



The XGTDL family of survival distributions

Gilbert MacKenzie¹ · Miliça Blagojevic-Bucknall² · Yasin Al-tawarah³ · Defen Peng⁴

Received: 8 March 2021 / Revised: 14 May 2021 / Accepted: 29 May 2021
© The Author(s) 2021

Abstract

Non-PH parametric survival modelling is developed within the framework of the multiple logistic function. The family considered comprises three basic models: (a) a PH model, (b) an accelerated life model and (c) a model which is non-proportional hazards and non-accelerated life. The last model, the generalised time-dependent logistic model was described first by the author in 1996 and this model gives its name to the entire family. The family is generalised by means of a Gamma frailty extension which is shown to accommodate crossing hazards data. A further generalisation is the inclusion of a dispersion model. These extensions lead naturally to the concept of a multi-parameter regression model described by Burke and MacKenzie in which the scale and shape parameters are modelled simultaneously as functions of covariates. Where possible, we include the MPR extension in the XGTDL family. Following a simulation study, the new models are used to analyse two sets survival data and the methods are discussed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42081-021-00129-9>.

✉ Gilbert MacKenzie
gilbert.mackenzie@ul.ie
Miliça Blagojevic-Bucknall
m.bucknall@keele.ac.uk
Yasin Al-tawarah
yaltaw@hotmail.com
Defen Peng
DPeng@providencehealth.bc.ca

- ¹ Centre of Biostatistics, University of Limerick, Limerick, Ireland
- ² Centre for Primary Care, University of Keele, Keele, UK
- ³ Royal Jordanian Army Medical Core, Amman, Jordan
- ⁴ BC Centre for ICV Health, University of British Columbia, Vancouver, Canada

Keywords Accelerated life · Crossing hazards · Cure fraction · Dispersion model · Frailty · MPR models · Non-PH model · PH model · XGTDL family

1 Introduction

Prior to the development of the proportional hazards (PH) survival model, epidemiologists worked with the multiple logistic function in prospective longitudinal studies where the follow-up duration was fixed, e.g. in relation to disease incidence (Cox 1970). Not all incidence data conform to the PH assumption and MacKenzie (1996) introduced a family of survival models based on the multiple logistic function which specifically recognised this fact. The generalised time-dependent logistic family (GTDL) comprised three models (a) a PH model, (b) an accelerated life (AL) model and (c) a model which is non-PH and non-AL. Development was focussed on model (c), designated the GTDL model, and a Gamma extension demonstrated its ability to deal with crossing hazards survival data (MacKenzie and Ha 2007). This example involved the creation of a covariate-dependent shape parameter. Accordingly, the idea of modelling the shape parameter more generally intrudes and this has motivated the development of multi-parameter regression (MPR) survival models (Burke and MacKenzie 2013, 2017). In this paper, we extend the GTDL family to the XGTDL family ('X' for extended) by including: (a) Gamma frailty, (b) a dispersion model for the frailty variance, and where possible, (c) a MPR model to model non-PH survival data.

2 The GTDL family

The GTDL family of regression survival models was predicated on the multiple logistic function and the defining hazard functions of the three models in the family are as follows:

- (a) The logistic PH model (LPH)

$$\lambda(t; x) = \pi(t\alpha + \gamma_0)\varphi, \quad (1)$$

where $\pi(\cdot) = \exp(\cdot)/[1 + \exp(\cdot)]$ and the hazard function is $\lambda(\cdot) > 0$ and the following scalars are defined as: $\alpha, \gamma_0 \in (-\infty, +\infty)$, and $\varphi = \exp(x'\beta) > 0$. There is no intercept term in the linear predictor in φ , in keeping with Cox's PH model (Cox 1972).

- (b) The logistic AL model (LAL)

$$\lambda(t; x) = \lambda_0\pi(\alpha\varphi t), \quad (2)$$

where $\pi(\cdot)$, α and φ are defined above and $\lambda > 0$ is a scalar.

- (c) The GTDL model (TDL when $\lambda_0 = 1$)

$$\lambda(t; x) = \lambda_0\pi(t\alpha + \gamma), \quad (3)$$

where $\pi(\cdot)$ and α are defined above and $\lambda_0 > 0$ is a scalar. Here, $\gamma = x'\beta$ and x' includes an intercept term x_0 . This model is neither PH nor AL.

2.1 Some remarks

Remark 1 Whilst we are principally interested in non-PH survival, model (a) is included for the sake of completeness. It has a very simple interpretation in which the baseline hazard is a linear logistic function of time and the hazard function is clearly unbounded above.

Remark 2 Model (b) is rather more complicated and it was developed specifically to avoid having a bounded hazard function (Al-tawarah and MacKenzie, 2003). It was derived by accelerating the baseline survivor function, which is obtained from equation (2), by first setting $x' = \text{zero vector}$ (whence $\varphi = 1$) and integrating to obtain the baseline integrated hazard. In particular, the scalar α parameter was introduced to allow for negative trends.

Remark 3 In many ways, model (c) is the most interesting, but its development was not straightforward. The constant λ_0 , introduced to avoid having a bounded hazard function, transpired to be inestimable in maximum likelihood estimation (MLE) analysis as it is aliased with the intercept term β_0 in the linear predictor γ . One approach was to drop the intercept term (MacKenzie 2002; Blagojevic et al. 2003), in which case all of the parameters are estimable in MLE. This model was termed the canonical form and designated CTDL.

This idea seemed to work, but in fact it led to a more insidious problem, namely the lack of invariance of the model to the choice of reference subclass when fitting categorical covariates. Moreover, this problem was undetectable when the total sample size was less than 1000 and might well have been overlooked completely. Surprisingly, the solution to this problem (which does not occur with continuous covariates) is to permit λ_0 to vary, for example, with individuals as a Gamma frailty term. Prior to this solution, the GTDL model was used successfully with $\lambda_0 = 1$, when it was designated the TDL model. However, today, the Gamma frailty extension is preferred (Blagojevic et al. 2003; MacKenzie and Ha 2007; Ha and MacKenzie 2010). \square

Later, we extend the entire family to incorporate Gamma frailty in which case all the models will be non-PH.

2.2 Survival functions

Corresponding to Eqs. (1)–(3), we have the survivor functions:

(a) The LPH model

$$S(t; x) = \left[\frac{1 + \exp(t\alpha + \gamma_0)}{1 + \exp(\gamma_0)} \right]^{-\varphi/\alpha}, \quad (4)$$

where $\varphi = \exp(x'\beta)$.

(b) The LAL model

$$S(t; x) = \left[\frac{1 + \exp(\alpha \varphi t)}{c} \right]^{-\lambda/\alpha}, \quad (5)$$

where $c = 1/2$ and $\varphi = \exp(x'\beta)$.

(c) The GTDL model

$$S(t; x) = \left[\frac{1 + \exp(t\alpha + \gamma)}{1 + \exp(\gamma)} \right]^{-\lambda_0/\alpha}, \quad (6)$$

where $\gamma = x'\beta$ (including an intercept) is a scale parameter. Again when fitting this model we shall specialise $\lambda_0 = 1$, the TDL.

2.3 Cure fractions

It may be verified that the behaviour of the survival functions is determined by the value of α . Typically, for $\alpha > 0$, $S(0|x) = 1$ and $S(\infty|x) = \lim_{t \rightarrow \infty} S(t|x) = 0$, in other words, the survival function is proper, but for $\alpha < 0$, $S(0|x) = 1$ and $\lim_{t \rightarrow \infty} S(\infty|x) \neq 0$. Then, the survival function is improper, and in each case, we have a natural tail-deficit or cure rate model. We do not pursue the technical details of cure models in the XGTDL family here, but draw attention to related work by Milani et al. (2015), who studied the tail-deficit GTDL and the tail-deficit GTDL with Gamma frailty, while Rasouli et al. (2016) studied a GTDL model as a classical cure rate mixture model and Calsavara et al. (2020) analysed a melanoma dataset.¹

3 Frailty extension

Standard arguments involving an unobservable, multiplicative, random effect, u (> 0), on the hazard function yields the general formulae for the marginal survivor and hazard functions, viz.

$$S_m(t) = \mathcal{L}[\Lambda(t)] = [1 + \phi \Lambda(t)]^{-1/\phi}, \quad (7)$$

and

$$\lambda_m(t) = \lambda(t) \frac{-\mathcal{L}'[\Lambda(t)]}{\mathcal{L}[\Lambda(t)]} = \frac{\lambda(t)}{1 + \phi \Lambda(t)}, \quad (8)$$

respectively.

Here $S(t)$, $\Lambda(t)$ and $\partial \Lambda(t) \partial t = \lambda(t)$ are the basic survival quantities. Moreover, $\mathcal{L}[s] = E(e^{-sU}) = \mathcal{L}[\Lambda(t)]$ with $s = \Lambda(t)$ is the Laplace transform and

¹ Surprisingly, this latter paper omits key earlier references covered here.

U is the random variable generating the random effect, u , with density $g(u) = [\phi^{\frac{1}{\phi}} \Gamma(1/\phi)]^{-1} u^{\frac{1}{\phi}-1} \exp(-u/\phi)$, where $E(U) = 1$ and $V(U) = \phi$. It is convenient to write $\phi = \exp(\psi) > 0$.

We use these formulae to incorporate a Gamma frailty (GF) extension in all three members of the GTDL family in turn.

(a) LPH with Gamma frailty

The marginal survivor function is given by

$$S_m(t; x) = [1 + \phi \alpha^{-1} \varphi \log_e R(t)]^{-\frac{1}{\phi}} \quad (9)$$

with corresponding marginal hazard function

$$\lambda_m(t; x) = \frac{\pi(t\alpha + \gamma_0)\varphi}{[1 + \phi \alpha^{-1} \varphi \log_e R(t)]}, \quad (10)$$

where $R(t) = [1 + \exp(t\alpha + \gamma_0)]/[1 + \exp(\gamma_0)]$ and regression parameter $\varphi = \exp(x'\beta)$.

(b) LAL with Gamma frailty

The marginal survivor function is given by

$$S_m(t; x) = \left[1 + \phi \frac{\lambda}{\alpha} \log_e \left(\frac{1 + \exp(\alpha\varphi t)}{2} \right) \right]^{-\frac{1}{\phi}}, \quad (11)$$

with corresponding marginal hazard function

$$\lambda_m(t; x) = \frac{\lambda\varphi\pi(\alpha\varphi t)}{[1 + \phi \frac{\lambda}{\alpha} \log_e \left(\frac{1 + \exp(\alpha\varphi t)}{2} \right)]}, \quad (12)$$

where, as before, the regression is held in $\varphi = \exp(x'\beta)$.

(c) GTDL with Gamma frailty

The marginal survivor function is given by

$$S_m(t; x) = \left[1 + \frac{\phi}{\alpha} \log_e R(t; x) \right]^{-1/\phi}, \quad (13)$$

with corresponding marginal hazard function

$$\lambda_m(t; x) = \frac{\pi(t\alpha + \gamma)}{[1 + \frac{\phi}{\alpha} \log_e R(t; x)]}, \quad (14)$$

where $R(t; x) = [1 + \exp(t\alpha + \gamma)]/[1 + \exp(\gamma)]$ and $\gamma = x'\beta$. Here λ_0 does not appear as it was replaced by the frailty term, u , and integrated out (see Remark 3 in Sect. 2.1 for more explanation). \square

Thus, all three members of the GTDL family now incorporate frailty and are fully parametric, non-PH, survival models capable of further extension.

4 Frailty dispersion model

The use of Gamma frailty can be justified on the grounds of mathematical convenience (the probability density function of the marginal distribution can always be found analytically) and on the flexible range of shapes allowed by the Gamma distribution. However, it is sometimes argued that a single Gamma distribution does not afford sufficient flexibility. Accordingly, the flexibility can be extended by allowing the frailty variance to become a person-specific quantity, via another regression model, i.e.

$$\phi_i = \exp(w_i^T \psi), \quad (15)$$

where $w_i^T = (w_{0i}, w_{1i}, w_{2i}, \dots, w_{ri})$ and $\psi = (\psi_0, \psi_1, \psi_2, \dots, \psi_r)^T$ are vectors (of order $r + 1$) of person-specific covariates and regression coefficients, respectively, and $w_{0i} = 1$, for $i = 1, 2, \dots, n$, where n is the sample size.

Thus, the dispersion model (DM) is regression-based and hence driven by the data. When a DM is required, it relaxes the assumption that a single Gamma frailty distribution is sufficient. For example, if a binary covariate is supported by the data in the DM we shall have two GF distributions and when the frailty variance does not depend on covariates the model reduces to the standard frailty model viz: $\phi_i = \exp(w_{0i} \times \psi_0) = \exp(\psi_0)$ the same constant for each person.

The concept of modelling the structure of the dispersion has been addressed in the joint mean-dispersion modelling literature (Smyth 1989; Lee and Nelder 2001; Pan and MacKenzie 2003), but its adoption in the frailty paradigm, in survival analysis, is more recent (Lynch and MacKenzie 2014). Furthermore, the use of frailty dispersion modelling in combination with an underlying GTDL family MPR model (below) is novel in the survival literature.

With the dispersion model in place we have in effect created the core of the XGTDL family. In the next section we extend the family further using the ideas of MPR modelling.

5 MPR modelling

Following Burke and MacKenzie (2017), the key idea is to model the scale and shape parameters simultaneously in terms of the covariates. Again, this idea is relatively new in the survival literature and is best illustrated by a familiar example. We start with the MPR Weibull (WB) model which is non-PH and used later as a comparator for the MPR models which we develop later in the XGTDL family.

5.1 The Weibull MPR model

The Weibull hazard function (Collett 2003) is $\lambda(t) = \lambda \gamma t^{\gamma-1}$ with $(\lambda > 0$ and $\gamma > 0)$ whence the MPR form is simply: $\log_e \gamma = z' \alpha$ (shape) and $\log_e \lambda = x' \beta$ (scale)

resulting in the MPR hazard function

$$\lambda(t; x, z) = \exp(x'\beta) \exp(z'\alpha) t^{\exp(z'\alpha)-1},$$

a non-PH model in which $x = (x_0, x_1, x_2, \dots, x_p)$ and $z = (z_0, z_1, z_2, \dots, z_q)$. When x and z contain the same covariates, including intercepts, $q = p$. When $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_q) = 0$ the model reduces to an Exponential model (PH) with parameter φ . And when $\alpha = (\alpha_0, 0, 0, \dots, 0)$ the model is a standard Weibull model (PH) in which α_0 is the shape parameter. Inference in the MPR Weibull model is discussed in detail by Burke and MacKenzie (2017) and a wide range of two-parameter MPR survival models can be fitted using the `mpr` package developed by Burke (2016) in R (Core 2018).

With these arrangements in place, we can apply the MPR concept to existing models in the XGTDL family. It is worth noting that any model with more than one linear predictor is an MPR model, e.g. a Weibull regression model with a constant shape parameter and a Gamma frailty dispersion model.

5.2 A motivating example

We can gain some insight into the flexibility of MPR models by analysing the data provided by the Gastrointestinal Tumor Study Group (Schein and Lombard et al. 1982). This was a randomised controlled trial of the effects of two treatment regimens: chemotherapy *vs* combined chemotherapy and radiotherapy on the survival times of 90 (45 per regimen) gastric cancer patients. The resulting Kaplan and Meier (1958) survival curves are shown in Fig. 1. The survival curves cross indicating the presence of crossing hazards (violating the Cox (1972) PH assumption) and a significant difference in survival between groups (MacKenzie and Ha 2007). Early survival (0–2 years) is better in the chemotherapy group while longer term survival is better in the combined therapy group.

Table 1 shows a number of different parametric models (one PH and the rest non-PH) fitted to the data. Inspecting the treatment effects (the $\hat{\beta}_1$ s) in Table 1, the TDL + MPR + GF model with Gamma frailty and separate shape parameters (the $\hat{\alpha}$ s) shows a clear and statistically significant treatment effect. The WB + MPR + GF model shows a marginal result ($t = -0.869/0.432 = -2.01$) for the treatment effect, but the AIC for the TDL + MPR + GF model is much superior, suggesting that a non-PH base survival function is also required.

It also clear that Gamma frailty is important—the WB + GF model almost detects a difference in treatment effect ($t = -1.006/0.508 = -1.98$) but the TDL + GF result is unequivocal ($t = -3.288/0.822 = -4.0$). However, neither of these models can fit the crossing hazards data in Fig. 1. That requires an MPR model with separate tracks for each group.

Looking at the other models in Table 1 would lead one to conclude, erroneously, that there was no statistically significant difference between survival in the two treatment groups. This finding highlights the need for graphical confirmation of tableau results.

Fig. 1 GTSG crossing hazards data: KM survival by treatment group. Chemo. = lower-tailed curve (red) and Chemo.+ radiotherapy = higher tailed curve (black)

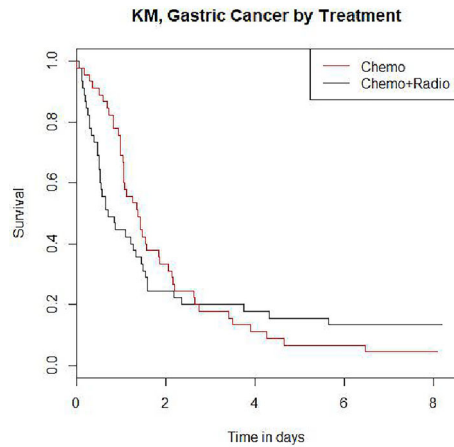


Figure 2 shows the resulting goodness-of-fit of the TDL + MPR + GF model. The model fits the survival data in both treatment groups well and captures the crossing point of the survival functions. It is apparent that the goodness-of-fit lies, in large part, in modelling the shape parameters in each treatment arm. Accordingly, it is but a short step to consider that modelling the shape parameters may be a useful stratagem in general.

6 XGTDL-MPR GF models

We have already seen above that the standard two-parameter Weibull survival model can be extended to MPR form (Burke and MacKenzie 2017) and that the TDL + GF model can be extended (MacKenzie and Ha 2007). However, in general, it is still an open question as to whether all two-parameter survival models permit this extension. Accordingly, we turn now to systematically develop the MPR models in the XGTDL family. For convenience we assume that x contains the same covariates as z , whence $q = p$.

(a) The LPH + MPR + GF model

The LPH model with Gamma frailty is easily extended to MPR form by modelling the α parameter in (9). We choose

$$\begin{aligned}\alpha &= \alpha_0 z_0 + \alpha_1 z_1 + \alpha_2 z_2 + \dots + \alpha_p z_p, & (\text{shape}) \\ \varphi &= \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p), & (\text{scale})\end{aligned}\quad (16)$$

where, in the shape parameter, α_0 can be negative, representing a declining hazard function (MacKenzie 1996). Note that the scale parameter φ is unchanged.

(b) The LAL + MPR + GF model

Table 1 GTSG crossing hazards data: Weibull and TDL models fitted; ML estimates, standard errors (S.E.) and AICs

Model	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$var(U)$	AIC
WB	-0.158 (0.086)	- -	-0.612 (0.181)	0.031 (0.223)	- -	290.36
WB+MPR	-0.357 (0.125)	0.448 (0.171)	-0.449 (0.181)	-0.373 (0.273)	- -	285.62
WB+GF	0.502 (0.164)	- -	0.577 (0.467)	-1.006 (0.508)	1.489 (0.244)	275.30
WB+MPR+GF	0.259 (0.208)	0.381 (0.194)	0.311 (0.392)	-0.869 (0.432)	1.249 (0.755)	273.53
TDL	-0.482 (0.139)	- -	0.737 (0.431)	0.304 (0.495)	- -	277.88
TDL+MPR	-0.832 (0.242)	-2.175 (2.057)	1.494 (0.666)	-1.380 (0.822)	- -	273.10
TDL+GF	3.910 (1.475)	- -	0.813 (1.559)	-3.288 (0.822)	0.691 (0.122)	268.56
TDL+MPR+GF	-0.789 (0.326)	4.288 (1.447)	2.380 (1.414)	-4.612 (1.676)	1.802 (0.572)	265.78

NB: AIC = Akaike Information Criterion.

Scale parameters: β_0 (combined therapy) & β_1 (chemo therapy).Shape parameters: α_0 (combined therapy) & α_1 (chemo therapy).

It may be thought that the LAL model with Gamma frailty could be extended to MPR form in a similar fashion, but in this AL model the parameter

$$\varphi = \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (\text{shape and scale})$$

influences both shape and scale. From (10), we see that the only other free parameter which could be given a linear predictor is λ , but this, too, influences the scale. Attempts to implement such a MPR model produced unstable results and we omit this development.

(c) The TDL + MPR + GF model

The TDL model with Gamma frailty is extended to MPR form as follows

$$\begin{aligned} \alpha &= \alpha_0 z_0 + \alpha_1 z_1 + \alpha_2 z_2 + \dots + \alpha_p z_p, & (\text{shape}) \\ \gamma &= \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p. & (\text{scale}) \end{aligned} \quad (17)$$

and λ_0 has been replaced by the frailty term and integrated out so we refer to this model as TDL rather than GTDL. Note also that the definition of γ is unchanged.

□

With these arrangements, the α parameter has become person-specific via the regression so that each subject has a separate covariate-dependent shape parameter. In this way, shape is modelled symmetrically with scale. Clearly, the same scheme applies equally to MPR models without Gamma frailty e.g. the LPH + MPR model.

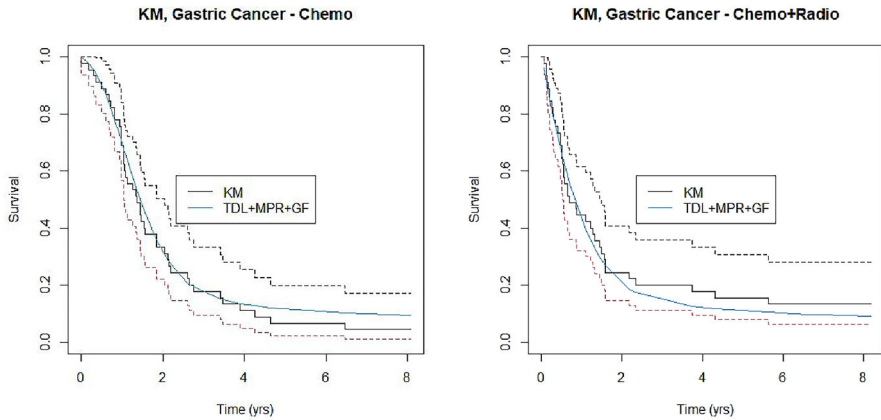


Fig. 2 GTSG crossing hazards data: KM and model-based survival by treatment groups; KM = step function, Model = smooth curve, KM confidence interval = dots. The fitted model is TDL + MPR + GF, Table 1

7 Estimation and fitting

Model formulation may be quite detailed in the XGTDL family, but, fortunately, Maximum Likelihood (ML) estimation in parametric survival models is classical and straightforward. We assume that the observed survival time, $t = \min(t_1, t_2)$, where t_1 is distributed as a XGTDL random variable, T_1 , and T_2 is a chosen censoring distribution such that T_1 is statistically independent of T_2 . For data (t_i, d_i, x_i) , where $i = 1, 2, \dots, n$ subjects and with independent censoring, the likelihood for the parameters, say θ , is

$$L(\theta) = \prod [\lambda(t_i; x_i, \theta)^{d_i} S(t_i; x_i, \theta)], \quad (18)$$

where d_i is the censoring indicator with $d_i = 1$ for an event and $d_i = 0$ for a censored observation. The hazard functions and survival functions for all of the models in the XGTDL family have been provided in foregoing sections.

We fitted the models using a specially written R programme (Core 2018) which called the numerical optimisation procedure `nlm` to maximise the log-likelihood function, compute the maximum likelihood estimators and their standard errors from the observed information matrix.

8 Model selection

In the XGTDL family, there are three model classes, LPH, LAL and TDL. For the purpose of selecting amongst models within a class, standard information criteria may be used. We opted for the use of the Akaike (1974) information criterion, $AIC = -2 \log(\hat{\theta}) + 2 \dim(\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator and $\dim(\hat{\theta})$ is the number of parameters to be estimated. There are two levels of model selection, both

of which can be handled by the AIC, namely: (a) covariate selection within a single model type (within a class) and (b) model type selection across classes.

In the three classes, there are $6 + 3 + 6 = 15$ model types in all. To see this, let the base = LPH or LAL or TDL basic regression survival models. Then, for the LPH and TDL classes, we have: (1) base, (2) base + GF, (3) base + DM, (4) base + MPR, (5) base + MPR + GF, (6) base + MPR + DM. And for the LAL class, we have the first three model types only. To this, we can add another 6 models with base = WB (Weibull) as comparators making 21 models in all, and not space enough to present the findings for the whole family in the paper. Accordingly, we present only an example simulation.

The use of the AIC may pre-suppose that one is prepared, *á priori*, to fit several model classes. However, a great deal may be done, initially, by classical plotting methods: for example, by plotting model-based survival functions and/or the cumulative hazard functions against time and comparing with their non-parametric counterparts. In this way, the adequacy of the functional form of the base regression model may be assessed. Next, a check for possible shape effects would seem reasonable. Finally, the addition of gamma frailty might be considered on the grounds that the current regression may not be completely comprehensive. At each stage the AIC and the fit of the survival function can be checked for improvement.

9 Simulation study

We conducted the simulation with: two covariates x_1 (continuous, $N(0, 1)$) and x_2 (binary, 50% split); three sample sizes ($n = 200, 500$ and 1000); three censoring rates (10, 20 and 50%); three values of the frailty variance ϕ (0.5, 1 and 1.5) and finally for three different parameter vectors. Thus, there were 81 scenarios in principle for each model and the number of replications was $m = 1000$.

In the simulations, the proportion of right (random) censoring, p_c , was controlled using an exponential distribution where the value of the controlling parameter, ϑ , was obtained from the $J(\cdot)$ function approach. Suppose the independent censoring times follow an exponential distribution with pdf $g(t; \vartheta) = \vartheta \exp(-\vartheta t)$. Let

$$J(\vartheta) = \left[\int_0^\infty S(t; \theta) g(t; \vartheta) dt - p_c \right]^2$$

for fixed θ and a required censoring proportion, p_c . Then, we find $\hat{\vartheta} = \arg \min[J(\vartheta)]$. These minimizations were carried out in the R statistical package (Core (2018)) using the `optimize` procedure. The function $S(t; \theta)$ is, XGTDL family, model-specific.

We emphasise that the purpose of these simulations is simply to ensure that the parameters are recoverable in the scenarios studied and to ascertain the level of bias (if any). The true values of the parameters are shown at the bottom of the table.

Table 2 shows the results for TDL + MPR + GF model. The bias is generally small and decreases with increasing sample size in the scale and shape parameters. The frailty variance is $\phi = \exp(\psi_0 = -0.7) = 0.5$. However, for the smallest sample size

Table 2 Simulation: ML estimates, estimated standard errors (S.E.) and percentage bias for the TDL, MPR, Gamma frailty model with different sample sizes and censoring rates

Details	$\hat{\beta}$	Scale s.e.	%B	$\hat{\alpha}$	Shape s.e.	%B	$\hat{\psi}_0$	Frailty s.e.	%B
$p_c = 0.1$									
$n = 200$									
x_0	-2.053	0.402	2.6	1.028	0.308	2.8	-0.775	0.332	10.7
x_1	0.573	0.262	14.5	0.218	0.150	-12.6	-	-	-
x_2	0.285	0.522	-5.1	-0.092	0.342	-8.3	-	-	-
$n = 500$									
x_0	-2.003	0.243	0.2	1.002	0.152	0.8	-0.730	0.144	4.2
x_1	0.525	0.155	5.0	0.238	-4.611	-0.4	-	-	-
x_2	0.295	0.310	-1.6	-0.104	4.548	1.0	-	-	-
$n = 1000$									
x_0	-2.011	0.170	0.5	1.006	0.124	0.5	-0.713	0.099	1.8
x_1	0.518	0.103	3.6	0.243	0.054	-2.8	-	-	-
x_2	0.298	0.216	-0.8	-0.100	0.116	-0.2	-	-	-
$p_c = 0.2$									
$n = 200$									
x_0	-2.062	0.405	3.1	1.014	0.317	1.4	-0.785	0.340	12.2
x_1	0.554	0.245	8.1	0.230	0.144	-7.9	-	-	-
x_2	0.316	0.516	5.3	-0.102	0.299	2.4	-	-	-
$n = 500$									
x_0	-2.025	0.249	1.2	1.015	0.019	1.5	-0.719	0.166	2.7
x_1	0.537	0.140	7.3	0.240	0.064	-3.8	-	-	-
x_2	0.303	0.314	1.1	-0.100	0.177	-0.5	-	-	-
$n = 1000$									
x_0	-2.007	0.174	0.3	0.999	0.132	-0.1	-0.714	0.115	2.1
x_1	0.518	0.109	3.7	0.243	0.060	-2.9	-	-	-
x_2	0.302	0.223	0.6	-0.094	0.126	-5.7	-	-	-
$p_c = 0.5$									
$n = 200$									
x_0	-2.111	0.457	5.5	1.093	0.421	9.3	-0.811	1.111	15.8
x_1	0.543	0.280	8.5	0.245	0.200	-1.9	-	-	-
x_2	0.350	0.585	16.7	-0.127	0.391	26.8	-	-	-
$n = 500$									
x_0	-2.015	0.272	0.8	1.010	0.236	1.0	-0.758	0.391	8.3
x_1	0.534	0.162	6.9	0.235	0.102	-6.1	-	-	-
x_2	0.272	0.345	-9.2	-0.090	0.210	-9.6	-	-	-
$n = 1000$									
x_0	-2.017	0.192	0.8	1.009	0.168	0.9	-0.716	0.218	2.3
x_1	0.517	0.116	3.5	0.245	0.073	-1.8	-	-	-
x_2	0.310	0.242	3.3	-0.098	0.148	-1.6	-	-	-

NB: The true model generating the data has parameters, $\beta_0 = -2.0, \beta_1 = 0.5, \beta_2 = 0.3$ for the scale and $\alpha_0 = 1.0, \alpha_1 = 0.25, \alpha_2 = -0.1$ for the shape: the frailty variance is $\text{var}(u) = \exp(\psi_0) = \exp(-0.7) = 0.5$ and %B = percentage bias.

($n = 200$), some estimates can be biased, but the bias is much reduced for increasing sample sizes and small when $n = 1000$. Of course, we expect that more complicated models tend to require larger sample sizes, but, overall, our simulation suggests that the model parameters are recoverable for reasonable sample sizes.

We also computed the probabilities that the intervals $I_j = [a_j, b_j]$, for $j = 1, \dots, m (= 1000)$ replications, covered the true value of the corresponding parameter, say θ_j , where $a_j, b_j = \hat{\theta}_j \mp 2 \times \text{se}(\hat{\theta}_j)$, respectively. Setting $c_j = 1$ when I_j covers θ_j and $c_j = 0$, otherwise, the estimated coverage probability for θ_j is then

$Pr(I_j \curvearrowright \theta_j) = \sum c_j/m$. (The symbol \curvearrowright reads 'covers'). For all of the parameters in Table 2 the coverage percentage lay between 92.5 and 97.5% (not shown).

9.1 Some remarks on model fitting

Remark 1. Throughout the simulation study (and subsequent data analysis), a user-written R script was used which called the `nlm` procedure to fit the model. This is a numerical optimization routine with which we are familiar. It can use analytical first and second derivatives, but we allowed the routine to compute numerical derivatives and an approximation to the hessian. This approach has worked well for us in the past. As with any new model, it is important to be careful in the selection of the true values, to be simulated from, which must generate $0 < t < \infty$ via the inverse distribution function transform ($t = F^{-1}(u)$, where $u \sim \text{Uniform}(0, 1)$). Similarly, with the selection of starting values which must generate a valid interior point in the parameter space. A poor choice of starting values may lead `nlm` to a false start with no progress from the starting values. This situation was trapped in the user-written R script. At the time of writing, we do not have experience of other methods of fitting these models.

Remark 2. In the case of the TDL + MPR + GF models, the simulation study suggests that at least 20 observations per parameter are required for successful fitting and estimation of the corresponding variance covariance matrix. In general, studies of survival data with small sample sizes are unlikely to be of great interest. The minimum TDL + MPR + GF model with a single covariate in the scale and shape parameters requires 5 parameters. In the Gastric Tumour Study (1982), there were 90 observations (45 per group) and the 5 parameter TDL + MPR + GF model fitted easily. However, the censoring rates were low: 4.4% in one group and 13.2% in the other.

Remark 3. The distribution of the estimated parameters appeared as a uni-modal, symmetric, histogram—approximately Normal in shape, regardless of the sample size studied. Exceptions occurred when the simulation produced outliers due to non-convergence of the numerical optimization routine. Such outliers were produced more often in scenarios where the sample size was small and the censoring rate high. When the outliers were removed, the Normal-like distribution re-emerged. \square

The simulation study included all of the models listed in Table 3 and the results for other models (not shown) follow a similar pattern.

10 Analysis of lung cancer data

10.1 Data analysed

We turn now to analyse the Lung Cancer dataset (Wilkinson 1995). This was a multi-source, population-based, study of 855 incident cases diagnosed in Northern Ireland, UK, between October 1st, 1991 and September 30th, 1992. The patients were followed

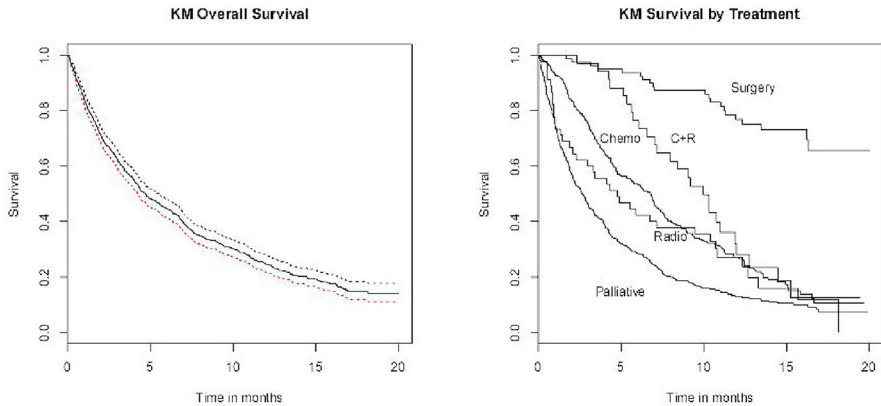


Fig. 3 NI lung cancer study: overall Kaplan–Meier survival curve (left panel) and by treatment (right panel)

for almost 2 years and survival time was computed as the time from first diagnosis to death or censoring. Some 673 (77%) patients had died by the censoring date (30th May, 1993), leaving 182 (23%) censored. We adopt the conventional 5% level of statistical significance in what follows.

10.2 Analysis of treatment

Figure 3 (left panel) shows the Kaplan–Meier (KM) estimator of overall survival from incident lung cancer. Half the patients were dead by 6 months and more than half (51.6%) were allocated to the palliative care group at the time of diagnosis, indicating that no curative treatment was planned. Typically, such poor prognosis results from late diagnosis.

The oncologists were principally interested in the effect of treatment and in which model(s) fitted the treatment survival data, as later, these would be adjusted for the effects of the other measured covariates.

The right panel of Fig. 3 shows the KM estimator by treatment category (Surgery, Chemotherapy, Radiotherapy, Chemo. + Radio., Palliative care). Surgical treatment has the best survival pattern and palliative care the worst. The observed survival tracks do not appear to follow a proportional hazards model—notice the late crossing hazards between chemotherapy and radiotherapy and the convergent behaviour of the combined therapy group.

As a first step, we illustrate fitting the models with only the treatment covariate. Table 3 shows the results of fitting all 21 models in the XGTDL family in terms of the type of parameters fitted, the log-likelihood, the AICs obtained, and the number of parameters fitted. Within a model class (e.g. Weibull) the various model types form a hierarchy in terms of the number of parameters fitted, which is the same across the classes (Weibull, LPH and TDL); LAL being the exception.

Inspection of Table 3 shows that including Gamma frailty in the models results in statistically significant decreases in the observed AICs throughout. For example, in the TDL class a decrease (dAIC) of 22.99 units is obtained between the basic Weibull

model (3932.32) and the same model with Gamma frailty included (3888.34), for the addition of just 1 additional parameter. Increasing complexity, by including a dispersion model, also improves model fit throughout the table, although the gains, while still statistically significant, are more modest.

Moving to the MPR modelling scheme (without Gamma frailty) yields models which are close to the basic models with Gamma frailty. For example, in the TDL class the TDL + DM model has an AIC of 3887.20 while the AIC of the TDL + MPR model is 3888.95. In the table, only in the LPH class, is the LPH + MPR model superior to the LPH + DM ($\text{dAIC} = 2.78$). Both types model shape: the MPR model by person-specific covariates influencing the shape parameter and the DM model by Gamma frailty distributions stratified by covariates. That these are different measures of shape is evidenced in the table by the models which support Gamma frailty *and* MPR parameters. In fact the best model in 3 of the 4 classes is the MPR + GF model. Note that the addition of a dispersion model to a non-MPR model reduces the AICs by an significant quantity. But when the model is MPR, the reduction is not statistically significant. Overall, the conclusion is that the WB + MPR + GF model is a good working assumption.

In these data, the optimum AIC ($= 3875.89$) is provided by the WB + MPR + GF model which is substantially better than the AIC ($= 3896.24$) produced by the WB + MPR model which was used to analyse treatment in Burke and MacKenzie (2017). We note that none of the new models proposed fit the data as well as the WB + MPR + GF comparator. The nearest contender is the TDL + MPR + GF model with an AIC of 3883.70 which, however, is also significantly better than the WB + MPR model. We conclude that while the Gastric Cancer data required a non-PH base hazard function the Lung Cancer data do not.

10.3 Covariate adjustment

The NI Lung Cancer Study is an observational, population-based, study, and therefore, we adjust the treatment effects for the other measured covariates, using the models which performed well in the previous section, namely, the WB + MPR + GF and TDL + MPR + GF models.

The results are shown in Table 4 for fitting the full and reduced MPR + GF models in the Weibull and TDL classes. There are several interesting findings. First, the two models are re-assuringly broadly similar in relation to the magnitude and interpretation of the effects in the scale and shape regressions. Second, we note the large improvement in AIC when compared to any model involving treatment alone (cf. Table 4), indicating that non-treatment covariates are clearly important. Third, we note that more covariates influence the scale than the shape—twice as many in the Weibull case. This is a common finding in MPR modelling. On eliminating the redundant effects from the full models in Table 5, we reach dAICs of 23.36 and 20.42 for the Weibull and TDL models, respectively, indicating that the reduced models have significantly improved fit. We note, however, that in these reduced models the shape parameters cannot be ignored, $\text{dAIC} = 9.80$ for the Weibull and $\text{dAIC} = 4.11$ for the TDL, respectively, and thus the shape effects are less influential in the TDL + MPR + GF model.

Table 3 AICs for XGTDL family and Weibull comparator models fitted using the treatment covariate

Model	Scale	Shape	DM	$\ell(\hat{\theta})$	AIC	npars
WB	c,rx	c	-	-1960.75	3933.51	6
WB+GF	c,rx	c	c	-1943.57	3901.13	7
WB+DM	c,rx	c	c,rx	-1937.63	3895.38	10*
WB+MPR	c,rx	c,rx	-	-1938.12	3896.24	10
WB+MPR+GF	c,rx	c,rx	c	-1926.93	3875.89	11
WB+MPR+DM	c,rx	c,rx	c,rx	-1925.77	3883.86	15
LPH	c,rx	c	-	-1959.16	3930.33	6
LPH+GF	c,rx	c	c	-1942.36	3891.51	7
LPH+DM	c,rx	c	c,rx	-1934.76	3898.71	11
LPH+MPR	c,rx	c,rx	-	-1931.98	3883.96	10
LPH+MPR+GF	c,rx	c,rx	c	-1932.23	3886.47	11
LPH+MPR+DM	c,rx	c,rx	c,rx	-1931.46	3892.92	15
LAL	c	c,rx	-	-1949.90	3911.80	6
LAL+GF	c	c,rx	c	-1944.95	3903.90	7
LAL+DM	c	c,rx	c,rx	-1936.44	3894.88	11
TDL	c,rx	c	-	-1960.16	3932.32	6
TDL+GF	c,rx	c	c	-1937.17	3888.34	7
TDL+DM	c,rx	c	c,rx	-1932.60	3887.20	11
TDL+MPR	c,rx	c,rx	-	-1934.48	3888.95	10
TDL+MPR+GF	c,rx	c,rx	c	-1930.85	3883.70	11
TDL+MPR+DM	c,rx	c,rx	c,rx	-1930.12	3890.24	15

NB: WB = Weibull, c=constant (x_0), rx=Treatment (4 dummy variables), GF=Gamma frailty, DM= Dispersion Model (DM includes GF), *one term could not be fitted.

Table 4 Comparison of Weibull and TDL (MPR + GF) models fitting all nine covariates

Covariate	Weibull		TDL	
	Scale	Shape	Scale	Shape
Treatment	✓	✓	✓	✓
Age	x	x	x	x
Sex	x	x	x	x
WHO status	✓	x	✓	x
Cell type	✓	x	✓	x
Sodium	✓	✓	✓	x
Albumen	✓	x	✓	x
Metastases	✓	✓	✓	✓
Smoking status	x	x	x	x
Full Model				
$\ell(\hat{\theta})$	-1793.66		-1811.76	
AIC	3681.32		3717.51	
Reduced Model				
$\ell(\hat{\theta})$	-1800.98		-1822.54	
AIC	3657.96		3697.09	

NB: ✓ = Significant, x = Not significant.

Table 5 Unadjusted and adjusted treatment comparisons showing mles (S.E.): adjusted for covariates in the reduced models

Treatment	Weibull		TDL	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Surgery	-3.25 (0.34)	-1.88 (0.81)	-4.37 (0.43)	-1.79 (0.36)
Chemotherapy	-0.64 (0.28)	-0.56 (0.39)	-0.90 (0.39)	-0.90 (0.44)
Radiotherapy	-1.10 (0.17)	-1.15 (0.22)	-1.53 (0.20)	-1.01 (0.22)
Chemo. + Radio.	-1.73 (0.33)	-4.87 (1.11)	-2.45 (0.38)	-3.18 (0.71)

NB: Reference category (Palliative care) not shown.

Table 5 compares unadjusted and adjusted treatment effects in the MPR + GF models. The adjustment is for the effect of the covariates (case-mix) included in the reduced models (see Table 4). The pattern is the same for both models, namely, adjustment for the effects of case-mix reduces the beneficial effects of surgery, chemotherapy and radiotherapy observed in the unadjusted analysis, but enhances the effect of the combined therapy. However, since only 34 cases received combined therapy the overall effect of adjustment is to reduce the benefit attributable to treatment. That the large benefit in the combined therapy group is an early effect, attenuated later as follow-up progresses, is evidenced by the large positive effect in the shape regression (coefficients not shown).

11 Discussion

The non-PH GTDL family is of course not the only way to model non-PH data, but it was the use of a member of this family in the crossing hazards example that led, in part, to the broader idea of modelling the shape parameter systematically with the usual scale parameter in survival distributions. A parallel idea was to routinely incorporate frailty because, realistically, survival data do not always conform to the proportional hazards assumption and important covariate information may be missed, even in randomised controlled trials using minimization. Accordingly, the extension to the XGTDL family is a natural statistical modelling development.

Whilst frailty modelling is now relatively well established, the idea of modelling shape symmetrically with scale (the MPR idea) is less well accepted in the survival literature. This paper demonstrates the potential benefits of adopting a MPR approach and corroborates our previous work in this area (Burke and MacKenzie 2017; Peng et al. 2020). Moreover, we have also demonstrated the importance of including both the frailty concept and the MPR approach, a finding which suggests that the MPR model captures a form of time dependence which classical frailty models cannot.

Although this is only one set of data, our experience with other data leads us to suggest that the addition of Gamma frailty and MPR regression to a base model will often be a good general starting place from which to analyse non-PH survival data.

When we embarked on this extension the question, as to whether all survival distributions could be extended easily to MPR form, was open. With the logistic accelerated

life (LAL) class, we encountered problems when developing the MPR extension. The difficulty may result simply from the functional form of the hazard in AL models and we continue to investigate this topic in our laboratory. Meanwhile, we hope the models developed thus far in the XGTDL family will be of use to the survival modelling community.

Acknowledgements We thank the Editors of the Special Edition for encouragement and two anonymous referees whose comments improved the paper.

Funding Open Access funding provided by the IReL Consortium

Declarations

Conflict of interest There are no known conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Al-tawarah, Y. & MacKenzie, G. (2003). A non-PH Accelerated Hazard Model for Analyzing Clinical Trial Data. In: *Proceedings of the 18th International Workshop on Statistical Modelling*. Leuven, Belgium. pp 27–32.
- Blagojevic, M., MacKenzie, G., & Ha, I. D. (2003). A comparison of PH and non-PH survival models with Gamma frailty. In: *Proceedings of the 18th. International Workshop on Statistical Modelling*, Leuven, Belgium, pp. 39–44.
- Burke, K. (2016). *mpr*: it multi-parameter regression (MPR). R package version 1.0.2 (cran.r-project.org/package=mpr).
- Burke K., & MacKenzie G. (2013). Multi-parameter regression survival models. In: *Proceedings of the 28th. International Workshop on Statistical Modelling*, Palermo, Sicily, Italy, pp. 93–96.
- Burke, K., & MacKenzie, G. (2017). Multi-parameter regression survival modelling—An alternative to proportional hazards. *Biometrics*, 73, 361–716.
- Calsavara, V. F., Milani, E. A., Bertolli, E., & Tomazella, V. L. (2020). Long-term frailty modeling using a non-proportional hazards model: Application with a melanoma dataset. *Statistical Methods in Medical Research*, 29(8), 2100–2118.
- Collett, D. (2003). *Modelling survival data in medical research by David Collett*. New York: Chapman & Hall-Crc.
- Cox, D. R. (1970). *The analysis of binary data*. London: Methuen.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B*, 34, 187–220.
- Ha, I. D., & MacKenzie, G. (2010). Robust frailty modelling using non-proportional hazards models. *International Journal of Statistical Modelling*, 10(3), 315–332.
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. <https://doi.org/10.2307/2281868>.
- Lee, Y., & Nelder, J. (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random effect models and structured dispersions. *Biometrika*, 88(4), 987–1006.

- Lynch, J., & MacKenzie, G. (2014). Frailty models with structural dispersion. In G. MacKenzie & D. Peng (Eds.), *Statistical modelling in biostatistics and bioinformatics*. Munich: Springer. ISBN= 978-3-319-04578-8.
- MacKenzie, G. (1996). Regression models for survival data: The generalized time dependent logistic family. *Journal of the Royal Statistical Society, Series D*, 45, 21–34.
- MacKenzie, G. (2002). A logistic regression model for survival data. In M. Stasinopoulos, & G. Touloumi (Eds.), *Proceedings of the 17th. International Workshop on Statistical Modelling*, Chania, Crete, Greece, pp. 431–437.
- MacKenzie, G., & Ha, I. D. (2007). Modelling survival data with crossing hazards. In *Proceedings of the 22nd. International Workshop on Statistical Modelling*, Barcelona, Spain, pp. 416–420.
- Milani, E. A., Tiomazella, V. L., Dias, T. C. M., & Louzada, F. (2015). The generalized time-dependent logistic frailty model: An application to a population-based prospective study of incident cases of lung cancer diagnosed in Northern Ireland. *Brazilian Journal of Probability and Statistics*, 29(1), 132–144.
- Pan, J., & MacKenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika*, 90(1), 239–244.
- Peng, D., MacKenzie, G., & Burke, K. (2020). A multiparameter regression model for interval-censored survival data. *Statistics in Medicine*, 39(14), 1903–1918.
- R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rasouli, R., Ghadimi, M., Shohani, M., Ashouri, A., Hosseinzadeh, S., Mohammad, K., et al. (2016). Cure mixture model based on hazard modeling non-proportional, application to esophageal cancer. *Journal of Biostatistics and Epidemiology*, 2(2), 52–9.
- Schein, P. S., Lombard, V. T., et al. (1982). A comparison of combination chemotherapy and combined modality therapy for locally advanced gastric carcinoma. *Cancer*, 49, 1771–1777.
- Smyth, G. (1989). Generalized linear models with varying dispersion. *JRSS, Series B Methodological*, 51(1), 47–60.
- Wilkinson, P. (1995). *Lung cancer in Northern Ireland 1991–1992*. PhD thesis, Queen’s University, Belfast, Northern Ireland, UK.