



Newcastle University ePrints

Barrett J, Diggle P, Henderson R, Taylor-Robinson D. [Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference](#). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2015, 77(1), 131-148.

Copyright:

© 2014 The Authors Journal of the Royal Statistical Society: Series B (Statistical Methodology) 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.

Link to published article:

<http://dx.doi.org/10.1111/rssb.12060>

Date deposited: 9th January 2015



This work is licensed under a [Creative Commons Attribution 3.0 Unported License](#)

ePrints – Newcastle University ePrints

<http://eprint.ncl.ac.uk>



J. R. Statist. Soc. B (2015)
77, Part 1, pp. 131–148

Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference

Jessica Barrett,

University of Cambridge, UK

Peter Diggle,

Lancaster University and University of Liverpool, UK

Robin Henderson

University of Newcastle, UK

and David Taylor-Robinson

University of Liverpool, UK

[Received September 2012. Final revision October 2013]

Summary. Random effects or shared parameter models are commonly advocated for the analysis of combined repeated measurement and event history data, including dropout from longitudinal trials. Their use in practical applications has generally been limited by computational cost and complexity, meaning that only simple special cases can be fitted by using readily available software. We propose a new approach that exploits recent distributional results for the extended skew normal family to allow exact likelihood inference for a flexible class of random-effects models. The method uses a discretization of the timescale for the time-to-event outcome, which is often unavoidable in any case when events correspond to dropout. We place no restriction on the times at which repeated measurements are made. An analysis of repeated lung function measurements in a cystic fibrosis cohort is used to illustrate the method.

Keywords: Cystic fibrosis; Dropout; Joint modelling; Repeated measurements; Skew normal distribution; Survival analysis

1. Introduction

There is a close relationship between modelling longitudinal data subject to dropout and modelling survival time data in the presence of imprecisely observed time varying covariates. In both cases we have a vector of repeated measurements Y and a time to event S . In the survival context S would normally be considered to be measured in continuous time, though possibly right censored. In the dropout context S usually corresponds to a discrete interval between scheduled measurement times. Typically the occurrence of the event terminates observation of the repeated measurements.

A common approach for both the survival time and the dropout problems is to assume

Address for correspondence: Jessica Barrett, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Wort's Causeway, Cambridge, CB1 8RN, UK.
E-mail: jkb23@medschl.cam.ac.uk

© 2014 The Authors Journal of the Royal Statistical Society: Series B (Statistical Methodology) 1369–7412/15/77131
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.

conditional independence between Y and S given underlying random effects U . An important early reference is Wulfsohn and Tsiatis (1997), who assumed that Y was linear in a Gaussian random effect U and took a proportional hazards model for S , conditional on U . Tsiatis and Davidian (2004) have given an excellent review of work in the area to that date, and Albert and Follmann (2009) and Diggle *et al.* (2009) have given additional references including later developments. More recent contributions include Geskus (2012), Gueorguieva *et al.* (2012), Proust-Lima *et al.* (2012) and Rizopoulos (2011, 2012).

The random-effects or shared parameter approach to jointly modelling repeated measurement and event time data is conceptually attractive in many settings, but its routine application is hampered by computational difficulties. Fast but approximate methods have been developed for some forms of joint model (see for example Rizopoulos (2012) and references therein) but implementation remains difficult unless the random-effects component is of low dimension. Wulfsohn and Tsiatis (1997), for instance, assumed that U was bivariate Gaussian and adopted a Laird and Ware (1982) random intercept and slope model for the longitudinal trajectory. Henderson *et al.* (2000) argued that when follow-up is relatively long then it is unreasonable to assume a sustained trend in the trajectory of Y and advocated inclusion of an unobserved stationary Gaussian process $W(t)$ in a linear predictor for Y to bring more flexibility. In principle this assumes the presence of an infinite dimensional random effect but, under either a discrete dropout or a semiparametric proportional hazards model for S , likelihood inference requires the value of $W(t)$ at only measurement times or event times. Hence $W(t)$ can be represented by a finite dimensional vector U . The dimension increases with sample size and there is no generally available software for this type of model.

In this paper we propose an approach which admits exact likelihood inference for a wide range of random-effect specifications. The key is to consider events S to occur only at a discrete set of potential times. In principle, the discretization can be made arbitrarily fine, but at the expense of increasing computational effort. Hence, the practical advantage of our approach relies on its remaining computationally feasible for a discretization that is sufficiently fine to capture the essential features of its continuous time limit. When the events correspond to dropout, their recorded times of occurrence are typically confined to a discrete set of scheduled measurement times. When the event is a survival time then our approach is a form of coarsening at random (Heitjan and Rubin, 1991). For example, in Section 6 we describe an example involving survival in a cystic fibrosis cohort, where we choose to measure survival by calendar year of death. With 10 years of follow-up and no evidence of significant local variation in hazard rates this gives adequate granularity. There is no need for the timescale to be discretized in the submodel for the repeated measurements Y under our approach, which therefore allows for irregular and subject-specific measurement timings.

The general model and some special cases are set out in Section 2. In Section 3 we derive the key likelihood-based methods on the basis of recent work on the extended skew normal family of distributions. Simulations to assess the performance of the method are described in Section 4. An examination of efficiency is presented in Section 5 and in Section 6 we describe application to the cystic fibrosis cohort study. Concluding discussion is presented in Section 7.

The programs that were used to analyse the data can be obtained from

<http://wileyonlinelibrary.com/journal/rss-datasets>

2. Model

We take a shared parameter approach, with the common assumption that event times S and longitudinal data Y are independent conditionally on random effects U .

2.1. General formulation

We assume that subject i provides repeated measurements $Y_i = \{Y_{ij} : i = 1, 2, \dots, N; j = 1, \dots, n_i\}$ at follow-up times t_{ij} , together with a time-to-event outcome S_i , which terminates the observation of Y . Time varying covariates are allowed. We write x_{ij} for covariates that are operative on Y_{ij} , and \tilde{x}_{is} for covariates that are operative at time s on the event process. For the remainder of this section we drop the subscript i and consider a generic subject. It will be implicit that the number and timing of measurements can vary between subjects.

We assume two timescales: a discrete timescale for events, which without loss of generality we can label as $\mathcal{S} = \{1, 2, \dots, m\}$, and a continuous timescale $\mathcal{T} = [0, \tau]$ for measurements. We also assume that there is a surjection $s(t)$ from \mathcal{T} to \mathcal{S} . For instance, \mathcal{S} might be a partition of \mathcal{T} . We consider \mathcal{S} to represent a series of time intervals; for example if the timescale \mathcal{T} is years then \mathcal{S} might contain yearly or 6-monthly intervals. Alternatively, \mathcal{S} might contain time intervals of different lengths, or it might represent intervals between scheduled measurement times. We denote by $t^*(s)$ the midpoint of the set of times in \mathcal{T} that map to $s \in \mathcal{S}$.

2.1.1. Random effects

We model the random effects as a p -vector $U = (U_1, \dots, U_p)^T$, assuming a Gaussian distribution $U \sim N(0, \Sigma)$ with a general covariance structure.

2.1.2. Event times

With the skew normal results in mind, we adopt a probit model for the discrete hazard function (equivalently the dropout model), which is very similar to the more widely used logistic model. For the most general model we define $W(s) = B(s)U$ to be a \tilde{p} -vector of linear combinations of the random effects U , where $B(s)$ is a $\tilde{p} \times p$ matrix which may depend on the time interval s . Then we assume that

$$P(S = s | S > s - 1, U) = 1 - \Phi \left\{ \tilde{x}_s^T \tilde{\beta} + \sum_{k=1}^{\tilde{p}} \gamma_{sk} W_k(s) \right\}, \quad s \in \mathcal{S}, \tag{1}$$

where $\Phi(\cdot)$ is the standard normal distribution function. In equation (1) the survival model is allowed to depend on time through covariates that are contained in a p_1 -vector \tilde{x}_s . In principle a non-linear function of time could be used, and often a separate intercept would be used for each time interval. In our examples we shall focus on models for which the probit probability of surviving a time interval is linearly dependent on time. The association between survival and the random effects can also vary with time, although all examples in this paper will assume that γ_{sk} does not depend on s .

Note that the model thus defined is a sequential probit model because the probability of surviving a time interval is conditional on having survived all previous time intervals. Albert and Chib (2001) have provided discussion on the application of sequential models in survival.

2.1.3. Repeated measurements

We consider a linear mixed effects Gaussian model for the sequence of repeated measurements, $Y = \{Y_j : j = 1, \dots, n\}$, at measurement times t_j . At time t_j we assume

$$Y_j = x_j^T \beta + a_j^T U + Z_j, \quad j = 1, 2, \dots, n, \tag{2}$$

where x_j and a_j are a p_2 -vector and a p -vector respectively, of possibly time-dependent covari-

ates. The $\{Z_j\}$ are mutually independent measurement errors, taken as $Z \sim \text{MVN}(0, \nu^2 I)$ with I an identity matrix. We can write expression (2) in the vector form

$$Y = X\beta + AU + Z \quad (3)$$

where $X = (x_1, x_2, \dots, x_n)^T$ and $A = (a_1, a_2, \dots, a_n)^T$.

2.2. Examples

2.2.1. Random intercept

Our formulation includes the random-intercept model. Assume that U is a scalar random effect with $N(0, \sigma^2)$ distribution. Then set $a_j = 1$ for all j and $W_s = U$, with $\gamma_s = \gamma$ for all $s \in \mathcal{S}$.

2.2.2. Random intercept and slope

Let $(U_1, U_2)^T$ be zero mean bivariate normal with variance matrix

$$\Sigma^{\text{IS}} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

and set $a_j = (1, t_j)^T$. For $s \in \mathcal{S}$ set $W(s) = U_1 + U_2 t^*(s)$ and $\gamma_s = \gamma$ for all $s \in \mathcal{S}$. Thus we can model a sustained trend in the survival hazard, albeit with a piecewise constant form. Given that we work with events in discrete time this is unavoidable. Alternatively we can adjust the model to allow the intercept and slope random effects to enter the event times model directly, setting $W(s) = (U_1, U_2)^T$ and $\gamma_s = (\gamma_1, \gamma_2)^T$.

2.2.3. Stationary Gaussian process

For the stationary Gaussian process model we assume that U is an m -vector with variance matrix

$$(\Sigma^{\text{SGP}})_{jk} = \sigma^2 \rho^{|t^*(j) - t^*(k)|},$$

so that the random effects correspond to a discretely observed stationary Gaussian process. Here $W(s) = U_s$ and $\gamma_s = \gamma$ for all $s \in \mathcal{S}$. We assume that measurements Y_j and Y_k with $s(t_j) = s(t_k)$ share the same U_s . We define a_j to be an m -vector with value 1 at element $s(t_j)$ and 0 elsewhere. If fluctuation at a very short timescale is thought to be materially important then a fine discretization \mathcal{S} is needed to capture this behaviour. Note that the discretization does not need to scale with sample size. By extending $W(s)$ we can allow survival or dropout to depend on the prior history of random effects; for example for one time lag set $W(s) = (U_s, U_{s-1})^T$ and $\gamma_s = (\gamma, \gamma_{\text{lag}})^T$.

2.2.4. Stationary Gaussian process plus random intercept and slope

We can combine a stationary Gaussian process with a random intercept and slope by defining U to be an $(m+2)$ -vector with variance matrix

$$\Sigma = \begin{pmatrix} \Sigma^{\text{SGP}} & 0 \\ 0 & \Sigma^{\text{IS}} \end{pmatrix}$$

Now $W(s) = (U_s, U_1, U_2)$ and $\gamma_s = (\gamma, \gamma_1, \gamma_2)$ for all $s > 1$.

3. Inference

We now set out the general form of the likelihood for the class of models that was defined in Section 2.1. For simplicity we shall continue with the contribution of a single individual only.

We shall make use of results concerning the properties of the skew normal distribution (Azzalini, 1985, 2005; Arnold and Beaver, 2000; Arnold, 2009) to obtain a closed form for the likelihood. In particular, a multivariate Gaussian hidden truncation distribution that was considered by Arnold (2009) leads to the result

$$\Phi^{(k_1)}(\omega_1; 0, \Sigma_{22}) = \int \Phi^{(k_1)}(\omega_1 - \Sigma_{12}^T \Sigma_{11}^{-1}(\omega_2 - \eta); 0, \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}) \phi^{(k_2)}(\omega_2; \eta, \Sigma_{11}) d\omega_2 \quad (4)$$

where ω_1, ω_2 and η are $k_1 \times 1, k_2 \times 1$ and $k_2 \times 1$ vectors respectively, and Σ_{11}, Σ_{22} and Σ_{12} are concordant variance and covariance matrices. Here $\phi^{(k)}(x; \mu, \Sigma)$ is the k -dimensional Gaussian density with mean vector μ and covariance matrix Σ and $\Phi^{(k)}(x; \mu, \Sigma)$ is the corresponding distribution function. By rearranging the likelihood so that it takes the form of the integrand on the right-hand side of equation (4), we can use this result to integrate out the random effects to arrive at a closed form expression.

3.1. Preliminaries

We begin by rearranging the longitudinal and random-effects components of the likelihood so that they take the form of the Gaussian density function under the integrand in equation (4).

Assume that the dropout or survival time is s and introduce an indicator δ of an event being observed ($\delta = 1$) or right censoring ($\delta = 0$). The random-effects and longitudinal components of the likelihood are

$$f(u) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{u^T \Sigma^{-1} u}{2}\right)$$

and

$$f(y|u) = \frac{1}{(2\pi)^n n! \nu^n} \exp\left\{-\frac{(y - X\beta - Au)^T (y - X\beta - Au)}{2\nu^2}\right\}.$$

Note that $f(y|u) f(u)$ contains two terms under the exponential which are quadratic in u . We can complete the square in u so that u appears in a single quadratic term (see Appendix A for details). It will be convenient to collect all unknown parameters into vector θ and to write

$$f(y|u) f(u) = L_1(\theta; y) \phi^{(p)}(u; h, H^{-1}), \quad (5)$$

where

$$H = (A^T A / \nu^2 + \Sigma^{-1}),$$

$$h = H^{-1} A^T (y - X\beta) / \nu^2.$$

Note that $\phi^{(p)}(u; h, H^{-1})$ depends on a subvector of the parameter vector θ through h and H . The expression for $L_1(\theta; y)$ is

$$L_1(\theta; y) = \frac{1}{(2\pi)^n n! \nu^n |\Sigma H|^{1/2}} \exp\left\{-\frac{(y - X\beta)^T (y - X\beta)}{2\nu^2} + \frac{h^T H h}{2}\right\}.$$

Turning to the event times, we have

$$f(s, \delta|u) = \left\{ \prod_{v=1}^{s-1} \Phi\left(\tilde{x}_v^T \tilde{\beta} + \sum_{k=1}^{\tilde{p}} \gamma_{vk} w_k\right) \right\} \Phi\left(\tilde{x}_s^T \tilde{\beta} + \sum_{k=1}^{\tilde{p}} \gamma_{sk} w_k\right)^{1-\delta} \left\{ 1 - \Phi\left(\tilde{x}_s^T \tilde{\beta} + \sum_{k=1}^{\tilde{p}} \gamma_{sk} w_k\right) \right\}^\delta \quad (6)$$

where $w = B(s)u$ and $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function. Now define $\tilde{X}_v = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_v)^\top$, and an $s \times \tilde{p}$ matrix Γ such that $\Gamma_{vk} = \gamma_{vk}$. Then equation (6) can be written as

$$f(s, \delta|u) = \begin{cases} \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)u\} & \delta = 0, \\ \Phi^{(s-1)}\{\tilde{X}_{s-1}\tilde{\beta} + \Gamma_{s-1} B(s-1)u\} - \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)u\} & \delta = 1, \end{cases} \quad (7)$$

where Γ_s denotes the first s rows of Γ . Here $\Phi^{(s)}(\cdot) = \Phi^{(s)}(\cdot; 0, I)$ is the standard s -dimensional multivariate Gaussian cumulative distribution function.

3.2. Likelihood

Combining expressions (5) and (7) and integrating out the random effects leads to the likelihood

$$L(\theta; y, s, \delta) = \int f(s, \delta|u) f(y|u) f(u) du$$

$$= \begin{cases} L_1(\theta; y) \int \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)u\} \phi^{(p)}(u; h, H^{-1}) du & \delta = 0, \\ L_1(\theta; y) \int [\Phi^{(s-1)}\{\tilde{X}_{s-1}\tilde{\beta} + \Gamma_{s-1} B(s-1)u\} \\ - \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)u\}] \phi^{(p)}(u; h, H^{-1}) du & \delta = 1. \end{cases} \quad (8)$$

We can convert the right-hand side of expression (4) to the form of the integrals in expression (8) by defining the parameters $\omega_1, \omega_2, \eta, \Sigma_{11}, \Sigma_{12}$ and Σ_{22} appropriately (see Appendix A for details). Using expression (4) to obtain closed form expressions for the integrals then gives

$$L(\theta; y, s, \delta) = L_1(\theta; y) L_2(\theta; y, s, \delta) \quad (9)$$

where

$$L_2(\theta; y, s, \delta) = \begin{cases} \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)h; 0, I + \Gamma_s B(s)H^{-1}B(s)^\top \Gamma_s^\top\} & \delta = 0, \\ \Phi^{(s-1)}\{\tilde{X}_{s-1}\tilde{\beta} + \Gamma_{s-1} B(s-1)h; 0, I \\ + \Gamma_{s-1} B(s-1)H^{-1}B(s-1)^\top \Gamma_{s-1}^\top\} \\ - \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)h; 0, I + \Gamma_s B(s)H^{-1}B(s)^\top \Gamma_s^\top\}. & \delta = 1. \end{cases}$$

Numerical evaluation of equation (9) is now straightforward by using available software to calculate multivariate normal probabilities; we used the R package `mnormt` (Azzalini, 2013). Parameter estimation can be conducted by numerical maximization of the likelihood, and candidate models can be compared by exact likelihoods or information criteria. Having specified a fully parametric model, the maximum likelihood estimator $\hat{\theta}$ is asymptotically Gaussian, centred at the true value θ_0 , and with variance given by the inverse Fisher information. Standard regularity conditions apply, with some supplementation to ensure identifiability of parameters when data can be missing. Details are provided in Appendix B.

4. Simulation studies

We simulated repeated measurements with dropout data from the models specified by equations (1) and (2). Random effects were generated either as a stationary Gaussian process or as a random intercept and slope. Longitudinal measurements took place at n measurement times randomly distributed over m time intervals, so that individuals could have a varying number of visits in each time interval. A uniform distribution was chosen for the measurement times, motivated by our cystic fibrosis application. We assumed that dropout could occur during any

time interval. For the stationary Gaussian process models we generated m -dimensional random effects $U = (U_1, \dots, U_m)^T$ and Gaussian repeated measurements

$$Y_j = x_j^T \beta + U_{s(t_j)} + Z_j, \quad j = 1, 2, \dots, n,$$

where t_j are the measurement times, and

$$P(S = s | S > s - 1, U) = 1 - \Phi(\tilde{x}_s^T \tilde{\beta} + \gamma U_s), \quad s = 1, 2, \dots, m.$$

For the random intercept and slope simulations we generated bivariate random effects $U = (U_1, U_2)^T$ and Gaussian repeated measurements

$$Y_j = x_j^T \beta + U_1 + U_2 t_j + Z_j, \quad j = 1, 2, \dots, n,$$

and

$$P(S = s | S > s - 1, U) = 1 - \Phi(\tilde{x}_s^T \tilde{\beta} + \gamma_1 U_1 + \gamma_2 U_2), \quad s = 1, 2, \dots, m.$$

For each simulation study 500 data sets were generated; each of 1000 individuals followed over $m = 5$ time intervals, with 1–10 repeated measurements per person. Survival parameters

Table 1. Simulation results from a stationary Gaussian process model†

Parameter	Results for simulation SGP I				Results for simulation SGP II			
	True value	Mean	Coverage	MSE	True value	Mean	Coverage	MSE
<i>Longitudinal</i>								
Intercept	90	89.947 (2.556)	0.959	6.521	90	90.005 (0.127)	0.944	0.016
Time	-1.7	-1.700 (0.272)	0.941	0.074	-1.7	-1.700 (0.016)	0.944	0.000
Age at t_0	-1.7	-1.696 (0.115)	0.957	0.013	-1.7	-1.700 (0.006)	0.942	0.000
Sex (males)	2	1.943 (1.343)	0.943	1.803	2	2.001 (0.061)	0.948	0.004
<i>Survival</i>								
Intercept	2	1.998 (0.210)	0.949	0.044	2	2.013 (0.214)	0.942	0.046
Time	0.01	0.009 (0.031)	0.952	0.001	0.01	0.009 (0.036)	0.948	0.001
Age at t_0	0.01	0.011 (0.009)	0.949	0.000	0.01	0.010 (0.009)	0.946	0.000
Sex (males)	0.1	0.098 (0.099)	0.957	0.010	0.1	0.097 (0.105)	0.952	0.011
γ	0.05	0.050 (0.003)	0.961	0.000	0.05	0.049 (0.072)	0.966	0.005
<i>Others</i>								
ν	7	7.005 (0.104)	0.959	0.011	1	1.000 (0.014)	0.944	0.000
σ_u	25	24.985 (0.450)	0.959	0.203	1	0.999 (0.025)	0.948	0.001
ρ	0.7	0.699 (0.015)	0.947	0.000	0.7	0.700 (0.025)	0.946	0.001

†Sample size 1000 and 500 replicates. Shown are the mean (with standard deviations in parentheses) parameter estimates, empirical coverage probabilities of nominal 95% confidence intervals and mean-squared errors MSE.

were chosen such that the death rate per time interval was 1–2%, which is similar to that observed in the cystic fibrosis application. Censoring took place at the end of the fifth time interval. Covariates that were included in both models were age, with initial values generated uniformly from the interval (15,30), and a binary covariate, which was either 0 or 1 with probability $\frac{1}{2}$. These covariates were similar to the observed covariate structure in the cystic fibrosis data.

Table 1 summarizes the results of two studies with stationary Gaussian process random effects. For the first simulation (SGP I), the repeated measures and random effects were taken to have large standard deviations, similar to those observed for the cystic fibrosis patients. The simulation was repeated (SGP II) with smaller standard deviations. For both scenarios the means of the parameter estimates are close to the true values. For simulation SGP I the standard deviations of parameter estimates are relatively large for the longitudinal parameters owing to the high

Table 2. Simulation results from a random intercept and slope model†

Parameter	Results for simulation IS I				Results for simulation IS II			
	True value	Mean	Coverage	MSE	True value	Mean	Coverage	MSE
<i>Longitudinal</i>								
Intercept	90	89.842 (2.781)	0.940	7.743	90	89.994 (0.141)	0.950	0.020
Time	-1.7	-1.701 (0.109)	0.950	0.012	-1.7	-1.701 (0.036)	0.952	0.001
Age at t_0	-1.7	-1.693 (0.125)	0.956	0.016	-1.7	-1.700 (0.006)	0.950	0.000
Sex (males)	2	2.056 (1.357)	0.952	1.840	2	2.003 (0.071)	0.956	0.005
<i>Survival</i>								
Intercept	2	2.051 (0.226)	0.946	0.054	2	2.015 (0.187)	0.962	0.035
Time	0.01	0.009 (0.037)	0.956	0.001	0.01	0.010 (0.037)	0.946	0.001
Age at t_0	0.01	0.010 (0.010)	0.932	0.000	0.01	0.010 (0.009)	0.948	0.000
Sex (males)	0.1	0.112 (0.100)	0.966	0.010	0.1	0.113 (0.107)	0.938	0.012
γ_1	0.01	0.011 (0.005)	0.970	0.000	0.01	0.019 (0.107)	0.950	0.011
γ_2	0.1	0.112 (0.106)	0.970	0.011	0.1	0.107 (0.084)	0.946	0.007
<i>Others</i>								
ν	7	6.997 (0.080)	0.952	0.006	1	1.000 (0.011)	0.972	0.000
σ_1	25	24.993 (0.640)	0.948	0.409	1	0.994 (0.042)	0.980	0.002
σ_2	2	1.989 (0.112)	0.940	0.012	1	0.999 (0.027)	0.952	0.001
ρ	-0.6	-0.600 (0.038)	0.950	0.001	-0.6	-0.599 (0.032)	0.958	0.001

†Sample size 1000 and 500 replicates. Shown are the mean (with standard deviations in parentheses) parameter estimates, empirical coverage probabilities of nominal 95% confidence intervals and mean-squared errors MSE.

level of measurement error. For simulation SGP II the corresponding standard deviations are smaller, but the standard deviation of γ is larger. This is because the variance of the random effects is smaller, so we are effectively regressing on a covariate with a smaller range of values. Coverage probabilities for both simulation SGP I and simulation SGP II are good, with all empirical values within simulation noise of the nominal level and no consistent overestimation or underestimation of the coverage.

Table 2 shows the results of two simulation studies with random intercepts and slopes. Again, for one study (IS I) data were generated with large measurement error, very high variance in the random intercept U_1 and high variance for the random slope U_2 , as observed for cystic fibrosis patients. For the second simulation study (IS II) the variance parameters were reduced. Means of parameter estimates are once more close to the true values. The standard deviations of longitudinal parameter estimates are large for simulation IS I and small for IS II, as expected, whereas the standard deviation of estimates of γ_1 is much higher for simulation IS II than for IS I. Interestingly, the standard deviation for γ_2 is approximately the same in both scenarios. Coverage probabilities are once more close to the nominal levels.

In a second simulation study data were generated from a Weibull model and analysed by using four methods. To reduce the computation time a simpler model was chosen with a random intercept only and using a single binary covariate. The Weibull parameters were based on a Weibull fit to cystic fibrosis survival, with intercept -4.2 and shape 1.2 . Other aspects of the model that was used to generate data were the same as in the previous simulation study, except times were censored at $t = 6$ to enable us straightforwardly to carry out different discretizations of the timescale. Each simulated data set was analysed by using

- (a) a longitudinal model with a random intercept (so ignoring the survival data),
- (b) the R package `joiner` to fit a joint model with a shared random intercept and proportional hazards survival model,
- (c) the discrete time method that is described in this paper with six time intervals and
- (d) the discrete time method with three time intervals (note that our sequential probit model is misspecified for this scenario).

Table 3. Simulation results from a Weibull model†

Parameter	True value	Results for longitudinal model			Results for <code>joiner</code>		
		Mean	Coverage	MSE	Mean	Coverage	MSE
<i>Longitudinal</i>							
Intercept	90	90.213 (1.133)	0.948	1.326	89.982 (1.106)	0.954	1.222
Time	-1.7	-1.628 (0.061)	0.794	0.009	-1.698 (0.061)	0.954	0.004
Sex (males)	2	1.806 (1.578)	0.964	2.522	1.944 (1.581)	0.946	2.497
<i>Survival</i>							
Sex (males)	-0.3	—	—	—	-0.300 (0.196)	0.958	0.038
γ_1	-0.1	—	—	—	-0.100 (0.005)	0.926	0.000
<i>Others</i>							
ν	7	7.007 (0.086)	0.928	0.007	6.999 (0.082)	0.946	0.007
σ_1	25	24.789 (0.577)	0.824	0.377	24.977 (0.547)	0.956	0.300

†Sample size 1000 and 500 replicates. Shown are the mean (with standard deviations in parentheses) parameter estimates, empirical coverage probabilities of nominal 95% confidence intervals and mean-squared errors MSE. For the longitudinal model standard errors were calculated for variance parameters by using bootstrapping.

Table 4. Simulation results from a Weibull model†

Parameter	True value	Results for discrete time model, 6 intervals			Results for discrete time model, 3 intervals		
		Mean	Coverage	MSE	Mean	Coverage	MSE
<i>Longitudinal</i>							
Intercept	90	90.034 (1.108)	0.958	1.225	90.041 (1.105)	0.960	1.221
Time	-1.7	-1.700 (0.061)	0.956	0.004	-1.697 (0.061)	0.954	0.004
Sex (males)	2	1.952 (1.584)	0.947	2.505	1.952 (1.581)	0.948	2.497
<i>Survival</i>							
Intercept	—	2.150 (0.122)	—	—	1.813 (0.132)	—	—
Time	—	-0.038 (0.024)	—	—	-0.039 (0.029)	—	—
Sex (males)	—	0.163 (0.110)	—	—	0.183 (0.126)	—	—
γ_1	—	0.054 (0.004)	—	—	0.061 (0.004)	—	—
<i>Others</i>							
ν	7	6.998 (0.082)	0.945	0.007	7.000 (0.082)	0.948	0.007
σ_1	25	24.932 (0.548)	0.958	0.303	24.908 (0.546)	0.958	0.305

†Sample size 1000 and 500 replicates. Shown are the mean (with standard deviations in parentheses) parameter estimates, empirical coverage probabilities of nominal 95% confidence intervals and mean-squared errors MSE.

Table 3 shows the results for the longitudinal model and the joint model fitted by using `joiner`. Fitting the longitudinal model separately gives slightly biased estimates for the longitudinal slope parameter and binary covariate. The joint model fit using `joiner`, however, gives parameter estimates with less bias and with good coverage, as would be expected because the data were generated under the same model as used in fitting. Table 4 shows results by using the discrete time model with six and three time intervals. Here the model that was used to generate the survival data and the model used to fit the survival data are not equivalent, and so survival parameters estimated by the discrete time model cannot be compared with true values of the survival parameters. The coverage probabilities and mean-squared error could not therefore be calculated for these parameters. We can, however, see that the directions of the survival parameter estimates agree with those of the true model; positive parameter estimates in the discrete time model are in accordance with negative estimates in the true model because the former are linked to the probability of survival rather than the hazard of an event. Comparing longitudinal parameters, the standard errors are similar to the `joiner` results. Again, survival parameters are not directly comparable between the two models because the survival parameters of the six-interval model relate to the probability of surviving 1 year, and the parameters of the three-interval model to the probability of surviving 2 years. We would expect covariate effects to be similar, however, as is indeed the case. Comparing standard errors of survival parameter estimates, we find that the standard errors are slightly larger for the coarser time discretization.

5. Efficiency under coarsening at random

Although we allow the longitudinal measurements to be obtained in continuous or discrete time, we have required the event time data either originally to be measured on a discrete scale or to be placed on a discrete scale through coarsening at random via artificial interval censoring. We assume that our discrete time model is correct and so—because we use exact likelihood—our

estimates are consistent and fully efficient given the data that we have chosen to use. However, discretization will affect the efficiency of an analysis, which is considered briefly in this section and expanded on in an on-line supplementary document. Here we concentrate only on survival analysis and omit much of the detail: this is provided together with further examples in the supplementary material. The supplementary material also includes a separate study into loss of information caused by discretization in the presence of random effects.

Let T be the continuous event time. Assume type 1 censoring at a maximum follow-up time τ . The follow-up interval $(0, \tau]$ is partitioned into m disjoint intervals, with boundaries $0 = t_0 < t_1 < t_2 < \dots < t_m = \tau$. Let S denote the interval within which T falls, with $S = m + 1$ if T is censored at τ . Define $W = (T - t_{s-1}) / (t_s - t_{s-1})$, which is the within-interval information on a $(0, 1)$ scale. Note that there is a one-to-one correspondence between T and (S, W) . We shall investigate the loss of efficiency that is caused by ignoring W .

The sequential probit model (1) is assumed for event probabilities within each interval j , for $j = 1, 2, \dots, m$, with time constant covariates and covariate effects, but possibly time varying intercepts:

$$P\{S > j | S > (j - 1), \tilde{x}\} = \Phi(\tilde{\beta}_{0j} + \tilde{x}^T \tilde{\beta}).$$

We assume for simplicity that the conditional within-interval distribution of event times is the same for all intervals. Let the corresponding probability density function be $h(w|\tilde{x})$, which will usually depend on $\tilde{\beta}$ (and perhaps other parameters). Information on $\tilde{\beta}$ from the within-interval distribution of event times provides the extra efficiency for the complete-data analysis. To illustrate, assume that

$$h(w|\tilde{x}) = \frac{r\psi}{1 - (1 - \psi)^r} (\psi w + 1 - \psi)^{r-1},$$

where $r = r(\tilde{x}^T \tilde{\beta}) = \exp(\tilde{x}^T \tilde{\beta})$ and $0 < \psi < 1$. This is the within-interval distribution that arises if a Weibull distribution is discretized.

The information on $\tilde{\beta}$ that is brought by W depends on how strongly $h(w|\tilde{x})$ depends on $\tilde{\beta}$.

Table 5. Efficiency of the coarsened-at-random analysis compared with the complete-data analysis†

ψ	$c(\tilde{x})$	Results for the following values of m :			
		$m = 3$	$m = 5$	$m = 7$	$m = 9$
0.9	9.75	0.919	0.930	0.936	0.939
0.8	2.28	0.958	0.964	0.967	0.969
0.7	0.91	0.976	0.979	0.981	0.982
0.6	0.44	0.986	0.988	0.989	0.990
0.5	0.22	0.992	0.993	0.994	0.994
0.4	0.11	0.996	0.996	0.997	0.997
0.3	0.05	0.998	0.998	0.998	0.998
0.2	0.02	0.999	0.999	0.999	0.999
0.1	0.00	1.000	1.000	1.000	1.000
$S(\tau \tilde{x} = 1)$		0.201	0.152	0.123	0.104

†Values in the main block are the ratios of asymptotic variance estimators for $\tilde{\beta}$ without and with W .

We can measure this through the curvature

$$c(\tilde{x}) = \int_0^1 h'(w|\tilde{x})^2 dw.$$

Given this set-up, the information that is associated with the marginal likelihood based only on S can be derived and compared with that from the full likelihood given both S and W . Table 5 provides some numerical values for the ratio of asymptotic variances of maximum likelihood estimators. For this example we took a single binary covariate and set $S(\tau|\tilde{x}=0) = 0.7$ and $\tilde{\beta} = -1$. Since $\tilde{\beta}$ is fixed, changing m changes $S(t|\tilde{x}=1)$. The final row in Table 5 gives $S(\tau|\tilde{x}=1)$ for each m . Intercepts $\tilde{\beta}_{0j}$ were chosen by assuming equal failure probabilities within each interval at $\tilde{x}=0$.

Some calibration based on curvature of discretized Weibull distributions is provided in the on-line supplementary material. The curvature in Table 5 at $\psi > 0.7$ is higher than anything seen for the Weibull discretizations that we considered. Even so, the efficiency was more than 90% in all simulation scenarios and was more than 97% in more realistic cases. We obtained similar results for other $\tilde{\beta}$, and when we fixed $S(t|\tilde{x}=1)$ and allowed $\tilde{\beta}$ to vary with m .

6. Application: disease progression and survival in cystic fibrosis patients

We now apply our methods to data on repeated measurements of lung function in cystic fibrosis patients, some of whom died during follow-up. The data were from the UK Cystic Fibrosis registry and cover the years 1999–2010. Cystic fibrosis is the commonest serious inherited disease among Caucasian populations, and most patients die as a result of progressive respiratory disease (Davies and Alton, 2009). Previous examples of longitudinal modelling applied to cystic fibrosis data include Schluchter *et al.* (2002), van Diemen *et al.* (2011) and Taylor-Robinson *et al.* (2012). Here, we fit a range of joint models to data from the 1980–1984 birth cohort of the UK registry, conditional on survival to capture in the registry in 1999. The data set included 1231 patients who were alive in 1999, of whom 230 died during the course of follow-up to 2010. Our repeated measurement outcome is per cent predicted forced expiratory volume in 1 s, %FEV1, which is used as a measure of lung function and is recognized as a key outcome measure in cystic fibrosis (Rosenfeld *et al.*, 2005; Davies and Alton, 2009; Orens, 2006). Measurements were taken approximately annually, with some variation between patients. The number of measurements per person varied from 1 to 17. The covariates used were sex and age at initial visit. The timescale was the number of years since the initial visit, with initial age fitted as a separate covariate to allow for left truncation and cohort effects. For the survival model the timescale was discretized into yearly intervals. At the initial visit 54.2% of patients were male, the mean age was 18.9 years (standard deviation SD 2.8) and the mean %FEV1 was 67.1 (SD 25.3).

Table 6 shows the results from fitting four random-effects models to the data: a stationary Gaussian process SGP, a stationary Gaussian process with one time lag in the survival model, lagged SGP, a random intercept and slope model IS and a stationary Gaussian process plus random intercept and slope, SGP + IS. The positive effect of the sex covariate in the survival model indicates that males have a significantly better probability of survival than females. The estimated covariate effects are in agreement with expectations for cystic fibrosis patients, with older patients tending to have poorer lung function and survival than younger patients and males tending to have better lung function and to survive longer than females.

For the SGP model, the positive estimate of γ means that better lung function is associated with improved survival in cystic fibrosis patients. For the lagged SGP model there is no evidence

Table 6. Joint model fits of the data from cystic fibrosis patients†

Parameter	Results for the following models:			
	SGP	Lagged SGP	IS	SGP + IS
<i>Longitudinal</i>				
Intercept	74.899 (4.868)	74.904 (4.948)	76.683 (5.069)	76.669 (4.406)
Time	-1.502 (0.074)	-1.502 (0.074)	-1.762 (0.082)	-1.652 (0.080)
Age at t_0	-0.454 (0.250)	-0.454 (0.256)	-0.577 (0.262)	-0.568 (0.218)
Sex (males)	0.786 (1.415)	0.785 (1.377)	1.759 (1.420)	1.506 (1.143)
<i>Survival</i>				
Intercept	2.964 (0.344)	2.962 (0.355)	3.290 (0.422)	3.441 (0.276)
Time	-0.023 (0.011)	-0.023 (0.012)	-0.050 (0.014)	-0.052 (0.013)
Age at t_0	-0.021 (0.017)	-0.021 (0.018)	-0.034 (0.020)	-0.034 (0.013)
Sex (males)	0.234 (0.084)	0.234 (0.084)	0.268 (0.091)	0.287 (0.097)
γ	0.037 (0.002)	0.037 (0.003)	—	0.040 (0.009)
γ_{lag}	—	0.000 (0.001)	—	—
γ_1	—	—	0.035 (0.002)	0.036 (0.004)
γ_2	—	—	0.270 (0.032)	0.418 (0.084)
<i>Others</i>				
ν	7.235 (0.100)	7.235 (0.100)	8.806 (0.084)	7.231 (0.108)
σ_u	25.081 (0.485)	25.081 (0.485)	—	12.832 (1.655)
σ_1	—	—	24.240 (0.523)	20.486 (1.211)
σ_2	—	—	2.152 (0.074)	1.249 (0.108)
ρ_{SGP}	0.969 (0.002)	0.969 (0.002)	—	0.890 (0.031)
ρ_{IS}	—	—	-0.066 (0.042)	0.218 (0.086)
AIC	49551.96	49553.96	49878.47	49522.25

†The random-effects models fitted were a stationary Gaussian process SGP, a stationary Gaussian process with one time lag in the survival model, lagged SGP, a random intercept and slope model IS and a stationary Gaussian process plus random intercept and slope, SGP + IS. For each model estimated parameter values are presented with standard errors in parentheses, and the Akaike information criterion AIC.

that lung function during the previous time interval affects survival, after adjusting for current lung function. The estimated parameters γ_1 and γ_2 from model IS indicate that patients with higher intercepts and less negative slopes of %FEV1 are more likely to survive. Comparing the fit to the data of all four models by using the Akaike information criterion AIC suggests that the model providing the best fit is the model combining both a stationary Gaussian process and a random intercept and slope.

One way to facilitate interpretation of the parameters of the probit model is suggested in Table 7. Here we explore the effect of changing a parameter value on the probability of death in a time interval, while all other parameters are fixed to their default (i.e. mean or baseline) values. For example, the probability of death in year 1 for a woman with default covariate values is 0.003, compared with a probability of 0.001 for a man with the same characteristics.

7. Discussion

We have described a method that combines flexibility of model specification with tractability of likelihood. It can be applied to repeated measurement data with dropout occurring between scheduled measurement times, or to the joint analysis of longitudinal and survival time data,

Table 7. Interpretation of probit parameters†

<i>Survival parameter</i>	<i>Default value</i>	<i>Test value</i>	<i>Probability of death</i>
<i>Time effects</i>			
Year 1			0.003
Year 2			0.003
Year 3			0.004
Year 4			0.004
Year 5			0.005
Year 6			0.006
Year 7			0.006
Year 8			0.007
Year 9			0.009
Year 10			0.010
<i>Covariate effects</i>			
Age at t_0 (years)	18.9	23.9	0.004
Sex	Female	Male	0.001
Stationary Gaussian process U	0	-10	0.008
Random intercept U_1	0	-10	0.007
Random slope U_2	0	-2	0.025

†Default values are the mean or baseline value of each parameter. For the effect of time, the probabilities of death in each time interval are given, at default values of other parameters. For other covariate effects the probability of death in the first year is given, when the parameter indicated is changed to the test value, and all other parameters take their default values.

provided that the survival timescale is discrete, or can realistically be discretized. It avoids the need for numerical approximation of an integral over the random effects or EM methods.

The method allows the fitting of models with more complex random effects because the number of random-effects terms in the survival model is not constrained by computational time. This may come at the cost of discretizing the timescale, but simulation studies and analysis of special cases suggest that, although some information is inevitably lost, parameter estimation may not be greatly influenced by discretization. In the on-line supplementary material we show that there is no loss of information if the survival functions are linear between discrete time points. Hence a discretization that keeps approximate linearity is recommended. Our evidence shows that there can be little loss of efficiency even in the presence of quite strong non-linearity. In practice there may often be a natural discrete timescale. For example the cystic fibrosis patients had visits around once a year, and an annual discretization seems suitable because shorter-term fluctuations in the underlying continuous time hazard will be poorly identified.

Computational time is driven primarily by the need to calculate multivariate normal probabilities, which can be time consuming for high dimensional data. But the intercept and slope model fitted in around a third of the time required by the R package `joiner` using 200 bootstraps to calculate standard errors. Alternative approaches to maximizing the likelihood may enable computation time to be further reduced, e.g. by iterating between a Newton–Raphson step for covariate parameters and numerical maximization over variance parameters. Alternatively a prespecified number of steps of an EM algorithm could be used to obtain initial values, as in the R package `JM` (Rizopoulos, 2010). Even with our current procedure we could fit fairly high dimensional random-effects models in simulations and the cystic fibrosis application (i.e. model SGP, which has a random effect associated with each time interval), whereas current

software (R packages JM and joiner) is limited to simple random-effects models or relies on rough approximations of multiple integrals over random effects. Inclusion of stationary Gaussian process random effects led to a marked improvement in AIC for the cystic fibrosis data (Table 6).

Without readily available software it is unlikely that any reasonably sophisticated statistical methodology will find wide use in applications. Our intention is to develop R software to implement the technique and so to supplement the joint models that can be fitted in JM and joiner. We have provided R code to calculate the likelihood as supplementary material available from

<http://wileyonlinelibrary.com/journal/rss-datasets>

Acknowledgements

We thank John Matthews for drawing our attention to the skew normal results. We thank Adelchi Azzalini, Torsten Hothorn and Alan Genz for advice on the numerical integration of multivariate Gaussian cumulative distribution functions. JB was supported by Medical Research Council grant G0902100. DTR was supported by a Medical Research Council Population Health Scientist Fellowship to him (G0802448). We thank the UK Cystic Fibrosis Trust for access to the UK cystic fibrosis registry and all the staff who input cystic fibrosis data to the UK cystic fibrosis registry. We are grateful for the helpful comments of the Joint Editor and three referees.

Appendix A: Evaluation of the likelihood

A.1. Evaluation of L₁

The longitudinal and random-effects components of the likelihood together are

$$f(y|u) f(u) = \frac{1}{(2\pi)^{(p+n)/2} \nu^n |\Sigma|^{1/2}} \exp\left\{-\frac{(y - X\beta - Au)^T (y - X\beta - Au)}{2\nu^2}\right\} \exp\left(-\frac{u^T \Sigma^{-1} u}{2}\right).$$

This can be rearranged as

$$f(y|u) f(u) = \frac{1}{(2\pi)^{(p+n)/2} \nu^n |\Sigma|^{1/2}} \exp\left\{-\frac{(y - X\beta)^T (y - X\beta)}{2\nu^2} + \frac{h^T Hh}{2}\right\} \exp\left\{-\frac{(u - h)^T H(u - h)}{2}\right\},$$

where

$$H = (A^T A / \nu^2 + \Sigma^{-1}),$$

$$h = H^{-1} A^T (y - X\beta) / \nu^2.$$

In turn

$$f(y|u) f(u) = \frac{1}{(2\pi)^{n/2} \nu^n |\Sigma H|^{1/2}} \exp\left\{-\frac{(y - X\beta)^T (y - X\beta)}{2\nu^2} + \frac{h^T Hh}{2}\right\} \phi^{(p)}(u; h, H^{-1}) \tag{10}$$

where $\phi^{(p)}(\cdot; h, H^{-1})$ denotes the p -dimensional Gaussian density with mean vector h and variance matrix H^{-1} . Collecting all unknown parameters into vector θ , equation (10) can be written as

$$f(y|u) f(u) = L_1(\theta; y) \phi^{(p)}(u; h, H^{-1}), \tag{11}$$

where

$$L_1(\theta; y) = \frac{1}{(2\pi)^{n/2} \nu^n |\Sigma H|^{1/2}} \exp\left\{-\frac{(y - X\beta)^T (y - X\beta)}{2\nu^2} + \frac{h^T Hh}{2}\right\}.$$

A.2. Evaluation of L_2

The full likelihood, integrating over random effects, is

$$L(\theta; y, s, \delta) = \begin{cases} L_1(\theta; y) \int \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)u\} \phi^{(p)}(u; h, H^{-1}) du & \delta = 0, \\ L_1(\theta; y) \int [\Phi^{(s-1)}\{\tilde{X}_{s-1}\tilde{\beta} + \Gamma_{s-1} B(s-1)u\} - \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)u\}] \phi^{(p)}(u; h, H^{-1}) du & \delta = 1. \end{cases} \tag{12}$$

Consider the integral in likelihood (12) at $\delta = 0$, i.e.

$$L_2(\theta; y, s, \delta = 0) = \int \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)u\} \phi^{(p)}(u; h, H^{-1}) du. \tag{13}$$

To perform the integration we use the skew normal result (4), which is

$$\Phi^{(k_1)}(\omega_1; 0, \Sigma_{22}) = \int \Phi^{(k_1)}\{\omega_1 - \Sigma_{12}^T \Sigma_{11}^{-1}(\omega_2 - \eta); 0, \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}\} \phi^{(k_2)}(\omega_2; \eta, \Sigma_{11}) d\omega_2. \tag{14}$$

We can convert the right-hand side of equation (14) to the form (13) by sequentially setting

$$k_1 = s, \quad k_2 = p, \quad \omega_2 = u, \quad \eta = h, \quad \Sigma_{11} = H^{-1}, \quad \Sigma_{12}^T = -\Gamma_s B(s) \Sigma_{11};$$

then

$$\begin{aligned} \Sigma_{22} &= I + \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} \\ &= I + \Gamma_s B(s) H^{-1} B(s)^T \Gamma_s^T \end{aligned}$$

and finally $\omega_1 = \tilde{X}_s\tilde{\beta} + \Gamma_s B(s)h$. Hence

$$L_2(\theta; y, s, \delta = 0) = \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)h; 0, I + \Gamma_s B(s) H^{-1} B(s)^T \Gamma_s^T\}.$$

The contribution of the integral $L_2(\theta; y, s, \delta = 1)$ in expression (12) at $\delta = 1$ can be obtained similarly, to give

$$\begin{aligned} L_2(\theta; y, s, \delta = 1) &= \Phi^{(s-1)}\{\tilde{X}_{s-1}\tilde{\beta} + \Gamma_{s-1} B(s-1)h; 0, I + \Gamma_{s-1} B(s-1) H^{-1} B(s-1)^T \Gamma_{s-1}^T\} \\ &\quad - \Phi^{(s)}\{\tilde{X}_s\tilde{\beta} + \Gamma_s B(s)h; 0, I + \Gamma_s B(s) H^{-1} B(s)^T \Gamma_s^T\}. \end{aligned}$$

Appendix B: Regularity conditions

Collect all unknown parameters into a vector θ , with true value θ_0 in the interior of a compact set Ω . Let the observed data on individual i be z_i , consisting of longitudinal responses, event times, censoring indicator and covariates. Write $f(z_i|\theta)$ for the likelihood contribution of subject i , for $i = 1, 2, \dots, n$.

From Cramér’s theorem (Ferguson, 1996; Kotz and Johnson, 1985) the following conditions are sufficient for the maximum likelihood estimator θ to converge in distribution to a Gaussian variable, centred at θ_0 and with variance given by the inverse expected information.

Condition 1. Z_1, Z_2, \dots, Z_n are independent and identically distributed.

Condition 2. There is an open subset ω of Θ such that for all $\theta \in \omega$ and almost all z the second partial derivatives $\partial^2 f(z|\theta)/\partial\theta \partial\theta^T$ exist and may be passed under the integral sign in

$$\int f(z|\theta) d\nu(z),$$

where $\nu(z)$ is the measure associated with Z .

Condition 3. The Fisher information

$$I(\theta) = -E[\partial^2 f(Z|\theta)/\partial\theta \partial\theta^T]$$

is positive definite for all θ in ω .

Condition 4. There are functions $K(z)$, independent of θ , such that each component of $\partial^2 f(z|\theta)/\partial\theta\partial\theta^T$ is bounded by $K(z)$ uniformly for $\theta \in \omega$.

Condition 5. $f(z|\theta)$ is identifiable, i.e. if $f(z|\theta_1) = f(z|\theta_2)$ almost everywhere then $\theta_1 \neq \theta_2$.

Condition 6. The support of $f(z|\theta)$ does not change with θ .

Conditions 1–4 are standard given our fully parametric Gaussian random-effects model, provided that we assume (as we do) that the longitudinal measurement schedule $\{t_j\}$, the number of discrete intervals m and the selection of covariates are all fixed and ancillary. Further, all variance parameters are finite and strictly positive.

Conditions 5 and 6 need more attention in the presence of missing data. We need to ensure that as the sample size increases there are sufficient observations to identify the survival probability for each discrete time interval, and all parameters in the longitudinal model. Hence we make the following additional assumptions.

Assumption 1. Let θ_Y be the subset of θ that appears in the marginal distribution $f(y|\theta_Y)$ of Y implied by expressions (2) and (3). If $\gamma = 0$, condition 5 holds for $f(y|\theta_Y)$ and θ_Y .

Assumption 2. Any right censoring of event times is independent (Andersen *et al.* (1993), definition II.2.1).

Assumption 3. There exists $c > 0$ such that $P(\Delta = 0) < 1 - c$, where Δ is the censoring indicator.

Assumption 4. There exist $c_1 > 0$ and $c_2 > 0$, independent of θ , such that

$$c_1 < E_X[\Phi(\tilde{X}_s^T \tilde{\beta})] < 1 - c_2$$

for each $s \in \mathcal{S}$ and all $\tilde{\beta}$ in the appropriate subset of ω .

Assumption 5. For all $s \in \mathcal{S}$ and all U , the second-partial-derivative matrix

$$E_{U,X}[\partial^2 \Phi(\tilde{X}_s^T \tilde{\beta} + \Gamma_s B_s U) / \partial\theta_s \partial\theta_s^T]$$

is of full rank, where $\theta_s \in \omega$ denotes the combined parameters $(\tilde{\beta}, \gamma)$ appearing in the event time model (7).

Recalling that the parameter vector γ links the longitudinal and event processes, assumption 1 ensures that the longitudinal model is well defined. The independent censoring assumption referred to in assumption 2 is essentially a requirement that there is no prognostic information in knowing that an event time is censored. Assumption 3 ensures that all event times in \mathcal{S} are observable and assumption 4 makes sure that in large samples there are both events and survivors for each interval. Finally assumption 5 ensures identifiability of $\tilde{\beta}$ and γ .

In assumptions 4 and 5 we have assumed a structural model in which the covariates are considered to be independent and identically distributed random variables. In the functional case we need alternatives of the form

$$c_1 < \frac{1}{n} \sum_{i=1}^n \Phi(\tilde{X}_{is}^T \tilde{\beta}) < 1 - c_2$$

and

$$\frac{1}{n} \sum_{i=1}^n E_U \left[\frac{\partial^2 \Phi(\tilde{X}_{is}^T \tilde{\beta} + \Gamma_s B_s U)}{\partial\theta_s \partial\theta_s^T} \right]$$

is of full rank, in each case for $n > n_0$ say.

References

Albert, J. H. and Chib, S. (2001) Sequential ordinal modelling with applications to survival data. *Biometrics*, **57**, 829–836.

- Albert, P. S. and Follmann, D. A. (2009) Shared-parameter models. In *Longitudinal Data Analysis* (eds G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs), pp. 349–366. Boca Raton: Chapman and Hall–CRC.
- Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993) *Statistical Models based on Counting Processes*. New York: Springer.
- Arnold, B. C. (2009) Flexible univariate and multivariate models based on hidden truncation. *J. Statist. Plannng Inf.*, **139**, 3741–3749.
- Arnold, B. C. and Beaver, R. J. (2000) Hidden truncation models. *Sankhya A*, **62**, 22–35.
- Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171–178.
- Azzalini, A. (2005) The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.*, **32**, 159–200.
- Azzalini, A. (2013) mnormt: the multivariate normal and t distributions. *R Package Version 1.4-7*. (Available from <http://CRAN.R-project.org/package=mnormt>.)
- Davies, J. C. and Alton, E. W. (2009) Monitoring respiratory disease severity in cystic fibrosis. *Respir. Care*, **54**, 606–617.
- van Diemen, C. C., Postma, D. S., Siedlinski, M., Blokstra, A., Smit, H. A. and Boezen, H. M. (2011) Genetic variation in TIMP1 but not MMPs predict excess FEV1 decline in two general population-based cohorts. *Respir. Res.*, **12**, article 57.
- Diggle, P. J., Henderson, R. and Philipson, P. (2009) Random effects models for joint analysis of repeated-measurement and time-to-event outcomes. In *Longitudinal Data Analysis* (eds G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs), pp. 349–366. Boca Raton: Chapman and Hall–CRC.
- Ferguson, T. S. (1996) *A Course in Large Sample Theory*. London: Chapman and Hall.
- Geskus, R. B. (2012) Which individuals make dropout informative? *Statist. Meth. Med. Res.*, to be published.
- 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.
- Heitjan, D. F. and Rubin, D. B. (1991) Ignorability and coarse data. *Ann. Statist.*, **19**, 2244–2253.
- Henderson, R., Diggle, P. and Dobson, A. (2000) Joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465–480.
- Kotz, S. and Johnson, N. L. (eds) (1985) *Encyclopedia of Statistical Sciences*, vol. 5. New York: Wiley.
- Laird, N. M. and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Orens, J. B., Estenne, M., Arcasoy, S., Conte, J. V., Corris, P., Egan, J. J., Egan, T., Keshavjee, S., Knoop, C., Kotloff, R., Martinez, F. J., Nathan, S., Palmer, S., Patterson, A., Singer, L., Snell, G., Studer, S., Vachieri, J. L. and Glanville, A. R. (2006) International guidelines for the selection of lung transplant candidates: 2006 update—a consensus report from the Pulmonary Scientific Council of the International Society for Heart and Lung Transplantation. *J. Hrt Lung Transplnt*, **25**, 745–755.
- Proust-Lima, C., Séne, 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.*, to be published.
- Rizopoulos, D. (2010) JM: an R Package for the joint modelling of longitudinal and time-to-event data. *J. Statist. Softwr.*, **35**, 1–33.
- Rizopoulos, D. (2011) Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, **67**, 819–829.
- Rizopoulos, D. (2012) Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computnl Statist Data Anal.*, **56**, 491–501.
- Rosenfeld, M., Coates, A., Corey, M., Goss, C. H., Howard, M. B., Morgan, W., Quinton, H. and Schluchter, M. (2005) Task Force to Evaluate Choice of Spirometric Reference Equations for the National Patient Registry: summary and recommendations. *Report*. Cystic Fibrosis Foundation Registry Committee.
- Schluchter, M. D., Konstan, M. W. and Davis, P. B. (2002) Jointly modelling the relationship between survival and pulmonary function in cystic fibrosis patients. *Statist. Med.*, **21**, 1271–1287.
- Taylor-Robinson, D., Whitehead, M., Diderichsen, F., Olesen, H. V., Pressler, T., Smyth, R. L. and Diggle, P. (2012) Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax*, **67**, 860–866.
- Tsiatis, A. A. and Davidian, M. A. (2004) Joint modelling of longitudinal and time-to-event data: an overview. *Statist. Sin.*, **14**, 809–834.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997) A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.

Supporting information

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

'Supplementary material 1: efficiency under coarsening at random'.