Inclusion of time-varying covariates in cure survival models with an application in fertility studies

Philippe Lambert[1,2,*] and Vincent Bremhorst[2]

Abstract

Cure survival models are used when one desires to explicitly acknowledge that an unknown proportion of the studied population will never experience the event-of-interest. An extension of the promotion time cure model enabling to include time-varying covariates as regressors when modelling (simultaneously) the probability and the timing of the monitored event is presented. Our proposal enables to handle non monotone population hazard function without specific parametric assumption on the baseline hazard. This extension is motivated and illustrated on data from the German Socio-Economic Panel (GSOEP) by studying the transition to second and third births in West Germany.

Keywords: Bayesian P-splines; Cure survival model; Fertility studies; Frailty ; Promotion time model ; Time-varying covariates.

1 Introduction

In classical survival analysis under a sufficiently long follow-up time, the event of interest is assumed to be, later or sooner, observed for each individual under study, leading to a proper population survival function (i.e. $\lim_{t \to +\infty} S_p(t) = 0$). In fertility studies, when studying the transition to second birth for example, this assumption is not reasonable. Indeed, it is well-known that an unknown proportion of one-child mothers will never have a second child for personal or medical reasons. Cure survival models define an improper

---

[1]Université de Liège, Institut de Recherche en Sciences Sociales (IRSS), Méthodes quantitatives en sciences sociales, Place des Orateurs, 3 - Quartier Agora (Bât. B31), B-4000 Liège, Belgium.

[2]Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

[*]Corresponding author: p.lambert@uliege.be

population survival function with a limiting value ($\lim_{t \to +\infty} S_p(t) = p > 0$) corresponding to the probability of being *cured*. This class of models enables to disentangle the subject characteristics influencing the probability to experience the event-of-interest from those influencing its timing. The first cure survival model, named the *mixture cure model* as the global population was seen as a mixture of cured and susceptible sub-populations, was proposed by Boag (1949) and Berkson and Gage (1952):

$$S_p(t) = p + (1-p)S_u(t), \tag{1}$$

where p denotes the probability to be cured and $S_u(t)$ the proper survival function of susceptible individuals. During the last decades, the mixture cure model has been extensively studied in the statistical literature, see among others Kuk and Chen (1992); Taylor (1995); Peng and Dear (2000); Sy and Taylor (2000); Li and Taylor (2002); Zhang and Peng (2007); Lu (2010); Wang et al. (2012); Zhang et al. (2013); Zhou et al. (2016) and López-Cheda et al. (2017).

In the late nineties, a new family of cure survival models was developed: the *promotion time cure model* (Yakovlev and Tsodikov, 1996; Tsodikov, 1998; Chen et al., 1999). This model, first presented to analyze survival data in cancer studies, argues that the observed time-to-event time is defined as the minimum time for one of $N \sim \mathscr{P}(\theta)$ (Poisson distributed) latent factors to become detectable. These latent factors $(Y_1, \ldots, Y_N)$ are assumed to be directly active at the beginning of the follow-up, independent and identically distributed (with a proper CDF $F(t)$ independent of N). In the realm of fertility studies a latent factor might be interpreted as a potential decisive argument to decide to have an additional child and the 'time for its detection' as the time required for it to be convincing (Bremhorst et al., 2016).

The population survival, hazard and density functions of the promotion time model can be shown to be

$$S_p(t|\theta, F) = \exp\left(-\theta F(t)\right) \; ; \; h_p(t|\theta, F) = \theta f(t) \; ; \; f_p(t|\theta, F) = h_p(t|\theta)S_p(t|\theta), \tag{2}$$

where $f(t) = \frac{\partial F(t)}{\partial t}$ is the latent density. If no latent factor is active at the beginning of the follow-up (i.e. if $N = 0$), the subject is *cured* and the cure fraction is given by

$$P[N = 0] = \exp(-\theta) = \lim_{t \to +\infty} S_p(t|\theta, F). \tag{3}$$

Independent baseline covariates, denoted by $\boldsymbol{x}$ (including an intercept) and $\boldsymbol{z}$ (without intercept), may enter, for example, the model through a log-link on parameter θ and through a Cox model for $F(t)$, respectively:

$$\theta(\boldsymbol{x}) = \exp(\boldsymbol{\alpha}^T \boldsymbol{x}), \tag{4}$$
$$F(t|\boldsymbol{z}) = 1 - S_0(t)^{\exp(\boldsymbol{\beta}^T \boldsymbol{z})}. \tag{5}$$

Tsodikov (2002) proposed, in a frequentist framework, a nonparametric estimation of the baseline survival function $S_0(t)$, while Yin and Ibrahim (2005), using a piecewise exponential distribution, and Bremhorst and Lambert (2016), using P-splines, suggested

a flexible specification of $S_0(t)$ in a Bayesian framework, see Gressani and Lambert (2018) for fast inference in that Bayesian setting. Recently, several extensions of the promotion time model were proposed in the literature, see for example Liu and Shen (2009); Kim et al. (2009); Lopes and Bolfarine (2012); Li and Lee (2017) and Bremhorst et al. (2019).

Our article is motivated by the analysis of data from the German Socio-Economic Panel (Wagner et al., 2007) studying the transition to second and third births. Some of the women/family characteristics, such as the educational attainments of the mother and of her partner for example, may vary over time. Therefore, regression models enabling to deal with categorical time-varying covariates need to be specified. Unfortunately, the probability to have an additional child cannot be directly connected to time-varying covariates in the promotion time model since that chance is a one-to-one function of the mean number of activated latent factors, assumed to be fixed (and unknown) for a given woman directly at the beginning of the follow-up study. Chi and Ibrahim (2006) proposed an extension of that family of models to handle such covariates by allowing the latent factors to turn active at any time during the follow-up. However, the assumptions of their extension imply that the population hazard function is monotonically increasing with time which can be unrealistic in specific settings such as fertility studies.

Brown and Ibrahim (2003), Chen et al. (2004) and Kim et al. (2013) developed joint models for longitudinal and cure survival data in the context of the promotion time model. Brown and Ibrahim (2003) proposed a longitudinal model for the immunologic response to vaccination over time, specified the latent hazard as a function of the trajectory of the immunologic marker and modelled the probability of observing the monitored event using baseline covariates only. Chen et al. (2004) and Kim et al. (2013) suggested that the (continuous) longitudinal biomarker has an impact on the probability of being cured only through the random effect of the longitudinal model.

In none of the abovementioned references are time-varying covariates directly included as regressors when modelling the event risk. In the present article, we propose an extension of the promotion time model allowing categorical time-varying covariates to have (simultaneously) an impact on the probability and on the timing of the event-of-interest.

The paper is organized as follows: Section 2 begins with an illustrative example motivating the proposed extension of the promotion time model dealing with categorical time-varying covariates and ends with the theoretical notations and properties of the model. The Bayesian model specification and the MCMC algorithm are described in Section 3. Section 4 is devoted to a simulation study assessing the accuracy of the methodology. Two analyses on real data from the German Socio-Economic Panel studying the transition to second and third births are presented in Section 5. A discussion concludes the paper.
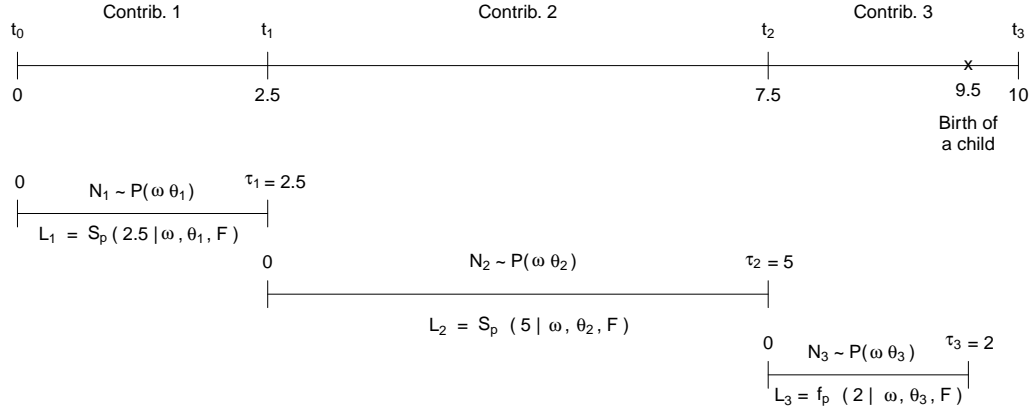
Figure 1: Likelihood contribution for a woman with two variations of a single time-varying categorical covariate (her education level for example) influencing the probability of having an additional child. As an example, θ_1 may denote the expected mean number of potential decisive arguments for a less educated one-child woman to have a second child after her first delivery, while θ_2 and θ_3 may denote that expected mean number directly after getting a secondary school diploma or a university degree, respectively. The unobserved (inter-subject) heterogeneity in the underlying risk is quantified by a multiplying random effect ω with mean 1.0.

2 Extension of the promotion time model

2.1 Motivation

For simplicity, the principles underlying the extended promotion time model are explained on an example, pictured in Figure 1, dealing with a single woman and her educational attainment assumed to affect the probability of pregnancy (but not its timing). Assume that a woman gave birth to her first child at 16 when she was still a student at the secondary school. She is, therefore, reported as a *less educated* one-child woman. As motivated by the promotion time model, it is assumed that she is directly exposed to $N_1 \sim \mathscr{P}(\omega\theta_1)$ potential decisive arguments to decide to have a second child, where ω is a random effect (with density function $g(\omega)$ and mean 1.0) accounting for the unobserved (inter-subject) heterogeneity in the underlying risk. After τ_1 years (2.5, say), she graduated from secondary school without getting a new child. The contribution of this first period to the conditional likelihood is $S_p(\tau_1|\omega,\theta_1,F)$ and her level of education is updated to *medium educated*. The proposed extension of the promotion time model assumes that when the characteristics of the woman change, the N_1 preceding available potential decisive arguments to initiate a second pregnancy are replaced by N_2 new ones where $N_2 \sim \mathscr{P}(\omega\theta_2)$. Five years ($\tau_2 = t_2 - t_1 = 5$) after secondary school, she gets a university degree, again without intermediate pregnancy. The contribution of the com-

pleted period to the conditional likelihood is therefore $S_p(\tau_2|\omega, \theta_2, F)$. As assumed by our proposed extension of the promotion time model, the available potential decisive arguments, corresponding to this second period, to decide to have a second child are replaced by N_3 new ones with $N_3 \sim \mathscr{P}(\omega\,\theta_3)$. Two years later ($\tau_3 = 2$), she gave birth to her second child. The contribution to the conditional likelihood of this final event is $f_p(\tau_3|\omega, \theta_3, F)$. Thus, the contribution to the marginal likelihood of this woman is given by

$$L_{\text{birth}} = \int_0^{+\infty} S_p(\tau_1|\omega, \theta_1, F)S_p(\tau_2|\omega, \theta_2, F)f_p(\tau_3|\omega, \theta_3, F)g(\omega)d\omega.$$

Assume, now, that another woman gets a university degree τ_1 years after entering the study (i.e. after the birth of her first child). Moreover, assume that τ_2 years later, she left the study without getting a second child. Then, her contribution to the marginal likelihood is given by

$$L_{\text{right cens.}} = \int_0^{+\infty} S_p(\tau_1|\omega, \theta_2, F)S_p(\tau_2|\omega, \theta_3, F)g(\omega)d\omega.$$

2.2 General setting

For clarity of exposition, we first assume, in Section 2.2.1, that (time-dependent) co-variates only affect the probability of being cured. That restriction will be dropped in Section 2.2.2 by allowing such covariates to also have an impact on the event timing.

2.2.1 Effect of time-varying covariates on cure probability

Let $\mathscr{X}_i = \{\boldsymbol{x}_{i1}(t_{i0}), \ldots, \boldsymbol{x}_{iJ_i}(t_{i(J_i-1)})\}$ be the history of the vector of covariates influencing the cure probability of subject i $(i = 1, \ldots, I)$, where $\boldsymbol{x}_{i1}(t_{i0})$ denotes the covariate values at the beginning of the first period in the follow-up. As soon as one of the covariates changes, the covariate vector needs to be updated. Assume that those variations occurred at $(J_i - 1)$ occasions $t_{i1}, \ldots, t_{i(J_i-1)}$ such that $t_{i0} = 0 < t_{i1} < \ldots < t_{i(J_i-1)} < t_{iJ_i} = t_i \leq +\infty$ yielding the following sequence of covariate vectors $\boldsymbol{x}_{i1}(t_{i0}), \boldsymbol{x}_{i2}(t_{i1}), \ldots, \boldsymbol{x}_{iJ_i}(t_{i(J_i-1)})$. For $j = 1, \ldots J_i$, let $\tau_{ij} = t_{ij} - t_{i(j-1)}$, be the time spent by the ith subject in his/her jth covariate configuration $\boldsymbol{x}_{ij}(t_{i(j-1)})$ and assume the following regression model $\theta_{ij} = \exp\left(\boldsymbol{\alpha}^T \boldsymbol{x}_{ij}\left(t_{i(j-1)}\right)\right)$ for the expected mean number of latent factors for subject i at the beginning of the jth covariate configuration $\boldsymbol{x}_{ij}(t_{i(j-1)})$. The population survival and hazard functions of the *extended* promotion time model are, respectively, given by

$$S_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) = \left(\prod_{j=1}^{j_i(t)-1} S_p(\tau_{ij}|\omega_i, \theta_{ij}, F) \right) S_p \left(t - \sum_{j=1}^{j_i(t)-1} \tau_{ij} \middle| \omega_i, \theta_{ij_i(t)}, F \right)$$

$$= \exp \left[-\omega_i \left\{ \sum_{j=1}^{j_i(t)-1} \theta_{ij} F(\tau_{ij}) + \theta_{ij_i(t)} F \left(t - \sum_{j=1}^{j_i(t)-1} \tau_{ij} \right) \right\} \right] ; \ (6)$$

$$h_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) = \frac{-d \left(\log \left(S_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) \right) \right)}{dt}$$

$$= h_p \left(t - \sum_{j=1}^{j_i(t)-1} \tau_{ij} \middle| \omega_i, \theta_{ij_i(t)}, F \right)$$

$$= \omega_i \theta_{ij_i(t)} f \left(t - \sum_{j=1}^{j_i(t)-1} \tau_{ij} \right), \qquad (7)$$

where $\mathscr{F}_i = \{ \mathscr{X}_i, t_{ij} : j = 0, \ldots, J_i \}$ is the set of the follow-up history of subject i, $j_i(t) = \max \{ j | t < t_{ij}, j = 1, \ldots, J_i \}$, $\prod_{j=1}^{0}(\cdot) = 1$ and $\sum_{j=1}^{0}(\cdot) = 0$.

Note that the survival function defined in (6) is an improper survival function and its limit given by

$$\lim_{t \to +\infty} S_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) = \exp \left[-\omega_i \left(\sum_{j=1}^{J_i-1} \theta_{ij} F(\tau_{ij}) + \theta_{iJ_i} \right) \right] > 0 \qquad (8)$$

defines the conditional probability that subject i is cured.

Given the follow-up history $\mathscr{F}_i$ of subject i, if we assume that he/she belongs to the susceptible sub-population after the last update of his/her covariate configuration, while his/her status (susceptible or cured) was unknown during the first $J_i - 1$ periods of his/her follow-up, his/her survival and hazard functions are, respectively, defined as

$$S_u^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) = \begin{cases} S_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) & \text{if } t \leq t_{i(J_i-1)} \\ S_p^{\text{ext}}(t_{i(J_i-1)}|\omega_i, \mathscr{F}_i, F) S_u(\tau_{it}|\omega_i, \theta_{iJ_i}, F) & \text{if } t > t_{i(J_i-1)} \end{cases} ;$$

$$h_u^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) = \begin{cases} h_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) & \text{if } t \leq t_{i(J_i-1)} \\ \frac{S_p(t|\omega_i, \theta_{iJ_i}, F)}{S_p(t|\omega_i, \theta_{iJ_i}, F) - \exp[-\theta_{iJ_i}]} h_p(\tau_{it}|\omega_i, \theta_{iJ_i}, F) & \text{if } t > t_{i(J_i-1)} \end{cases} ,$$

where $\tau_{it} = t - \sum_{j=1}^{J_i-1} \tau_{ij}$ and $S_u(t|\omega, \theta, F) = \frac{S_p(t|\omega, \theta, F) - \exp[-\omega\theta]}{1 - \exp[-\omega\theta]}$ is the susceptible survival function of the classical promotion time model. Integrating out the unobservable heterogeneity ω_i, the marginal population survival and hazard functions are, respectively, given by

$$S_p^{\text{ext}}(t|\mathscr{F}_i, F) = E_{\omega_i} \left[S_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) \right] ; \ h_p^{\text{ext}}(t|\mathscr{F}_i, F) = E_{\omega_i} \left[h_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F) \right], \quad (9)$$

while the expressions for a susceptible subject in the last period are

$$S_u^{\text{ext}}(t|\mathscr{F}_i, F) = E_{\omega_i}\left[S_u^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F)\right] \ ; \ h_u^{\text{ext}}(t|\mathscr{F}_i, F) = E_{\omega_i}\left[h_u^{\text{ext}}(t|\omega_i, \mathscr{F}_i, F)\right]. \quad (10)$$

These expressions will be found useful in the application of Section 5 to compare the evolution of the hazard function at the population level (mixing 'cured' and susceptible subjects) with its behavior over time in the subgroup of susceptible subjects, see Figures 5 and 6. Their baseline values can be obtained by setting each covariate at its reference value.

2.2.2 Effect of time-varying covariates on event timing

In the previous section, the history of the covariate vector of subject i influencing his/her probability to be cured was used to split the follow-up into J_i periods of respective length τ_{ij} ($j = 1, \ldots, J_i$). Since the time-varying covariates influencing the probability of being cured may differ from those influencing the event timing, one needs to determine the history of the covariate vector entering the model through the latent distribution $F(t)$ during each of the J_i periods discussed in Section 2.2.1: it yields $\mathscr{Z}_{ij} = \{z_{ij1}(t_{ij0}), \ldots, z_{ijL_{ij}}(t_{ij(L_{ij}-1)})\}$, where $t_{i(j-1)} = t_{ij0} < \ldots < t_{ij(L_{ij}-1)} < t_{ijL_{ij}} = t_{iJ_i}$ denote the changing times of z_i's within period j ($j = 1, \ldots J_i$) for subject i. Using this new set of covariates, the extended population survival function defined in (6) becomes:

$$S_p^{\text{ext}}(t|\omega_i, \mathscr{F}_i) = \left(\prod_{j=1}^{j_i(t)-1} S_p(\tau_{ij}|\omega_i, \theta_{ij}, F_{ij})\right) S_p\left(t - \sum_{j=1}^{j_i(t)-1} \tau_{ij} \middle| \omega_i, \theta_{ij_i(t)}, F_{ij_i(t)}\right), \quad (11)$$

where $\mathscr{F}_i : \{\mathscr{X}_i, t_{i0}, t_{ij}, \mathscr{Z}_{ij}, t_{ijl} : j = 1, \ldots, J_i; l = 0, \ldots, L_{ij}\}$ is the set of the follow-up history of the ith subject and F_{ij} is specified according to the Andersen-Gill model (Andersen and Gill, 1982; Fleming and Harrington, 1991; Andersen et al., 1993):

$$
\begin{aligned}
F_{ij}(t|\mathscr{Z}_{ij}, t_{ij0}, \ldots, t_{ijL_{ij}}) &= 1 - S_{ij}(t|\mathscr{Z}_{ij}, t_{ij0}, \ldots, t_{ijL_{ij}}) \\
&= 1 - \exp\left(-\sum_{l=1}^{L_{ij}} \exp\left(\boldsymbol{\beta}^T \boldsymbol{z}_{ijl}\left(t_{ij(l-1)}\right)\right) H_{ijl}\right), \quad (12)
\end{aligned}
$$

where $H_{ijl} = \int_{t_{ij(l-1)}}^{\min(t, t_{ijl})} h_0(u)du$. If $L_{ij} = 1$, (12) defines the classical Cox proportional hazards model. As suggested by Bremhorst and Lambert (2016), in the context of the promotion time model, the logarithm of the baseline hazard function, $\log(h_0(t))$, can be modelled using a linear combination of (a large number K of) cubic B-splines :

$$h_0(t) = \exp\left(\sum_{k=1}^{K} b_k(t)\phi_k\right),$$

where $\{b_k(.), k = 1, \ldots, K\}$ denotes the cubic B-splines basis associated to a predefined number of equidistant knots on $[0, \max_{ij}(\tau_{ij})]$. As proposed by Eilers and Marx (1996,

2010), a roughness penalty on rth order finite differences of adjacent B-spline parameters $(\lambda_1 \sum_k (\Delta^r \phi_k)^2$, where λ_1 is the penalty parameter) is subtracted from the log-likelihood to ensure the smoothness of the baseline hazard function.

2.3 Likelihood and identification issues

Denote by δ_i the event indicator for subject i and let $\boldsymbol{\Phi}$ be the set of all model parameters. In a right censored setting, the likelihood is obtained by integrating out the unobserved individual random effects from the complete data likelihood:

$$
\begin{aligned}
L\left(\boldsymbol{\Phi} | \mathscr{F}, \boldsymbol{\delta}\right) &= \prod_{i=1}^{I} E_{\omega_i} \left[\left(S_p^{\text{ext}} \left(t_i | \omega_i, \mathscr{F}_i \right) \right)^{1-\delta_i} \left(f_p^{\text{ext}} \left(t_i | \omega_i, \mathscr{F}_i \right) \right)^{\delta_i} \right] \\
&= \prod_{i=1}^{I} \left(\mathscr{L}_{\omega_i} \left(\mathscr{H} \left(t_i | \mathscr{F}_i \right) \right) \right)^{1-\delta_i} \left(-\mathscr{L}_{\omega_i}^{(1)} \left(\mathscr{H} \left(t_i | \mathscr{F}_i \right) \right) \right)^{\delta_i},
\end{aligned}
\tag{13}
$$

where $f_p^{\text{ext}} \left(t_i | \omega_i, \mathscr{F}_i \right) = -\frac{dS_p^{\text{ext}}(t_i | \omega_i, \mathscr{F}_i)}{dt}$ is the population density function associated to (11), $\mathscr{H} \left(t_i | \mathscr{F}_i \right) = -E_{\omega_i} \left[\log \left(S_p^{\text{ext}} \left(t_i | \omega_i, \mathscr{F}_i \right) \right) \right]$, $\mathscr{L}_{\omega}$ and $\mathscr{L}_{\omega}^{(1)}$ respectively denote the Laplace transform and the first derivative of the Laplace transform of the individual random effect density $g(\omega)$.

The unobserved inter-subject heterogeneity in the event (conditional) probability is quantified by ω and assumed to have a Power Variance Function (PVF) density function $g(\omega)$ (Hougaard, 1986). The Laplace transform and the first derivative of the Laplace transform are, respectively, given by

$$
\begin{aligned}
\mathscr{L}_{\omega}(s) &= \exp \left[\frac{\tilde{\eta}}{\exp(\xi)(1-\tilde{\eta})} \left(1 - \left(1 + \frac{\exp(\xi+\mu) s}{\tilde{\eta}} \right)^{1-\tilde{\eta}} \right) \right] ; \\
\mathscr{L}_{\omega}^{(1)}(s) &= -\exp(\mu) \left(1 + \frac{\exp(\xi+\mu) s}{\tilde{\eta}} \right)^{-\tilde{\eta}} \mathscr{L}_{\omega}(s),
\end{aligned}
\tag{14}
$$

where $\mu, \xi, \eta \in \mathbb{R}$ and $\tilde{\eta} = \frac{\exp(\eta)}{1+\exp(\eta)}$, see Aalen (1992).

This flexible family of distributions contains the gamma (when $\mu = 0$ and $\eta \to +\infty$), the inverse Gaussian (when $\mu = \eta = 0$) and the positive stable (when $\mu \to +\infty$ and $\exp(\xi) = \tilde{\eta}(1-\tilde{\eta})^{-1/\tilde{\eta}} \exp \left(\mu(\tilde{\eta}^{-1} - 1) \right)$ as limiting cases (Duchateau and Janssen, 2008, Section 4.5).

In the proposed extension of the promotion time model, ω can be seen as a frailty term in a survival framework (Clayton, 1978; Liang et al., 1995; Hougaard, 1995; Govindarajulu et al., 2011). To ensure identifiability of the random effect distribution parameters, one usually constrains the mean frailty to be one, or, alternatively, the mean of the log frailty to be zero. In our setting, we decided to set $\log \left(E_\omega[\omega] \right) = \mu = 0$.

A crucial assumption to ensure identifiability in cure survival models (when only baseline covariates are taken into account) is to impose the zero tail constraint on the latency part of the model (Taylor, 1995; Zeng et al., 2006). Practically, it is assumed

8

that the latent distribution $F(\cdot)$ in the promotion time model is 1.0 beyond the last event time. When the latent distribution is specified using B-splines, Bremhorst and Lambert (2016) suggest to fix the last spline coefficient ϕ_K to a large enough value (say, 10) to force $S(t)$ to smoothly decrease to 0 by the end of the follow-up, see Lambert and Bremhorst (2019) for a broader perspective on identification issues. In practice, this sufficiently long follow-up assumption looks reasonable if the Kaplan Meier estimate of the survival function shows a plateau in the right tail of the distribution.

In the proposed extension of the promotion time model, the identification will be ensured by assuming that some subjects are observed for a sufficiently long time τ_{ij} with the same combination of covariate values $\boldsymbol{x}_{ij}$ in the cure probability specification in (4) and (6). In practice, the sufficiently long assumption will be considered reasonable if $\tau_0 > \tau_1$, where τ_0 denotes the maximum time spent within a covariate configuration without having had the event-of-interest and τ_1 the maximum observed event time within a covariate configuration. However, when τ_0 is much larger than τ_1, the estimation of the spline parameters can be numerically instable since no information is available to estimate the spline parameters supported by knots located in the interval $[\tau_1, \tau_0 = \max_{ij}(\tau_{ij})]$. Therefore, all the spline parameters associated to knots located over τ_1, the maximum observed failure time within a covariate configuration, are set to an arbitrary large value (10, say) and not only the last one as previously suggested.

3 Bayesian inference

3.1 Prior distributions

As suggested by Lang and Brezger (2004) in the Bayesian framework, the penalty introduced by Eilers and Marx (1996, 2010) to force smoothness (here, in the baseline hazard) is translated into a multivariate normal prior distribution for the spline parameters:

$$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K) \sim N_K(0, (\lambda_1 P)^{-1}),$$

where λ_1 is the penalty parameter, $P = D^\top D + \epsilon I_k$ is a full rank matrix for some small quantity ϵ (10^{-6}, say) and D is the rth order difference matrix.

A robust prior distribution is considered for the penalty parameter λ_1 (Jullion and Lambert, 2007):

$$\lambda_1 | \lambda_2 \sim \mathscr{G}\left(\frac{\nu}{2}, \frac{\nu \lambda_2}{2}\right) \; ; \; \lambda_2 \sim \mathscr{G}\left(a_{\lambda_2} = 10^{-4}, b_{\lambda_2} = 10^{-4}\right),$$

where $\mathscr{G}(a, b)$ denotes a Gamma distribution with mean a/b. Jullion and Lambert (2007) showed that whatever the chosen noninformative prior distribution for ν, no relevant information can be obtained from its posterior distribution since it is close to the uniform. Therefore, fixing $\nu = 2$, for example, will not influence the shape of the estimated hazard function.

Non-informative normal prior distributions are considered for all regression parameters and the PVF distribution parameters η and ξ in (14). Prior knowledge, when available, can be injected through specific priors.

3.2 Posterior sampling using MCMC

Using Bayes' theorem, the joint posterior distribution is given by

$$\pi(\mathbf{\Phi}|\mathscr{F},\boldsymbol{\delta}) = L(\mathbf{\Phi}|\mathscr{F},\boldsymbol{\delta})\pi(\boldsymbol{\phi}|\lambda_1)\pi(\lambda_1|\lambda_2)\pi(\lambda_2)\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\xi)\pi(\eta). \qquad (15)$$

Only the conditional posterior distribution of the penalty parameters λ_1 and λ_2 belong to a known family of distribution :

$$\lambda_1|\lambda_2,\boldsymbol{\phi},\mathscr{F},\boldsymbol{\delta} \sim \mathscr{G}\left(\frac{\nu+K}{2},\frac{\nu\lambda_2}{2}\boldsymbol{\phi}^\top P\boldsymbol{\phi}\right) \; ; \; \lambda_2|\lambda_1,\mathscr{F},\boldsymbol{\delta} \sim \mathscr{G}\left(a_{\lambda_2}+\frac{\nu}{2},b_{\lambda_2}+\frac{\nu\lambda_1}{2}\right).$$

Therefore, a Metropolis-within-Gibbs algorithm will be used to sample from the joint posterior distribution. The proposals for the spline parameters ($\boldsymbol{\phi}$) and the regression parameters related to the probability of the event ($\boldsymbol{\alpha}$) and its timing ($\boldsymbol{\beta}$) are generated in multivariate Metropolis steps while univariate ones are used for the PVF distribution parameters. As suggested by Haario et al. (2001) and by Atchadé and Rosenthal (2005), for each Metropolis step, the standard deviation of the proposal distribution is updated during the burnin period to reach the targetted acceptance rate (44% and 23% in the univariate and multivariate cases, respectively), see Roberts and Rosenthal (2001). Reparametrizing the conditional posterior distribution, using a rough frequentist estimation of the posterior correlation structure of the parameters, could reduce the autocorrelation in the posterior chains (Lambert, 2007). Moreover, the convergence of the chain could be speeded up by a clever choice of the initial value of the model parameters, the mode of the joint posterior distribution in (15), for example. It can be computed using a non-linear optimizer with as a by-product minus the inverse of the Hessian matrix for further use as the variance-covariance matrices in the preceding (multivariate) Metropolis steps.

4 Simulation study

The behaviour of the extended promotion time model, dealing with categorical time-varying covariates defined in (11), combined with the Bayesian inference techniques described in Section 3 are assessed through an extensive simulation study. Two different proportions of cured individuals (25% and 50%) without and with (15% of) random right censoring among the susceptible population were considered. The simulation results are based on $R = 500$ replications of sample size $n = 500$ and $n = 1000$ with an average of 1.5 or 3 covariate changes per subject in the cure probability model (cf. Section 2.2), leading to 16 different settings. For each of them, the baseline distribution $S_0(t)$ of the Andersen-Gill model, defined in (12), is specified as a Weibull distribution with mean 8.0 and standard deviation 4.18. For numerical purposes, the maximum time spent by a subject in the jth covariate configuration $\boldsymbol{x}_j$ ($j = 1, \ldots, J$) was set to 25.0. If the simulated time exceeds 25.0 under the last covariate configuration $\boldsymbol{x}_J$, the corresponding subject is considered *cured*. This is not a restriction since less than 0.1% of the events are expected to occur after that time under the chosen Weibull distribution.

Three independent covariates will simultaneously influence the probability of being cured and the distribution of the event timing for susceptible subjects: one time-varying binary covariate ($W_1 \sim$ Bernoulli(0.5) at the beginning of the follow-up), one fixed binary ($W_2 \sim$ Bernoulli(0.5)) and one fixed continuous ($W_3 \sim N(0, 0.25)$) covariates. If any, the variation times $t_1, \ldots, t_{J-1}$ of the time-dependent covariate W_1 are controlled using an exponential distribution with rate ψ. The values of the regression parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ and of ψ in Table 2 are tuned to reach the desired percentage of cured individuals, while the expected number of changes per subject in the time-varying W_1 is denoted by E(J). One additional time-varying binary covariate ($W_4 \sim$ Bernoulli(0.5) at the beginning of the follow-up and independent of W_1, W_2, W_3) with $L \sim \mathscr{P}(2)$ (Poisson distributed) variations within each covariate configuration $\boldsymbol{x}_j$ is assumed in the Andersen-Gill model describing the event timing for the susceptible sub-population. In each setting, the changing times of W_4 ($t_{j1}, \ldots, t_{jL}$ for $j = 1, \ldots, J$), if any, are controlled using an exponential distribution with mean 10.0 and the values of the regression parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$, in the Andersen-Gill model, are set to $(-0.5, 0.7, 0.5, -0.4)$.

The results presented in Table 1 and in Figures 2, 3 and 4 are based on one posterior chain of length 70 000 including a burnin period of length 35 000. Convergence diagnostics tools such as those proposed in Geweke (1992) are used to assess the convergence of the MCMC chain.

Table 1 reports the coverage probabilities of the 95% credible intervals for all regression parameters ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) under each of the 16 different simulation settings. Except for β_4 (the regression parameter associated to the time-varying categorical covariate influencing only the timing of the event for susceptible subjects), the coverage probabilities are all close to the 95% nominal value whatever the simulation setting, even with a moderate sample size ($n = 500$) and a high percentage of right censoring (65%). Figure 2 shows the errors (evaluated as the posterior median minus the true value) of the regression parameters $\boldsymbol{\beta}$ in the Andersen-Gill model, related to the timing of the event for susceptible subjects. The fourth row suggests a small negative bias for β_4 usually associated to slightly undercovering credible intervals (cf. Table 1). A small negative bias, unaffected by the proportion of cured subjects, decreasing with sample size and increasing with the amount of right censored observations (in the susceptible population), is also revealed for the intercept α_0 (the first row of Figure 3). However, as mentioned previously, since the coverage probabilities of α_0 always coincide with their nominal value, it suggests that the uncertainty and the plausible values for that intercept are properly quantified. The other regression parameters are estimated without bias.

As expected, Figures 2 and 3 also indicate that the uncertainty in the parameter estimates increases with the proportion of cured individuals and with the percentage of random right censored susceptible subjects, but decreases with sample size.

The baseline distribution is estimated without bias as illustrated in Figure 4 for datasets of sample size $n = 500$ with an average of 3 covariate changes per subject in the cure probability model. Not surprisingly, compared to settings where no susceptible subject are censored, a growth in the uncertainty is observed when right censoring is

introduced. From non-reported results, one can also conclude that the uncertainty of the baseline distribution estimates decreases with sample size.

Finally, the simulation results suggest that the number of covariate changes per subject in the cure probability model does not markedly affect the accuracy of the model parameter estimates.

Table 1: Simulation results for $R = 500$ replicates: coverage probabilities (in %) of the 95% credible intervals of each regression parameter. The true values of the regression parameters $\boldsymbol{\alpha}$ (tuned to reach the target percentage of cured individuals) are given in Table 2 while the true values of the regression parameters $\boldsymbol{\beta}$ are set to (-0.5,0.7,0.5,-0.4).

n	E(J)	Cured	Censored	Probability				Timing			
				α_0	α_1	α_2	α_3	β_1	β_2	β_3	β_4
500	1.5	25%	25%	94.8	93.4	94.4	95.0	94.8	93.4	96.0	92.6
		25%	40%	95.2	96.2	92.8	94.8	96.0	92.2	93.4	89.6
		50%	50%	97.6	94.2	97.0	95.6	93.4	96.2	93.0	94.0
		50%	65%	95.0	94.2	93.8	95.2	92.2	93.8	93.2	89.0
	3	25%	25%	95.4	94.8	95.6	94.6	92.6	95.4	93.4	91.6
		25%	40%	93.6	91.8	93.8	92.4	91.8	94.8	92.4	89.2
		50%	50%	97.8	95.6	94.6	93.6	92.6	94.4	95.0	91.2
		50%	65%	95.6	94.4	93.0	95.4	92.2	92.4	92.0	90.6
1000	1.5	25%	25%	97.0	95.2	93.8	95.2	94.8	93.6	95.0	93.2
		25%	40%	95.8	95.0	94.6	94.4	95.6	94.6	93.6	91.4
		50%	50%	97.0	96.0	95.6	94.4	93.0	95.6	94.4	91.8
		50%	65%	96.0	94.4	94.0	96.2	92.2	92.4	95.2	88.4
	3	25%	25%	96.0	95.2	95.2	94.6	95.6	94.4	93.2	87.4
		25%	40%	96.0	95.0	95.2	94.2	94.6	93.8	92.8	84.8
		50%	50%	97.2	95.6	94.4	95.6	94.0	94.8	93.6	90.0
		50%	65%	96.6	95.0	94.2	94.4	94.6	92.2	93.4	87.0

Table 2: True values for the regression parameter vector $\boldsymbol{\alpha}$ influencing the probability of the event and for the rate parameter ψ controlling the number of contributions J per subject to the likelihood.

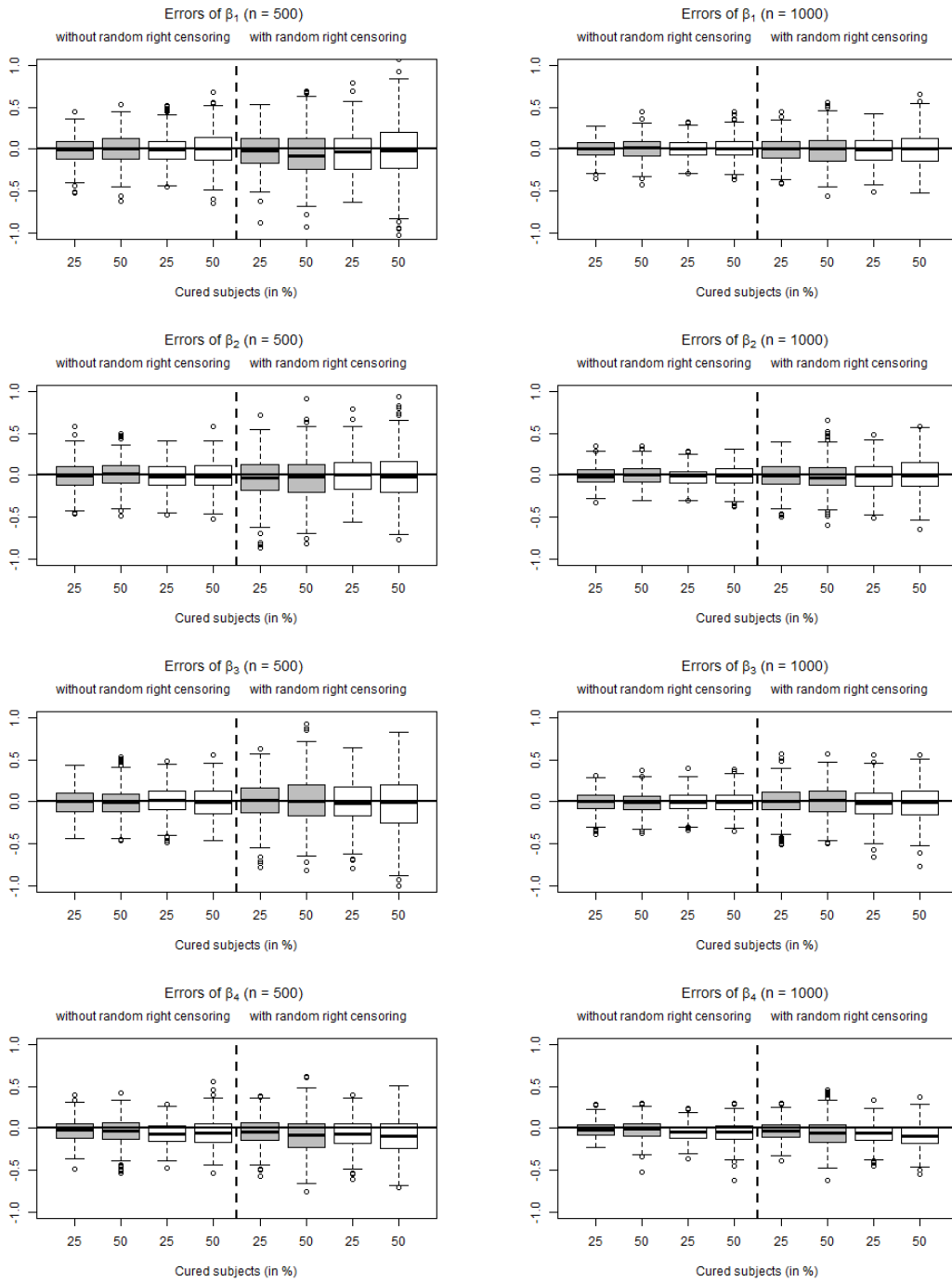| E(J)	Cured	Censored	α_0	α_1	α_2	α_3	ψ
1.5	25%	25%	0.25	-0.50	0.40	0.60	0.033
	25%	40%	0.30	-0.50	0.40	0.60	0.050
	50%	50%	-0.20	-0.40	-0.40	0.60	0.024
	50%	65%	-0.10	-0.40	-0.40	0.60	0.128
3	25%	25%	0.15	-0.50	-0.50	0.60	0.077
	25%	40%	0.25	-0.40	-0.40	0.60	0.128
	50%	50%	-0.60	-0.50	-0.50	0.60	0.059
	50%	65%	-0.40	-0.50	-0.40	0.60	0.115

Figure 2: Simulation results for $R = 500$ replicates: Boxplots of the errors of the regression parameter estimates (the posterior medians) in the regression model for the timing of the event for susceptible subjects under each of the 16 simulation settings (see Table 1). Grey boxes: $\mathrm{E}(J) = 1.5$. White boxes: $\mathrm{E}(J) = 3$.
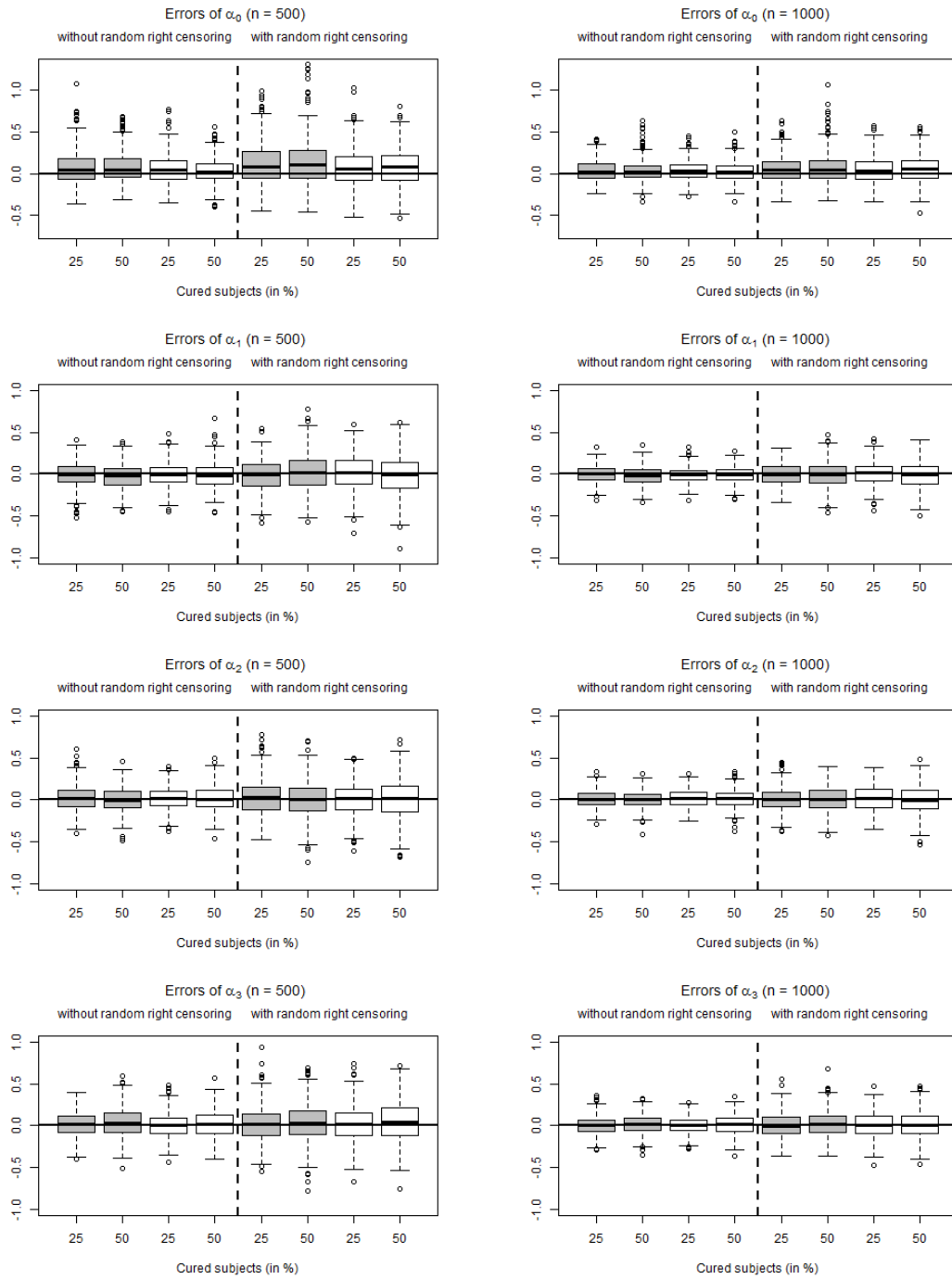
Figure 3: Simulation results for $R = 500$ replicates: Boxplots of the errors of the regression parameter estimates (the posterior medians) influencing the probability of being cured. Grey boxes: $E(J) = 1.5$. White boxes: $E(J) = 3$.
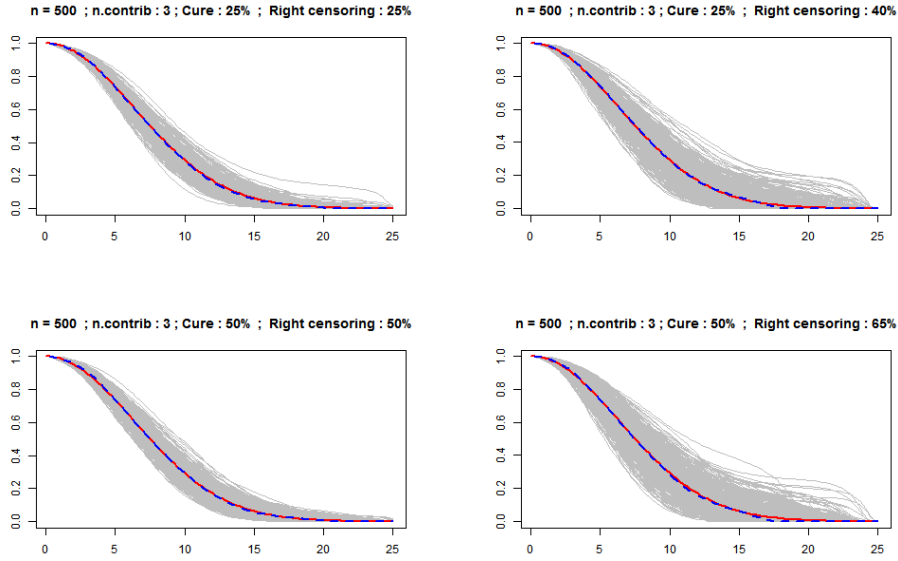
14

Figure 4: Simulation results for $R = 500$ replicates: estimates of the baseline distribution $S_0(t)$ (one gray curve per data set), sample size $n = 500$ and an average of 3 likelihood contributions per subject with 25% (row 1) and 50% (row 2) of cured individuals. The solid line corresponds to the true function and the dashed line is the pointwise median of the 500 estimated curves. Left : All right censored individuals are cured. Right : 15% of the susceptible subjects are right censored.

5 Application to fertility studies

The German Socio-Economic Panel (GSOEP) is a representative panel for Germany. More information on this survey, launched in 1984 and updated on an annual basis since then, can be found in Wagner et al. (2007). Using version v31.0, Bremhorst et al. (2016) studied the transition to second (resp. third) birth. Only one child (resp. two child) German women living in West Germany who were still of childbearing age (17-49 years) between 1984 and 2013 were considered in the sample. The main interest of the authors was to study the effect of the educational attainments of the women and of their partners on the probability and on the timing of an additional child. In their analyses, the educational levels were frozen at the onset of the process (i.e. directly after the first or second birth) since time-dependent covariates cannot enter the probability of the event in the promotion time model. Using our proposed extension of the promotion time model, results accounting for possible evolution of the mother and father educational levels are reported in this section. As in Bremhorst et al. (2016), the educational attainments are divided into three categories : *low* (less than a vocational training degree), *medium* (at most vocational training degree), and *high* (a university or a college degree). Regarding the partner's educational level, an additional category named *no partner* was introduced

15

for single women. Given that a degree cannot be lost, cases where a decreasing level of education were reported were deleted from the sample. The final sample contains 1352 women (resp. 1128) of whom 48% (resp. 19%) reported having had a second (resp. third) child.

Age at first birth and the calendar period at the beginning of the follow-up (for identification purposes) are considered as independent baseline continuous covariates. Moreover, the time elapsed between the first two births and the sex composition of prior kids will also enter the model for third birth. All independent variables will influence simultaneously the probability and the timing of getting a new child.

It should however be noted that although one would expect that a high age at first birth would have a stronger negative effect on the probability of having a second or third child for low-educated women than for highly educated ones (assuming that highly educated women regularly postpone childbirth in order to advance in their career, while women with a low level of education who postpone parenthood are a select group, possibly with impaired abilities to have children), in our particular application, the interaction terms between age at first birth and mother's education were not found to be statistically significant and are, therefore, not considered in the final model. The descriptive statistics for the covariates are presented in Tables 3, 4 and 5. For the partner's education level, it should be noted that a divorce or break up with her co-residential partner is considered as a change in the covariate (then, updated to *no partner*). The numerical results presented in Tables 6 and 7 are based on one posterior chain of length $150\,000$ including a burnin of length $50\,000$. For given initial values, our code required about 8 minutes to produce that long chain for the second birth dataset, and about one minute less for the third birth data. Convergence of the chains was checked by an examination of the trace plots and using the z-score of the Geweke convergence diagnostics (Geweke, 1992).

Table 3: Descriptive statistics for the number J of changes in the education level for the 48% of one-child ($n = 1352$) and 19% of two-child ($n = 1128$) mothers (and their possible partner) reported to have had an extra child before censoring.

	Number J of changes in Mother's education level			
Extra child	0	1	2	Total
Second child	95.6%	4.2%	0.2%	100%
Third child	97.9%	2.1%	0.0%	100%

	Number J of changes of partners or in Partner's education level					
Extra child	0	1	2	3	4	5	Total
Second child	81.3%	15.2%	2.5%	0.6%	0.2%	0.2%	100%
Third child	83.1%	12.6%	3.4%	0.5%	0.3%	0.1%	100%

Table 4: Descriptive statistics of the categorical covariates at the end of the follow-up for the 48% of one-child ($n = 1352$) and 19% of two-child ($n = 1128$) mothers reported to have had an extra child before censoring.

| | Mother's education | | | |
Extra child	Low	Medium	High	Total
Second child	15.5%	68.2%	16.3%	100%
Third child	12.8%	70.7%	16.5%	100%

| | Partner's education level | | | | |
Extra child	Low	Medium	High	No partner	Total
Second child	8.4%	51.8%	17.8%	22.0%	100%
Third child	8.0%	49.9%	21.0%	21.1%	100%

| Sex of the first two kids for 3-child mothers | | | |
2 boys	2 girls	Boy & girl	Total
25.3%	25.4%	49.3%	100%

Table 5: Descriptive statistics of the baseline continuous covariates for the 48% of one-child ($n = 1352$) and 19% of two-child ($n = 1128$) mothers reported to have had an extra child before censoring.

	Mean	Median	SD
Calendar period			
Second child	1998	1999	8.04
Third child	2000	2001	8.39
Age at 1st birth			
Second child	28.3	28.0	4.93
Third child	27.4	27.0	4.58
Time between			
first 2 kids	3.44	2.83	2.27

5.1 Second birth

Without using cure survival models and with time-constant covariates, Kreyenfeld (2002) and Bartus *et al.* (2013) studied the transition to second birth in West Germany and in Hungary, respectively. Both analyses pointed out that high educated women space the first and the second births closer together than women with lower education. The interpretation was that, compared to low educated women, those who have a university or a college degree generally tend to have their first child later and are, therefore, closer to the end of their reproductive span, which reduces the waiting time to the second birth. Using the promotion time cure model with covariates set at their values directly after

Table 6: Second birth. Estimate of the posterior median and of the posterior standard deviation for each regression parameter using MCMC. (Signif. codes : * = 0.1 ; ** = 0.05 ; *** = 0.01)

	Probability			Timing		
	Est	sd$_{post}$		Est	sd$_{post}$	
Intercept	0.441	0.203		–	–	
Education						
Low	-0.468	0.195	**	0.344	0.217	
Medium (Ref.)	0.000			0.000		
High	1.076	0.324	***	-0.921	0.339	***
Partner's education						
No partner	-1.116	0.200	***	0.230	0.225	
Low	-0.012	0.257		-0.061	0.288	
Medium (Ref.)	0.000			0.000		
High	0.233	0.195		0.253	0.212	
Calendar Period (std)	0.105	0.088		-0.092	0.102	
Age at first birth (std)	-0.636	0.128	***	0.264	0.129	**

first birth, Bremhorst et al. (2016) concluded that the effect of the education level of West German women on the timing of a second birth was not (statistically) significant (at the 5 percent credibility level), see Table 3 in that paper.

The inclusion of time-varying covariates in our extended promotion time model (cf. Section 2) for the second birth leads to other conclusions, see Table 6. Indeed, while the (significantly) increasing trend in parameter estimates (from -0.468 to 1.076) suggests (like in the basic promotion model) that the probability to have a second child increases with the mother's education level, we found in addition that a woman with a high education level willing to have a second child tends to wait (significantly) longer than others to experience her pregnancy (see the -0.921 estimate in the right part of the table). On the other hand, and as expected, while the probability to have a second child (significantly) decreases with the mother's age at first birth, the time to the second child (significantly) decreases with that age for a mother deciding to have an extra one. Finally, while the education level of partners does not (significantly) affect the probability and the timing of a second delivery, their absence not surprisingly (significantly) decreases the chance of a family extension.

Figure 5 pictures the population baseline hazard function (left) and the baseline hazard function of susceptible women (right) with their 95% pointwise credible intervals. These baseline values are obtained by setting each covariate at its reference value in Equations (9) and (10).

Since the population is a mixture of susceptible and non-susceptible women, it was expected that the instantaneous risk of having a second child is smaller for the whole population than for susceptible women. It should be noted that the shapes of the two functions slightly differ: the baseline hazard function for susceptible women peaks 3.5 years after the last update of their education level or of their partner (or after the first
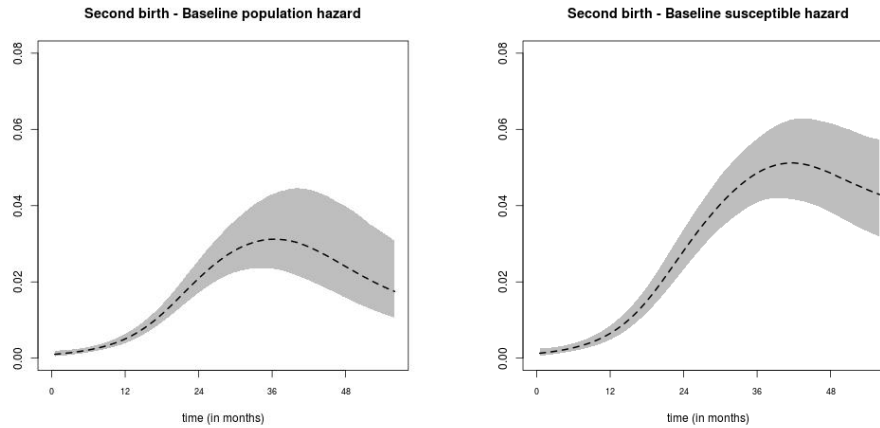
Figure 5: Second birth - Fitted baseline population hazard (left) and fitted baseline hazard for susceptible women (right) with 95% pointwise credible intervals.

birth if no change occurred in their education levels), then tends to slightly decrease afterwards. On the other hand, the population hazard function shows a peak sooner (3 years after after the update of the parent education levels or after the first birth if no change occurred in their education levels) and decreases thereafter. A more detailed explanation of the differences between the population and the susceptible hazards is available in Bremhorst et al. (2016).
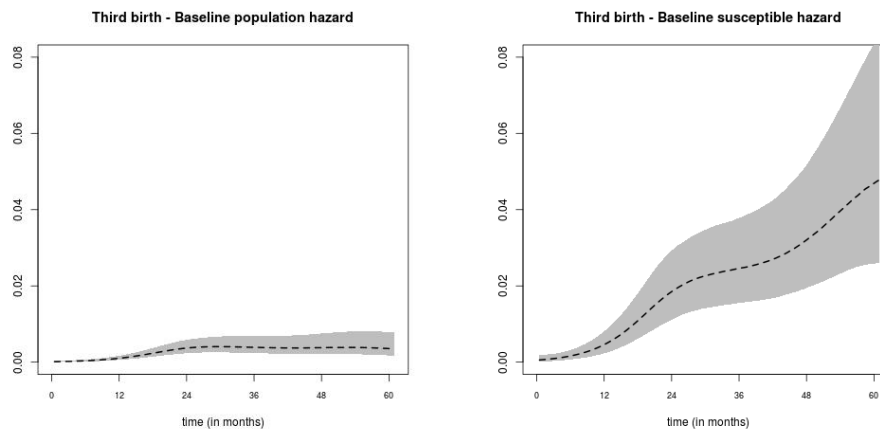
5.2 Third birth

Figure 6: third birth - Fitted baseline population hazard (left) and fitted baseline hazard for susceptible women (right) with 95% pointwise credible intervals.

Table 7: Third birth. Estimate of the posterior median and of the posterior standard deviation for each regression parameter using MCMC. (Signif. codes : * = 0.1 ; ** = 0.05 ; *** = 0.01)

	Probability			Timing		
	Est	sd_{post}		Est	sd_{post}	
Intercept	-1.367	0.268		–	–	
Sex. comp. of prior kids						
Two boys	0.313	0.292		-0.261	0.309	
Boy & Girl (Ref.)	0.000			0.000		
Two girls	0.213	0.286		-0.147	0.293	
Education						
Low	0.479	0.371		-0.072	0.385	
Medium (Ref.)	0.000			0.000		
High	0.087	0.370		0.070	0.382	
Partner's education						
No partner	-0.241	0.350		0.765	0.411	*
Low	0.295	0.498		0.072	0.541	
Medium (Ref.)	0.000			0.000		
High	1.056	0.354	***	0.166	0.352	
Calendar Period (std)	-0.026	0.146		-0.202	0.146	
Age at first birth (std)	-0.875	0.205	***	0.251	0.194	
Time between first 2 kids (std)	-0.557	0.144	***	0.232	0.148	

Table 7 presents the posterior median (as point estimate) and the posterior standard deviation of each regression parameter in the extended promotion time model (cf. Section 2) fitted on the data for third birth. This time, the education level of the mother does not have any significant effect on the probability or the timing of a third pregnancy. However, having a highly educated partner significantly increases the probability of having a third child (but not its timing). As expected, significant negative effects were found for the mothers's age at first birth and for the time elapsed between the first two births on the third birth probability. The absence of any significant effect on the pregnancy timing might be explained by a lack of power following from the small number of observed third births.

Figure 6 shows that the population hazard function remains rather small after the last update of the parent education levels or after the second birth if no change in education levels occurred, while the hazard function of the susceptible mothers increases directly after it.

6 Discussion

This paper was motivated by the the desire to include time-varying categorical covariates in cure survival models. Our extension of the promotion time model enables to simultaneously include such covariates in the regression sub-models for the probability

of the monitored event and its timing for susceptible subjects. A random effect with a flexible parametric distribution, the Power Variance Function distribution (containing the gamma and the inverse Gaussian as particular cases), is also added to account for the unobserved heterogeneity. This new model family enables to handle non monotone population hazard functions and does not require restrictive or arbitrary parametric assumptions for the baseline hazard. It is assumed that the changing times of the (categorical) covariates are exactly known leading to a piecewise constant time-varying covariate framework. Introducing continuous (non piecewise constant) time-varying covariates directly as regressors in the probability model is not straightforward and would require further research. Chi and Ibrahim (2006) already proposed such a model, but their assumptions lead to a monotonically increasing population hazard function, which can be unrealistic in many contexts, including fertility studies. An extensive simulation study revealed that (nearly) all the regression parameters were properly estimated using our Bayesian strategy: biases were non significant with a frequentist coverage of Bayesian credible intervals coherent with their nominal value. The exception concerns regression parameters associated to a time-varying covariate only influencing the event timing for susceptible individuals. Then, a moderate bias was revealed together with slightly undercovering credible intervals. However, no bias was found for the model parameter estimates of time-varying covariates with a simultaneous impact on the probability and on the timing of the event for susceptible subjects. Therefore, when using this model, we suggest to enter time-varying covariates simultaneously in the two regression sub-models as done in the application. Further research is required to understand the cause of this moderate bias.

Regarding the demographic context, Bremhorst et al. (2016) studied the effects of the educational attainments of women and their partners, fixed at previous birth, on second and third parity progressions. Using our model extension, the possible progression in the mother and partner educational levels during the follow-up can now be taken into account. As a main result, the opposite of the 'work-accelerated childbearing assumption' suggested by Ní Bhrolcháin (1986); Kreyenfeld (2002) and Bartus et al. (2013) was revealed for women with a high education level living in West Germany. Indeed, our analysis not only pointed that highly educated woman were significantly more likely to have a second kid, but also less hurried than others to have their second pregnancy when that decision is taken. It should be contrasted to the corresponding statistically non significant effect in Bremhorst et al. (2016) where the education attainment was fixed at its value directly after the first birth, thereby ignoring the possibility that these women were still studying. A possible further study could be devoted to the analysis of the effects of (time-varying) women/family characteristics on the probability and the timing of a first pregnancy.

Acknowledgments

References

Aalen, O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, 2: 951–972.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10: 1100–1120.

Atchadé, Y. and Rosenthal, J. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5): 815–828.

Bartus, T., Murinkó, L., Szalma, I., and Szél, B. (2013). The effect of education on second births in Hungary: A test of the time-squeeze, self-selection, and partner-effect hypotheses. *Demographic Research*, 28(1): 1–32.

Berkson, J. and Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47: 501–515.

Boag, J. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1): 11–53.

Bremhorst, V., Kreyenfeld, M., and Lambert, P. (2016). Fertility progression in Germany: An analysis using flexible non parametric cure survival models. *Demographic Research*, 35(18): 505–534.

Bremhorst, V., Kreyenfeld, M., and Lambert, P. (2019). Nonparametric double additive cure survival models: an application to the estimation of the nonlinear effect of age at first parenthood on fertility progression. *Statistical Modelling*, 19(3): 248–275.

Bremhorst, V. and Lambert, P. (2016). Flexible estimation in cure survival models using Bayesian P-splines. *Computational Statistics and Data Analysis*, 93: 270–284.

Brown, E. and Ibrahim, J. (2003). Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*, 59: 686–693.

Chen, M., Ibrahim, J., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94: 909–919.

Chen, M., Ibrahim, J., and Sinha, D. (2004). A new joint model for longitudinal and survival data with a cure fraction. *Journal of Multivariate Analysis*, 91: 18–34.

Chi, Y. and Ibrahim, J. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 7: 432–445.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65: 141–151.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer, New York.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, 11: 89–121.

Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6): 637–653.

Fleming, T. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments (with discussion). In *Bayesian statistics 4*, Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (Eds.), 169–193. Oxford University Press, Oxford.

Govindarajulu, U. S., Lin, H., Lunetta, K. L., and D'Agostino, R. B. (2011). Frailty models: Applications to biomedical and genetic studies. *Statistics in Medicine*, 30: 2754–2764.

Gressani, O. and Lambert, P. (2018). Fast Bayesian inference in semi-parametric P-spline cure survival models using Laplace approximations. *Computational Statistics and Data Analysis*, 124: 151–167.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2): 223–242.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73: 387–396.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, 1: 255–273.

Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics and Data Analysis*, 51: 2542–2558.

Kim, S., Xi, Y., and Chen, M. (2009). A new latent cure rate marker model for survival data. *Annals of Applied Statistics*, 3: 1124–1146.

Kim, S., Zeng, D., Li, Y., and Spiegelman, D. (2013). Joint modelling of longitudinal and cure survival data. *Journal of Statistical Theory and Practice*, 7(2): 324–344.

Kreyenfeld, M. (2002). Time squeeze, partner effect or self-selection? An investigation into the positive effect of women's education on second birth risks in West Germany. *Demographic Research*, 7(2): 15–48.

Kuk, A. and Chen, M. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79: 531–541.

Lambert, P. (2007). Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics and Data Analysis*, 51: 6307–6320.

Lambert, P. and Bremhorst, V. (2019). Estimation and identification issues in the promotion time cure model when the same covariates enter the cure probability and time-to-event model components. *Biometrical Journal*, 61(2): 275–279.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13:183–212.

Li, C. and Taylor, J. (2002). A semi-parametric accelerated failure time cure model. *Statistics in Medicine*, 21: 3235–3247.

Li, L. and Lee, J. (2017). A latent promotion time cure rate model using dependent tail-free mixtures. *Journal of the royal statistical society. Serie A*.

Liang, K.-Y., Self, S. G., Bandeen-Roche, K. J., and Zeger, S. L. (1995). Some recent developments for regression analysis of multivariate failure time data. *Lifetime Data Analysis*, 1: 403–415.

Liu, H. and Shen, Y. (2009). A semi-parametric accelerate failure time cure model. *Journal of the American Statistical Association*, 104: 1168–1178.

Lopes, C. and Bolfarine, H. (2012). Random effects in promotion time cure rate models. *Computational Statistics and Data Analysis*, 56: 75–87.

López-Cheda, A., Cao, R., Jácome, M., and Van Keilegom, I. (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics and Data Analysis*, 105: 144–165.

Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica*, 20: 661–674.

Ní Bhrolcháin, M. (1986). Womens paid work and the timing of births: Longitudinal evidence. *European Journal of Population*, 2(1): 43–70.

Peng, Y. and Dear, K. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56: 237–243.

Roberts, G. and Rosenthal, J. (2001). Optimal scaling for various Metropolis Hasting algorithms. *European Journal of Population*, 16: 357–367.

Sy, J. and Taylor, J. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, 56: 227–236.

Taylor, J. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, 51: 899–907.

Tsodikov, A. (1998). A proportional hazard model taking account of long-term survivors. *Biometrics*, 54: 1508–1516.

Tsodikov, A. (2002). Semi-parametric model of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Statistic in Medicine*, 21: 895–920.

Wagner, G., Frick, J., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): Scope, evolution and enhancements. *Schmollers Jahrbuch*, 127: 139–169.

Wang, L., Du, P., and Liang, H. (2012). Two-component mixture cure rate model with spline estimated nonparametric components. *Biometrics*, 68: 726–735.

Yakovlev, A. and Tsodikov, A. (1996). *Stochastic Models for Tumor of Latency and Their Biostatistical Applications*. World Scientific Publishing, Singapore.

Yin, G. and Ibrahim, J. (2005). Cure rate models : a unified approach. *Canadian Journal of Statistics*, 33: 559–570.

Zeng, D., Yin, G., and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101: 670–684.

Zhang, J. and Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine*, 26: 3157–3171.

Zhang, J., Peng, Y., and Li, H. (2013). A new semiparametric estimation method for accelerated hazards mixture cure model. *Computational Statistics and Data Analysis*, 59: 95–102.

Zhou, J., Zhang, J., A.C., M., and Cai, B. (2016). A multiple imputation approach for semiparametric cure model with interval censored data. *Computational Statistics and Data Analysis*, 99: 105–114.