (Li *et al.*, 2012). The competing risks submodels proposed include a two-part mixture model (Elashoff *et al.*, 2007), the cause-specific hazards model with either unspecified (Elashoff *et al.*, 2008; Williamson *et al.*, 2008), spline (Rizopoulos, 2012; Andrinopoulou *et al.*, 2017) or piecewise constant baseline hazards (Andrinopoulou *et al.*, 2014), the subdistribution hazard model (Deslandes and Chevret, 2010) and a family of parametric cause-specific hazards models (Gueorguieva *et al.*, 2012) that admits the Weibull, log-logistic and an approximate log-normal distribution as special cases (Sparling *et al.*, 2006). The association between submodels has generally been captured through different functions of shared random effects (e.g. Li *et al.* (2012)) or correlated random effects and frailty terms (e.g. Elashoff *et al.* (2008)).

Radically different approaches have also been proposed for joint modelling of longitudinal and competing risks outcomes. Yu and Ghosh (2010) proposed a joint model comprised of a random-change-point submodel with a mixture time-to-event distribution for the competing risks, which also allowed for interval censoring on one of the event times. Proust-Lima *et al.* (2016, 2017) proposed the use of joint latent class models (JLCMs) instead of the classical shared random-effects models. Hu *et al.* (2016) proposed a multiple-imputation technique allowing for longitudinal and time-to-event data to be jointly imputed, thus permitting valid separate model inferences, which in principle is computationally simpler.

Research into joint modelling of longitudinal data and competing risks data has now changed direction away from modelling towards issues such as misspecification of distributional assumptions, model diagnostics and prediction assessment (Li *et al.*, 2012; Ko, 2014; Andrinopoulou *et al.*, 2017; Blanche *et al.*, 2015). However, penetration of this relatively new modelling framework into practical use is precluded by a lack of awareness and understanding between the different existing models that are available and the ability to fit the models by using standard statistical software packages—an issue pertinent to univariate joint modelling as well (Gould *et al.*, 2015). In this paper, we juxtapose four joint models of longitudinal and competing risks outcomes that have been published in recent years. In particular, we examine the submodel specification, latent association structure, model fitting algorithm, parameter uncertainty estimation and software implementation. We focus on these four particular models because each has available software, which enables them to be readily applied to alternative clinical data sets. To the best of our knowledge, this is the first time that multiple joint models for competing risks and longitudinal data have been presented side by side with the objective of enhancing the practical researcher's awareness and understanding between the models and their fitting options. To illustrate each model in turn, we fit them to a data set from one of the largest trials to date comparing AEDs, which was previously analysed by Williamson *et al.* (2008). In keeping with the analysis that was undertaken by Williamson and colleagues, the data set will be analysed with the objective of investigating the effect of drug titration on the relative effects of a standard and new AED on initial treatment failure, with interest lying in the two competing reasons for failure: a UAE or ISC.

The paper is structured as follows. We describe the data in Section 2. In Section 3 we describe the separate models. In Section 4 we describe the four joint models from the literature and fit these to the AED trial data in Section 5. We conclude with a discussion in Section 6.

2. Application: epilepsy drug trial data

The 'Standard and new anti-epileptic drugs' (SANAD) study was a non-blinded randomized controlled trial recruiting patients with epilepsy to test AEDs (Marson *et al.*, 2007). This study was registered as an international standard randomized controlled trial, number ISRCTN383 54748 (<http://www.isrctn.com/>; digital object identifier 10.1186/ISRCTN38354748). Patients were randomized to either the standard treatment (carbamazepine, CBZ) or to lamotrigine, LTG, gabapentin, oxcarbazepine or topiramate. Time to treatment failure was a primary outcome of the trial. For the study here, we consider only patients who were assigned to either CBZ or LTG. An important secondary objective was the analysis of competing risks for treatment failure: treatment failure due to either ISC or UAE. The primary analysis of patients with partial epilepsy concluded that the newer drug LTG was superior in terms of treatment failure to CBZ, which had been the standard for many years (Williamson *et al.*, 2007a). LTG was significantly better in terms of withdrawal for a UAE and not significantly worse in terms of withdrawal for ISC. Following publication of the trial results, it became evident that there was concern that differential titration rates may have been to the disadvantage of CBZ (Cross *et al.*, 2007). An AED that is titrated more quickly may bring benefits in terms of seizure control but be more likely to cause adverse effects.

The data set that is analysed here includes 605 patients, comparing CBZ ($n = 292$) and LTG ($n = 313$). This data set was first analysed by using a joint model framework by Williamson *et al.* (2008) and is publicly available in the R package `joiner` (Philipson *et al.*, 2017). During a maximum follow-up time of 6.6 years (median = 2.9 years), 94 patients withdrew from the randomized drug due to a UAE whereas 120 withdrew due to ISC. Withdrawals due to other

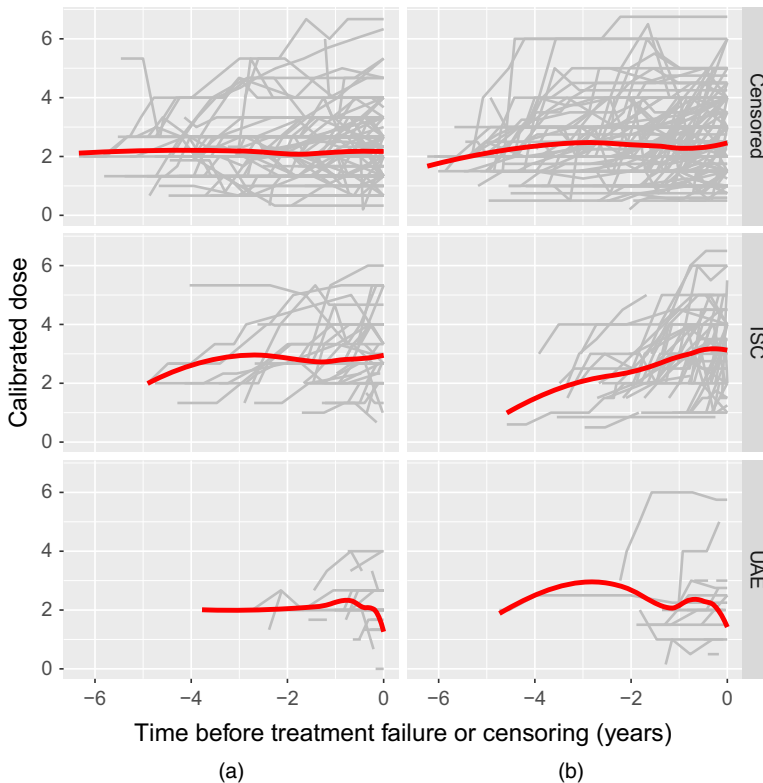


Fig. 1. Profile plots for calibrated dose against *reverse* time stratified by drug type and event time mechanism (the origin of the time axis is the treatment failure or censoring time; the bold curves are fitted LOESS curves): (a) CBZ; (b) LTG

reasons were treated as non-informative, and patients were censored at this time. To compare the two AEDs after adjusting for titration rate, dose calibration was first undertaken by standardizing the dose of both drugs relative to the midpoint of its maintenance dose range (Faught, 2007). The maintenance dose recommended in the SANAD trial was independently deemed to be reasonable and the approach to calibration considered sensible. These calibrated doses are taken to be the longitudinal measurements within the competing risks joint model. Patients had an average of 4.6 longitudinal measurements recorded, ranging from 1 to 15.

Fig. 1 shows the subject-specific longitudinal profiles for calibrated dose of drugs against *reverse* time, stratified by drug type and event time mechanism. A large variation across patients in the initial calibrated dose was observed. Additionally, treatment failure due to ISC was seen to occur in patients with increasing calibrated doses, whereas treatment failure due to a UAE seemed to happen following a dose reduction.

3. Separate submodels

We consider a repeated continuous longitudinal outcome $y_i(t)$ measured at time t for subject $i = 1, \dots, m$, where m is the total number of subjects in the study, and $t = 0$ denotes the time of randomization. These measurements are obtained at specific times t_{ij} which can vary between subjects, yielding data $y_i = \{y_i(t_{ij}), j = 1, 2, \dots, n_i\}$. We let T_i denote the observed failure time for subject i , taken as $T_i = \min(T_{1i}^*, T_{2i}^*, \dots, T_{Gi}^*, C_i)$, with T_{gi}^* denoting the true failure time of

subject i for each event type $g = 1, 2, \dots, G$, and C_i the censoring time. In addition to observing T_i , we also observe the event indicator δ_i , which equals 0 if censored, or g indicating that subject i failed from cause g . We assume that the censoring mechanism is independent of the event time.

3.1. Longitudinal submodel

The standard model for a continuous longitudinal outcome is the Laird and Ware (1982) linear mixed model

$$y_i(t_{ij}) = \mu_i(t_{ij}) + \varepsilon_{ij}, \quad (1)$$

where $\mu_i(t_{ij})$ is the linear predictor term (or mean response):

$$\mu_i(t_{ij}) = X_i^{(1)}(t_{ij})^T \beta^{(1)} + Z_i(t_{ij})^T b_i \quad (2)$$

and $\varepsilon_{ij} \sim N(0, \sigma^2)$ are the measurement errors, which are assumed to be independently and identically distributed. We let $X_i^{(1)}(t_{ij})$ and $Z_i(t_{ij})$ be time varying vectors of covariates for subject i associated with fixed, $\beta^{(1)}$, and subject-specific random, b_i , effects respectively. The standard model assumes that random effects are distributed as multivariate normal with mean 0 and covariance matrix Σ . Extensions to include a separate Gaussian process have been proposed; for example, Henderson *et al.* (2000) proposed a stationary Gaussian process, and Proust-Lima *et al.* (2016) proposed including either a Brownian motion or auto-regressive process. Additionally, we assume that b_i and ε_{ij} are independent.

3.2. Competing risks submodel

The standard approach of modelling competing risks data includes cause-specific hazards (Putter *et al.*, 2007), in which, for subject i , the instantaneous rate for failure of type g at time t is given by

$$\begin{aligned} \lambda_{ig}(t) &= \lim_{\delta t \rightarrow 0} \frac{P\{t \leq T_i < t + \delta t, \delta_i = g | T_i \geq t, X_i^{(2)}(t)\}}{\delta t} \\ &= \lambda_{0g}(t) \exp\{X_i^{(2)}(t)^T \beta_g^{(2)}\}, \end{aligned}$$

where $X_i^{(2)}(t)$ is a (possibly time varying) vector of covariates, and $\beta_g^{(2)}$ is a vector of fixed effect parameters for the g th event. The baseline hazard function $\lambda_{0g}(t)$ is cause specific and can either be left unspecified or modelled parametrically. In the extension to joint modelling, we consider a model of the form

$$\lambda_{ig}(t) = \lambda_{0g}(t) \exp\{X_i^{(2)T} \beta_g^{(2)} + W_{ig}(t)\}, \quad (3)$$

where $W_{ig}(t)$ denotes a mean 0 latent term and $X_i^{(2)}$ is a vector of covariates known at baseline.

4. Joint models

The basic premise of joint models is that the separate longitudinal and time-to-event models that are described in equations (1) and (3) are correlated. The four models that we consider in this paper are those of Williamson *et al.* (2008) (model 1), Elashoff *et al.* (2008) (model 2), Rizopoulos (2012) (model 3) and Proust-Lima *et al.* (2016) (model 4). In all models, except model 4, we assume the same submodel for the longitudinal data. Therefore, most of the joint models that are described here can be distinguished in terms of

- (a) the specification of the baseline hazard functions $\lambda_{0g}(t)$,
- (b) specification of the latent processes $W_{ig}(t)$,
- (c) the estimation approach (including parameter uncertainty estimation) and
- (d) software implementation.

A summary of the various model properties is given in Table 1.

Table 1. Summary of the models compared†

Model	Baseline hazard	Latent association
1	Non-parametric (unspecified)	Current value of latent process parameterization $W_{ig}(t) = \alpha_g Z_i(t)^T b_i$
2	Non-parametric (unspecified)	Correlated random-effects parameterization $W_{ig}(t) = \alpha_g \theta_i$, where $\begin{pmatrix} b_i \\ \theta_i \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{bb} & \Sigma_{b\theta} \\ \Sigma_{b\theta}^T & \Sigma_{\theta\theta} \end{pmatrix} \right\}$ For identifiability, we fix $\alpha_1 = 1$
3a	B-spline basis (on log-hazard scale) basis for cubic splines	Current value parameterization $W_{ig}(t) = \alpha_g \mu_i(t)$
3b		Time-dependent slopes parameterization $W_{ig}(t) = \alpha_g^{(1)} \mu_i(t) + \alpha_g^{(2)} \frac{d}{dt} \mu_i(t)$
3c		Lagged effects parameterization $W_{ig}(t) = \alpha_g \mu_i(\max\{t - c, 0\})$ c is a known (specified) lag constant
3d		Cumulative effects parameterization $W_{ig}(t) = \alpha_g \int_0^t \mu_i(s) ds$
3e		Weighted cumulative effects parameterization $W_{ig}(t) = \alpha_g \int_0^t w(t - s) \mu_i(s) ds$ $w(\cdot)$ is a known (specified) weighting function
3f		Special case of the random-effects parameterization (with fixed component) $W_{ig}(t) = \alpha_g (\beta_1^{(1)} + b_{i1})$ $\beta_1^{(1)}$ and b_{i1} denote the fixed and random-effect coefficients associated with the linear time covariate in a random-intercepts and random-slopes linear mixed model
4a	Weibull	Association between submodels accounted entirely for by latent classes
4b	Piecewise constant	Each subject i belongs to a single latent class, $a_i \in \{1, 2, \dots, R\}$, with submodels $\mu_i(t a_i = r) = \tilde{X}_i^{(1)}(t)^T \beta^{(1)} + \tilde{X}_i^{(1)}(t)^T \gamma_r^{(1)} + Z_i(t)^T b_{ir}$, $\lambda_{ig}(t a_i = r) = \lambda_{0g}(t a_i = r) \exp(\tilde{X}_i^{(2)T} \beta_g^{(2)} + \tilde{X}_i^{(2)T} \gamma_{gr}^{(2)})$,
4c	Cubic M -splines	where the covariate vector $X_i^{(1)}(t)$ is decomposed into two pieces: $\tilde{X}_i^{(1)}(t)$ and $\tilde{X}_i^{(1)}(t)$ associated with common $\beta^{(1)}$ and class-specific $\gamma_r^{(1)}$ fixed effects respectively, and similarly for $X_i^{(2)}$; the random effects in the longitudinal submodel have distribution $b_{ir} \sim N(0, \Sigma_r)$

†Notation: for models 1–3, α_g , $\alpha_g^{(1)}$ and $\alpha_g^{(2)}$ denote scalar association coefficient parameters corresponding to cause g .

4.1. Baseline hazard functions

Models 1 and 2 leave the baseline hazard functions completely unspecified, opting for a semi-parametric joint model, which is consistent with the classical Wulfsohn and Tsiatis (1997) joint model approach. Models 3 and 4 propose spline models for the baseline hazard, namely B -splines (on log-hazard) and either M -splines or piecewise constant (on the untransformed hazard scale). Provided that a sufficient number of knots are specified, each model in principle offers good flexibility to capture a wide range of hazard shapes (Rutherford *et al.*, 2015). Model 4 also considers a third specification for the hazard: the Weibull distribution. Although less flexible than the semiparametric and spline models, it can potentially reduce the computational time to fit the model. See Table 1 for a summary of the model specifications.

4.2. Latent association structure

Different characteristics of the longitudinal profile may influence the cause-specific hazards submodel. Within the joint model literature, several association structures have been proposed, but little attention has been given to the challenge of choosing the most appropriate structure for a given data set (Andrinopoulou and Rizopoulos, 2016). Among the models that are examined here, various latent association structures have been considered, including several alternatives for some models (Table 1). Model 1 assumes that $W_{ig}(t)$ is proportional to the latent term in the linear mixed model, $Z_i(t)^T b_i$, which extends the model that was proposed by Henderson *et al.* (2000) (Table 1, model 1). This is a form of the current values parameterization, whereby the ‘value’ is the subject-specific deviation from the average trajectory. Model 2 applies $W_{ig}(t) = \alpha_g \theta_i$ where θ_i is a zero-mean subject-specific random effect (Table 1, model 2). The correlation between the submodels is induced by assuming that $(\theta_i^T, \theta_i)^T$ are correlated. Unlike model 1, where association between submodels is tested on the basis entirely of the hypothesis of $\alpha_g = 0$, the null hypothesis of $\Sigma_{b\theta} = 0$ provides a useful assessment of association between outcomes for model 2.

Model 3 (Rizopoulos, 2012) has been developed with several useful structures, including the current value parameterization (Table 1, model 3a), the time-dependent slopes parameterization (Table 1, model 3b), lagged effects parameterization (Table 1, model 3c), cumulative effects parameterization (Table 1, model 3d), weighted cumulative effects parameterization (Table 1, model 3e) and a special case of the random-effects parameterization (Table 1, model 3f).

Model 4 is a JLCM (Proust-Lima *et al.*, 2016), which is fundamentally different from models 1–3. This model assumes that each subject i belongs to a single latent class, $a_i \in \{1, 2, \dots, R\}$. It is assumed that subjects within the same latent class share the same mean longitudinal trajectory and hazard risk (Proust-Lima *et al.*, 2012). Hence, the linear mixed model is class specific and the baseline hazard function is assumed to be class and cause specific. From this model, the association between the longitudinal and time-to-event submodels is entirely captured by the latent classes. Each class membership probability is modelled by a multinomial logit:

$$P(a_i = r) = \frac{\exp(X_i^{(3)T} \varphi_r)}{1 + \sum_{l=1}^{R-1} \exp(X_i^{(3)T} \varphi_l)},$$

where $X_i^{(3)}$ is a vector of baseline covariates corresponding to class-specific coefficient parameters, φ_r .

4.3. Estimation

All models here are fitted by maximum likelihood estimation using different algorithms. Different approaches to standard error estimation are also taken. Details about the estimation

algorithms including numerical quadrature approaches are given in the on-line supplementary data Table S1.

4.4. Software

Models 1 and 2 are available in the form of scripts for the languages R and C respectively (see Section 8 for details). Models 3 and 4 can be fitted by using R packages. Details on the general software implementation are given in the on-line Table S1. User instructions are provided with the source code for models 2–4. We have also integrated the scripts for model 1 into the R package `joiner` (Philipson *et al.*, 2017) including a complete manual. It is beyond the scope of this paper to detail every step of the software implementation including data preprocessing; however, as a demonstration, we have made software code that was used to fit the models to SANAD data available on line (see the data sharing details in Section 8).

5. Application of models to epilepsy drug trial data

To allow for comparison, we assumed the same functional form for each model as that in the original analysis undertaken by Williamson *et al.* (2008). Namely, the following linear mixed model for the longitudinal measurements submodel was assumed:

$$y_i(t_{ij}) = \beta_0^{(1)} + b_{i0} + (\beta_1^{(1)} + b_{i1})t_{ij} + \beta_2^{(1)}X_i + \beta_3^{(1)}X_it_{ij} + \varepsilon_{ij}, \quad (4)$$

with X_i denoting a binary time-independent treatment effect ($X_i = 1$ if patient i was randomized to LTG, and 0 if randomized to CBZ), and ε_{ij} and $(b_{i0}, b_{i1})^T$ are distributed $N(0, \sigma_\varepsilon^2)$ and $N_2(0, \Sigma)$ respectively. Although Williamson *et al.* (2008) also considered a piecewise linear spline model, for our comparison, we focused only on the above simple linear model. They further specified a cause-specific hazard model for the competing risks data by

$$\lambda_{ig}(t) = \lambda_{0g}(t) \exp\{X_i\beta_g^{(2)} + W_{ig}(t)\},$$

where $(\beta_{ISC}^{(2)}, \beta_{UAE}^{(2)})$ are the cause-specific treatment effects for the AED. For fitting each model, we used time measured in units of years.

All models were fitted by using the same computing environment with predominantly default settings, except where reported otherwise. Technical details for implementation of each model are given in the on-line supplementary data.

5.1. Interpretation of the treatment effects

The direct treatment effect on UAEs, $\beta_{UAE}^{(2)}$, was statistically significant for all joint models considered (Table 2). The direct treatment effect on ISC, $\beta_{ISC}^{(2)}$, was non-significant for all models except model 3b. Hence, the results from all models (with the single exception of model 3b, which we discuss below) suggest that, if LTG is titrated at the same rate as CBZ, the beneficial effect of LTG on a UAE would still be evident, and there remains no evidence of a difference in seizure control between the two drugs. This conclusion is consistent with the original trial analysis (Williamson *et al.*, 2007a), which we demonstrate by fitting the separate models for longitudinal outcome (4) and competing risks ((3), without adjustment for titration) (Table 2, model: separate).

The estimated treatment effect on calibrated dose, $\beta_2^{(1)}$, from models 1, 2 and 4a, indicates that patients who were assigned to LTG were titrated significantly more slowly than those allocated to CBZ. The estimated effects from model 3 were generally closer to the separate linear mixed

Table 2. Treatment effect and association parameter estimates for the competing risks submodel†

Model	Results for ISC		Results for UAE		Computation time
	Treatment (LTG), $\beta_{ISC}^{(2)}$ (95% confidence interval)	Association, α_{ISC} (95% confidence interval)‡	Treatment (LTG), $\beta_{UAE}^{(2)}$ (95% confidence interval)	Association, α_{UAE} (95% confidence interval)‡	
Separate	0.015 (−0.344, 0.374)	Not applicable	−0.608 (−1.102, −0.192)	Not applicable	< 1 s
1	0.028 (−0.329, 0.366)	0.590 (0.425, 0.768)	−0.660 (−1.090, −0.221)	−0.925 (−1.378, −0.519)	17 s (maximum likelihood estimates) 45 min (standard errors)
2	−0.306 (−0.744, 0.131)	−1.502 (−1.941, −1.062)	−0.543 (−0.997, −0.089)	1.000 Reference	5 h 22 min
3a	−0.119 (−0.482, 0.244)	0.598 (0.448, 0.747)	−0.625 (−1.044, −0.207)	−0.926 (−1.246, −0.607)	54 s
3b	−0.592 (−1.036, −0.148)	0.120 (current value) (−0.138, 0.377) 2.334 (slope) (1.360, 3.308)	−1.212 (−1.832, −0.593)	−1.239 (current value) (−1.642, −0.836) 2.724 (slope) (1.002, 4.447)	52 s
3c	−0.055 (−0.417, 0.306)	0.591 (0.426, 0.756)	−0.696 (−1.118, −0.274)	−1.016 (−1.347, −0.684)	52 s
3d	−0.035 (−0.395, 0.326)	0.212 (0.133, 0.291)	−0.612 (−1.027, −0.196)	−0.156 (−0.381, 0.070)	56 s
3e	−0.074 (−0.436, 0.288)	1.495 (1.095, 1.895)	−0.613 (−1.029, −0.196)	−0.869 (−1.848, 0.110)	51 s
3f	−0.090 (−0.497, 0.317)	2.619 (2.027, 3.212)	−0.868 (−1.446, −0.290)	−8.558 (−10.143, −6.972)	53 s
4a	−0.366 (−0.866, 0.134)	Not applicable	−0.876 (−1.391, −0.360)	Not applicable	3 min 34 s§
4b	−0.142 (−0.597, 0.314)	Not applicable	−0.693 (−1.178, −0.207)	Not applicable	2 min 50 s§

†Additional parameters (where relevant) are reported in the main text.

‡The association parameters, although capturing the strength of association between the submodels, have widely different interpretations between models and so should not be directly contrasted. See the main text for an explanation of each model. Note also that models 3b and 4 have two association parameters per cause.

§Time does not take into account the grid search algorithm or fitting times of models with other numbers of classes.

model submodel than those of models 1, 2 and 4 (Table 3). Treatment-and-time interaction effects $\beta_3^{(1)}$ were broadly similar for models 1 and 3, but attenuated towards 0 for model 2. The estimated fixed effects for time were also similar, confirming that the average calibrated dose was increased over the study period with the exception of model 3b, which was larger. Model 4 was fitted to allow for class-specific intercept, time and treatment-and-time interaction terms; hence, the estimates are not comparable with the average effects of the other models. We note that the parameter estimates from the longitudinal submodel were not of primary interest in the SANAD clinical trial; however, this example has highlighted that, among the joint models fitted, the association structures and method of estimation and software used to fit the joint models can have different influences on the longitudinal submodel estimates.

Table 3. Parameter estimates and 95% confidence intervals (CIs) for the longitudinal submodel and the random-effects variance–covariance matrix

Model	Intercept, $\beta_0^{(1)}$ (95% CI)	Time (years), $\beta_1^{(1)}$ (95% CI)	Treatment (LTG), $\beta_2^{(1)}$ (95% CI)	Time:treatment, $\beta_3^{(1)}$ (95% CI)	Error, σ_ε (95% CI)	Random-effects covariance matrix		
						Σ_{11} (95% CI)	Σ_{22} (95% CI)	Σ_{12} (95% CI)
Separate	1.932 (1.826, 2.039)	0.151 (0.085, 0.217)	−0.087 (−0.236, 0.061)	0.214 (0.124, 0.303)	0.446 (0.392, 0.497)	0.711 (0.615, 0.859)	0.150 (0.105, 0.231)	0.062 (0.022, 0.127)
1	1.973 (1.871, 2.084)	0.129 (0.080, 0.192)	−0.139 (−0.286, −0.001)	0.226 (0.126, 0.324)	0.443 (0.392, 0.497)	0.732 (0.615, 0.859)	0.162 (0.105, 0.231)	0.068 (0.022, 0.127)
2	1.927 (1.827, 2.027)	0.149 (0.092, 0.205)	−0.142 (−0.276, −0.007)	0.123 (0.054, 0.191)	0.429 (0.424, 0.434)	0.660 (0.579, 0.741)	0.223 (0.175, 0.270)	0.120 (0.078, 0.161)
3a	1.929 (1.822, 2.036)	0.141 (0.073, 0.209)	−0.091 (−0.239, 0.058)	0.227 (0.135, 0.319)	0.444 (0.442, 0.446)	0.715 (0.615, 0.859)	0.160 (0.105, 0.231)	0.067 (0.022, 0.127)
3b	1.878 (1.779, 1.977)	0.226 (0.184, 0.268)	−0.073 (−0.211, 0.064)	0.197 (0.139, 0.254)	0.442 (0.445, 0.446)	0.737 (0.615, 0.859)	0.168 (0.105, 0.231)	0.022 (0.022, 0.127)
3c	1.927 (1.82, 2.034)	0.146 (0.079, 0.213)	−0.087 (−0.235, 0.062)	0.220 (0.129, 0.311)	0.445 (0.446, 0.446)	0.717 (0.615, 0.859)	0.154 (0.105, 0.231)	0.066 (0.022, 0.127)
3d	1.932 (1.825, 2.038)	0.151 (0.085, 0.217)	−0.088 (−0.236, 0.059)	0.215 (0.126, 0.305)	0.446 (0.445, 0.445)	0.707 (0.615, 0.859)	0.149 (0.105, 0.231)	0.064 (0.022, 0.127)
3e	1.930 (1.823, 2.036)	0.153 (0.087, 0.220)	−0.089 (−0.237, 0.059)	0.218 (0.127, 0.309)	0.445 (0.447, 0.447)	0.707 (0.615, 0.859)	0.153 (0.105, 0.231)	0.063 (0.022, 0.127)
3f	1.967 (1.867, 2.068)	0.111 (0.053, 0.169)	−0.126 (−0.262, 0.011)	0.217 (0.145, 0.289)	0.447 (0.432, 0.462)	0.684 (0.579, 0.741)	0.200 (0.175, 0.270)	0.117 (0.078, 0.161)
4a†	1.771 (r=1) (1.569, 1.946) (0.241 (r=2) (−0.172, 0.654) (1.886 (r=3) (1.785, 1.988) (2.209 (r=4) (1.927, 2.492) (3.896 (r=5) (3.526, 4.267)	0.619 (r=1) (0.384, 0.854) (−0.454 (r=2) (−2.799, 1.891) (0.056 (r=3) (0.007, 0.105) (1.682 (r=4) (1.247, 2.116) (0.103 (r=5) (−0.098, 0.305)	−0.157 (−0.275 −0.039) (−0.262, 0.011) (−0.172, 0.654) (1.886 (r=3) (1.785, 1.988) (2.209 (r=4) (1.927, 2.492) (3.896 (r=5) (3.526, 4.267)	0.326 (0.072, 0.569) (0.145, 0.289) (−0.019, 0.114) (0.746 (0.222, 1.270) (0.013, 0.469)	0.432 (0.418, 0.446)	0.284 (0.418, 0.446)	0.031 (0.418, 0.446)	0.032 (0.418, 0.446)

(continued)

Table 3 (continued)

Model	Intercept, $\beta_0^{(1)}$ (95% CI)	Time (years), $\beta_1^{(1)}$ (95% CI)	Treatment (LTG), $\beta_2^{(1)}$ (95% CI)	Time:treatment, $\beta_3^{(1)}$ (95% CI)	Error, σ_ε (95% CI)	Random-effects covariance matrix		
						Σ_{11} (95% CI)	Σ_{22} (95% CI)	Σ_{12} (95% CI)
4b†	0.253 ($r=1$)	−0.492 ($r=1$)	−0.164	1.237	0.432	0.286	0.028	0.027
	−0.132, 0.638)	(−2.832, 1.849)	(−0.283, 0.045)	(−0.961, 3.435)	(0.418, 0.446)	Not estimated‡	Not estimated‡	Not estimated‡
	1.849 ($r=2$)	0.516 ($r=2$)		0.432				
	(1.658, 2.039)	(0.339, 0.693)		(0.228, 0.637)				
	1.887 ($r=3$)	0.038 ($r=3$)		0.065				
	(1.787, 1.988)	(−0.012, 0.087)		(−0.002, 0.132)				
	2.212 ($r=4$)	1.673 ($r=4$)		0.761				
	(1.927, 2.497)	(1.234, 2.112)		(0.231, 1.291)				
	3.871 ($r=5$)	0.106 ($r=5$)		0.265				
	(3.490, 4.252)	(−0.094, 0.306)		(0.024, 0.507)				

† r denotes the class.
‡CIs are reported for the Cholesky decomposition of the variance–covariance matrix, rather than directly for the covariance parameters. We have not attempted to back-transform these parameters.

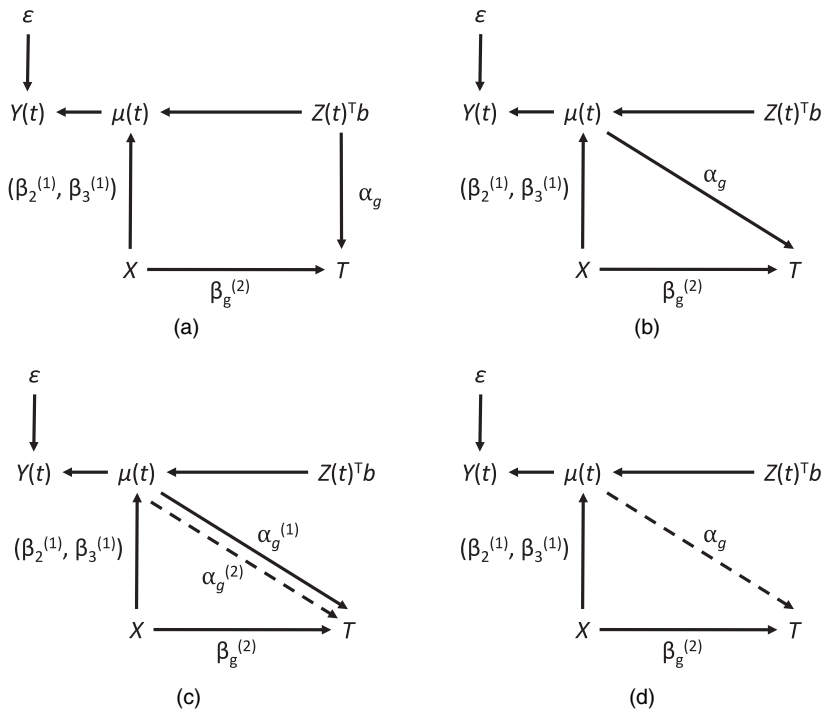


Fig. 2. Causal diagrams for models 1 and 3 ($Y(t)$, observed longitudinal data; $\mu(t)$, mean trajectory function; ε , error; T , failure time; X , treatment; $Z(t)^T b$, mean 0 latent process; $(\beta_2^{(1)}, \beta_3^{(1)})$, treatment effects on longitudinal processes; α_g , effect of longitudinal process on treatment failure (superscripts denote multiple effects); $\beta_g^{(2)}$, treatment effect on failure time; $-\rightarrow$, effect via a transformation or operator function; for model 3c, the log-hazard of treatment failure is associated with a lagged value of the mean trajectory function in addition to the direct treatment effect; for models 3b, 3d and 3e, the log-hazard of treatment failure is associated with a linear sum of $\mu(t)$ and the first-order derivate or (weighted) integral of $\mu(t)$, in addition to the direct treatment effect; for model 3f, the log-hazard depends on the subject-specific random slope for time in addition to the direct treatment effect; in models 1 and 3f, the treatment effect is both direct and overall): (a) model 1; (b) models 3a and 3c; (c) models 3b, 3d and 3e; (d) model 3f

5.2. Interpretation of latent association parameters

A summary of the time-to-event submodel parameter estimates is given in Table 2. We strongly emphasize that caution is required when contrasting treatment and association parameter effects *between* models as each has a different latent association structure, despite the shared common notation (for example, see Fig. 2 for a contrast of models 1 and 3). We describe the interpretation for each model in turn below.

5.2.1. Model 1

The parameter estimates that are associated with changes in $W_{ig}(t) = b_{i0} + b_{i1}t$ were significant for both events, leading to the inference that calibrated drug dose is associated with time to treatment failure for both failure reasons. Moreover, the parameters had opposite signs. The clinical interpretation, as discussed in Williamson *et al.* (2008), was that patients on higher doses may be more likely to withdraw from treatment because of ISC since the reason that they are on higher doses is that they are having continued seizures. Also, patients on higher doses may be less likely to be withdrawn from treatment because of a UAE since the reason that a required dose increase is possible is because no such events have occurred.

5.2.2. Model 2

The shared random effect (frailty term) in the competing risks submodel had an estimated variance $\sigma_\theta^2 = 0.791$ (95% confidence interval (CI), from 0.571 to 1.010). The two covariance terms between longitudinal submodel random effects and the frailty term, $\Sigma_{b\theta}^T$, were both negative: -0.093 (95% CI, from -0.187 to 0.001) and -0.402 (95% CI, from -0.495 to -0.310) respectively, implying that an increase in calibrated dose relative to the population average (which corresponds to an increase in the components of b_i) is associated with a decrease in frailty θ_i . The coefficient α_{ISC} multiplies the frailty for the event due to ISC relative to a UAE (and α_{UAE} multiplies the frailty for the event due to a UAE, but is constrained to 1 for identifiability). As $\hat{\alpha}_{ISC}$ was negative, when interpreted alongside the negative correlation between b_i and θ_i , it implies that an increase in calibrated dose is associated with a decrease in risk of a UAE (due to decreasing frailty), and an increase in risk of ISC. This is consistent with the original analysis conclusions.

5.2.3. Model 3a

Model 3a assumes that the hazard is associated with the current expected value of the calibrated drug dose at time t , i.e. $\beta_0^{(1)} + b_{i0} + (\beta_1^{(1)} + b_{i1})t + \beta_2^{(1)}X_i + \beta_3^{(1)}X_it$, whereas model 1 assumed that the association depended on only the random subject-specific deviation from the current expected value, i.e. $b_{i0} + b_{i1}t$. Hence, the overall treatment effect on the event hazard is decomposed into the direct effect $\beta_g^{(2)}$ and the indirect effect $\alpha_g(\beta_2^{(1)} + \beta_3^{(1)}t)$. Therefore, the direct treatment effect must be interpreted by also adjusting for treatment-specific intercept and slope of dose titration in the hazard model.

5.2.4. Model 3b

Rapid changes in titration might be indicative of risk of treatment failure, i.e. any two patients having the same mean value at a given time t might have very different slopes if their random intercepts also vary. Therefore, a separate term that captures the changes in slope might be required. Contrary to the other models and to our clinical understanding, this results in a statistically significant treatment effect for failure due to ISC. The model assumes that the instantaneous hazard of dropout is associated with a linear combination of current mean value and its first-order derivative, namely $W_{ig}(t) = \alpha_g^{(1)}\mu_i(t) + \alpha_g^{(2)}\mu'_i(t)$, where $\mu'_i(t) = \beta_1^{(1)} + b_{i1} + \beta_3^{(1)}X_i$, and $(\alpha_g^{(1)}, \alpha_g^{(2)})$ are a pair of association parameters corresponding to the current value and first-derivate terms for cause g . Expanding $W_{ig}(t)$, we can decompose the treatment effects (Fig. 2) in the cause-specific hazards submodel into a direct effect $\beta_g^{(2)}$ and an indirect effect,

$$\alpha_g^{(2)}\beta_3^{(1)} + \alpha_g^{(1)}(\beta_2^{(1)} + \beta_3^{(1)}t).$$

Hence, the disparity of direct treatment effect from other models is purely an artefact caused by indirect treatment effect adjustments.

5.2.5. Model 3c

The risk of treatment failure might depend, not on the current value of the marker, but on a historical calibrated dose, e.g. if the AED takes a long period to have a physiological effect on patients. In this case, we might consider the calibrated dose at a previous time point, $t - c$ (for some choice of c , taken to be $c = 6$ months in this example for illustration). Inferences remained consistent with the non-lagged model, model 3a.

5.2.6. Models 3d and 3e

Expanding on the clinical rationale for models 3a and 3c, similar arguments might lead us to hypothesize that the complete history of the calibrated dose up to time t affects the risk of treatment failure, where for example pharmacokinetic–pharmacodynamic properties can lead to cumulative toxicity effects. Hence, α_g captures the association between the risk of event and the (weighted) cumulative effect up to time t . The choice of weight function will be application specific. Here, we use a Gaussian density function, as in Rizopoulos (2012), which gives less weight to points the further away from the current time they are. For both models, cumulative (calibrated) drug dose was significantly associated with ISC, which is consistent with the other models. However, cumulative (calibrated) drug dose was not significantly associated with a UAE, although the direction of effects remained consistent.

5.2.7. Model 3f

A simple time-independent association is specified, whereby the risk of treatment failure is associated with the subject-specific random slope for t . This is a special case of random-effects parameterization. Both latent association parameters were significantly different from 0 and in the same direction as for other models; for example model 3a indicates that patients who are titrated more quickly are at increased risk of ISC but reduced risk of a UAE. However, the direct treatment effect for treatment failure also coincides with the overall treatment effect in this model. It was also observed that the latent association parameters were much larger than for other models. This is explained because the variances of the random effects are relatively small, here 0.684 and 0.200 respectively, and hence a unit change in slope is quite extreme. Rizopoulos (2012), page 114, therefore recommended scaling $W_{ig}(t)$ by this standard deviation to facilitate interpretation. However, to avoid confusion in the comparison between models, we report unscaled estimates.

5.2.8. Models 4a–4c

There are no parameter estimates for inference about the degree of association between the submodels. Instead, the patients are assumed to be members of one of five latent classes. For example, model 4a comprised five classes with 22.8%, 6.6%, 58.3%, 7.4% and 4.8% of subjects allocated, with good discrimination (the mean of the posterior probabilities in each class ranged from 76% (class 1) to 93% (class 2)). The treatment effects were modelled as cause specific only (i.e. independent of class), which led to consistent inferences with the other models. Figs S1 and S2 in the on-line supplementary material summarize the fitted submodels for each class. Model 4b also yielded similar effect sizes, albeit smaller in absolute magnitude. However, model 4b offered increased flexibility in the baseline hazard function compared with model 4a, which was constrained to follow a Weibull model. We were unable to fit model 4c because of lack of convergence.

5.3. Model assessment and comparison

Software implementations of models 1 and 2 do not incorporate any features that facilitate model assessment or comparison. However, we have recently updated the R package `joiner` (Philipson *et al.*, 2017) to report the log-likelihood for model 1.

The software implementation of model 3 reports several statistics that are useful for model comparison, including the log-likelihood at the maximum likelihood estimate, the Akaike information criterion (AIC) and the Bayes information criterion (BIC). For models 3a–3f, the AIC

values were 7099.34, 7061.75, 7106.35, 7163.33, 7139.52 and 7029.09 respectively, thus suggesting that model 3f is the preferred model. The same conclusion is obtained when contrasting the BIC for each model (7161.25 for model 3f), and the log-likelihood (−3484.55). In addition, several residual diagnostics for the fitted longitudinal data submodel can be readily viewed from a fitted model (e.g. Fig. S3 in the on-line supplementary material). Although the residual plots indicated some systematic trend, the non-random dropout mechanism can mean that such plots are potentially misleading (Rizopoulos (2012), page 157). However, the solution that was proposed by Rizopoulos *et al.* (2010) to overcome this—multiple-imputation residuals—is not available in the software for competing risks models. Moreover, residuals for the event time submodel appear to be unavailable for competing risks joint models.

The software implementation of model 4 (Proust-Lima *et al.*, 2017) provides a score test for the independence assumption (Jacqmin-Gadda *et al.*, 2010) between the submodels (conditional on the latent classes)—both globally and cause specific—which JLCMs depend on. Model 4a yielded a statistically significant score test statistic ($P = 0.012$); however, the cause-specific score tests were non-significant ($P = 0.23$ (UAE); $P = 0.13$ (ISC)). A six-class model (the results are not shown here) did not reject the hypothesis, and estimated treatment effects $\hat{\beta}_{\text{UAE}}^2 = -0.839$ (95% CI, from −1.522 to −0.157) and $\hat{\beta}_{\text{ISC}}^{(2)} = 0.018$ (95% CI, from −0.563 to 0.599), which yield broadly similar inferences as per the estimates from the five-class model. Therefore, the rejection of the independence assumption does not appear to have an effect on the inferences in this instance. Model 4b did not reject the conditional independence assumption ($P = 0.12$). However, it should be noted that score tests are distribution dependent, which might explain the different test conclusions between models 4a and 4b. Posterior class membership tables can also be extracted from fitted models, which enable the user to assess the goodness of fit of the model and model discrimination (Proust-Lima *et al.*, 2012). As per model 3, the AIC, BIC and log-likelihood statistics are given as output. The five-class Weibull model had the lowest AIC (6787.68 *versus* 6805.78) and BIC (6955.08 *versus* 6990.80), and maximum log-likelihood (−3355.84 *versus* −3360.89) relative to the five-class piecewise constant model. Although model 4a had a lower AIC than did model 3f, caution should be taken in choosing one over the other, as each offers a fundamentally different structure. Finally, the software for model 4 offers several residual diagnostics (e.g. Fig. S4 in the on-line supplementary material) from the fitted longitudinal data submodel, which can be readily viewed from a fitted model.

6. Discussion

Over the past decade, research into joint modelling of longitudinal and competing risks data has grown. We described four models (Elashoff *et al.*, 2008; Williamson *et al.*, 2008; Rizopoulos, 2012; Proust-Lima *et al.*, 2017), each of which has provided software to enable model fitting.

The models that were explored here encompassed several association structures. Misspecification of the association structure might lead to biased effects. Ideally, a pragmatic balance between biologically plausible associations and a parsimonious model should be used; however, this might still be challenging. To date, there is limited research on this particular issue (Andrinopoulou and Rizopoulos, 2016). Although each model can be used for inference, if interest lies in risk prediction for personalized treatments, then model 4 would perhaps be most appropriate. The reasons for this are twofold: first, no consideration on the correct functional form of the latent association structure is required; second, the software implementation (Proust-Lima *et al.*, 2017) can readily produce predictions. Although the JLCM is potentially robust to misspecification of the latent association structure, which might be exploitable for inference, it conversely precludes inference on that aspect of the joint model, which might be of interest

to researchers. At the expense of the flexibility that is offered by model 2, the interpretation of the latent association is also less intuitive. Models 2, 3f and 4 also assume time-independent associations, whereas models 1 and 3a–3e consider different time-dependent association structures.

Among the proposals for the cause-specific hazards competing risks submodel, all models allowed for the specification of flexible baseline hazard functions, including spline and semi-parametric models. In principle therefore, if the hazard is void of abrupt changes and a sufficient number of knots are specified for the spline models (Rutherford *et al.*, 2015), then the differences in hazards should only subtly affect the model estimates. Notwithstanding calls for both cause-specific hazard and cumulative incidence function models to be used in analyses (Latouche *et al.*, 2013), the latter have received relatively little attention in the joint model framework. Deslandes and Chevret (2010) proposed modelling of competing risks data using the subdistribution hazard model of Fine and Gray (1999), and Proust-Lima *et al.* (2016) constructed cumulative incidence graphs from fitted JLCMs.

Estimation approaches and software implementations also varied substantially. A summary of the merits and limitations of each model implementation is given in Table S1 in the on-line supplementary material. One very practical limitation is that the software for models 1 and 2 permits only two causes of failure, although the models are presented in full generality. Despite some variation in the model fits to the SANAD trial data note that we did not ‘validate’ the software for each model, as this has already been done separately in each of the individual methodological research papers. Moreover, a grand simulation analysis would be conflated by the different ‘true’ association structures. Nonetheless, understanding whether there is a gain in efficiency by using one particular model, or whether any are more robust to misspecification than others, would be a useful direction of future research. We are also aware that two new software applications have recently been published: Andrinopoulou *et al.* (2014) has recently made available a WinBUGS script for fitting a joint model of bivariate longitudinal data (including an ordinal outcome) and competing risks data, and Wang *et al.* (2017) published an SAS macro for fitting parametric joint models involving competing risks data.

We illustrated the model fits and differences in interpretation by using a real world data set from an AED trial. Unlike research in other disease areas, there is very limited use of joint modelling in the disease area of epilepsy. The focus of this study was on the direct treatment effects, $\beta_{ISC}^{(2)}$ and $\beta_{UAE}^{(2)}$, following a clinical hypothesis that differential titration between the two AED drugs, LTG and CBZ, resulted in an unfair comparison in a separate model competing risks analysis (Williamson *et al.*, 2008). We note that a fundamental complication in contrasting models and treatment effects is that some models estimate an overall treatment effect, whereas others estimate separate direct and indirect treatment effects (Ibrahim *et al.*, 2010). Titration of AEDs is a complex issue that has far-ranging implications for the design and interpretation of drug trials. Joint modelling provides a sensible analytical method to handle this issue. We found that adjustment for calibrated dose in a joint model framework resulted in similar conclusions to those found in the original analysis of the SANAD trial data.

To allow for the routine adoption of joint modelling of competing risks and longitudinal data in the analysis of clinical data sets, further availability of model diagnostics, model comparisons and statistical tools for study design are necessary. We noted that some diagnostic and goodness-of-fit assessment methods are available in the JM (Rizopoulos, 2010) and l_{cm} (Proust-Lima *et al.*, 2017) R packages (see models 3 and 4 respectively). Prediction methods and assessment tools are also emerging. Blanche *et al.* (2015) have proposed dynamic area under the receiver operating characteristic curve and Brier score statistics applicable to these models, and the l_{cm} R package can generate dynamic predictions (see model 4). Finally, in our application

the measurement schedule was fixed by the trial protocol; however, in other cases, particularly observational studies, the measurement times may be informative. One possible future solution would be the addition of a third submodel to capture the recurrent events process (Liu *et al.*, 2008).

7. Conclusion

Although the models that were investigated here represent progress within the field, joint modelling of competing risks and longitudinal data is still very much in its infancy in terms of methodological development and is particularly restricted in terms of its ease of application to analyse clinical data sets. There is a real need to increase the availability and accessibility of software to allow the routine adoption of the methodologies when designing and analysing clinical studies. Further work is also required on the development of model diagnostics to aid model choice. Despite the vastly different association structures, the inference remains consistent across fitted models when applied to the SANAD data. The association between AED titration and treatment failure was significant in most models, and the estimated treatment effects implied a beneficial effect of LTG on UAEs over CBZ, but the two drugs were similar in terms of the hazard for ISC.

8. Data sharing

The data that are analysed in this paper are freely available from the R package `joiner`. The software and code that were used to fit the models are available from <https://github.com/graemeleehickey/comprisk>.

Acknowledgements

This research was funded by Medical Research Council grant MR/M013227/1. We are grateful to the Joint Editor, the Associate Editor and two referees for their incredibly helpful comments that have helped us to improve this manuscript.

References

- Andrinopoulou, E.-R. and Rizopoulos, D. (2016) Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Statist. Med.*, **35**, 4813–4823.
- Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J. M. and Lesaffre, E. (2014) Joint modeling of two longitudinal outcomes and competing risk data. *Statist. Med.*, **33**, 3167–3178.
- Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J. M. and Lesaffre, E. (2017) Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statist. Meth. Med. Res.*, **26**, 1787–1801.
- Asar, O., Ritchie, J., Kalra, P. and Diggle, P. J. (2015) Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *Int. J. Epidem.*, **4**, 334–344.
- Bakoyannis, G. and Touloumi, G. (2012) Practical methods for competing risks data: a review. *Statist. Meth. Med. Res.*, **21**, 257–272.
- Blanche, P., Proust-Lima, C., Loubère, L., Berr, C., Dartigues, J.-F. and Jacqmin-Gadda, H. (2015) Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, **71**, 102–113.
- Chen, L. M., Ibrahim, J. G. and Chu, H. (2011) Sample size and power determination in joint modeling of longitudinal and survival data. *Statist. Med.*, **30**, 2295–2309.
- Commission on Antiepileptic Drugs (1998) Considerations on designing clinical trials to evaluate the place of new antiepileptic drugs in the treatment of newly diagnosed and chronic patients with epilepsy. *Epilepsia*, **39**, 799–803.
- Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187–220.

- Cross, H., Ferrie, C., Lascelles, K., Livingston, J. and Mewasingh, L. (2007) Old versus new antiepileptic drugs: the SANAD study. *Lancet*, **370**, 314–316.
- Deslandes, E. and Chevret, S. (2010) Joint modeling of multivariate longitudinal data and the dropout process in a competing risk setting: application to ICU data. *BMC Med. Res. Methodol.*, **10**, article 69.
- Elashoff, R. M., Li, G. and Li, N. (2007) An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statist. Med.*, **26**, 1999–2011.
- Elashoff, R. M., Li, G. and Li, N. (2008) A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, **64**, 762–771.
- Faught, E. (2007) Epilepsy drugs: getting it right the first time. *Lancet Neurol.*, **6**, 476–478.
- Fine, J. P. and Gray, R. J. (1999) A proportional hazards model for the subdistribution of a competing risk. *J. Am. Statist. Ass.*, **94**, 496–509.
- Gould, A. L., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S. and Bois, F. Y. (2015) Joint modeling of survival and longitudinal non-survival data: current methods and issues: Report of the DIA Bayesian joint modeling working group. *Statist. Med.*, **34**, 2181–2195.
- Gueorguieva, R., Rosenheck, R. and Lin, H. (2012) Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *J. R. Statist. Soc. A*, **175**, 417–433.
- Henderson, R., Diggle, P. J. and Dobson, A. (2000) Joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465–480.
- Hickey, G. L., Philipson, P., Jorgensen, A. and Kolamunnage-Dona, R. (2016) Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med. Res. Methodol.*, **16**, 1–15.
- Hogan, J. W. and Laird, N. M. (1997) Model-based approaches to analysing incomplete longitudinal and failure time data. *Statist. Med.*, **16**, 259–272.
- Hu, B., Li, L. and Greene, T. (2016) Joint multiple imputation for longitudinal outcomes and clinical events that truncate longitudinal follow-up. *Statist. Med.*, **35**, 2991–3006.
- Hu, W., Li, G. and Li, N. (2009) A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statist. Med.*, **28**, 1601–1619.
- Hu, B., Li, L., Wang, X. and Greene, T. (2012) Nonparametric multistate representations of survival and longitudinal data with measurement error. *Statist. Med.*, **31**, 2303–2317.
- Huang, X., Li, G. and Elashoff, R. M. (2010) A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statist. Interfc.*, **3**, 185–195.
- Huang, X., Li, G., Elashoff, R. M. and Pan, J. (2011) A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetim. Data Anal.*, **17**, 80–100.
- Ibrahim, J. G., Chu, H. and Chen, L. M. (2010) Basic concepts and methods for joint models of longitudinal and survival data. *J. Clin. Oncol.*, **28**, 2796–2801.
- Jacqmin-Gadda, H., Proust-Lima, C., Taylor, J. M. G. and Commenges, D. (2010) Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics*, **66**, 11–19.
- Kalbfleisch, J. and Prentice, R. (2002) *The Statistical Analysis of Failure Time Data*, 2nd edn. Hoboken: Wiley.
- Ko, F.-S. (2014) Identification of longitudinal biomarkers for survival by a score test derived from a joint model of longitudinal and competing risks data. *J. Appl. Statist.*, **41**, 2270–2281.
- Laird, N. M. and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Latouche, A., Allignol, A., Beyersmann, J., Labopin, M. and Fine, J. P. (2013) A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J. Clin. Epidemiol.*, **66**, 648–653.
- Li, N., Elashoff, R. M. and Li, G. (2009) Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biometr. J.*, **51**, 19–30.
- Li, N., Elashoff, R. M., Li, G. and Saver, J. (2010) Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statist. Med.*, **29**, 546–557.
- Li, N., Elashoff, R. M., Li, G. and Tseng, C. H. (2012) Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Statist. Med.*, **31**, 1707–1721.
- Liu, L., Huang, X. and O’Quigley, J. (2008) Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics*, **64**, 950–958.
- Marson, A. G., Al-Kharusi, A. M., Alwaidh, M., Appleton, R., Baker, G. A., Chadwick, D. W., Cramp, C., Cockerell, O. C., Cooper, P. N., Doughty, J., Eaton, B., Gamble, C., Goulding, P. J., Howell, S. J. L., Hughes, A., Jackson, M., Jacoby, A., Kellett, M., Lawson, G. R., Leach, J. P., Nicolaidis, P., Roberts, R., Shackley, P., Shen, J., Smith, D. F., Smith, P. E. M., Tudur-Smith, C., Vanoli, A. and Williamson, P. R. (2007) The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. *Lancet*, **369**, 1000–1015.
- Philipson, P., Sousa, I., Diggle, P. J., Williamson, P. R., Kolamunnage-Dona, R., Henderson, R. and Hickey, G. L. (2017) Package “joineR”. R Foundation for Statistical Computing, Vienna.
- Proust-Lima, C., Dartigues, J.-F. and Jacqmin-Gadda, H. (2016) Joint modelling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Statist. Med.*, **35**, 382–398.

- Proust-Lima, C., Philipps, V. and Liqueur, B. (2017) Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *J. Statist. Softw.*, **78**, no. 2, 1–56.
- Proust-Lima, C., Sene, M., Taylor, J. M. G. and Jacqmin-Gadda, H. (2012) Joint latent class models for longitudinal and time-to-event data: a review. *Statist. Meth. Med. Res.*, **23**, 74–90.
- Putter, H., Fiocco, M. and Geskus, R. B. (2007a) Tutorial in biostatistics: competing risks and multi-state models. *Statist. Med.*, **26**, 2389–2430.
- Rizopoulos, D. (2010) JM: an R package for the joint modelling of longitudinal and time-to-event data. *J. Statist. Softw.*, **35**, 1–33.
- Rizopoulos, D. (2012) *Joint Models for Longitudinal and Time-to-event Data, with Applications in R*. Boca Raton: Chapman and Hall–CRC.
- Rizopoulos, D., Verbeke, G. and Molenberghs, G. (2010) Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics*, **66**, 20–29.
- Rutherford, M. J., Crowther, M. J. and Lambert, P. C. (2015) The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *J. Statist. Comput. Simuln.*, **85**, 777–793.
- Sparling, Y. H., Younes, N., Lachin, J. M. and Bautista, O. M. (2006) Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*, **7**, 599–614.
- Tsiatis, A. A. and Davidian, M. (2004) Joint modeling of longitudinal and time-to-event data: an overview. *Statist. Sin.*, **14**, 809–834.
- Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S. (1995) Modeling the relationship of survival to longitudinal data measured with error—applications to survival and CD4 counts in patients with AIDS. *J. Am. Statist. Ass.*, **90**, 27–37.
- Wang, W., Wang, W., Mosley, T. H. and Griswold, M. E. (2017) A SAS macro for the joint modeling of longitudinal outcomes and multiple competing risk dropouts. *Comput. Meth. Progrms Biomed.*, **138**, 23–30.
- Williamson, P. R., Kolamunnage-Dona, R., Philipson, P. and Marson, A. G. (2008) Joint modelling of longitudinal and competing risks data. *Statist. Med.*, **27**, 6426–6438.
- Williamson, P. R., Kolamunnage-Dona, R. and Tudur-Smith, C. (2007a) The influence of competing-risks setting on the choice of hypothesis test for treatment effect. *Biostatistics*, **8**, 689–694.
- Williamson, P. R., Tudur-Smith, C., Sander, J. W. and Marson, A. G. (2007b) Importance of competing risks in the analysis of anti-epileptic drug failure. *Trials*, **8**, article 12.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997) A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.
- Yu, B. and Ghosh, P. (2010) Joint modeling for cognitive trajectory and risk of dementia in the presence of death. *Biometrics*, **66**, 294–300.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary material for “A comparison of different joint models for longitudinal and competing risks data: with application to an epilepsy drug randomised control trial”’.