MEGH: A parametric class of general hazard models for clustered survival data

Francisco Javier Rubio[1] (iD) **and Reza Drikvandi[2]** (iD)

Abstract

In many applications of survival data analysis, the individuals are treated in different medical centres or belong to different clusters defined by geographical or administrative regions. The analysis of such data requires accounting for between-cluster variability. Ignoring such variability would impose unrealistic assumptions in the analysis and could affect the inference on the statistical models. We develop a novel parametric mixed-effects general hazard (MEGH) model that is particularly suitable for the analysis of clustered survival data. The proposed structure generalises the mixed-effects proportional hazards and mixed-effects accelerated failure time structures, among other structures, which are obtained as special cases of the MEGH structure. We develop a likelihood-based algorithm for parameter estimation in general subclasses of the MEGH model, which is implemented in our R package MEGH. We propose diagnostic tools for assessing the random effects and their distributional assumption in the proposed MEGH model. We investigate the performance of the MEGH model using theoretical and simulation studies, as well as a real data application on leukaemia.

Keywords

Accelerated failure time, diagnostic tool, proportional hazards, random effects, survival data

1 Introduction

In the analysis of time-to-event data, it is common to obtain clustered samples. For instance, in medical statistics, one may be interested in modelling the survival times of a sample of patients who receive attention in different health centres with varying characteristics or when the patients live in different geographic or administrative regions. This clustering or grouping information needs to be incorporated in any survival model to avoid under-estimation of the standard errors of the parameter estimators. Moreover, ignoring unobserved heterogeneity, due to clustering, in hazard regression models has an effect on model misspecification as the marginal model obtained by integrating out random effects may not coincide with the model ignoring clustering.[1–3] A simple way to account for clustered survival data consists of fitting a stratified survival model. This is typically done by fitting a proportional hazards (PHs) model for each strata level with different baseline hazards but sharing the same regression coefficients.[4] Alternatively, when the number of clusters is small, the cluster indicator can be coded as a categorical variable. The main disadvantage of these approaches is that they do not account for between-cluster variability. Modelling the clustering effect is more formally done by incorporating random effects, which are random variables that capture the hierarchical structure of the data and account for unobserved variability or heterogeneity between clusters. Regression models, including survival models, that incorporate random effects are often referred to as mixed-effects models (we refer the reader to McCulloch et al.[5] and the references therein for an extensive review of this kind of models). The most common mixed-effects models for time-to-event data correspond to extensions of common regression survival models. For example, the mixed-effects PHs (MEPHs) model[6,7] is an extension of the semi-parametric Cox PHs models. This model is obtained by adding random effects, indexed by the cluster indicator, to the

[1]Department of Statistical Science, University College London, London, UK
[2]Department of Mathematical Sciences, Durham University, Durham, UK

Corresponding author:
Reza Drikvandi, Department of Mathematical Sciences, Durham University, Durham, UK.
Email: reza.drikvandi@durham.ac.uk

linear predictor function in the Cox model. Another common strategy to capture clustering consists of adding random effects to the linear predictor function in parametric and semiparametric accelerated failure time (AFT) models, leading to the so-called mixed-effects AFT (MEAFT) model (see Rubio and Steel[8] and Do Ha and Lee[9] for a review on these models). A related class of mixed-effects models is constructed by adding random effects or Gaussian processes to the linear predictor function in hazard-based (parametric and semiparametric) regression models.[10,11] Some examples of these kinds of models are the mixed-effects spline regression models formulated at the cumulative hazard level by Crowther et al.,[12,13] Wood,[14] or those formulated at the hazard level by Vaida and Xu[15] and Charvat et al.[16] In a related vein, Zhou and Hanson[17] proposed mixed-effects models by adding random effects to PHs, proportional odds and AFT models in a Bayesian semiparametric framework. These classes of models are thus characterised by aiming at capturing the clustering effect via the incorporation of random effects into the linear predictor of different sorts of fixed effects hazard-based regression models.

A natural question about mixed-effects models is the effect of misspecification of the random effects structure on the estimation of quantities of interest. It has been found that misspecifying the random-effects distribution has a negligible effect on the point estimation of the fixed effects parameters of some classes of random effects models.[18] However, there are more tangible effects of this misspecification on interval estimation of the parameters and predictions[8,19] as well as point and interval estimation of the parameters of the baseline hazard.[20] More crucially, ignoring a significant random effect could affect both parameter estimation and inference.[21] Tools for diagnosing misspecification of the random-effects part have been developed for mixed-effects models (without censoring) in recent years (see the authors in[22–24]).

We develop a novel mixed-effects general hazard (MEGH) structure which incorporates random effects into a hazard-based regression model[25,26] at the hazard scale and the time scale. The resulting model can capture clustering affecting the hazard scale and the time scale, and contains the MEPH and MEAFT models as particular cases. We focus our attention on two tractable subclasses of mixed-effects models that generalise the MEPH and MEAFT models. The specification of the MEGH model requires modelling a baseline hazard, for which we employ several flexible parametric distributions. We provide diagnostic tools for the random-effects part of the proposed classes of models in the presence of censoring. We conduct extensive simulation studies that reveal that not only misspecifying the random-effects distribution is problematic, but also misspecifying the role of the random effects in the hazard structure has adverse effects.

The remainder of the paper is organised as follows. In Section 2, we describe the proposed MEGH model and discuss two submodels of interest. In Section 3, we present the expressions for the conditional and marginal likelihood functions, and define the marginal maximum likelihood estimator (MMLE), which defines the estimation approach followed in this paper. We also present theoretical results regarding the consistency and asymptotic normality of the MMLE under standard regularity conditions. Section 4 presents a test for the inclusion of random effects, as well as a graphical diagnostic tool to assess the goodness-of-fit of the random-effects distribution. Section 5 presents an extensive simulation study which illustrates the performance of the proposed methodology as well as an exploration of the effects of misspecifying the distribution and the role of the random effects. Section 6 presents an application with real data in the context of cancer survival with a spatial component. We conclude with a general discussion and possible extensions of our work in Section 7.

2   MEGH: Mixed-effects general hazard models

In this section, we describe the proposed MEGH model, and discuss two general subclasess of interest. To this end, let us first introduce some notation. Let $o_{ij} \in \mathbb{R}_+$, $i = 1, \ldots, r$ indicates the cluster, $j = 1, \ldots, n_i$ denotes the individuals, be a sample of times-to-event of interest, and $\mathbf{x}_{ij} \in \mathbb{R}^p$ be a vector of covariates associated with the ijth individual. Let $c_{ij} \in \mathbb{R}_+$ be the right-censoring times, and $t_{ij} = \min\{o_{ij}, c_{ij}\}$ be the observed survival times. Let $d_{ij} = I(o_{ij} < c_{ij})$ be the vital status indicators (1 - dead, 0 - alive), and $n = \sum_{i=1}^{r} n_i$ be the total sample size across r clusters. We propose the MEGHs model, which is defined through the individual hazard function:

$$h(t_{ij} \mid \mathbf{x}_{ij}, u_i, \tilde{u}_i) = h_0\left(t_{ij} \exp\left\{\tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + \tilde{u}_i\right\}\right) \exp\left\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} + u_i\right\}, \tag{1}$$

where $h_0(\cdot)$ is a baseline hazard, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{\tilde{p}})^\top \in \mathbb{R}^{\tilde{p}}$, $\mathbf{x}_{ij} \in \mathbb{R}^p$ is a vector of covariates affecting the hazard scale, $\tilde{\mathbf{x}}_{ij} \in \mathbb{R}^{\tilde{p}}$ is a vector of covariates affecting the time scale (typically, $\tilde{\mathbf{x}}_{ij} \subset \mathbf{x}_{ij}$), and $(u_i, \tilde{u}_i) \overset{iid}{\sim} G$, where G is a continuous distribution with support on $\mathbb{R}^2$ and zero mean. Denote by $\mathbf{X} = \{\mathbf{x}_{ij}\}$, $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_{ij}\}$, the design matrices, and by $\mathbf{X}_o, \tilde{\mathbf{X}}_o$ the sub-matrices with the rows associated with the uncensored individuals. To avoid collinearity problems, we assume that the matrices associated with uncensored individuals are full column rank. The MEGH model represents an extension of the general hazard (GH) structure by Chen and Jewell,[25] which is obtained by removing the random effects from (1). Clearly, the GH structure contains the PHs model for $\boldsymbol{\alpha} = \boldsymbol{0}$, the AFT model

for $\boldsymbol{\alpha} = \boldsymbol{\beta}$, and the accelerated hazards (AHs) model for $\boldsymbol{\beta} = \boldsymbol{0}$.[25] Consequently, we also need to impose that the baseline hazard h_0 does not belong to the Weibull family, as in this case the GH structure, and consequently the MEGH structure, becomes non-identifiable.[25] Random effects are typically incorporated at the hazard scale (MEPH) or simultaneously at the hazard and time scales (MEAFT). The MEGH structure allows for incorporating random effects in both scales separately. The MEGH structure can also be seen as a generalisation of shared frailty survival models, which aim at accounting for unobserved heterogeneity between clusters.[27,16] Another appealing feature of the MEGH formulation is that the cumulative hazard function can be written in closed-form as follows

$$H(t_{ij} \mid \mathbf{x}_{ij}, u_i, \tilde{u}_i) = H_0\left(t_{ij} \exp\left\{\tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + \tilde{u}_i\right\}\right) \exp\left\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} - \tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + u_i - \tilde{u}_i\right\}.$$

The MEGH structure (1) is a very rich mixed hazard structure that contains, as particular cases, a number of hazard models of interest. In this paper, we particularly focus on the following two subclasses.

2.1 MEGH-I: Mixed Structure I

The first subclass, MEGH-I, is defined by the hazard and cumulative hazard functions

$$
\begin{aligned}
h_1(t_{ij} \mid \mathbf{x}_{ij}, u_i) &= h_0\left(t_{ij} \exp\left\{\tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha}\right\}\right) \exp\left\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} + u_i\right\}, \\
H_1(t_{ij} \mid \mathbf{x}_{ij}, u_i) &= H_0\left(t_{ij} \exp\left\{\tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha}\right\}\right) \exp\left\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} - \tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + u_i\right\}.
\end{aligned}
\tag{2}
$$

The mixed hazard structure (2) contains the MEPH model ($\boldsymbol{\alpha} = \boldsymbol{0}$),[6] while allowing for the inclusion of time-dependent effects through the parameter $\boldsymbol{\alpha}$.

2.2 MEGH-II: Mixed Structure II

The second subclass, MEGH-II, is defined by the hazard and cumulative hazard functions

$$
\begin{aligned}
h_2(t_{ij} \mid \mathbf{x}_{ij}, u_i) &= h_0\left(t_{ij} \exp\left\{\tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + u_i\right\}\right) \exp\left\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} + u_i\right\}, \\
H_2(t_{ij} \mid \mathbf{x}_{ij}, u_i) &= H_0\left(t_{ij} \exp\left\{\tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + u_i\right\}\right) \exp\left\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} - \tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha}\right\},
\end{aligned}
\tag{3}
$$

The MEGH-II structure (3) generalises the MEAFT structure ($\boldsymbol{\alpha} = \boldsymbol{\beta}$).
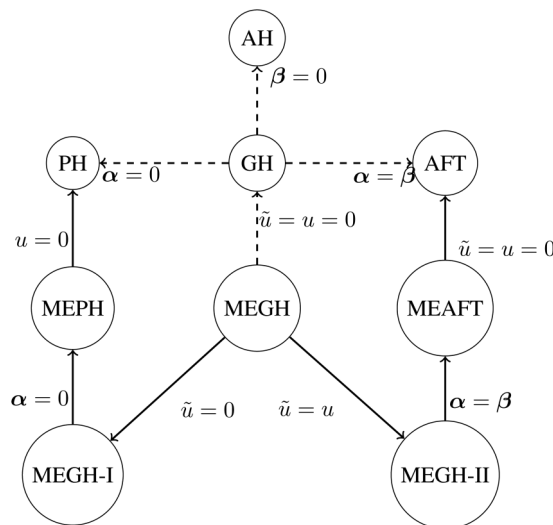

Figure 1. Diagram of the link between the MEGH and some particular hazard structures of interest. (AH: accelerated hazards; AFT: accelerated failure time; PH: proportional hazards; GH: general hazard; MEPH: mixed-effects proportional hazards; MEAFT: mixed-effects accelerated failure time; MEGH: mixed-effects general hazard).

Figure 1 shows the link between the MEGH and some submodels of interest, while the proposed MEGH structure is much richer than illustrated in this figure.

We will adopt parametric baseline hazards $h_0(\cdot \mid \boldsymbol{\theta})$ for models (2)–(3), based on flexible parametric distributions. This is an important step as the choice of the parametric baseline hazard determines the hazard shapes one can obtain. For instance, the log-normal hazard function is unimodal (up-then-down), while the Gamma hazard function can be increasing, decreasing or flat. There exists other (three-parameter) distributions that can capture the basic shapes of the hazard (increasing, decreasing, unimodal, and bathtub), such as the Exponentiated Weibull, Generalised Gamma and Power Generalised Weibull (PGW). However, it is important to keep in mind that an efficient estimation of the parameters of more flexible distributions typically requires larger sample sizes.[26,28] On the other hand, parametric models have been found to perform well, compared to semiparametric counterparts, in the context of clustered survival data.[29]

3 Likelihood function and parameter estimation

We here describe the conditional and marginal likelihood functions, which will be used to calculate the marginal maximum likelihood estimators as well as to develop the diagnostic tools presented in the next section. We discuss the details behind the calculation of the marginal likelihoods and the optimisation methods. Finally, we present a theoretical result on the consistency and asymptotic normality of the marginal maximum likelihood estimators.

3.1 Likelihood and marginal likelihood functions

Let $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})$ be the sample associated with the ith cluster and $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$ denote the corresponding matrix of covariates. The cluster-specific marginal likelihood function associated with the ith cluster, after integrating out the random effects u_i and $\tilde{u}_i$, is given by

$$\int_{\mathbb{R}^2} \prod_{j=1}^{n_i} h(t_{ij} \mid \mathbf{x}_{ij}, u_i, \tilde{u}_i)^{d_{ij}} \exp\left\{-H(t_{ij} \mid \mathbf{x}_{ij}, u_i, \tilde{u}_i)\right\} dG(u_i, \tilde{u}_i).$$

For the case where the random-effects distribution G is assumed to be a multivariate distribution with mean $\mathbf{0}$ and finite variance, with parameters $\boldsymbol{\xi}$, we can write down the marginal likelihood function of the MEGH model (1) in terms of the hazard and cumulative hazard functions as follows. To simplify notation, let us denote $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\xi})$ and let Γ be the parameter space. The marginal likelihood function is

$$m(\boldsymbol{\eta}) = \prod_{i=1}^{r} m_i(\boldsymbol{\eta}), \tag{4}$$

where m_i represents the cluster-specific marginal likelihood associated with the ith cluster

$$m_i(\boldsymbol{\eta}) = \int_{\mathbb{R}^2} \exp\left\{\ell_i(\boldsymbol{\eta}, u_i, \tilde{u}_i)\right\} dG(u_i, \tilde{u}_i; \boldsymbol{\xi}), \tag{5}$$

and

$$\ell_i(\boldsymbol{\eta}, u_i, \tilde{u}_i) = \sum_{j=1}^{n_i} d_{ij} \log h(t_{ij} \mid \mathbf{x}_{ij}, u_i, \tilde{u}_i) - \sum_{j=1}^{n_i} H(t_{ij} \mid \mathbf{x}_{ij}, u_i, \tilde{u}_i),$$

denotes the log-likelihood function conditional on the random effects u_i and $\tilde{u}_i$. The marginal maximum likelihood estimator (MMLE) is then defined as $\widehat{\boldsymbol{\eta}} = argmax_\Gamma m(\boldsymbol{\eta})$.

3.2 Computations

To calculate the MMLE $\widehat{\boldsymbol{\eta}}$, we need to evaluate the marginal likelihood function (4), which is a product of integrals that define the cluster-specific marginal likelihoods (5). The integrand in (5) can be small when the number of individuals belonging to that cluster is moderate or large. To deal with this numerical challenge, we scale the integrand by using

$$m_i(\boldsymbol{\eta}) = K_i \int_{\mathbb{R}^2} \exp\left\{\ell_i(\boldsymbol{\eta}, u_i, \tilde{u}_i) - \log K_i\right\} dG(u_i, \tilde{u}_i),$$

where $\log K_i = \max_{u_i, \tilde{u}_i} \ell_i(\boldsymbol{\eta}, u_i, \tilde{u}_i)$, for fixed values of $\boldsymbol{\eta}$. This allows for a more stable numerical integration, as the integrand is now bounded between $(0, 1]$ (always taking the value 1 at the maximum). Maximisation over the random effects is simpler as their variance is rarely large. Numerical integration in our empirical examples, which include only one hierarchical level, is done by using the R command `integrate` with default options. Nonetheless, other methods, such as Laplace approximations or Monte Carlo integration, could also be used. The optimisation step is carried out using standard R routines, such as those implemented in the commands `optim` and `nlminb`. The R package `MEGH` (https://github.com/FJRubio67/MEGH) implements the MEGH-I and `MEGH`-II models under these specifications.

3.3 Theory

Here, we prove the consistency and asymptotic normality of the MMLE, when $r \to \infty$ (which implicitly implies $n \to \infty$), under Assumptions A1–A7 given in the online Supplemental Material. We denote by $\boldsymbol{\eta}^\star = (\boldsymbol{\beta}^\star, \boldsymbol{\alpha}^\star, \boldsymbol{\theta}^\star, \boldsymbol{\xi}^\star)$ the unknown true value of the parameters, which is assumed to be an interior point of the parameter space Γ.

Theorem 1 Consider the `MEGH` model (1). Let $h_0(\cdot \mid \boldsymbol{\theta})$ be a pre-specified parametric baseline hazard as in Section 2. Under Assumptions A1–A7 given in the online Supplemental Material, it follows that

(i) Consistency: $\widehat{\boldsymbol{\eta}} \xrightarrow{P} \boldsymbol{\eta}^\star$ as $r \to \infty$,
(ii) Asymptotic normality: $\sqrt{r}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^\star) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}(\boldsymbol{\eta}^\star)^{-1})$ as $r \to \infty$,

where $\mathbf{I}(\boldsymbol{\eta}^\star)$ is the expectation matrix $\mathbf{I}(\boldsymbol{\eta}) = cov[\nabla_{\boldsymbol{\eta}} \log m_1(\boldsymbol{\eta})]$ evaluated at $\boldsymbol{\eta}^\star$.

Regarding the required Assumptions A1–A7, we note that Assumption A1 is a restriction to a compact parameter space, which can be assumed to be a large compact set. Assumption A2 concerns the censoring mechanism, which is assumed to be non-informative. Assumption A3 guarantees identifiability and continuity of the `MEGH` model. Assumption A4 togheter with Assumptions A5–A7 represent standard integrability and differentiability conditions on the marginal likelihood function, which implicitly impose conditions on the choice of the baseline hazard. Assumption A6 assumes the existence and non-singularity of the information matrix. Verifying such condition in practice for specific choices of the parametric baseline hazard is complicated. However, it has been shown that certain smoothness and identifiability conditions, which are easier to verify in practice, guarantee the non-singularity of this matrix (see Asgharian[30]). Similar asymptotic results have also been obtained for parametric copula survival models for clustered data in Prenen et al.,[29] where parametric models were found to perform better than their semiparametric counterparts.

4 Diagnostic tools for random effects in the `MEGH` model

As already discussed, the proposed `MEGH` model (1) incorporates random effects to capture the unknown variability between clusters or groups in the survival data. Since random effects are latent and unobservable variables, their assumed distribution can be subject to misspecification, so it is important to check their distributional assumption to identify potential model misspecification. It is also important to test which random effects are statistically significant and should be included in the model. These two problems are crucial especially for practical use, so we study them for the `MEGH` model in the following two subsections, respectively.

4.1 Detecting misspecification of the random-effects distribution

We here present a diagnostic tool for detecting misspecification of the random-effects distribution in the `MEGH` model. The assumed random-effects distribution G is typically a multivariate normal distribution. To check the appropriateness of an assumed random-effects distribution,[31] suggested to use the so-called gradient function for models with random effects. Their diagnostic tool based on a gradient function is not directly applicable to the `MEGH` model (1). We extend the gradient function approach for the `MEGH` model. For this purpose, we use the key idea of the gradient function, that is, to check if the marginal log-likelihood of the model can be increased considerably by replacing G with another random-effects distribution, say W. If so, then the assumed distribution G may not be adequate for the random effects because another random-effects distribution produces a considerably larger marginal likelihood.

Recall that $\eta = (\beta, \alpha, \theta, \xi)$ is the combined vector of the model parameters. We here write the marginal likelihood as $m(G)$, instead of $m(\eta)$, to emphasise its dependence on the assumed random-effects distribution G. To check if the marginal log-likelihood can be increased considerably by replacing G with another random-effects distribution W, we use the directional derivative of the marginal log-likelihood $\log m(.)$ at G into the direction W as follows

$$\Phi(G, W) = \lim_{\varepsilon \,\substack{> \\ \to}\, 0} \frac{\log m\big((1-\varepsilon)G + \varepsilon W\big) - \log m(G)}{\varepsilon}$$

$$= \frac{\partial \log m\big((1-\varepsilon)G + \varepsilon W\big)}{\partial \varepsilon}\Big|_{\varepsilon=0}.$$

Then, there is no better random-effects distribution than G if $\Phi(G, W) \le 0$ for all W. We can write that

$$\frac{1}{r}\Phi(G, W) = \frac{1}{r}\frac{\partial \sum_{i=1}^{r} \log\left[(1-\varepsilon)m_i(G) + \varepsilon m_i(W)\right]}{\partial \varepsilon}\Big|_{\varepsilon=0}$$

$$= \frac{1}{r}\sum_{i=1}^{r} \frac{m_i(W) - m_i(G)}{m_i(G)}$$

$$= \frac{1}{r}\sum_{i=1}^{r} \frac{\int_{\mathbb{R}^2} \exp\{\ell_i(u, \tilde{u})\}dW(u, \tilde{u})}{m_i(G)} - 1$$

$$= \int_{\mathbb{R}^2}\Big[\frac{1}{r}\sum_{i=1}^{r} \frac{\exp\{\ell_i(u, \tilde{u})\}}{m_i(G)}\Big]dW(u, \tilde{u}) - 1$$

$$= \int_{\mathbb{R}^2} \Delta\big(G, (u, \tilde{u})\big)dW(u, \tilde{u}) - 1,$$

in which $\Delta(G, (u, \tilde{u}))$ is the gradient function for the MEGH model, defined as

$$\Delta\big(G, (u, \tilde{u})\big) = \frac{1}{r}\sum_{i=1}^{r} \frac{\exp\{\ell_i(u, \tilde{u})\}}{m_i(G)}, \quad (u, \tilde{u}) \in \mathbb{R}^2. \tag{6}$$

Hence, we have $\phi(G, W) \le 0$ for all W, if $\Delta(G, (u, \tilde{u})) \le 1 \;\forall\,(u, \tilde{u}) \in \mathbb{R}^2$. The gradient function (6) has useful properties. The calculation of the gradient function does not require specifying an alternative distribution W. Also, since $\Phi(G, G) = 0$, we must have $\Delta(G, (u, \tilde{u})) = 1$ for all random effects points $(u, \tilde{u})$ in the support of G. As a graphical diagnostic tool, we plot the gradient function versus random effects points $(u, \tilde{u})$, and if the gradient plot does not exceed 1 then the assumed random-effects distribution G will be deemed to be adequate for random effects; otherwise, the random-effects distribution is deemed to be misspecified. See[31] and[32] who applied a similar approach to generalised linear and nonlinear mixed models. The calculation of the gradient function (6) is straightforward as it only requires the calculation of the marginal and conditional distributions for all r clusters. Note that the dependence of the gradient function on the parameters $\eta = (\beta, \alpha, \theta, \xi)$ is suppressed for simplicity of the notation, however, we replace these parameters with their estimates when calculating the gradient function.

4.2 Tests for the need of random effects in the MEGH model

Since random effects are latent and unobservable variables, it is not obvious in practice which random effects must be included in the model. As an initial check, one can use a plot of the survival curves for all clusters to get some understanding of the variability between clusters and the potential need of random effects (see an example in Figure 4). We provide formal tests for verifying which random effects are statistically significant and must be present in the MEGH model. If the test confirms that all the random effects are not significant, then the extended model (1) will not be essential and one could use the reduced models without random effects as in Chen and Jewell[25] and Rubio et al.[26] The test for inclusion or exclusion of random effects can be carried out via a test for zero random effects. From a statistical perspective, to test for zero random effects is equivalent to testing if their variance parameters are equal to 0. We want to test whether or not the random effects u_i and $\tilde{u}_i$ can be excluded from the MEGH model (1). This hypothesis testing problem can be expressed as follows

$$\begin{cases} H_0 : \sigma_u^2 = 0 \text{ and } \sigma_{\tilde{u}}^2 = 0 \\ H_1 : \sigma_u^2 > 0 \text{ or } \sigma_{\tilde{u}}^2 > 0. \end{cases} \tag{7}$$

This is a non-standard testing problem because the null hypothesis places the true variance parameters on the boundary of the parameter space.[33–36] Classical tests such as the likelihood ratio, Wald and score tests suffer from testing on the boundary of the parameter space, because the regularity conditions do not hold under such situations. As a consequence, the usual asymptotic chi-squared distribution of the likelihood ratio or score statistic is incorrect here. The correct asymptotic distribution is well understood to be a mixture of chi-squared distributions when $r \to \infty$ under some mild conditions.[33,34]

The hypotheses H_0 and H_1 in (7) correspond to Case 7 of Self and Liang.[33] Thus, under H_0, the correct asymptotic distribution of the likelihood ratio statistic for testing (7) would be $\frac{1}{4}\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2$, where χ_0^2 is a point mass at 0, and χ_1^2 and χ_2^2 are the chi-squared distributions with one and two degrees of freedom, respectively. The p-value of the likelihood ratio test for the hypothesis test (7) is as follows (see also Verbeke and Molenberghs[37])

$$p = \frac{1}{4}P(\chi_0 \geq R_{\text{obs}}) + \frac{1}{2}P(\chi_1 \geq R_{\text{obs}}) + \frac{1}{4}P(\chi_2 \geq R_{\text{obs}}), \tag{8}$$

where R_{obs} denotes the observed likelihood ratio statistic.

For the MEGH model with the mixed hazard structures (2)–(3), since u_i and $\tilde{u}_i$ are the same, the hypothesis test (7) collapses to

$$\begin{cases} H_0 : \sigma_u^2 = 0 \\ H_1 : \sigma_u^2 > 0. \end{cases}$$

The correct asymptotic distribution of the likelihood ratio statistic for this hypothesis test is $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, as it corresponds to Case 5 of Self and Liang.[33]

We will apply the above diagnostic tools to our case study in Section 6. We note that there are also non-asymptotic tests available for testing random effects, which avoid the boundary issue. Bootstrap and permutation tests as well as Bayesian tests do not rely on asymptotic results and hence avoid boundary issues when testing random effects. Those tests have been well studied for linear and generalised linear mixed models[24,36,38]; however, such tests may not be easily applied to the case of MEGH model, due to the presence of censored and clustered data which makes bootstrap or permutation challenging. Furthermore, Verbeke and Molenberghs[37] and Claeskens et al.[39] provide asymptotic score tests that can be applied in this context.

5 Simulation studies

In this section, we conduct simulations to evaluate the performance of the proposed MEGH model. In the simulations, we first examine the accuracy of parameter estimation as well as statistical inference with the MEGH model under different scenarios. We investigate how the MEGH model compares with the model ignoring random effects. Such comparisons are crucial in demonstrating the advantages of the proposed MEGH model over the models ignoring random effects. We then assess the impact of misspecifying the mixed structure as well as misspecifying the random-effects distribution. We also examine the finite-sample performance of the diagnostic tests for testing random effects in the MEGH model. Some of the simulation results are reported in the online Supplemental Material for the sake of space.

In the simulations, we study both mixed hazard structures (2)–(3) and consider the following MEGH model, according to our real data application in Section 6,

$$h(t_{ij} \mid \mathbf{x}_{ij}, u_i, \tilde{u}_i) = h_0\left(t_{ij} \exp\left\{\alpha_1 \text{age}_{ij} + \tilde{u}_i\right\}\right) \exp\left\{\beta_1 \text{age}_{ij} + \beta_2 \text{sex}_{ij} + \beta_3 \text{wbc}_{ij} + \beta_4 \text{tpi}_{ij} + u_i\right\}, \tag{9}$$

where the covariates $\mathbf{x}_{ij}^{\top} = (\text{age}_{ij}, \text{sex}_{ij}, \text{wbc}_{ij}, \text{tpi}_{ij})$ are in accordance with the leukaemia data in Section 6, and $h_0(\cdot)$ is the baseline hazard for which we here consider the PGW baseline hazard with parameters $\boldsymbol{\theta} = (\eta, \nu, \delta)$ as in the online Supplemental Material. We also conduct simulations with the log-logistic baseline hazard, which are presented in the Supplemental Material.

Recall that $\tilde{u}_i = 0$ for the mixed hazard structure MEGH-I, and $\tilde{u}_i = u_i$ for the mixed hazard structure MEGH-II. In the simulations, we choose the true parameter values according to the estimates obtained from the leukaemia data application. The number of clusters is 24. We simulate 250 data sets from the above model with each of the hazard structures MEGH-I and MEGH-II, each with the same size as of the real data application (i.e. $n = 1043$), using our R package MEGH available in the online Supplemental Material. The random effects are generated from a normal distribution with mean 0 and two standard deviation values of $\sigma_u = 0.5$ and $\sigma_u = 1$. The censoring rate is set to 25%.
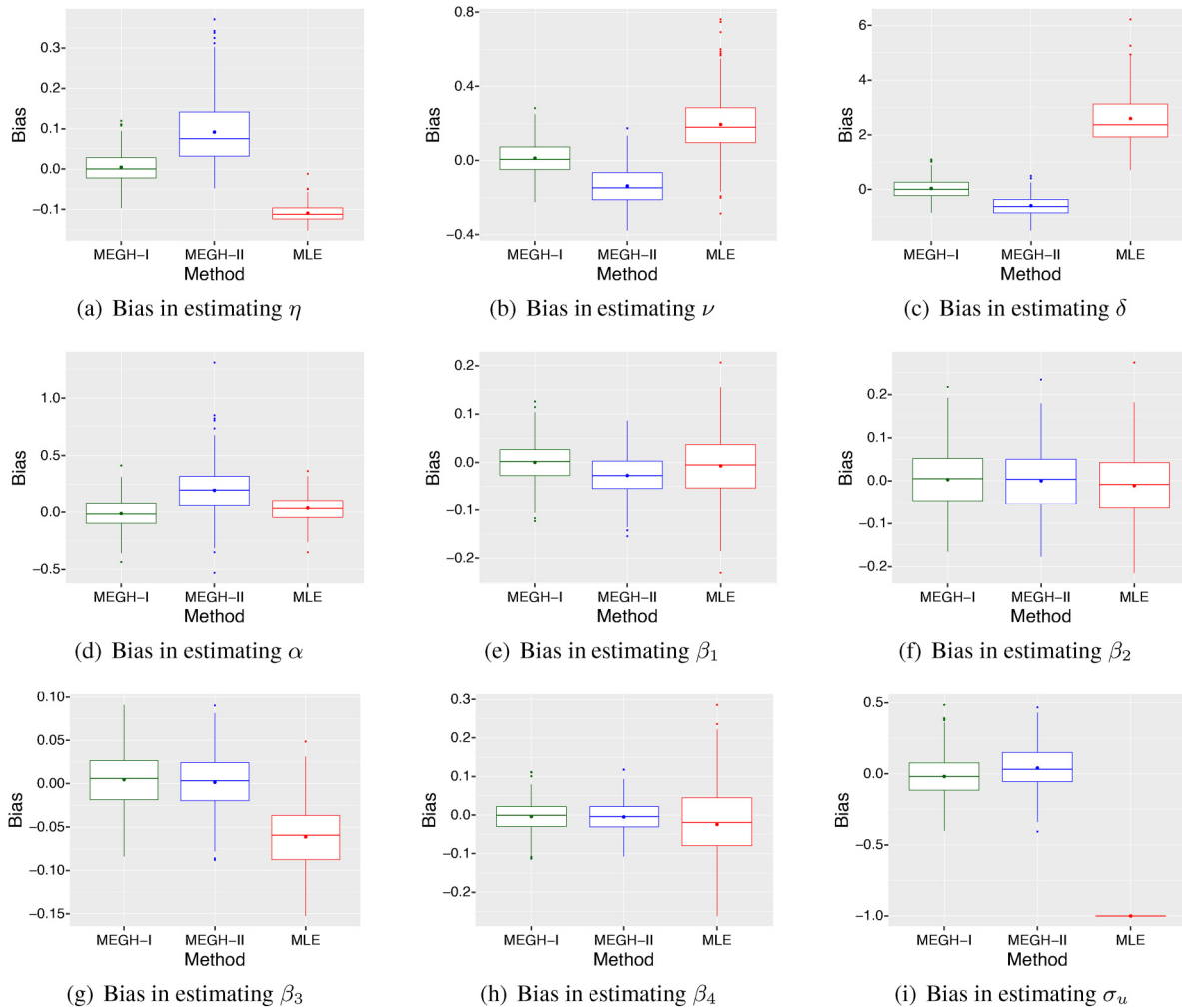
Figure 2. The bias of the estimates from the three methods: MEGH-I, MEGH-II and MLE, for all the parameters based on 250 simulation replications when the simulated data are generated from model (9) with the mixed structure I and PGW baseline hazard, and a normal distribution for the generated random effects with $\sigma_u = 1$.

We fit the MEGH model with both mixed hazard structures (2)–(3) separately to the 250 simulated data sets. We also fit the model ignoring random effects to the simulated data sets. We calculate the bias of the parameter estimates from each of these three models. Figures 2 and 3 show the estimation bias for these three models. It can be seen that both MEGH models with mixed hazard structures (2)–(3) produce estimates with considerably smaller bias compared to the model ignoring random effects. The bias of the model ignoring random effects for most parameters is higher due to neglecting the presence of random effects. However, we note that the estimates obtained from models with or without random effects are not generally comparable like with like, as discussed by Lee and Nelder.[40] This is because the model with random effects is conditional on random effects, while the model without random effects is marginal (or population-average).

We also observe from the simulation results presented in Figures 2 and 3 that when the mixed hazard structure is misspecified it leads to relatively higher bias compared to the model with the correct mixed hazard structure. This suggests that the choice of the mixed hazard structure is crucial for achieving a reliable model fit and results. We note that the bias of the MEGH model with the correct mixed hazard structure is very small in the simulations.

Table 1 presents the 95% confidence intervals for all the regression parameters obtained from the three models considered for the simulation case when the true mixed structure is MEGH-I. The results show that the model with the correct mixed structure MEGH-I produces reliable confidence intervals, but both the model ignoring random effects and the model with the incorrect mixed structure MEGH-II tend to produce inaccurate confidence intervals. In particular, the confidence intervals from the model ignoring random effects mostly do not cover the true parameter values.
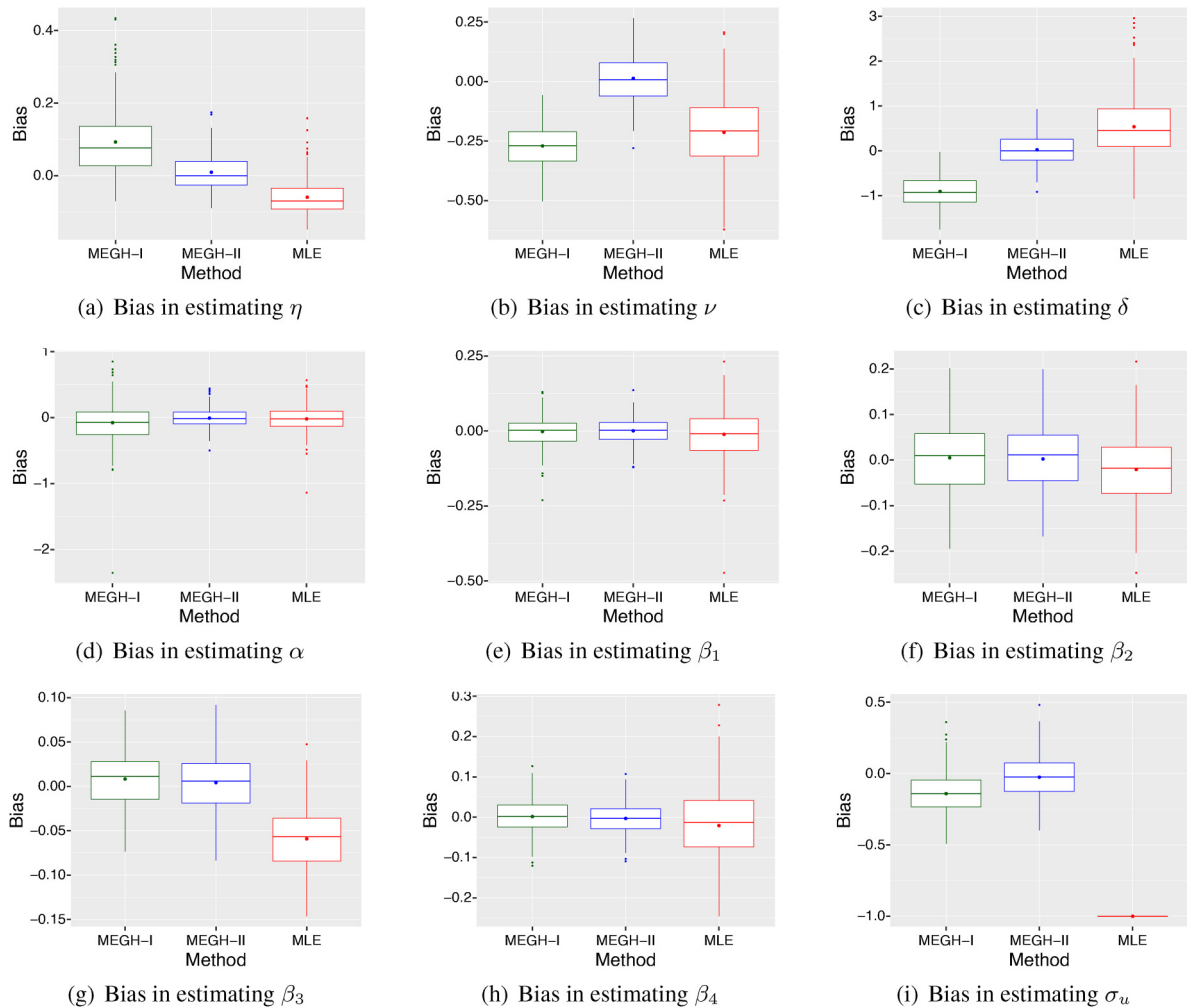
Figure 3. The bias of the estimates from the three methods: MEGH-I, MEGH-II and MLE, for all the parameters based on 250 simulation replications when the simulated data are generated from model (9) with the mixed structure II and PGW baseline hazard, and a normal distribution for the generated random effects with $\sigma_u = 1$.

Table 2 shows the average AIC of the model fit across the 250 simulation replications for the three models considered. The results indicate that the MEGH model with the correct mixed hazard structure has the smallest AIC, while the model ignoring random effects tends to produce a very large AIC. Table 2 also presents the average power of the likelihood ratio test for random effects across the 250 simulation replications, when the random effects are generated from a normal distribution with $\sigma_u = 1$. It can be seen that the test provides a high power with this sample size. Further simulation results on the performance of the test for random effects under different number of clusters, censoring rates and variance values can be found in the Supplemental Material.

6 Real data application

In this section, we apply the MEGH model to analyse the data set LeukSurv, which contains information on the survival of $n = 1043$ patients of acute myeloid leukaemia in northwest England, recorded between 1982 and 1998. This data set is available in the R package spBayesSurv.[17] Previous analyses of this data set have suggested that there is evidence of variation in survival across this region,[17] and we can see that this is also suggested by the nonparametric Kaplan-Meier estimators of the survival curves by district shown in Figure 4. We use information about the survival time (in years), vital status at the end of follow-up (0 – right-censored, 1 – dead), age (in years), sex (0 – female, 1 – male), white blood cell count at diagnosis (wbc, truncated at 500), the Townsend score (tpi, higher values indicates less affluent areas), and administrative district of residence (district, 24 districts). We fit the MEGH models with

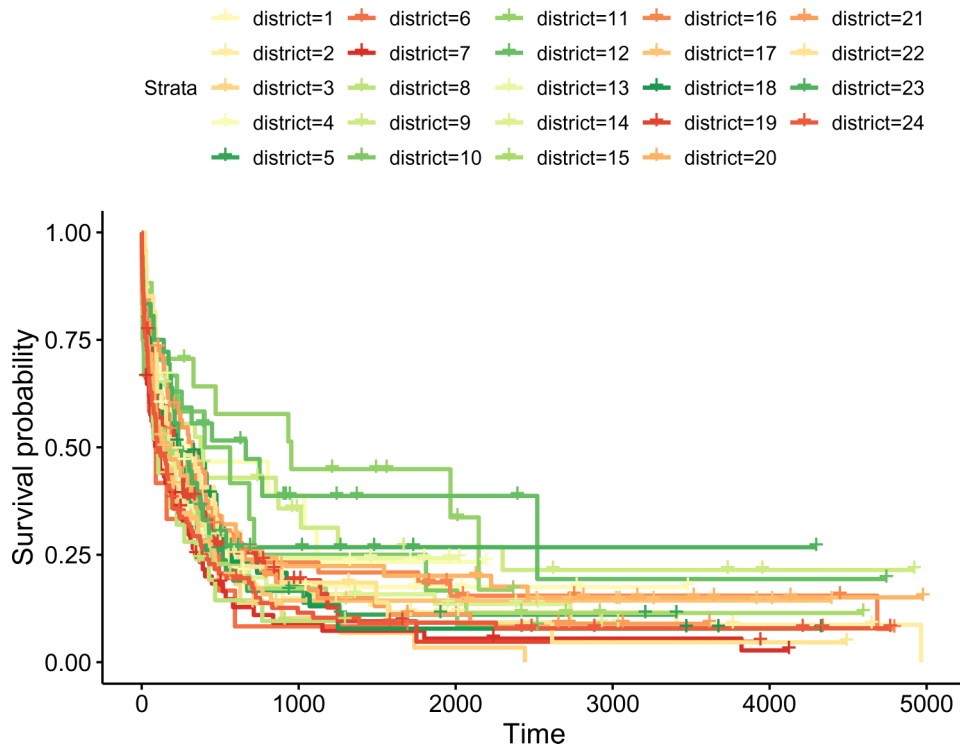**Figure 4.** Leukaemia data: Kaplan-Meier estimators of survival by district.

Table 1. The 95% confidence intervals for all the regression parameters from the three methods: MEGH-I, MEGH-II and MLE, for the case when the true mixed structure is MEGH-I and the random effects are generated from a normal distribution with $\sigma_u = 1$.

Parameter	True value	Method	95% confidence interval
		MLE	(0.0880, 0.0933)
η	0.20	MEGH-I	(0.1999, 0.2089)
		MEGH-II	(0.2815, 0.3017)
		MLE	(1.6734, 1.7146)
ν	1.50	MEGH-I	(1.5012, 1.5244)
		MEGH-II	(1.3492, 1.3762)
		MLE	(5.4804, 5.7141)
δ	3.00	MEGH-I	(2.9988, 3.0882)
		MEGH-II	(2.3658, 2.4574)
		MLE	(0.9812, 1.0106)
α	0.96	MEGH-I	(0.9307, 0.9643)
		MEGH-II	(1.1294, 1.1829)
		MLE	(0.9847, 1.0010)
β_1	1.00	MEGH-I	(0.9951, 1.0054)
		MEGH-II	(0.9676, 0.9782)
		MLE	(0.0596, 0.0785)
β_2	0.08	MEGH-I	(0.0741, 0.0919)
		MEGH-II	(0.0713, 0.0891)
		MLE	(0.1540, 0.1633)
β_3	0.22	MEGH-I	(0.2204, 0.2289)
		MEGH-II	(0.2172, 0.2258)
		MLE	(0.0641, 0.0872)
β_4	0.10	MEGH-I	(0.0914, 0.1009)
		MEGH-II	(0.0901, 0.1001)

Table 2. Average AIC of the model fit and average power of the test for random effects across 250 simulation replications, when the random effects are generated from a normal distribution with $\sigma_u = 1$.

True structure	Fitted model	AIC	Power
	MLE	1390.66	—
MEGH-I	MEGH-I	927.84	1.0
	MEGH-II	975.79	1.0
	MLE	1383.18	—
MEGH-II	MEGH-I	996.35	1.0
	MEGH-II	962.57	1.0

hazard structures MEGH-I (2) and MEGH-II (3), and with the PGW and log-logistic baseline hazards. For the random-effects distribution, we use the normal, Student-*t* and two-piece normal distributions (to account for heavier tails than normal and asymmetry). In all fitted models, we consider the time dependent effect ($\tilde{\mathbf{x}}_{ij}$) of age (standardised), and the hazard-level effects ($\mathbf{x}_{ij}$) of age (standardised), sex, white blood cell count at diagnosis (wbc, standardised) and the Townsend score (tpi, standardised). The variable district is used to define the random effects in the two mixed hazard structures considered. In addition, we fit the models ignoring random effects for comparison. We also assess the need for including random effects using the diagnostic tests in Section 4.2.

The best model selected using AIC (1553.725) is the model with the mixed structure MEGH-I, log-logistic baseline hazard (with log-location and scale parameters $(\theta_1, \theta_2) = (\mu, \tau)$) and normal random effects. The AIC for the model ignoring random effects with log-logistic baseline hazard is 1556.366, which is fairly close to the AIC value for the best model. The reason for this is that the estimate of the variance of random effects is relatively small ($\widehat{\sigma}_u = 0.144$), and the number of clusters ($r = 24$) is not large in this data set. However, the p-value for testing $H_0: \sigma_u = 0$ is 0.0156, which suggests that the random effect is significant and must be present in the model. This implies that there is significant between-cluster variability after incorporating the information on the observed covariates (see also the Box and Whisker plot in the online Supplemental Material). Figure 5 shows the maps of aggregated summaries at district level. We notice from Figure 5(c) and (d) that the marginal survival functions (which is obtained as the average of cluster-specific marginal survival functions using Monte Carlo integration) at $t = 1, 5$ years exhibit some variability by district. Figure 5(a) and (b) shows that the distribution of mean age and mean Townsend score is also quite different for the different districts. Consequently, survival gaps between different districts are explained by a combination of complex factors, including different age and Townsend distribution, a conclusion that could be harder to obtain with fully nonparametric methods that require data stratification.

The gradient function plot for the model MEGH-I fitted under normal random effects is presented in Figure 6(a), together with the 95% confidence bands similar to Verbeke and Molenberghs,[31] which indicates that the normality assumption on random effects is valid since the gradient values remain under or close to 1. To double check whether the fit can be improved by using a distribution for random effects which allows to capture asymmetry, we calculate the gradient function plot with the two-piece normal distribution for random effects. The gradient plot with the 95% confidence bands for the two-piece normal distribution, shown in Figure 6(b), is very similar, suggesting that the usual normal distribution is adequate for random effects.

7 Discussion

We have introduced a general mixed-effects hazard structure (MEGH), as well as two general subclasses of interest (MEGH -I and MEGH-II). The proposed structures generalise classical survival regression models of interest in practice, such as the MEPH and MEAFT. We have adopted a flexible parametric approach, which allowed us to prove asymptotic results under standard regularity conditions. Our simulation studies show that the estimates of the parameters of the proposed models, based on the marginal likelihood, have good frequentist properties. Another contribution of this work consists of evaluating the impact of misspecifying the role of the random effects and potentially the random-effects distribution. Our simulation study shows that not only misspecifying the distribution of the random effects has an impact on the quality of the inference on the parameters, but also misspecifying the role of the random effects in the hazard regression model, a topic that has received little attention in survival analysis. The routines used to produce our results are implemented in the R package MEGH which is available, along with the real data application, at https://github.com/FJRubio67/MEGH. In our simulations and applications, we have focused on the use of the MEGH-I and MEGH-II structures, as these represent generalisations of the most common
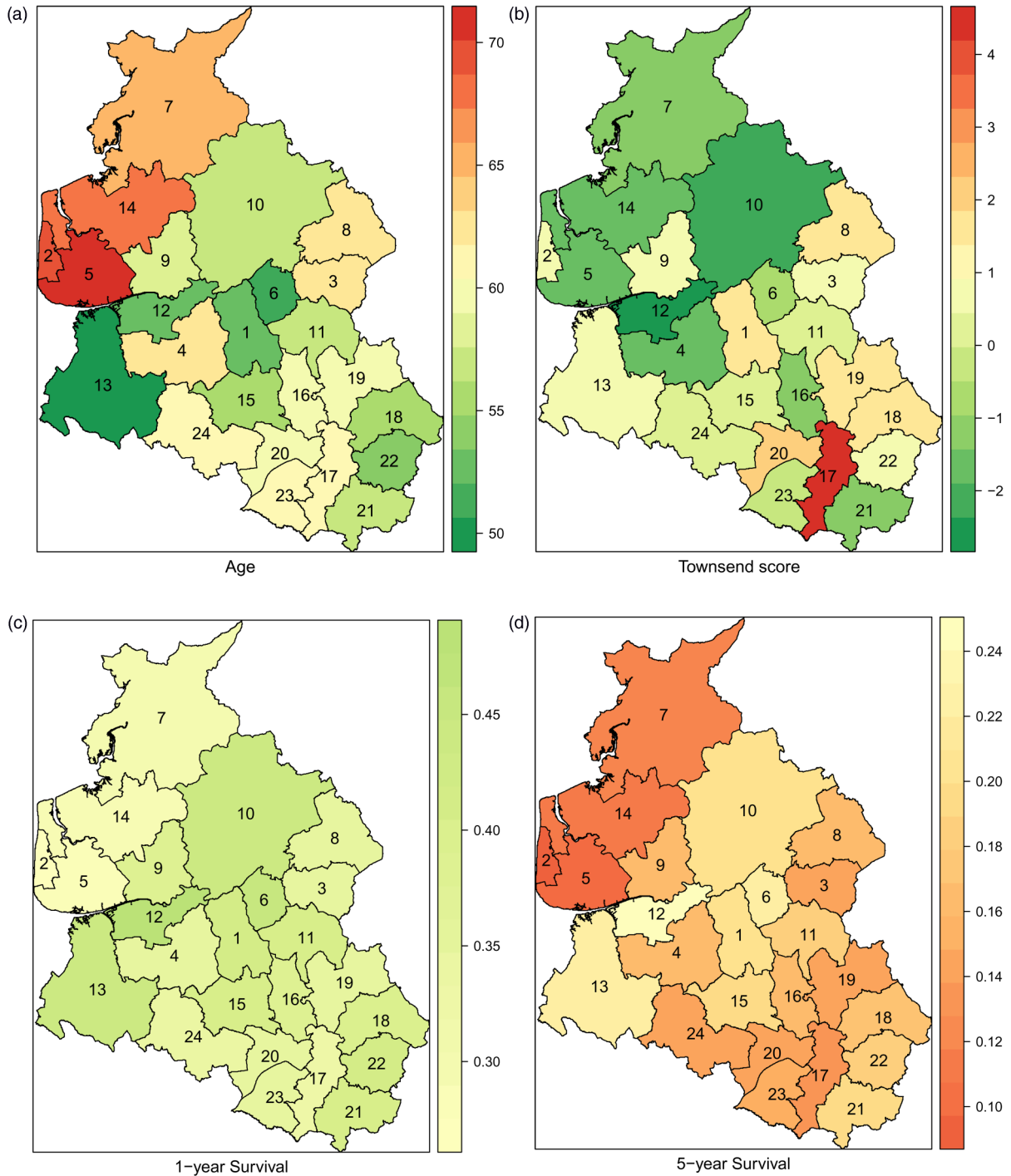
Figure 5. Leukaemia data: Maps showing the aggregated summaries for each district: (a) mean age, (b) mean Townsend score, (c) marginal 1-year survival and (d) marginal 5-year survival.

mixed models in survival analysis. However, the implementation of the general structure (1), with two random effects, represents a possible extension of this work. Such extension does not only bring computational challenges, as one requires double numerical integration, but also the modelling of the dependence between the two random effects. The latter can be explored by using copulas to capture different types of dependencies, and a variety of parametric marginal distributions.[8]
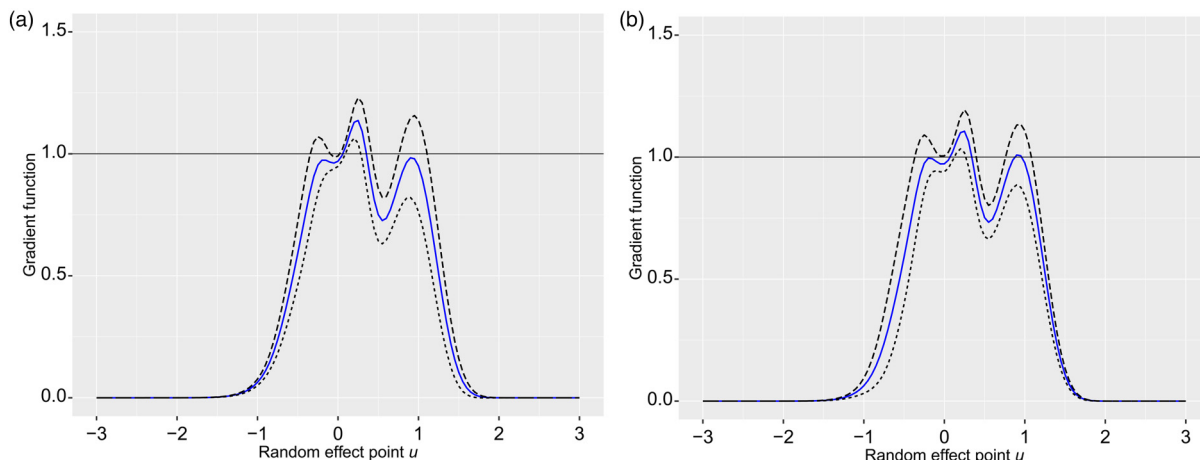
Figure 6. Leukaemia data: (a) Gradient function plot with the 95% confidence bands for the MEGH model (2) with normal random effects fitted to the leukaemia data. (b) Gradient function plot with the 95% confidence bands for the MEGH model (2) with a two-piece normal distribution for the random effects.

The tractability of the MEGH model allows for several extensions, such as accounting for left-censored or interval censored times. A more substantial extension corresponds to the case where the distribution of the random effects reflects a spatial multilevel structure.[41] Another possible extension consists of modelling the baseline hazard using nonparametric (frequentist or Bayesian) methods.[25] Extensions to the inclusion of more than one hierarchical level would also be of practical interest. This could be done simply by considering multivariate random effects

$$h(t_{ij} \mid \mathbf{x}_{ij}, \mathbf{u}_i, \tilde{\mathbf{u}}_i) = h_0 \left(t_{ij} \exp \left\{ \tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + \tilde{\mathbf{z}}_{ij}^\top \tilde{\mathbf{u}}_i \right\} \right) \exp \left\{ \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{u}_i \right\},$$

where $\mathbf{z}_{ij}$ and $\tilde{\mathbf{z}}_{ij}$ are random effects covariates associated with the hazard and time scales respectively, and $(\mathbf{u}_i, \tilde{\mathbf{u}}_i) \overset{iid}{\sim} G_{q+\tilde{q}}$, where $G_{q+\tilde{q}}$ is a continuous distribution with support on $\mathbb{R}^{q+\tilde{q}}$ and zero mean. We emphasise that, although interesting, this strategy requires a careful study of the identifiability of parameters. Our formulation remains valid for these extensions, but the main challenges relate to the development of computational methods for addressing those questions.

Declaration of conflicting interests

Funding

ORCID iDs

Francisco Javier Rubio https://orcid.org/0000-0001-7183-8407
Reza Drikvandi https://orcid.org/0000-0002-7245-9713

Supplemental material

Supplemental material for this article is available online.

References

1. Struthers CA and Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika* 1986; **73**: 363–369.
2. Omori Y and Johnson RA. The influence of random effects on the unconditional hazard rate and survival functions. *Biometrika* 1993; **80**: 910–914.
3. Hougaard P. Frailty models for survival data. *Lifetime Data Anal* 1995; **1**: 255–273.
4. Collett D. *Modelling survival data in medical research*. London: Chapman and Hall, 2015.
5. McCulloch C, Searle S and Neuhaus J. *Generalized, Linear, and Mixed Models*. Hoboken, NJ: John Wiley & Sons, 2008.
6. Therneau T and Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York: Springer, 2000.
7. Henderson R, Shimakura S and Gorst D. Modeling spatial variation in leukemia survival data. *J Am Stat Assoc* 2002; **97**: 965–972.

8. Rubio F and Steel M. Flexible linear mixed models with improper priors for longitudinal and survival data. *Electron J Stat* 2018; **12**: 572–598.
9. Do Ha I and Lee Y. Multilevel mixed linear models for survival data. *Lifetime Data Anal* 2005; **11**: 131–142.
10. Li Y and Ryan L. Modeling spatial survival data using semiparametric frailty models. *Biometrics* 2002; **58**: 287–297.
11. Austin P. A tutorial on multilevel survival analysis: methods, models and applications. *Int Stat Rev* 2017; **85**: 185–203.
12. Crowther M, Look M and Riley R. Multilevel mixed effects parametric survival models using adaptive Gaussian Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Stat Med* 2014; **33**: 3844–3858.
13. Crowther M. Multilevel mixed-effects parametric survival analysis: Estimation, simulation, and application. *Stata J* 2019; **19**: 931–949.
14. Wood S. *Generalized additive models: an introduction with R*. Boca Raton: CRC-Chapman and Hall, 2017.
15. Vaida F and Xu R. Proportional hazards model with random effects. *Stat Med* 2000; **19**: 3309–3324.
16. Charvat H, Remontet L, Bossard N, et al. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates. *Stat Med* 2016; **35**: 3066–3084.
17. Zhou H and Hanson T. A unified framework for fitting Bayesian semiparametric models to arbitrarily censored survival data, including spatially referenced data. *J Am Stat Assoc* 2018; **113**: 571–581.
18. Verbeke G and Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Comput Stat Data Anal* 1997; **23**: 541–556.
19. McCulloch C and Neuhaus J. Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Stat Sci* 2011; **26**: 388–402.
20. Rossell D and Rubio F. Additive Bayesian variable selection under censoring and misspecification. *Stat Sci* 2022; in press.
21. Heagerty P and Kurland B. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 2001; **88**: 973–985.
22. Drikvandi R, Verbeke G and Molenberghs G. Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics* 2017; **73**: 63–71.
23. Efendi A, Drikvandi R, Verbeke G, et al. A goodness-of-fit test for the random-effects distribution in mixed models. *Stat Methods Med Res* 2017; **26**: 970–983.
24. Rao K, Drikvandi R and Saville B. Permutation and Bayesian tests for testing random effects in linear mixed-effects models. *Stat Med* 2019; **38**: 5034–5047.
25. Chen Y and Jewell N. On a general class of semiparametric hazards regression models. *Biometrika* 2001; **88**: 687–702.
26. Rubio F, Remontet L, Jewell N, et al. On a general structure for hazard-based regression models: an application to population-based cancer research. *Stat Methods Med Res* 2019; **28**: 2404–2417.
27. Wienke A. *Frailty Models in Survival Analysis*. Boca Raton, FL: Chapman and Hall/CRC, 2010.
28. Alvares D and Rubio F. A tractable Bayesian joint model for longitudinal and survival data. *Stat Med* 2021; **40**: 4213–4229.
29. Prenen L, Braekers R and Duchateau L. Extending the archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. *J R Stat Soc B* 2017; **79**: 483–505.
30. Asgharian M. On the singularities of the information matrix and multipath change-point problems. *Theor Probab Appl* 2014; **58**: 546–561.
31. Verbeke G and Molenberghs G. The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics* 2013; **14**: 477–490.
32. Drikvandi R. Nonlinear mixed-effects models with misspecified random-effects distribution. *Pharm Stat* 2020; **19**: 187–201.
33. Self S and Liang K. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 1987; **82**: 605–610.
34. Stram D and Lee J. Variance components testing in the longitudinal mixed effects model. *Biometrics* 1994; **50**: 1171–1177.
35. Crainiceanu C and Ruppert D. Likelihood ratio tests in linear mixed models with one variance component. *J R Stat Soc B* 2004; **66**: 165–185.
36. Drikvandi R, Verbeke G, Khodadadi A, et al. Testing multiple variance components in linear mixed-effects models. *Biostatistics* 2013; **14**: 144–159.
37. Verbeke G and Molenberghs G. The use of score tests for inference on variance components. *Biometrics* 2003; **59**: 254–262.
38. Saville B and Herring A. Testing random effects in the linear mixed model using approximate Bayes factors. *Biometrics* 2009; **65**: 369–376.
39. Claeskens G, Nguti R and Janssen P. One-sided tests in shared frailty models. *Test* 2008; **17**: 69–82.
40. Lee Y and Nelder JA. Conditional and marginal models: another view. *Stat Sci* 2004; **19**: 219–238.
41. Zhou H, Hanson T and Zhang J. spBayesSurv: Fitting Bayesian spatial survival models using R. *J Stat Softw* 2020; **92**: 1–33.