ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Cure Models in Survival Analysis

Maïlis Amico and Ingrid Van Keilegom

Research Centre for Operations Research and Business Statistics (ORSTAT), KU Leuven, 3000 Leuven, Belgium; email: ingrid.vankeilegom@kuleuven.be

Annu. Rev. Stat. Appl. 2018. 5:311–42

First published as a Review in Advance on December 13, 2017

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-031017-100101

Keywords

survival analysis, mixture cure model, promotion time cure model, censoring, regression

Abstract

When analyzing time-to-event data, it often happens that a certain fraction of the data corresponds to subjects who will never experience the event of interest. These event times are considered as infinite and the subjects are said to be cured. Survival models that take this feature into account are commonly referred to as cure models. This article reviews the literature on cure regression models in which the event time (response) is subject to random right censoring and has a positive probability to be equal to infinity.

1. INTRODUCTION

In classical survival analysis, one of the main assumptions is to consider that all subjects will eventually experience the event of interest. But it often happens that a fraction of subjects will never experience it. These subjects are usually considered as having infinite survival times and are said to be cured. In order to take this feature into account, classical survival models have been extended to what are commonly referred to as cure models. These models borrow their name from their natural area of application, namely, from medical studies where one is interested in the time until recurrence of a certain disease. Some patients will never suffer a relapse of a given disease and are hence cured of their disease. Another prominent example comes from economics, where one is often interested in the time until an unemployed person finds a new job. Some unemployed people will actually never find a new job, so they have an infinite duration of unemployment. In economics, cure models are often called split population models (see Schmidt & Witte 1989), referring to the fact that the population is split in two groups, namely, the cured ones and those who are susceptible to the event of interest. A third example comes from the area of reliability, where engineers are interested in the time until a machine or device fails, and some of them never fail. Engineers often talk about limited-failure population life models (see Meeker 1987). Other examples can be found in finance (time until a bank goes bankrupt), marketing (time until someone buys a new product), insurance (time until warranty claim), demography (time until someone marries), sociology (time until a rearrest for released prisoners), and education (time taken to solve a problem), among others.

In all the above examples, the variable of interest is a nonnegative random variable T representing the time until the event of interest occurs. We suppose that this variable is subject to random right censoring, that is, instead of observing T, we observe $Y = \min(T, C)$ and $\Delta = I(T \leq C)$, where $I(.)$ is the indicator function and C is the random censoring time. When a cure fraction is present, the survival function $S(t) = P(T > t)$ of T is such that $\lim_{t\to\infty} S(t) > 0$, and this limiting value, denoted by $1 - p$, corresponds to the proportion of cured subjects, called the cure rate. Due to the presence of right censoring, we never observe T when it equals infinity. In fact, when $\Delta = 1$ (uncensored observation), we know for sure that the individual is susceptible (uncured), whereas when $\Delta = 0$ (censored observation) he or she can belong to either of the two subpopulations, and we do not know which one.

In order to illustrate the existence of a cure fraction, we simulate 300 data points from a model in which 32% of the observations are cured and 40% are censored, accounting not only for the cured observations but also for the censored uncured observations, which represent 8% of the population. In **Figure 1** we show the Kaplan & Meier (1958) estimator (hereafter the Kaplan-Meier estimator) of these 300 observations with different representations of the censoring times. As can be seen, there is a clear plateau in the right tail, of which the height is an estimator of the cure proportion $1 - p$. However, it is well known that the Kaplan-Meier estimator is inconsistent in the right tail when the last follow-up time is a censoring time. It is clear that certain conditions need to be fulfilled in order to be sure that the height of the plateau accurately estimates $1 - p$. In fact, it could happen that some of the observations in the plateau correspond to censored uncured observations, and in that case the height of the plateau will be larger than $1 - p$. So how can we know what the cure fraction is if we cannot distinguish cured observations from large censored uncured observations? Formal identifiability conditions are given later in this article, but informally speaking, we can say that if we have a long plateau that contains a large number of data points, we can be confident that (almost) all observations in the plateau correspond to cured observations, as in the simulated example. Often the context of the study also tells us whether there is a cure fraction or not.

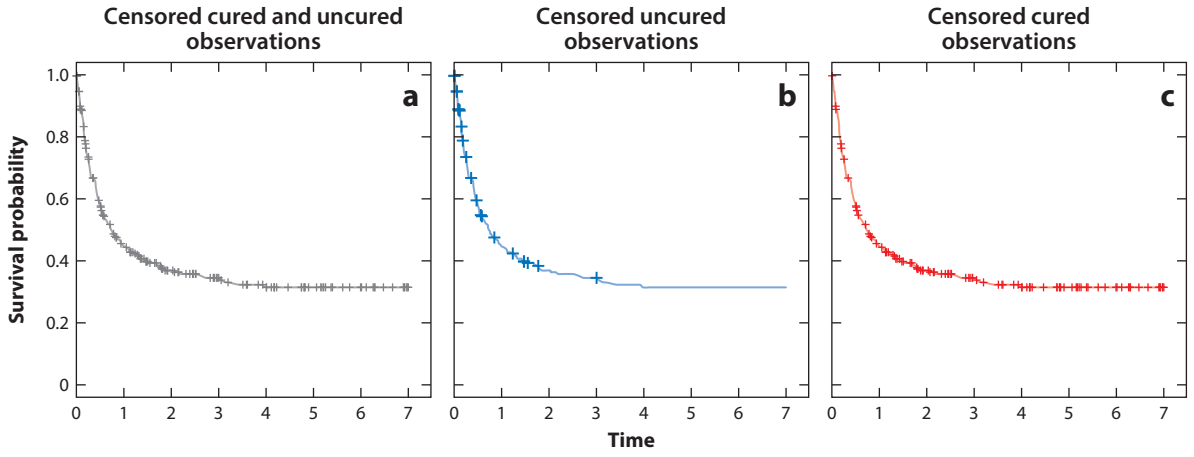

Figure 1

Kaplan & Meier (1958) estimator for 300 data points simulated from a model containing a cure fraction.

We suppose now that we observe a set of covariates $\mathbf{X}$, and another set of covariates $\mathbf{Z}$, which might be identical to $\mathbf{X}$ or partially or completely different from $\mathbf{X}$. In the literature, two main families of cure regression models have been proposed, called mixture cure models and promotion time cure models. The latter models are also called bounded cumulative hazard models or proportional hazards (PH) cure models. The mixture cure model was proposed by Boag (1949) and Berkson & Gage (1952) (see also Farewell 1982) and belongs to the class of two-part models that considers jointly the modeling of a response variable for two different groups identified by a binary variable. Such models have been investigated for semicontinuous data, for count data with the so-called zero-inflated Poisson models, and for longitudinal data (see Farewell et al. 2017 for an extensive review on two-part models for longitudinal data). In the context of survival data with a cure fraction, the mixture cure model writes the survival function $S_{\text{pop}}(t|\mathbf{x}, \mathbf{z}) = P(T > t|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})$ as

$$S_{\text{pop}}(t|\mathbf{x}, \mathbf{z}) = 1 - p(\mathbf{x}) + p(\mathbf{x})S_u(t|\mathbf{z}), \qquad 1.$$

where $p(\mathbf{x}) = P(B = 1|\mathbf{X} = \mathbf{x})$ is the probability of being susceptible (often called the incidence of the model), and $S_u(t|\mathbf{z}) = P(T > t|\mathbf{Z} = \mathbf{z}, B = 1)$ is the (proper) conditional survival function of the susceptibles (often called the latency of the model). Here, $B = I(T < \infty)$ is the latent binary variable indicating whether someone is cured or not. Equivalently, we can write the model as $F_{\text{pop}}(t|\mathbf{x}, \mathbf{z}) = p(\mathbf{x})F_u(t|\mathbf{z})$, where $F_{\text{pop}} = 1 - S_{\text{pop}}$ and $F_u = 1 - S_u$. Note that all that the model given in Expression 1 says is that the cure rate only depends on $\mathbf{X}$ (and not on $\mathbf{Z}$) and that the conditional survival function of the susceptibles only depends on $\mathbf{Z}$ (and not on $\mathbf{X}$). Further model assumptions can be made on $p(\mathbf{x})$ and on $S_u(t|\mathbf{z})$, leading to parametric, semiparametric, or nonparametric families of mixture cure models. For the cure rate $1 - p(\mathbf{x})$, the logistic model is commonly assumed, whereas for the survival function of the susceptibles, a variety of models have been proposed in the literature. We describe these models in detail in Section 2 and also discuss other points, such as the important issue of identifiability and the verification of this model.

We now briefly describe the second class of cure models, the promotion time cure model, which has a much shorter history than the mixture cure model and has therefore not been studied as much (yet) in the literature. It was proposed by Yakovlev et al. (1996) as an adaptation of the Cox (1972) PH model to allow for a cure fraction, and it supposes that the survival function

$S_{\text{pop}}(t|\mathbf{x}) = P(T > t|\mathbf{X} = \mathbf{x})$ can be written as

$$S_{\text{pop}}(t|\mathbf{x}) = \exp\left[-\theta(\mathbf{x})F(t)\right], \qquad 2.$$

where $\mathbf{X}$ now represents the complete vector of covariates, $F(\cdot)$ is a proper baseline distribution function and $\theta(\mathbf{x})$ captures the effect of the covariates $\mathbf{x}$ on the survival function $S_{\text{pop}}(t|\mathbf{x})$. Unlike the mixture cure model, this formulation has a PH structure. The cure proportion is now equal to $P(B = 0|\mathbf{X} = \mathbf{x}) = \exp[-\theta(\mathbf{x})]$. One often chooses $\theta(\mathbf{x}) = \exp(\mathbf{x}^T\beta)$, where the first component of the covariate $\mathbf{x}$ is supposed to be 1, in order to include an intercept in the model. Note that the Cox model without cure fraction does not include an intercept since it supposes that $F(t)$ tends to infinity when t tends to infinity, and an intercept would therefore not be identifiable. This formulation for the promotion time cure model is the most encountered in the literature. However, covariates may also be introduced in $F(t)$. In Section 3 we discuss several aspects related to this model (both when F depends on $\mathbf{X}$ and when it does not), such as the identifiability, the estimation, and the case where measurement errors are present.

There also exists a literature on models that unify the mixture cure model and the promotion time cure model into one single overarching model, hence avoiding the delicate task of choosing between these two models. These unifying models are described in Section 4.

Many other topics have been investigated in the framework of cure models—for example, the introduction of frailties, competing risks, quantile regression, and different types of censoring such as interval-censoring. However, we focus on a detailed description of the two main classes of cure models in a classical setting in this review to give the reader a better understanding of the basis of cure models.

In order to go further inside cure models, we illustrate the added value of cure models over classical survival models in the presence of a cure fraction by considering simulated data from a mixture cure model given in Expression 1. We assume the logistic regression model $p(\mathbf{x}) = [\exp(\gamma^T\mathbf{x})]/[1 + \exp(\gamma^T\mathbf{x})]$ for the probability of being uncured, where the vector of covariates is $\mathbf{X} = (1, X_1, X_2)^T$ and $\gamma = (\gamma_0, \gamma_1, \gamma_2)^T$ is a vector of regression coefficients associated with $\mathbf{X}$. X_1 and X_2 are independent and follow uniform distributions on $[-1, 1]$ and $\gamma = (1, 2, 0.5)^T$ in order to achieve a cured proportion of 32%. For the survival times, cured observations are associated with infinite survival times. We set their survival times equal to a very large value (for example, 10,000), and for uncured observations, we draw survival times from the exponential model $S_u(t|\mathbf{z}) = \exp[-\exp(\beta_0)\exp(\beta^T\mathbf{z})t]$, where $\beta = (\beta_1, \beta_2)^T$ is a vector of parameters associated with $\mathbf{Z}$, with $\mathbf{Z} = (X_1, X_2)^T$. For all observations, censoring times follow an exponential distribution with density $f(t) = (1/\mu_c)\exp[-(t/\mu_c)]$. We assume that $\beta_0 = 0.5$, $\beta = (0.5,\ d)^T$ and $\mu_c = 10$ in order to achieve a proportion of censoring of 34%, close to the cure proportion. A total of 24.5% of the observations (corresponding to 72% of the censored observations) are in the plateau, all of them being cured as desired for survival data with a cure fraction. We consider samples of size $n = 300$ and a total of 250 datasets.

For each dataset, we estimate a Cox PH model with survival function $S(t|\mathbf{z}) = S_0(t)^{\exp(\beta^T\mathbf{z})}$, where $S_0(t)$ is the survival function for $\mathbf{Z} = 0$ called the baseline survival function, which is totally unspecified, and a mixture cure model assuming a Cox PH model for $S_u(t|\mathbf{z})$. Baseline survival functions and parameter estimates for both models are shown in **Figure 2**. As we can see, if we do not take into account the presence of the cure fraction, the baseline survival function is overestimated under the Cox model (**Figure 2*a***), whereas the mixture cure model performs very well and is estimated with a lower variability. Likewise, the parameter estimates from the classical Cox PH model are largely biased, upward for $\hat{\beta}_1$ (**Figure 2*b***) and downward for $\hat{\beta}_2$ (**Figure 2*c***). In contrast, the bias is very small for both parameter estimates for the cure model. Based on this

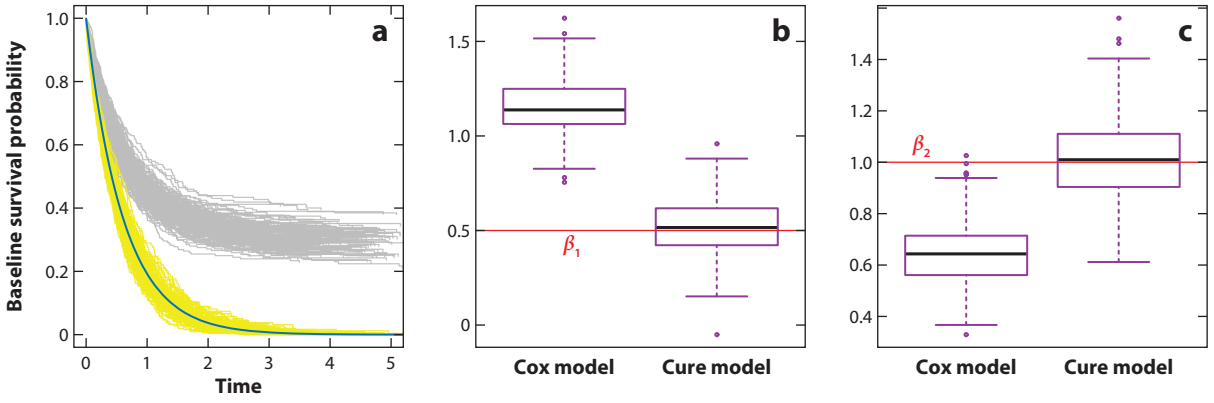

Figure 2

(*a*) True baseline survival function (*blue curve*) and estimated baseline survival function over 250 datasets from the classical Cox proportional hazards model (*gray curves*) and the mixture cure model (*yellow curves*). (*b*) Boxplots of $\hat{\beta}_1$ (where the *red line* is β_1). (*c*) Boxplots of $\hat{\beta}_2$ (where the *red line* is β_2).

simulated example, it is clear that not taking into account the presence of a cure fraction in survival data has important consequences that may lead to wrong conclusions.

Throughout the article, we illustrate models and methods that have been proposed in the literature on cure models with a dataset on breast cancer from Wang et al. (2005). The dataset consists of time to distant metastasis, expressed in days, for 286 patients who experienced a lymph-node-negative breast cancer between 1980 and 1995. Four covariates are considered: the age of the patient (ranging from 26 to 83 with a median of 52 years old), the estrogen receptor (ER) status [0 or 1, where 0 signifies ER−, defined as less than 10 fmol/mg protein (77 patients), and 1 signifies ER+, defined as at least 10 fmol/mg protein (209 patients)], the size of the tumor (ranging from 1 to 4 with a median of 1), and the menopausal status [where 0 signifies premenopausal (129 patients) and 1 signifies postmenopausal (157 patients)].

Figure 3 shows a graphical representation of the Kaplan-Meier estimator of the survival function. As can be seen, the curve levels off at a value greater than 0, approximately 60%, and there is a large plateau of approximately 2,770 days, a strong sign of the presence of a cure fraction. Moreover, among the 286 patients, 179 are censored, of which 88.3% are censored after the last observed event time. Hence, many of the censoring times are located in the plateau, indicating that a cure model can be considered. Finally, there is strong medical evidence for the presence of cured patients in breast cancer relapse. It turns out that this dataset is a perfect example of survival data with a cure fraction.

We end this section by briefly mentioning some other works on cure models. The textbook by Maller & Zhou (1996), which is completely devoted to the topic of cure models, gives a nice introduction to many of the specific aspects of cure models. Recently, Peng & Taylor (2014) wrote a review article on cure models in which they give a detailed overview of the existing cure models.

2. MIXTURE CURE MODELS

We start, in Section 2.1, with the most fundamental issue related to the definition of a model: its identifiability. Although this is often neglected in the statistical literature, it should be the first task when studying a new model. In Section 2.2, we examine several models for the components of the mixture cure model, and we see how they can be estimated, how the estimators can be

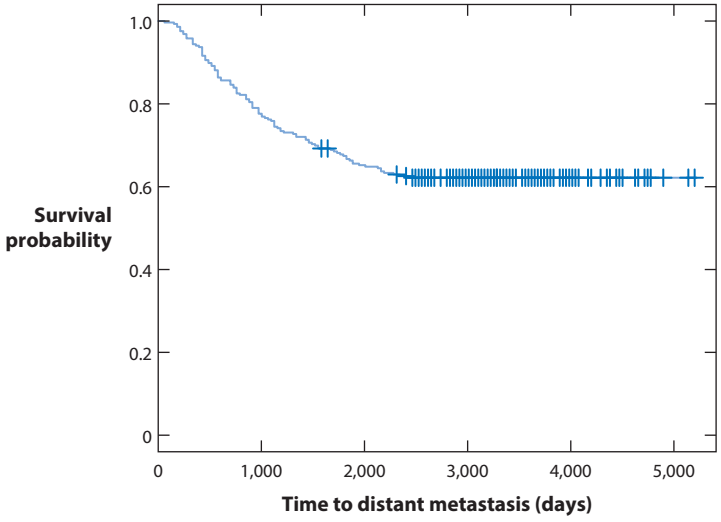

Figure 3

Kaplan & Meier (1958) estimator of the survival function for the breast cancer dataset of Wang et al. (2005) (*plus signs* are censored observations).

computed in practice, and how they behave asymptotically. Section 2.3 describes different issues related to the verification of the model, such as goodness-of-fit tests, variable selection, and model diagnostics. Finally, in Section 2.4, we apply the most common mixture cure model to the breast cancer data introduced in Section 1.

2.1. Identifiability

We have already mentioned the issue of identifiability of the model in a very informal way in the Introduction. Indeed, we said that in order to identify (in an informal way) the cure proportion, the plateau in the plot of the survival function for the whole population should only consist of cured subjects. When the plateau stays constant for a long time without decreasing even incrementally, we can be relatively confident that all uncured subjects had their event of interest before the start of the plateau, and hence the cure fraction corresponds to the height of the plateau. This informal analysis can be made more rigorous by saying that

$$\tau_{F_u} < \tau_G \qquad 3.$$

(we omit covariates here for simplicity), where $F_u = 1 - S_u$, G is the censoring distribution, and $\tau_F = \inf\{t : F(t) = 1\}$ for any distribution F. This assumption is crucial in most semi- and nonparametric papers on modeling of mixture cure models.

When we talk about the identifiability of a model, we should distinguish two common but different definitions of identifiability. The first definition (the weaker of the two) states that the mixture cure model given in Expression 1 is identifiable within families $\mathcal{P}$ and $\mathcal{S}$ of functions corresponding to the incidence and the latency respectively, if the equality

$$1 - p_1(\mathbf{x}) + p_1(\mathbf{x})S_{u1}(t|\mathbf{z}) = 1 - p_2(\mathbf{x}) + p_2(\mathbf{x})S_{u2}(t|\mathbf{z}) \qquad \text{for all } t, \mathbf{x}, \mathbf{z}, \qquad 4.$$

for some functions $p_1, p_2 \in \mathcal{P}$ and $S_{u1}, S_{u2} \in \mathcal{S}$, implies that $p_1(\mathbf{x}) = p_2(\mathbf{x})$ for all $\mathbf{x}$, and that $S_{u1}(t|\mathbf{z}) = S_{u2}(t|\mathbf{z})$ for all t and $\mathbf{z}$. This was studied in full detail and in a rigorous way by Hanin & Huang (2014), who considered several choices of the classes $\mathcal{P}$ and $\mathcal{S}$ under which the mixture cure

model given in Expression 1 is or is not identifiable. The paper by Hanin & Huang (2014) is an important improvement over earlier attempts to study the identifiability of the mixture cure model, in the sense that earlier papers contained mistakes in the proofs or did not study the problem in full generality. Note that this first issue of identifiability does not depend on the censoring mechanism, and hence the condition shown in Expression 3 does not play any role here.

A second type of identifiability is related to the uniqueness of the parameters of the model in another sense—namely, in the sense that there is a unique set of parameters for which the expected log-likelihood is maximal. So, instead of equating two models, which are unrelated to the type of data at hand, we look at the likelihood, which is based on the density of the observed variables (that are subject to random right censoring in our case) under the given model. The conditions under which there exists a unique $p \in \mathcal{P}$ that maximizes the expected log likelihood when $\mathcal{P}$ is a parametric class of probability functions coming, for example, from a logistic model, were rigorously studied by Patilea & Van Keilegom (2017) (see proposition 3.1), while Xu & Peng (2014) studied the case where $\mathcal{P}$ is nonparametric. In both papers, the assumption shown in Expression 3 turns out to be a crucial assumption to ensure identifiability of the model [although condition 7 of Xu & Peng (2014) could be relaxed to $\tau_{F_u}(\mathbf{z}) < \tau_G(\mathbf{z})$ for all $\mathbf{z}$].

We are now ready to study different models for the incidence $p(\cdot)$ and the latency $S_u(\cdot|\cdot)$, and their corresponding estimation procedures.

2.2. Modeling Approaches and Inference

The literature on mixture cure models offers a wide variety of modeling approaches ranging from fully parametric to completely nonparametric models. In what follows, we assume that we have independent and identically distributed (i.i.d.) data $(Y_i, \Delta_i, \mathbf{X}_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$, having the same distribution as $(Y, \Delta, \mathbf{X}, \mathbf{Z})$, where $Y = \min(T, C)$, $\Delta = I(T \leq C)$, $\dim(\mathbf{X}) = p$, with a first component equal to 1, $\dim(\mathbf{Z}) = q$, and for given values of $\mathbf{X}$ and $\mathbf{Z}$, the event time T follows the mixture cure model given in Expression 1. Let $Y_{(1)} \leq \cdots \leq Y_{(n)}$ be the order statistics of the observations $Y_1, \ldots, Y_n$, and let $Y^*_{(1)} < \cdots < Y^*_{(r)}$ be the distinct ordered uncensored observations, assuming there are $r \leq n$ in total.

2.2.1. Fully parametric models. The pioneer works on the mixture cure models are fully parametric approaches due to Boag (1949) and Berkson & Gage (1952). In both cases, the incidence is modeled as a constant, and the survival function for uncured observations takes the form of a log-normal model and an exponential model, respectively, not depending on covariates. Farewell (1977) introduced covariates in the incidence by assuming a logistic regression model for the probability of being uncured, that is, $p(\mathbf{x}) = [\exp(\gamma^T \mathbf{x})]/[1 + \exp(\gamma^T \mathbf{x})]$, and modeled the latency according to an exponential distribution, that is, $S_u(t) = \exp(-\lambda t)$. The introduction of covariates in the latency was proposed by Farewell (1982), who considered a Weibull model for the conditional survival function of the form $S_u(t|\mathbf{z}) = \exp(-\lambda \exp(\beta^T \mathbf{z}) t^\rho)$, where $\lambda > 0$ is a shape parameter and $\rho > 0$ is a scale parameter. Ghitany et al. (1994) proposed a logistic/exponential mixture cure model where the latency depends on covariates.

For these models, a maximum likelihood estimation method is proposed based on the likelihood function

$$\prod_{i=1}^{n} [p(\mathbf{X}_i) f_u(Y_i|\mathbf{Z}_i)]^{\Delta_i} \times \prod_{i=1}^{n} [1 - p(\mathbf{X}_i) + p(\mathbf{X}_i) S_u(Y_i|\mathbf{Z}_i)]^{1-\Delta_i}, \qquad 5.$$

where $f_u(t|\mathbf{z}) = -(d/dt) S_u(t|\mathbf{z})$. Derived as for classical survival models, the likelihood function for the mixture cure model is defined as the product of two different types of contributions—those from uncensored and censored observations. Uncensored observations contribute through the

density function, which is equal to $f(t|\mathbf{x},\mathbf{z}) = p(\mathbf{x})f_u(t|\mathbf{z})$, and censored observations contribute through the survival function given by the mixture cure model given in Expression 1. Note that no distinction is made between cured and uncured censored observations since the cure status is unknown. To estimate the logistic/Weibull mixture cure model, Farewell (1982) maximized this likelihood function numerically using the Newton-Raphson technique.

Other parametric mixture cure models include the accelerated failure time (AFT) model for the latency. Yamaguchi (1992) considered the extended family of generalized gamma models (Prentice 1974) for $\log(T^*) = \beta_0 + \beta^T\mathbf{Z} + \sigma\epsilon$, where T^* is the survival time for uncured observations, $\sigma > 0$ is a scale parameter, and ϵ is an error term with density function

$$f_\epsilon(t) = \begin{cases} \frac{|\lambda_\epsilon|}{\Gamma(\lambda_\epsilon^{-2})}(\lambda_\epsilon^{-2})^{\lambda_\epsilon^{-2}} \exp(\lambda_\epsilon t - e^{\lambda_\epsilon t}) & \text{if } \lambda_\epsilon \neq 0 \\ \frac{1}{(2\pi)^{1/2}} \exp(-t^2/2) & \text{if } \lambda_\epsilon = 0, \end{cases}$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\mathrm{d}x$ is the gamma function, and λ_ϵ is a shape parameter. Peng et al. (1998) proposed considering a generalized F distribution for T^* by assuming that $W = [\log(T^*) - \mu]/\sigma$, where μ is a location parameter and $\sigma > 0$ is a scale parameter, has density function

$$f_W(w) = \left(\frac{s_1 e^w}{s_2}\right)^{s_1} \left(1 + \frac{s_1 e^w}{s_2}\right)^{-(s_1+s_2)} B(s_1, s_2)^{-1},$$

where s_1 and s_2 are shape parameters and $B(\cdot,\cdot)$ is the beta function. In both cases, the choice of the distribution is motivated by its flexibility and because it embeds the exponential, Weibull (when $s_1 \to 1$ and $s_2 \to \infty$, we obtain the model proposed by Farewell 1982), log-normal, and gamma distributions as special cases, among others. For the incidence, both models assume a logistic regression model, as did Farewell (1982). Note that Yamaguchi (1992) also considered the case where the cure proportion is constant.

Yamaguchi (1992) and Peng et al. (1998) developed a maximum likelihood approach based on the likelihood given in Expression 5 for the two proposals in order to estimate these two models. First, the Newton-Raphson algorithm is used to maximize the likelihood function with respect to $(\gamma, \beta_0, \beta, \sigma)^T$. In a second step, a search for the value of the shape parameter(s) that maximize the likelihood is made (λ_ϵ for the extended family of generalized gamma models, s_1 and s_2 for the generalized F model).

2.2.2. Logistic/Cox proportional hazards mixture cure models. Semiparametric mixture cure models are a second class of mixture cure models that have been extensively studied in the literature. The main motivation is that they avoid the restrictions imposed by parametric conditional survival functions. Most of them focus on the latency while they keep the logistic regression form for the incidence. A first group of models is composed of mixture cure models assuming a Cox PH model for the conditional survival function, that is, $S_u(t|\mathbf{z}) = S_0(t)^{\exp(\beta^T\mathbf{z})}$, where the baseline survival function $S_0(t) = P(T > t|\mathbf{Z} = 0, B = 1)$ is totally unspecified. Introduced by Kuk & Chen (1992), this mixture cure model does not satisfy the PH assumption, contrary to the classical Cox PH model. As a consequence, the partial likelihood approach (Cox 1972) developed to estimate the Cox PH model cannot be applied for this model. Indeed, it is not possible to isolate the baseline survival function in the likelihood. Likewise, because the latency is defined conditionally on the uncured status, if one considers the baseline conditional survival function as a nuisance parameter, information about the cure status will be lost. The literature contains several proposals to estimate the model that take into account this situation. The approach of Kuk & Chen (1992), adapts a marginal likelihood approach developed by Kalbfleisch & Prentice (1973) for the classical Cox PH model. The marginal likelihood consists of integrating the likelihood function (Expression 5)

over $Y^*_{(j)}, j = 1, \ldots, r$. Estimators are obtained by maximizing the marginal likelihood function with respect to the parameters. In practice, it is not possible to compute this marginal likelihood function, and it is therefore approximated by Monte Carlo methods.

A second group of estimation approaches, proposed by Peng & Dear (2000) and Sy & Taylor (2000), is based on the expectation-maximization (EM) algorithm (Dempster et al. 1977). The choice for this methodology is justified by the fact that the model depends on a latent variable, the cure status. Another interesting argument lies in the fact that the principle of the EM algorithm is to approximately maximize Expression 5 from a complete-data likelihood, which takes the form

$$\mathcal{L}_c(\gamma, \beta, S_0) = \prod_{i=1}^{n} [p(\mathbf{X}_i)\lambda_u(Y_i|\mathbf{Z}_i)S_u(Y_i|\mathbf{Z}_i)]^{\Delta_i B_i} \times \prod_{i=1}^{n} [p(\mathbf{X}_i)S_u(Y_i|\mathbf{Z}_i)]^{(1-\Delta_i)B_i} \times \prod_{i=1}^{n} [1 - p(\mathbf{X}_i)]^{(1-\Delta_i)(1-B_i)}, \quad 6.$$

where $\lambda_u(t|\mathbf{z}) = f_u(t|\mathbf{z})/S_u(t|\mathbf{z})$ is the hazard function of the uncured observations. An interesting feature of Expression 6 is that it can be rewritten as the product of two elements,

$$\mathcal{L}_1(\gamma) = \prod_{i=1}^{n} p(\mathbf{X}_i)^{B_i} [1 - p(\mathbf{X}_i)]^{1-B_i} \quad \text{and} \quad 7.$$

$$\mathcal{L}_2(\beta, S_0) = \prod_{i=1}^{n} \left\{ [\lambda_u(Y_i|\mathbf{Z}_i)S_u(Y_i|\mathbf{Z}_i)]^{\Delta_i B_i} S_u(Y_i|\mathbf{Z}_i)^{(1-\Delta_i)B_i} \right\}, \quad 8.$$

each of them only containing the parameters of one of the two parts of the model. It is then possible to estimate separately the incidence and the latency. In such a case, it becomes possible to extend methods developed for the classical Cox PH model.

The implementation of the EM algorithm for the mixture cure model is as follows. The first step, the expectation step, consists of computing, at the mth iteration of the algorithm, the expectation of the complete-data likelihood given in Expression 6 given the current values of the parameters $\theta^{(m-1)} = (\gamma, \beta, S_0)^{(m-1)}$ and the observed data $V_i = (Y_i, \Delta_i, \mathbf{X}_i, \mathbf{Z}_i)$, with respect to the latent variable B_i. For the mixture cure model, it is the same as computing

$$E(B_i|V_i, \theta^{(m-1)}) = \Delta_i + (1 - \Delta_i)\frac{p(\mathbf{X}_i)S_u(Y_i|\mathbf{Z}_i)}{1 - p(\mathbf{X}_i) + p(\mathbf{X}_i)S_u(Y_i|\mathbf{Z}_i)} = W_i^{(m)}.$$

The expected complete-data likelihood is obtained by replacing B_i by its expectation $W_i^{(m)}$ in Expression 6.

The second step, the maximization step (M-step), consists of maximizing the expected complete-data likelihood with respect to the parameters of the model. For the incidence, Expression 7 is the same likelihood function as for a classical logistic regression model. The Newton-Raphson technique is applied to estimate the parameters. For the latency, three methods can be distinguished, all of them based on Expression 8:

- Sy & Taylor (2000) proposed a first approach based on the work of Breslow (1974). This two-step approach estimates nonparametrically the baseline conditional cumulative hazard function, $\Lambda_0(t) = -\log[S_0(t)]$, by

$$\hat{\Lambda}_0(t) = \sum_{j:Y^*_{(j)} \le t} \frac{D_{(j)}}{\sum_{k \in R_j} W_k^{(m)} \exp(\mathbf{Z}_k^T \beta)},$$

  where $D_{(j)}$ represents the number of observations experiencing the event of interest at time $Y^*_{(j)}$, and R_j is the set of observations that are still at risk just before the event time $Y^*_{(j)}$. This

estimator is then substituted in Expression 8, and the following partial likelihood is obtained (assuming no ties):

$$\tilde{\mathcal{L}}_2(\beta|W^{(m)}) = \prod_{i=1}^{n}\left[\frac{\exp(\mathbf{Z}_i^T\beta)}{\sum_{k\in R_i} W_k^{(m)}\exp(\mathbf{Z}_k^T\beta)}\right]^{\Delta_i}, \qquad 9.$$

where $\mathbf{W}^{(m)} = \{W_1^{(m)}, \ldots, W_n^{(m)}\}$. Note that when $W_i^{(m)} = 1$ for all $i = 1, \ldots, n$, Expression 9 is equal to the partial likelihood for the classical Cox PH model. The latency part is then estimated by maximizing Expression 9 with respect to β using the Newton-Raphson method.

- A second proposal from Sy & Taylor (2000) is a product-limit-type method in which the baseline conditional survival function is first estimated nonparametrically by a step function that takes a product-limit form:

$$S_0(t) = \prod_{j:Y^*_{(j)}\le t} \alpha_j,$$

where $\alpha_j = S_0(Y^*_{(j)})/S_0(Y^*_{(j-1)})$. In the absence of ties, the likelihood function given in Expression 8 is reparametrized in terms of α_j and the EM algorithm is applied in order to estimate α given β. In a second step, the estimator of α is substituted in the expected complete-data likelihood and a profile likelihood for β is obtained. We refer to Sy & Taylor (2000) for more details regarding the case with ties.

- A third approach is that of Peng & Dear (2000), who considered a marginal likelihood. As in Kuk & Chen (1992), the marginal likelihood function is obtained by integrating Expression 8 over $Y^*_{(j)}, j = 1, \ldots, r$. In the absence of ties, the following marginal likelihood is obtained:

$$\check{\mathcal{L}}_2(\beta|W^{(m)}) \approx \prod_{i=1}^{n}\left[\frac{\exp(\mathbf{Z}_i^T\beta)}{\sum_{k\in R_i} W_k^{(m)}\exp(\mathbf{Z}_k^T\beta)}\right]^{\Delta_i}.$$

Note that this marginal likelihood is approximately equivalent to the partial likelihood (Expression 9) obtained by Sy & Taylor (2000). For the case with ties, we refer the reader to Peng & Dear (2000).

Lu (2008) proposed another type of estimation method to estimate the logistic/Cox PH mixture cure model. Based on a nonparametric maximum likelihood function, the main idea is to consider a nonparametric estimator for $\Lambda_0(t)$, that is, a step function with jumps at all the event times, and to replace the baseline conditional hazard in Expression 5 by the size of the jump made by the cumulative baseline hazard at each event time. The likelihood function is then given by

$$\begin{aligned}\mathcal{L} = &\prod_{i=1}^{n}\left\{p(\mathbf{X}_i)[\Lambda_0(Y_i) - \Lambda_0(Y_i-)]\exp(\beta^T\mathbf{Z}_i)\exp[-\Lambda_0(Y_i)\exp(\beta^T\mathbf{Z}_i)]\right\}^{\Delta_i}\\ &\times\prod_{i=1}^{n}\left\{1 - p(\mathbf{X}_i) + p(\mathbf{X}_i)\exp[-\Lambda_0(Y_i)\exp(\beta^T\mathbf{Z}_i)]\right\}^{1-\Delta_i}.\end{aligned}$$

The major contribution of this mainly theoretical article is to show that the estimators of γ, β, and Λ_0 converge weakly to a zero-mean Gaussian process. Lu (2008) also provides an estimator of the asymptotic covariance function.

Finally, Corbière et al. (2009) proposed a penalized likelihood approach that has the advantage of producing a smooth estimator of the conditional hazard function. The method consists of considering the penalized likelihood

$$\log[\mathcal{L}(\gamma, \beta, \lambda_0)] - \kappa\int \lambda_0''(v)^2\,\mathrm{d}v, \qquad 10.$$

where

$$\mathcal{L}(\gamma,\beta,\lambda_0) = \prod_{i=1}^{n}\left\{p(\mathbf{X}_i)\lambda_0(Y_i)\exp(\beta^T\mathbf{Z}_i)\exp[-\Lambda_0(Y_i)\exp(\beta^T\mathbf{Z}_i)]\right\}^{\Delta_i} \times \prod_{i=1}^{n}\left\{1-p(\mathbf{X}_i)+p(\mathbf{X}_i)\exp[-\Lambda_0(Y_i)\exp(\beta^T\mathbf{Z}_i)]\right\}^{1-\Delta_i},$$

$\kappa\int\lambda_0''(v)^2dv$ is the penalization term, and $\kappa>0$ is a positive smoothing parameter balancing between the fit of the data and the smoothness of the function. The model is estimated by maximizing the likelihood function given in Expression 10 with respect to γ, β and λ_0. Since there is no explicit formula for the baseline conditional hazard that maximizes the likelihood, it is approximated by a linear combination of cubic normalized B-splines. They also provide a method to compute the variance of the parameter estimates based on the inverse of the matrix of the second derivatives of the penalized likelihood.

2.2.3. Logistic/semiparametric accelerated failure time models. In addition to the popular logistic/Cox PH mixture model, other papers, beginning with Li & Taylor (2002), focused on a semiparametric AFT model for the latency. They all consider that $\log(T^*)=\beta_0+\beta^T\mathbf{Z}+\epsilon$ and assume unspecified density and survival functions f and S, respectively, for the error term ϵ. As for the logistic/Cox PH mixture model, a logistic regression model is assumed for the incidence. Three different estimation approaches have been proposed, all based on the EM algorithm. Starting from the complete-data likelihood given in Expression 6, these methodologies are the same as for the logistic/Cox mixture cure model until the M-step for the latency estimation. For this latter part, they extend methods that were proposed for the classical semiparametric AFT models. Starting from the expected complete-data likelihood associated with the latency given by

$$\prod_{i=1}^{n}\left\{f_\epsilon\Big[\log(Y_i)-\beta_0-\beta^T\mathbf{Z}_i\Big]^{\Delta_i W_i^{(m)}} S_\epsilon\Big[\log(Y_i)-\beta_0-\beta^T\mathbf{Z}_i\Big]^{(1-\Delta_i)W_i^{(m)}}\right\}, \qquad 11.$$

- Li & Taylor (2002) proposed extending the work of Ritov (1990) based on M-estimators. Starting from the score equation for β given by

$$\sum_{i=1}^{n}\mathbf{Z}_i\left\{-W_i^{(m)}\Delta_i\frac{f_\epsilon'[\log(Y_i)-\beta_0-\beta^T\mathbf{Z}_i]}{f_\epsilon[\log(Y_i)-\beta^T\mathbf{Z}_i]}+W_i^{(m)}(1-\Delta_i)\frac{f_\epsilon[\log(Y_i)-\beta_0-\beta^T\mathbf{Z}_i]}{S_\epsilon[\log(Y_i)-\beta^T\mathbf{Z}_i]}\right\}=0,$$

  the principle consists of replacing $-f_\epsilon'/f_\epsilon$ in the score equation by an M-estimator and replacing the unknown survival function S_ϵ by its Kaplan-Meier estimator given β. Because the obtained score equation is not necessarily monotone and continuous, they propose estimating the parameters by using a grid search over the range of values of β.
- Zhang & Peng (2007) proposed rewriting Expression 11 as the likelihood function for a classical semiparametric AFT model. Using the fact that $\Delta_i=1$ and $W_i^{(m)}=1$ if the ith observation is uncensored, it turns out that $\Delta_i W_i^{(m)}\equiv\Delta_i$, and $\Delta_i\log W_i^{(m)}\equiv 0$. The likelihood function can be rewritten as

$$\prod_{i=1}^{n}\left\{W_i^{(m)}\lambda_\epsilon[\log(Y_i)-\beta_0-\beta^T\mathbf{Z}_i]\right\}^{\Delta_i}\times\prod_{i=1}^{n}\left\{S_\epsilon[\log(Y_i)-\beta_0-\beta^T\mathbf{Z}_i]\right\}^{W_i^{(m)}},$$

  where $\lambda_\epsilon=f_\epsilon/S_\epsilon$, which corresponds to the likelihood function of an AFT model with $\log(T_i^*)=\beta_0+\beta^T\mathbf{Z}_i+\epsilon_i^*$, where the hazard function of ϵ_i^* is $W_i^{(m)}\lambda_\epsilon(\epsilon_i^*)$, and $W_i^{(m)}$ is a constant. A rank estimation method proposed by Wei (1992) for classical semiparametric

AFT models is then used to estimate the latency. We refer readers to Zhang & Peng (2007) for more details.

- Lu (2010) proposed a profile likelihood approach. First, the hazard function given in Expression 11 is replaced by a piecewise constant hazard:

$$\lambda(t) = \sum_{j=1}^{J_n} \lambda_j I\left[t \in [x_{j-1}, x_j)\right],\ 0 \le t < M,$$

where the support $[0, M]$ of $\exp[\log(Y_i) - \beta^T \mathbf{Z}_i]$ is partitioned in J_n intervals of equal length and $I(\cdot)$ denotes the indicator function. The likelihood function is first maximized with respect to λ_j, $j = 1, \ldots, J_n$, given β. Then, the estimators of the λ_js are substituted in Expression 11. A profile likelihood is obtained. However, this profile likelihood is not smooth and presents local maxima. As a solution, a kernel-smoothed approximation is proposed, and the latency is estimated from this latter function.

2.2.4. Flexible semiparametric models. All the preceding models consider a logistic regression for the incidence. However, as mentioned by Peng (2003b) and proposed by Lam et al. (2005), other types of link functions can be considered. If the logit link is the canonical link function for binary response variables in the generalized linear model framework, one can also consider a probit or a complementary log-log link function, among others. These link functions only ask for a slight modification of the likelihood function. The EM algorithm can then be easily implemented to estimate these models. One possible limitation, however, is the lack of flexibility of parametric models. Even if parametric incidences offer some appealing characteristics, such as easy estimation and interpretation, one can question the quality of their fit. In order to widen the flexibility of the incidence of the mixture cure model, some semiparametric modeling approaches have been proposed. Wang et al. (2012) considered a smoothing splines analysis of variance (SS ANOVA) model for both the incidence and the latency. It consists of expressing the two parts of the model as $\log\{p(\mathbf{x})/[1 - p(\mathbf{x})]\} = \zeta(\mathbf{x})$ and $\lambda(t|\mathbf{x}) = \exp[\eta(\mathbf{x})]$, where $\zeta(\mathbf{x}) = \zeta_0 + \sum_{j=1}^{p} \zeta_j(x_j) + \sum_{j,k=1}^{p} \zeta_{jk}(x_j, x_k) + \cdots + \zeta_{1\ldots p}(x_1, \ldots, x_p)$ and $\eta(\mathbf{z}) = \eta_0 + \sum_{l=1}^{q} \eta_l(z_l) + \sum_{l,m=1}^{q} \eta_{lm}(z_l, z_m) + \cdots + \eta_{1\ldots q}(z_1, \ldots, z_q)$, and where all functions appearing in the formula of $\zeta(\mathbf{x})$ and $\eta(\mathbf{z})$ are unspecified. The SS ANOVA model is then estimated based on a penalized EM algorithm.

Another modeling approach was proposed by Amico et al. (M. Amico, C. Legrand, I. Van Keilegom, manuscript in preparation), who considered a single-index structure for the probability of being uncured, of the form $p(\mathbf{x}) = g(\gamma_*^T \mathbf{x}^*)$, where $g(\cdot)$ is a totally unspecified and not necessarily monotone link function, and the asterisk indicates that there is no intercept in the model. They considered a Cox PH model for the latency. They proposed an estimation method based on the EM algorithm, close to the proposal of Sy & Taylor (2000), but with an additional substep in the M-step added to estimate the unknown link function in the single-index. They considered a kernel estimator and proved the identifiability of the model.

In the above papers, the authors tried to relax the common logistic/Cox PH mixture cure model by replacing the logistic model with a more flexible model. Another way to make the latter model more flexible is by replacing the Cox PH model by a nonparametric model. This was proposed by Taylor (1995), who considered a fully nonparametric model for the latency that does not depend on any of the covariates, and a logistic regression model for the incidence. He developed an estimation method based on the EM algorithm where the latency is estimated from Expression 8 by a Kaplan-Meier-type estimator.

Another paper that considers a nonparametric model for the latency is Patilea & Van Keilegom (2017), but contrary to Taylor (1995), they allow the latency to depend on covariates, and they assume a parametric model for the incidence. So, no assumptions are made on the conditional survival function $S_u(\cdot|\mathbf{z})$ of the uncured subjects, except for smoothness and identifiability assumptions. They use a two-step procedure to estimate their model: In the first step, they fix the parameter vector coming from the incidence, and they estimate the survival function $S_u(\cdot|\mathbf{z})$ by means of a kernel approach. In the second step, they plug in this estimated function in the likelihood, which they maximize with respect to the parameters of the incidence. Patilea & Van Keilegom (2017) showed the weak consistency and the asymptotic normality of their model parameters, and they compared their estimated model with the logistic/Cox PH mixture cure model through finite sample simulations, which allowed them to study the sensitivity of the latter model with respect to the validity of the PH assumption.

Another approach is that of Lu & Ying (2004), who assume a semiparametric linear transformation model for the latency of the form $H(T^*) = -\beta^T \mathbf{Z} + \epsilon$, where H is an unknown monotone increasing function. Depending on the distribution of ϵ, different models are obtained, and two particular cases are mentioned. When an extreme value distribution is assumed, a PH model is obtained. In contrast, if ϵ follows a standard logistic distribution, the latency follows a proportional odds model. Lu & Ying (2004) proposed an estimation method based on counting processes and martingale theory. They derived estimating equations in order to estimate H, β, and γ, and they presented an iterative approach to solve them. The main objective of the paper is to derive the asymptotic normality of the proposed estimator and to obtain consistent variance estimates.

2.2.5. Nonparametric models. At least one part of each of the preceding models is modeled parametrically or semiparametrically. A last possibility is to assume a fully nonparametric mixture cure model. The main contribution on this topic comes from López-Cheda et al. (2017), who considered a nonparametric model for both the incidence and the latency, including covariates. They proposed estimating the two parts of the model based on the Beran (1981) estimator and proceeded as follows. For the incidence, they considered the cure rate estimator developed by Xu & Peng (2014):

$$1 - \hat{p}_b(x) = \prod_{j=1}^{n} \left[1 - \frac{\Delta_{(j)} B_{b(j)}(x)}{\sum_{k=j}^{n} B_{b(k)}(x)} \right],$$

where $B_{b(j)}(x) = K[(x - X_{(j)})/b] / \sum_{l=1}^{n} K[(x - X_{(l)})/b]$ are Nadaraya-Watson weights, b is a bandwidth, K is a kernel function, and $X_{(j)}$ and $\Delta_{(j)}$ are the values of the covariate and of the censoring indicator corresponding to the jth order statistic $Y_{(j)}$ (assuming no ties). The intuition behind this estimator is that the cure proportion corresponds to the value at which the survival function levels off, or equivalently, to the value of the survival function for the last uncensored event time.

For the latency, the idea is to rewrite the model given in Expression 1 assuming that $X = Z$, which gives $S_u(t|x) = \{S_{\text{pop}}(t|x) - [1 - p(x)]\}/p(x)$, and to use the following estimator:

$$\hat{S}_{u,b}(t|x) = \frac{\hat{S}_{\text{pop},b}(t|x) - (1 - \hat{p}_b(x))}{\hat{p}_b(x)},$$

where $\hat{S}_{\text{pop},b}(t|x)$ is the Beran (1981) estimator of the survival function $S_{\text{pop}}(t|x)$ given by $\prod_{j:Y_{(j)} \le t} \{1 - [\Delta_{(j)} B_{b(j)}(x) / \sum_{l=j}^{n} B_{b(l)}(x)]\}$, and $B_{b(j)}(x)$ are Nadaraya-Watson weights, with b a bandwidth not necessarily equal to h. They developed the asymptotic theory, as well as a bandwidth selection method based on bootstrap.

2.2.6. The zero-tail constraint and baseline conditional survival function tail estimation. Before ending this section on models and estimation methods for the mixture cure model, one issue still needs to be discussed. When a non- or semiparametric model is assumed for the latency, one issue arises from the conditional survival function estimation. Indeed, even if we suppose that $S_u(t) = 0$ when $t \to \infty$, the lack of information in the right tail of the survival function (with possibly a large number of censored observations after the last uncensored observation) may lead to a situation where $\hat{S}_u(Y^*_{(r)}) \neq 0$, inducing identifiability problems. In order to compensate for the lack of information in the right tail, Taylor (1995) proposed assuming that observations censored after $Y^*_{(r)}$ are cured. In practice, he proposed imposing in the expectation step of the EM algorithm that the weight W_i equals 0 if $Y_i > Y^*_{(r)}$. Known as the zero-tail constraint, this constraint has also been applied by Sy & Taylor (2000), Peng & Dear (2000), and Li & Taylor (2002), among others. However, this constraint may overestimate the number of cured observations. This motivated Peng (2003a) to propose another approach, which consists of parametrically estimating the tail of the conditional baseline survival function in a logistic/Cox PH mixture cure model, similar to what Moeschberger & Klein (1985) did for the classical Kaplan-Meier estimator. He proposed considering an exponential or a Weibull model for the baseline survival function $S_0(t)$ when $t > Y^*_{(r)}$. Using simulations, he showed that the proposed method reduced bias well compared with the zero-tail constraint.

2.3. Assessment of the Model

The literature on mixture cure models is quite extensive regarding not only model definition and inference but also the verification of the model where many different aspects have been investigated. In this section, we detail these aspects, ranging from testing for crucial hypotheses such as the presence of a sufficient follow-up or the presence of a cure fraction, to variable selection.

2.3.1. Testing for sufficient follow-up. As mentioned previously, a sufficient follow-up is an important element in order to consider a cure model. Heuristically, it consists of looking at the plateau of the Kaplan-Meier estimator and ensuring that it is long enough. To evaluate this formally, Maller & Zhou (1996) developed a test that consists of testing the null hypothesis $H_0 : \tau_{F_u} \geq \tau_G$ against the alternative hypothesis $H_1 : \tau_{F_u} < \tau_G$. Note that H_1 is exactly the identification condition given in Expression 3 that we argued is a crucial assumption in most semi- and nonparametric cure models. Intuitively, the main idea of the test is as follows: First, note that one can test H_0 by looking at the difference between the largest observation $Y_{(n)}$ and the largest uncensored observation $Y^*_{(r)}$. Indeed, if $Y_{(n)} - Y^*_{(r)}$ is large, the largest censoring time occurs well after the largest uncensored survival time, which is an indication that the follow-up is sufficiently long or that $\tau_{F_u} < \tau_G$. In contrast, if $\tau_{F_u} \geq \tau_G$, then $Y_{(n)} - Y^*_{(r)}$ will be close to zero. Based on a heuristic and informal reasoning, Maller & Zhou (1996) then used this information to propose the following test statistic:

$$q_n = \frac{N_n}{n},$$

where N_n represents the number of uncensored observations in the interval $[2Y^*_{(r)} - Y_{(n)}, Y^*_{(r)}]$. The decision rule is then as follows: H_0 is rejected if q_n exceeds a certain critical value, and the follow-up will be considered sufficiently long in that case. Since the distribution of q_n is not known, Maller & Zhou (1996) simulate the critical values when T follows an exponential distribution with mean 1 and C follows a uniform distribution on $[0, b]$, and they tabulated the critical values for several values of n, b and the cure rate $1 - p$. Although it is unlikely that H_0 can be tested in a general

setting, the proposed solution to find the critical value is very restrictive since it only holds for very specific parametric distributions.

2.3.2. Testing for the presence of a cure fraction. Having a sufficient follow-up is a necessary condition to consider a cure model. However, it does not necessarily imply the presence of a cure fraction. To evaluate whether a cure fraction exists or not, some authors have developed statistical tests. Zhao et al. (2009) proposed a score test in the setting of a logistic/Cox PH mixture cure model for the hypothesis $H_0 : p = 1$, or equivalently, $H_0 : \varphi = 0$ against $H_1 : \varphi > 0$, where $\varphi = (1 - p)/p$ and $0 \leq \varphi < \infty$. They assumed that p does not depend on any of the covariates. The test statistic is given by

$$S_n(\hat{\beta}) = U^T(\hat{\beta})\hat{\Gamma}^{-1}U(\hat{\beta}),$$

where $U(\hat{\beta})$ is the score vector of the logistic/Cox PH mixture cure model evaluated at $(\beta, \varphi) = (\hat{\beta}, 0)$, with $\hat{\beta}$ the estimator of the regression coefficients, and $\hat{\Gamma}$ the Fisher information matrix evaluated at $\hat{\beta}$ and $\varphi = 0$. Note that under H_0, the model reduces to a classical Cox PH model, and hence it is not necessary to estimate a mixture cure model to perform the test. Asymptotically, $S_n(\hat{\beta})$ converges under H_0 to a mixture of a χ_0^2 and a χ_1^2 distribution with equal probability. It is important to mention that the test is based on a parametric model for the conditional baseline hazard function in the latency.

Hsu et al. (2016) also developed a test for the presence of a cure fraction, but they allowed the cure fraction to depend on the covariates. Based on the model $p(\mathbf{x}) = [\exp(\alpha)\exp(\gamma^T\mathbf{x})]/[1 + \exp(\alpha)\exp(\gamma^T\mathbf{x})]$, or equivalently $p(\mathbf{x}) = [1 + \exp(-\alpha)\exp(-\gamma^T\mathbf{x})]^{-1}$, where α is an intercept and γ is a vector of slopes, the test looks for infinite values of the intercept by testing whether $H_0 : \psi^* = 0$ for all γ versus $H_1 : \psi^* > 0$ for some γ, where $\psi = \exp(-\alpha)$, and ψ^* is the true value of ψ. Note that this is equivalent to testing $H_0 : p(\mathbf{x}) = 1$ for all $\mathbf{x}$ versus $H_1 : p(\mathbf{x}) < 1$ for some $\mathbf{x}$. They derive a sup-score test statistic given by

$$T_n = \sup_{\gamma \in \mathcal{B}} S_n(\gamma),$$

where $\mathcal{B}$ is the support of γ and $S_n(\gamma)$ is a certain score test statistic obtained under H_0. Hsu et al. (2016) showed that under the null hypothesis and for fixed γ, the score $S_n(\gamma)$ converges in distribution to the mixture $(\chi_0^2 + \chi_1^2)/2$. However, the asymptotic distribution of $\sup_{\gamma \in \mathcal{B}} S_n(\gamma)$ is more complicated to obtain, and they proposed a simple resampling technique to approximate this distribution under the null hypothesis. Note that if the support of $\mathcal{B}$ equals 0, the test reduces to a test for a constant cure fraction. The test is derived assuming a logistic regression model for the incidence. However, the methodology can be implemented for any increasing, differentiable, and invertible link function. There are two main limitations: It assumes continuous covariates, and it considers a parametric form for the baseline conditional hazard (a Weibull or a log-logistic model).

2.3.3. Model diagnostics. Besides testing for sufficient follow-up and for the presence of a cure fraction, the literature on the mixture cure model also deals with model diagnostics. This topic was first investigated by Wileyto et al. (2013), who proposed deriving Schoenfeld residuals for parametric mixture cure models. Schoenfeld residuals are used to evaluate the departure from the PH assumption in a classical survival analysis context, and they are used here to evaluate the fit of the model. Indeed, Wileyto et al. (2013) proposed replacing the weight $\exp(\beta^T\mathbf{z})$ in the expected values of covariates by the hazard function of the entire population $\lambda_{\text{pop}}(t|\mathbf{x}, \mathbf{z})$. Since $\lambda_{\text{pop}}(t|\mathbf{x}, \mathbf{z})$ does not verify the PH assumption, these Schoenfeld residuals cannot be used to check for this property.

In order to complete the toolbox, other diagnosis tools were considered by Peng & Taylor (2017), who developed a series of residual-based model diagnostic tools for the overall mixture cure model and for the latency, including three types of model checking:

- To check for the functional form of covariates and to diagnose the presence of outliers, they developed martingale residuals for the overall model and modified martingale residuals for the latency.
- To evaluate the fit of the model, they developed Cox-Snell and modified Cox-Snell residuals for the overall model and for the latency, respectively. The Cox-Snell residuals for the mixture cure model are sampled from a mixture-type distribution, whereas a unit exponential distribution is used in the classical case. As explained by Peng & Taylor (2017), a unit exponential distribution can still be used in practice, and this has no impact on the analysis. Regarding the modified Cox-Snell residuals, they propose a Cramér–von Mises criterion to measure the distance between the estimated distribution and the unit exponential distribution.
- To evaluate the departure from the PH assumption for the latency, they proposed a score process from which they developed a Kolmogorov-type supremum test.

All these diagnostic tools can be used for mixture cure models with parametric and semiparametric latency, but some drawbacks have to be mentioned. First, the martingale residuals for the overall model are bounded from below, contrary to what happens in classical survival analysis, which may limit their application. They are also insensitive to the covariate effects in the incidence. Second, the Cox-Snell residuals have some difficulty detecting a misspecification in the incidence modeling, whereas they perform well for the latency. Finally, for detection of outliers, Peng & Taylor (2017) also mentioned that the modified martingale residuals are preferred to the martingale residuals because they are not bounded from below. However, they are not efficient at detecting outliers that are too large. As an alternative, they proposed considering deviance residuals.

2.3.4. Testing for the form of the incidence. The Cox-Snell and the martingale residuals proposed by Peng & Taylor (2017) have some difficulties evaluating the fit of the incidence. However, Müller & Van Keilegom (U.U. Müller & I. Van Keilegom, manuscript in preparation) developed a test for the parametric form of the incidence. Their test includes the special cases of a logistic model and the case where the cure rate does not depend on any covariates. The test statistic is a weighted L_2-distance between a nonparametric kernel estimator of the cure rate (obtained from Xu & Peng 2014) and a parametric estimator obtained under the null hypothesis. Although they proved the limiting distribution of their test statistic, they used a bootstrap procedure to calibrate the test, since the limiting distribution is only a reasonable approximation of the distribution of the test statistic for very large samples. The test can be used as a preliminary step before deciding on, for example, for a single-index model (M. Amico, C. Legrand & I. Van Keilegom, manuscript in preparation) or a completely nonparametric model (see Xu & Peng 2014) for the incidence.

2.3.5. Variable selection. Finally, some papers focused on the selection of relevant covariates. Liu et al. (2012) first developed a variable selection methodology for both parts of the logistic/Cox mixture cure model based on a penalized likelihood approach. Because of the interesting feature of the complete-data likelihood for such model, they proposed a penalized EM algorithm where two penalty terms are considered, one for γ and one for β, which is equivalent to maximizing a penalized logistic model and a penalized Cox PH model separately. They proposed using smoothly clipped absolute deviate (SCAD) penalties developed by Fan & Li (2001). This approach is only

possible because a logistic/Cox PH mixture cure model is assumed and because the EM algorithm is considered to estimate the model. However, when a parametric form is assumed for the latency, the Liu et al. (2012) approach is not natural because the complete-data likelihood is not used. For mixture cure models with a parametric latency, Scolas et al. (2016) proposed a method based on a penalized likelihood in the context of interval-censored cure data. Adaptive LASSO penalties are assumed, one for each part of the model, and the penalized likelihood is derived from Expression 5.

Dirick et al. (2015) developed an Akaike information criterion (AIC) to select the covariates in the incidence and the latency. They proposed constructing the criterion for a logistic/Cox PH mixture cure model from the complete-data likelihood used in the EM algorithm, and they considered two different approaches to compute the expectation of the complete-data likelihood.

Finally, a third approach to do variable selection in a logistic/Cox PH mixture cure model was proposed (G. Claeskens & I. Van Keilegom, manuscript in preparation). They considered a procedure based on the focused information criterion. This criterion selects the variables in the model in such a way that the resulting estimated model is the best possible model with respect to the estimation of a certain focus parameter. Here, "best possible model" should be understood in the sense that the mean squared error of the estimated focus parameter is the smallest among all candidate models. The focus parameter can be any parameter, depending on the latency and/or the incidence, for example, the cure rate, the regression parameters, or the conditional or unconditional survival or hazard function. Claeskens & Van Keilegom developed asymptotic theory for their proposed procedure, and they showed via simulations how the method works in practice.

2.4. Data Analysis

To illustrate the practical use of the mixture cure model, we estimate a logistic/Cox PH mixture cure model based on the Wang et al. (2005) dataset, assuming a Breslow (1974)-type estimator for the conditional survival function, and including all covariates in both parts of the model. The model is fitted with the `R` package `smcure` from Cai et al. (2012).

As can be seen from **Table 1**, two different groups of estimates are obtained, one for each part of the model. The table also shows the *p*-values for the parameters computed by bootstrap. If we first focus on the incidence, age and the menopausal status have a significant impact on the

Table 1 Parameter estimates from the mixture cure model together with their corresponding standard errors and *p*-values

Incidence	Estimate	Standard error	*Z* value	$P(>\|Z\|)$
Intercept	1.1110	0.8850	1.2555	0.2093
Age	−0.0382	0.0173	−2.2112	0.0270
ER+ versus ER−	0.1824	0.2704	0.6746	0.4999
Tumor size	−0.0784	0.2054	−0.3814	0.7029
Menopausal (post- versus pre-)	0.7721	0.4445	1.7371	0.0824
Latency	**Estimate**	**Standard error**	***Z* value**	$P(>\|Z\|)$
Age	−0.0127	0.0179	−0.7059	0.4802
ER+ versus ER−	−1.0365	0.2317	−4.4739	<0.0001
Tumor size	0.5203	0.2184	2.3820	0.0172
Menopausal (post- versus pre-)	0.0778	0.3970	0.1960	0.8446

Abbreviation: ER, estrogen receptor.

probability of being uncured at a 0.05 and 0.10 level of significance respectively. To interpret these effects, we proceed as for a classical logistic regression model: $\exp(-0.0382) = 0.9625$, and $1/0.9625 = 1.0390$, meaning that an additional year of age (at diagnosis) is associated with a 4% increase in the cure probability. Regarding the menopausal status, a postmenopausal woman has $\exp(0.7721) = 2.1643$ times more chance of being uncured than a premenopausal woman. For the latency, the ER status and the tumor size have significant effects on the conditional survival at a 0.05 level of significance. A Cox PH model is assumed for this part. The interpretation of the effect of the ER status, for example, is as follows: Among patients who experience metastasis, the hazard for ER+ patients is $1/\exp(-1.0365) = 2.82$ times smaller than the hazard for ER− patients. For the tumor size, the table shows that patients with a bigger tumor have a larger instantaneous risk than patients with a smaller tumor. As can be seen, covariates have different effects in the two parts of the model. This situation is representative of an interesting feature of mixture cure models. It is possible to distinguish long-term and short-term effects of covariates, a feature that is not present in classical survival analysis. Indeed, the incidence models the long-term effect of covariates on the cure status, which is something permanent, whereas the latency focuses on the short-term, time-dependent effect that only concerns uncured observations. More details about this point can be found in Sy & Taylor (2000).

3. PROMOTION TIME CURE MODELS

The promotion time cure model is the second main class of cure models. In this section, we first give some details about the definition of the model and its interpretation in Section 3.1. In Section 3.2, we detail the different modeling approaches that have been proposed and present the corresponding estimation methods in both frequentist and Bayesian settings. Measurement error is an important issue in medical studies and so also in the context of cure data. In Section 3.3, we present some work that has been done on this topic. We end this section with an application of the promotion time cure model to the breast cancer data used previously for the mixture cure model.

3.1. Model Justification and Its Interpretation

Proposed by Yakovlev et al. (1996), the promotion time cure model offers a different approach to model survival data with a cure fraction. Its mathematical definition comes from the assessment that, given that the survival function for the whole population is such that $\lim_{t\to\infty} S_{\text{pop}}(t) > 0$ when a cure fraction is present, one can define equivalently the cumulative hazard function as $\Lambda_{\text{pop}}(t) = \theta F(t)$, where $F(\cdot)$ is a proper distribution function and $\theta > 0$. In such a situation, the cumulative hazard is bounded, that is, $\lim_{t\to\infty} \Lambda_{\text{pop}}(t) = \theta < \infty$, taking into account a cure fraction. The (improper) survival function is then given by $S_{\text{pop}}(t) = \exp[-\theta F(t)]$. When θ is a function of covariates, we obtain Expression 2.

An interesting feature of the promotion time cure model is its biological interpretation. Indeed, in the particular context of cancer, the mathematical form described below can also be obtained by assuming that the survival time is the result of a latent process generating cancerous tumors. Originally proposed by Yakovlev et al. (1993) for modeling tumor latency, it was introduced for the promotion time cure model by Chen et al. (1999). The main idea is to assume that after a first cancer, a number $N \geq 0$ of carcinogenic cells can stay active in the organism of an individual, and that it will take a certain (latent) time $\tilde{T}_k$ for each cell $k = 1, \ldots, N$ to become an active tumor. For individuals for whom $N \geq 1$, that is, for uncured observations, the survival time T is defined as $\min\{\tilde{T}_k, k = 1, \ldots, N\}$. For cured individuals, no carcinogenic cells are still active, that is, $N = 0$, inducing that $T = \infty$. By assuming that N follows a Poisson distribution with parameter $\theta > 0$,

that the $\tilde{T}_k$ are i.i.d. with distribution function $F(\cdot)$, and that they are independent of N, we can derive the survival function for T in the following way:

$$\begin{aligned} P(T > t) &= P(N = 0) + P(\tilde{T}_1 > t, \ldots, \tilde{T}_N > t, N \geq 1) \\ &= P(N = 0) + P(\tilde{T}_1 > t) \times \cdots \times P(\tilde{T}_N > t) \times P(N \geq 1) \\ &= \exp(-\theta) + \sum_{N=1}^{\infty} S(t)^N \frac{\theta^N}{N!} \exp(-\theta) \\ &= \exp[-\theta F(t)], \end{aligned} \quad 12.$$

where $S(t) = 1 - F(t)$. The survival function given in Equation 12 corresponds to the survival function from the promotion time cure model.

The parameter θ represents the mean number of carcinogenic cells. In the presence of covariates, which are mainly introduced through θ, this parameter has a double interpretation. First, when θ is large, the mean number of carcinogenic cells is large and the probability of being cured is small. Second, a larger value for θ is also representative of a lower survival probability because a larger number of carcinogenic cells induces a smaller activation time. As can be seen, θ contains two types of effects, on the cure probability and on the survival, which cannot be separate. Such an interpretation can also be drawn from a mathematical point of view. As proposed by Chen et al. (1999), quantities for each type of observation can be derived from the model in the manner of the mixture cure model. For cured observations, the associated quantity is given by $\lim_{t\to\infty} S_{\text{pop}}(t) = \exp(-\theta)$. For uncured observations, Chen et al. (1999) considered the biological development of the model, and they proposed considering the survival function for uncured observations, which corresponds to observations with at least one carcinogenic cell, that is,

$$P(T > t | N \geq 1) = \frac{\exp[-\theta F(t)] - \exp[-\theta]}{1 - \exp[-\theta]}. \quad 13.$$

Note that there also exists a vast literature on tumor latency modeling from which the biological interpretation of the promotion time cure model is derived. Some references can be found in Tsodikov et al. (2003).

3.2. Modeling Approaches and Inference

3.2.1. Modeling approaches. The literature on the promotion time cure model contains two main types of modeling approaches, depending on how the covariates are introduced. The first group of models consists of introducing covariates only through θ, as defined in the Introduction. Proposed by Tsodikov (1998a), the survival function is given by Expression 2, where $\theta(\mathbf{x}) = \exp(\gamma^T \mathbf{x})$. Regarding $F(t)$, Tsodikov (1998a) and Tsodikov (2001) proposed letting the distribution function be totally unspecified. Some other forms have been proposed, such as a Weibull or a gamma distribution (Chen et al. 1999) or a semiparametric version $F(\cdot|\eta)$ for some parameter η (introduced by Ibrahim et al. 2001). An important characteristic of the model in Expression 2 is its PH property. Indeed, by assuming that $\mathbf{X}$ contains an intercept and that $\theta(\mathbf{x}) = \gamma^T \mathbf{x}$, Expression 2 can be rewritten as $S_{\text{pop}}(t|\mathbf{x}) = \exp[-\exp(\gamma_*^T \mathbf{x}^*)\Lambda_0^*(t)]$, where $\mathbf{x}^*$ is the vector of covariates $\mathbf{x}$ without an intercept, γ_* is the vector of parameters associated with $\mathbf{x}^*$, and $\Lambda_0^*(t) = \exp(\gamma_0)F(t)$ is the baseline cumulative hazard function.

As in the case of the mixture cure model, the identifiability of the promotion time cure model is an important first issue that needs attention before we can talk about the estimation of the model. Zeng et al. (2006) showed the strong identifiability of the model given in Expression 2 when $\theta(\mathbf{x}) = \eta(\gamma^T \mathbf{x})$ for a strictly increasing function η, for example, the exponential function,

and when $F(\cdot)$ is left unspecified. By strong identifiability, we mean that there are a unique vector γ and a unique function $F(\cdot)$ that maximize the expected log-likelihood under the model. Portier et al. (2017) improved the result of Zeng et al. (2006) by allowing the censoring time to be finite and by allowing the covariates to have noncompact support.

Tsodikov (2002) proposed a second modeling approach where covariates are introduced both in θ and $F(\cdot)$. The survival function is given by

$$S_{\text{pop}}(t|\mathbf{x},\mathbf{z}) = \exp[-\theta(\mathbf{x})F(t|\mathbf{z})], \qquad 14.$$

and the form $\theta(\mathbf{x}) = \exp(\gamma^T\mathbf{x})$ is usually assumed. Regarding $F(t|\mathbf{z})$, Tsodikov (2002) and Bremhorst & Lambert (2016) proposed the form $F(t|\mathbf{z}) = 1 - S_0(t)^{\exp(\beta^T\mathbf{z})}$, where $\mathbf{Z}$ does not contain an intercept. This model corresponds to a Cox PH model. Another form was proposed by Tsodikov (2002), $F(t|\mathbf{z}) = [\exp(\beta^T\mathbf{z})F_0(t)]/[1 - \exp(\beta^T\mathbf{z})F_0(t)]$, which corresponds to a proportional odds model. Contrary to the model given in Expression 2, this model does not respect the PH assumption because of the presence of covariates in $F(\cdot)$. In such a case, the hazard function is given by $h_{\text{pop}}(t|\mathbf{x},\mathbf{z}) = \theta(\mathbf{x})f(t|\mathbf{z})$, where $f(t|\mathbf{z}) = (d/dt)F(t|\mathbf{z})$.

The weak identifiability of this model was proved by Bremhorst & Lambert (2016) when $\theta(\mathbf{x}) = \exp(\gamma^T\mathbf{x})$ and $F(t|\mathbf{z}) = 1 - S_0(t)^{\exp(\beta^T\mathbf{z})}$. By weak identifiability, we mean that if $\exp\{-\exp(\gamma_1^T\mathbf{x})[1 - S_{01}(t)^{\exp(\beta_1^T\mathbf{z})}]\} = \exp\{-\exp(\gamma_2^T\mathbf{x})[1 - S_{02}(t)^{\exp(\beta_2^T\mathbf{z})}]\}$ for all $t \in [0,\infty]$ and all $\mathbf{x}$ and $\mathbf{z}$ in the support of $\mathbf{X}$ and $\mathbf{Z}$, then necessarily, $\gamma_1 = \gamma_2$, $\beta_1 = \beta_2$ and $S_{01} \equiv S_{02}$.

3.2.2. Frequentist estimation methods. Estimation methods for the promotion time cure model were first developed in a frequentist setting by Tsodikov (1998a, 2002). They consider the likelihood function given by

$$\mathcal{L} = \prod_{i=1}^{n}[\{\theta(\mathbf{X}_i)f(Y_i)\exp[-\theta(\mathbf{X}_i)F(Y_i)]\}^{\Delta_i} \times \{\exp[-\theta(\mathbf{X}_i)F(Y_i)]\}^{1-\Delta_i}] \qquad 15.$$

where $F(\cdot)$ may or may not depend on covariates, depending on the model we consider, and with $\theta(\mathbf{x}) = \exp(\gamma^T\mathbf{x})$.

Tsodikov (1998a) proposed a profile likelihood approach in order to estimate the semiparametric version of the model given in Expression 2, where $F(\cdot)$ is totally unspecified. In a first step, the distribution function $F(\cdot)$ is estimated by a nonparametric maximum likelihood estimator (NPMLE). It consists of maximizing the likelihood function given in Expression 15 with respect to $F(\cdot)$, where $F(\cdot)$ is replaced by a step function taking values at the failure times. In order to identify the model, he proposed performing this maximization under the constraint that $F_r = 1$, where $F_r = \sum_{j=1}^{r}\Delta F_j$ and ΔF_j represents the jump size of $F(\cdot)$ at time $Y_{(j)}^*$. This constraint maximization is similar to imposing the zero-tail constraint (Taylor 1995) in the mixture cure model. He derived score equations for the step sizes ΔF_j, $j = 1,\ldots,r$, and he developed his own maximization algorithm to avoid instability problems encountered with the Newton-Raphson technique when the number of parameters is very large. A profile likelihood for γ is obtained by substituting the NPMLE for $F(\cdot)$ in the likelihood function. Alongside the estimation method, Tsodikov (1998b) demonstrated the asymptotic relative efficiency of the semiparametric estimator of $\theta(\mathbf{x})$ in comparison with a parametric one. Rigorous asymptotic theory and efficiency results for the parametric component $\theta(\mathbf{x})$ and for the nonparametric component $F(\cdot)$ were developed by Portier et al. (2017). They also developed a weighted bootstrap procedure that allows for a consistent approximation of the asymptotic law of the estimators.

A frequentist approach was proposed for the model given in Expression 14 for $F(t|\mathbf{z}) = 1 - S_0(t)^{\exp(\beta^T\mathbf{z})}$, where $S_0(t) = 1 - F_0(t)$, and for $F(t|\mathbf{z}) = [\exp(\beta^T\mathbf{z})F_0(t)]/[1 - \exp(\beta^T\mathbf{z})F_0(t)]$

by Tsodikov (2002). He considered a profile likelihood method as for the model given in Expression 2, where the likelihood function (Expression 15), with F_0 replaced by a step function, is first maximized with respect to F_0 under the constraint that $F_{0r} = 1$. A NPMLE $\hat{F}_0$ for F_0 is obtained, and by substituting $\hat{F}_0$ in Expression 15, a profile likelihood is obtained depending of the vectors γ and β. Tsodikov (2002) proposed two methods to solve the score equations for F_0: an alternative approach to Newton-Raphson algorithm and a quasi-EM algorithm approach. Note that the quasi-EM algorithm requires untied data. When tied data are present, the first proposal is preferred.

3.2.3. Bayesian estimation methods. Bayesian inference was introduced by Chen et al. (1999). Let us denote by $F(\cdot|\eta)$ the distribution function F depending on some vector of parameters η. Chen et al. (1999) focused on the parametric version of the model given in Expression 2 where a Weibull distribution is considered for $F(\cdot|\eta)$, with $\eta = (\rho, \lambda)^T$. They proposed classes of both noninformative and genuine priors (based on historical data) for (γ, η), and they discussed some of their theoretical properties. The posterior distribution is given by

$$p(\gamma, \eta|\mathbf{D}_{\text{obs}}) \propto \mathcal{L}(\gamma, \eta)\pi(\gamma, \eta),$$

where $\pi(\cdot)$ represents the joint prior distribution, $\mathbf{D}_{\text{obs}} = (\mathbf{Y}, \Delta, \mathbf{X})$ are the observed data, and $\mathcal{L}(\gamma, \eta)$ is given by Expression 15. When historical data (obtained from a previous study) are available, genuine priors are defined as the joint posterior distribution from historical data:

$$\pi(\gamma, \eta, \alpha_0|\mathbf{D}_{0,\text{obs}}) \propto [\mathcal{L}(\gamma, \eta|\mathbf{D}_0)]^{\alpha_0}\, \pi_0(\gamma, \eta)\pi_0(\alpha_0),$$

where $\mathbf{D}_{0,\text{obs}} = (\mathbf{Y}_0, \Delta, \mathbf{X}_0)$ is the vector of the observed historical data, $\mathcal{L}(\gamma, \eta|\mathbf{D}_{0,\text{obs}})$ is the likelihood function given in Expression 15 from these historical data, $\pi_0(\gamma, \eta)$ represents the joint prior distribution considered for (γ, η) from historical data, and α_0 is a parameter taking values between 0 and 1 that controls the influence of the historical data on the current data.

Ibrahim et al. (2001) considered the semiparametric version of the model given in Expression 2 and proposed a piecewise constant hazard model for $F(\cdot|\lambda)$, where $\lambda = (\lambda_1, \ldots, \lambda_J)^T$ represents the vector of constant hazards associated with the J partitions of the time axis. When a semiparametric model is considered for F, nothing guarantees that $\hat{F}(Y^*_{(r)}) = 1$, as already explained for the mixture cure model. To overcome this difficulty, Ibrahim et al. (2001) proposed introducing a smoothing parameter in the prior distribution of λ_j, $j = 1, \ldots, J$, to control the degree of parametricity of the right tail of the survival function. In terms of inference, as in Chen et al. (1999), Ibrahim et al. (2001) discussed noninformative and informative priors as well as some of their properties. The posterior distributions were constructed assuming the same procedure as above. An extension of this method was proposed by Kim et al. (2007). Besides the consideration of the degree of parametricity for the right tail, they also proposed taking into account the correlation between $\log(\lambda_{j-1})$ and $\log(\lambda_j)$. As they explained, this latter proposal improves the right tail estimation because information in that part of the survival function can be borrowed from neighboring $\log(\lambda_j)$s. They proposed introducing a smoothing parameter and a correlation parameter in the prior distribution of $\log(\lambda_j)$ by considering a martingale-type process prior. Additionally, they allowed J to be random. Prior distributions were discussed as well as some of the properties of these priors. Because the dimensions of the posterior distribution can vary with random J, a reversible jump Metropolis-Hasting algorithm was proposed to sample from the posterior distribution (we refer the reader to the article for more details about this).

For the model given in Expression 14, Bremhorst & Lambert (2016) proposed a flexible semiparametric approach when $F(t|\mathbf{z}) = 1 - S_0(t)^{\exp(\beta^T \mathbf{z})}$, where the logarithm of the baseline hazard

function $\lambda_0(t) = -(\mathrm{d}/\mathrm{d}t)\log[1 - F_0(t)]$ is written as a linear combination of cubic B-splines. Bayesian inference was considered, where P-splines are assumed to estimate $\log[\lambda_0(t)]$. P-splines consist of taking a large number of B-splines and adding a penalty term, as proposed by Eilers & Marx (1996). In a Bayesian setting, the penalty term is taken into account through the specification of the prior distributions, as detailed by Lang & Brezger (2004). Following their proposal, Bremhorst & Lambert (2016) proposed prior distributions for the parameters, and they gave a Markov chain Monte Carlo (MCMC) algorithm to sample from the joint posterior distribution. For identifiability reasons, they also proposed setting the last spline parameter to a large value in order to guarantee that the baseline survival function is proper.

3.2.4. Some extensions. Under the biological perspective of the model, we assume that the latent times $\{\tilde{T}_1, \ldots, \tilde{T}_N\}$ are i.i.d. random variables. However, this assumption could be unrealistic in certain situations, since they concern the same individual. Zeng et al. (2006) introduced a subject-specific frailty term ξ_i in order to relax this assumption, and they obtained a more general class of cure models given by

$$S_{\text{pop}}(t|\mathbf{X}_i) = E_{\xi_i}\{\exp[-\theta(\mathbf{X}_i)F(t)\xi_i]\},\ i = 1, \ldots, n. \qquad 16.$$

Depending on the distribution for ξ_i, different models are obtained. They proposed considering a gamma distribution with mean 1 and obtained the model $S_{\text{pop}}(t|\mathbf{x}) = G_\eta[\theta(\mathbf{x})F(t)]$, where $G_\eta(u) = (1 + \eta u)^{1/\eta}$ when the transformation parameter η is positive, and $G_\eta(u) = \exp(-u)$ when η equals zero. They also mentioned that other distributions and transformations can be assumed. In terms of modeling, Zeng et al. (2006) assumed that $\theta(\mathbf{x}) = \exp(\gamma^T \mathbf{x})$, and they let $F(\cdot)$ be totally unspecified. They proposed a profile likelihood approach to estimate the model. As in Tsodikov (1998a), they proposed a NPMLE for $F(\cdot)$ obtained by maximizing the likelihood function under the constraint that $F_r = 1$. Score equations are obtained by making use of the Lagrange multiplier. The parameter γ is estimated from the profile likelihood obtained by substituting $\hat{F}(\cdot)$ in the likelihood function.

A second extension was proposed by Portier et al. (F. Portier, A. El Ghouch & I. Van Keilegom, manuscript in preparation), who considered a promotion time cure model of the form

$$S(t|\mathbf{x}) = \exp[-g(\gamma, \mathbf{x})\,\theta\,F(t)], \qquad 17.$$

where g is a given function (not necessarily monotone) depending on a parameter vector γ, $\theta > 0$, and $F(\cdot)$ is an unspecified proper distribution function. When $g(\gamma, \mathbf{x}) = \exp(\gamma_*^T \mathbf{x}^*)$, where the asterisk indicates that there is no intercept in the model, then Expression 17 reduces to the promotion time cure model given in Expression 2. For all other g functions, the two models are different. Portier et al. showed the identifiability of their model, developed the NPMLE of the model parameters, showed the asymptotics for their estimators with closed-form formulas of the variance of the limiting Gaussian distributions, and they considered a likelihood-based methodology to select an appropriate g function among a family of proposals.

Other extensions based on the biological interpretation of the promotion time cure model have been proposed in the literature. However, because they also embed the mixture cure model as a special case, they are detailed in Section 4.

3.3. Measurement Errors

In medical studies, it often happens that some variables in the model are measured with noise. For instance, the error can be caused by imprecise medical instruments, such as those for measuring blood pressure, weight, or cholesterol level. In economic studies, variables such as welfare or

income often cannot be measured in a precise way, in which case one has to work with approximate measures that might contain some error. Ignoring this measurement error can lead to wrong conclusions since the presence of measurement error leads to biased estimators (see, e.g., Carroll et al. 2006).

In the context of the promotion time cure model, several authors have considered the problem of estimating the model when one or several covariates in the model are measured with error. The model that is considered is the classical additive measurement error model of the form

$$\mathbf{W} = \mathbf{X} + \mathbf{U}, \qquad 18.$$

where $\mathbf{W}$ is the vector of observed covariates and $\mathbf{U}$ is the vector of measurement errors. We further assume that $\mathbf{U} \sim N_p(0, V)$, where V is known, and $\mathbf{U}$ is independent of $\mathbf{X}$. If some covariates are not subject to measurement error, then the corresponding elements of V are set to 0. It is also assumed that (T, C) and $\mathbf{W}$ are independent given $\mathbf{X}$.

Mizoi et al. (2007) first estimated the promotion time cure model given in Expression 2 in the presence of measurement error. The authors considered the case where only one covariate is measured with error (say X_1), and they assumed that the model is fully parametric, with $F(\cdot)$ equal to the distribution of a Weibull random variable. They used a corrected score approach to take the measurement error into account, which consists of replacing in the log-likelihood the unobserved covariate X_1 by a surrogate that depends on W_1 and the (known) variance of U_1. The form of the surrogate depends on the assumed normality of U_1, and hence the method cannot be extended in an obvious way to the case where other error distributions are assumed.

Ma & Yin (2008) extended the above paper to the case where the distribution F is unknown and possibly more than one covariate is subject to measurement error. They also used a corrected score approach and proved the asymptotic unbiasedness and the asymptotic normality of their estimators.

A third contribution comes from Bertrand et al. (2017a), who assumed the same model as Ma & Yin (2008), but they used a different approach to estimate the model parameters. Their method is based on the so-called SIMEX (simulation-extrapolation) approach that was proposed by Cook & Stefanski (1994). The SIMEX method consists of two steps. In the first step, increasing levels of measurement error variance are considered, and at each level a large number of datasets is generated. The idea is then to estimate at each level the vector of regression coefficients, ignoring the measurement error. In the second step, these estimators corresponding to the different levels of error are extrapolated to the situation where the covariates are observed without error. An important advantage of this approach is that the distribution of the error term $\mathbf{U}$ can be anything, as long as the covariates observed with error are continuous, so it does not restrict attention to the normal case. Bertrand et al. (2017a) established the asymptotic unbiasedness and the asymptotic normality of their estimators, under the assumption that the extrapolation function is correct, and they compared the finite sample performance of their method with that of the corrected score approach of Ma & Yin (2008) through a small simulation study.

Finally, Bertrand et al. (2017b) justified the need to take the measurement error into account via a theoretical study of the bias of the naive estimator that ignores the measurement error. They also performed an extensive simulation study investigating the robustness of both the corrected score and the SIMEX approach with respect to the model assumptions.

3.4. Data Analysis

To illustrate the implementation of the promotion time cure model, we fit the model on the Wang et al. (2005) data. We consider the semiparametric model proposed by Tsodikov (1998a) where covariates are introduced only through θ, assuming the form $\theta(\mathbf{x}) = \exp(\gamma^T \mathbf{x})$. We fit the model

Table 2 Parameter estimates from the promotion time cure model together with their corresponding standard errors and *p*-values

	Estimate	Standard error	Z value	$P(>\|Z\|)$	exp(estimate)
Intercept	0.6218	0.6926	0.8977	0.3693	1.8622
Age	−0.0326	0.0143	−2.2813	0.0225	0.9679
ER+ versus ER−	0.0285	0.2260	0.1260	0.8997	1.0289
Tumor size	0.0056	0.1659	0.0337	0.9731	1.0056
Menopausal (post- versus pre-)	0.6382	0.3503	1.8220	0.0684	1.8931

Abbreviation: ER, estrogen receptor.

using the estimation method proposed by Tsodikov (1998a). Parameter estimates are given in **Table 2**, together with their standard errors and their *p*-values.

In contrast to the mixture cure model, we only have one set of parameters, which influences both the cure probability and the survival as explained in Section 3.1. The interpretation of the parameter estimates is as follows. The quantity $\exp(\hat{\gamma}^T \mathbf{x})$ represents the estimated mean number of carcinogenic cells for a patient with covariates $\mathbf{x}$. It then encompasses two levels of interpretation:

- First, a larger value for $\exp(\hat{\gamma}^T \mathbf{x})$ is representative of a higher probability of not being cured. Indeed, the more carcinogenic cells are present, the more the observation has a chance to experience the event.
- Second, a larger value for $\exp(\hat{\gamma}^T \mathbf{x})$ is also associated with an earlier event because more carcinogenic cells means more chance of having a small activation time.

As can be seen, menopausal status has an impact on the survival probability at a 0.10 level of significance, and age has a significant impact at a 0.05 level of significance. This means that postmenopausal women have a higher risk of being uncured, and they experience the event earlier in comparison with premenopausal women. For age, the effect is reversed: Older women have a higher probability of being cured and they have a lower instantaneous risk of experiencing a relapse than younger women. Those results are different from what we obtained with the mixture cure model, where age and menopausal status have a significant impact only on the incidence, and tumor size and ER status significantly influence the latency. In the case of the promotion time cure model, it is not possible to distinguish and quantify the covariate effects related to the cure probability and those related to the survival of the uncured observations. We only have one global effect.

4. UNIFYING MODELS

The mixture cure model and the promotion time cure model represent two different modeling approaches for survival data with a cure fraction. Several differences have been underlined in the literature, but the two models are also related. Indeed, there exists a mathematical link between them, and they are equivalent in some situations. Moreover, shortly after the introduction of the promotion time cure model, the literature on cure models was broadened by unifying approaches embedding both the mixture cure model and the promotion time cure model. In this section, we first detail in Section 4.1 the differences between the mixture cure model and the promotion time cure model but also present their mathematical relationship. In Section 4.2, we review the unifying models that have been proposed in the literature and their estimation methods. We close by discussing model selection in Section 4.3.

4.1. Dissimilarities and Relationship Between the Mixture Cure Model and the Promotion Time Cure Model

Chen et al. (1999) distinguished three main differences between the mixture cure model and the promotion time cure model. First, in the presence of covariates, the mixture cure model does not respect the PH property. In contrast when we assume that only θ is a function of covariates, the promotion time cure model does respect this property, as already mentioned before. Second, Chen et al. (1999) argued that the promotion time cure model can be interpreted biologically in the context of cancer studies, which is not the case for the mixture cure model. However, this point of view was disputed by Peng & Xu (2012), who explained that by assuming a Bernoulli distribution for N with parameter p, a mixture cure model is obtained if the biological development described in Section 3 is followed. Note, however, that this biological interpretation is an oversimplification of tumor kinetics. Third, Bayesian techniques have been developed to estimate the promotion time cure model but not for the mixture cure model. As explained by Chen et al. (1999), the lack of such methods for the latter model is due to the necessity of taking proper priors for γ, both for the informative and for the noninformative case, in order to obtain proper posterior distributions, in contrast with the promotion time cure model.

All these elements favor the promotion time cure model. But an important difference in favor of the mixture cure model that was not mentioned by Chen et al. (1999) is in the interpretation of covariate effects. The mixture cure model distinguishes the effect of covariates on the probability of being uncured from the effect of covariates on the survival function of the uncured observations. It is then possible to consider different covariates in the two parts of the model and to evaluate the effect of the same covariate(s) on the two parts. For the promotion time cure model, the question is more delicate because, as detailed in Section 3.2, θ influences both the cure probability $\exp(-\theta)$ and the conditional survival function given in Expression 13. Even by introducing covariates in $F(\cdot)$, it is still not possible to separate the long-term (on the cure probability) and short-term (on the conditional survival function) effect of covariates.

Besides these differences, the two models are mathematically related, as explained by Chen et al. (1999). Given that the cure proportion $\exp(-\theta)$ in the promotion time cure model is equivalent to $1 - p$ in the mixture cure model, and considering the conditional survival function given by Expression 13, the promotion time cure model can be rewritten (we omit covariates for simplicity) as

$$S_{\text{pop}}(t) = \exp(-\theta) + [1 - \exp(-\theta)]\frac{\exp[-\theta F(t)] - \exp(-\theta)}{1 - \exp(-\theta)},$$

that is, as a mixture cure model. Furthermore, as discussed by Peng & Xu (2012), the two models are equivalent in some situations. First, when no covariates are considered at all and $S_u(\cdot)$ and $F(\cdot)$ are unspecified, the two models are obviously equivalent. Then, when p and θ are not a function of covariates, and when $F(\cdot|\mathbf{z})$ and $S_u(|\mathbf{z})$ are unspecified, they are also equivalent. They represent the same model in a different form. In contrast, when p and θ are a function of covariates the two models are different. In a first case, when $F(\cdot)$ does not depend on covariates, they will represent different models for different data structures (a model with the PH property for the promotion time cure model versus a mixture cure model that does not verify the property). In a second case, when $F(\cdot)$ is a function of covariates, they will be both flexible models for survival data with a cure fraction.

4.2. Unifying Models: Specification and Estimation

Two different streams have driven the development of unifying models. On one side, we have models that have a pure mathematical motivation. On the other side, some articles tried to extend the biological motivation of the promotion time cure model, and new classes of models appeared.

4.2.1. Mathematical perspective. Unifying models developed in a mathematical perspective are all based on Box-Cox transformations. Yin & Ibrahim (2005) proposed applying this transformation to the survival function $S_{\text{pop}}(t|\mathbf{x},\mathbf{z})$, such that

$$\begin{cases} \frac{S_{\text{pop}}(t|\mathbf{x},\mathbf{z})^{\alpha}-1}{\alpha} & = -\theta(\alpha,\mathbf{x})F(t|\mathbf{z}), \quad \text{if } 0<\alpha\leq 1 \\ \log[S_{\text{pop}}(t|\mathbf{x},\mathbf{z})] & = -\theta(0,\mathbf{x})F(t|\mathbf{z}), \quad \text{if } \alpha=0, \end{cases} \qquad 19.$$

where α is a transformation parameter. The survival function is then given by

$$S_{\text{pop}}(t|\mathbf{x},\mathbf{z}) = \begin{cases} [1-\alpha\theta(\alpha,\mathbf{x})F(t|\mathbf{z})]^{1/\alpha}, & \text{if } 0<\alpha\leq 1 \\ \exp[-\theta(0,\mathbf{x})F(t|\mathbf{z})], & \text{if } \alpha=0, \end{cases}$$

and the associated cure probability is $\lim_{t\to\infty} S_{\text{pop}}(t|\mathbf{x},\mathbf{z}) = [1-\alpha\theta(\alpha,\mathbf{x})]^{1/\alpha}$ for $0<\alpha\leq 1$, and $\exp[-\theta(0,\mathbf{x})]$ for $\alpha=0$. They proposed modeling $\theta(\alpha,\mathbf{x})$ as $\theta(\alpha,\mathbf{x}) = [\exp(\gamma^T\mathbf{x})]/[1+\alpha\exp(\gamma^T\mathbf{x})]$, and they assumed that $F(t|\mathbf{z}) = 1-S_0(t)^{\exp(\beta^T\mathbf{z})}$. The mixture cure model and the promotion time cure model are two special cases of this model when $\alpha=1$ and when $\alpha=0$, respectively. When $0<\alpha<1$, an intermediate model is obtained. Both Bayesian and frequentist estimation methods have been proposed. Yin & Ibrahim (2005) considered a Bayesian approach and assumed a piecewise exponential distribution for $S_0(\cdot)$. They considered α as random, and they proposed a uniform discrete prior for this parameter in order to guarantee stability of the model. Note that they mentioned that there is no advantage of a random α in comparison with a fixed one except that it facilitates Bayesian inference. Parameters are considered independent a priori, and they proposed noninformative prior distributions. Peng & Xu (2012) proposed a frequentist estimation method for the model given in Expression 19 based on a maximum likelihood approach. They considered a piecewise constant hazard model for $S_0(\cdot)$ defined as $S_0(t|\lambda) = \exp[-\int_0^t \lambda_0(s|\lambda)\mathrm{d}s]$, where $\lambda_0(t|\lambda) = \exp(\lambda_j)$, and $\lambda = (\lambda_1,\ldots,\lambda_J)$ is the vector of constant hazards corresponding to each of the J time intervals. Parameter estimates were obtained from the following likelihood function:

$$\mathcal{L}(\alpha,\gamma,\beta,\lambda) = \prod_{i=1}^{n}\left\{[\theta(\alpha,\mathbf{X}_i)f(Y_i|\mathbf{Z}_i)]^{\Delta_i}\times[1-\alpha\theta(\alpha,\mathbf{X}_i)F(Y_i|\mathbf{Z}_i)]^{1/\alpha-\Delta_i}\right\}.$$

Alternatively, Taylor & Liu (2007) proposed applying a Box-Cox transformation to both sides of the equation in order to obtain the unifying model

$$\begin{cases} \frac{S_{\text{pop}}(t|\mathbf{x},\mathbf{z})^{\alpha}-1}{\alpha} & = \frac{q(\mathbf{x})^{\alpha}-1}{\alpha}F(t|\mathbf{z}), \quad \text{if } 0<\alpha\leq 1 \\ \log[S_{\text{pop}}(t|\mathbf{x},\mathbf{z})] & = \log[q(\mathbf{x})]F(t|\mathbf{z}), \quad \text{if } \alpha=0, \end{cases}$$

where α is a transformation parameter and $q(\mathbf{x}) = 1-p(\mathbf{x})$. In that case, the survival function is given by

$$S_{\text{pop}}(t|\mathbf{x},\mathbf{z}) = \begin{cases} \{1+[q^{\alpha}(\mathbf{x})-1]F(t|\mathbf{z})\}^{1/\alpha}, & \text{if } 0<\alpha\leq 1 \\ \exp\{\log[q(\mathbf{x})]F(t|\mathbf{z})\}, & \text{if } \alpha=0, \end{cases}$$

with a cure probability equal to $q(\mathbf{x})$. As for the proposal of Yin & Ibrahim (2005), when $\alpha=0$ the model reduces to the promotion time cure model. When $\alpha=1$, it becomes the mixture cure model. In term of modeling, they proposed as an example a complementary log-log model for the cure probability $q(\mathbf{x})$ depending on a parameter vector γ, and they considered a Weibull model for $1-F(t|\mathbf{z}) = \exp[-\lambda t^{\rho}\exp(\beta^T\mathbf{z})]$. No formal estimation was proposed. However, Taylor & Liu (2007) defined the likelihood function for the model:

$$\mathcal{L}(\alpha,\gamma,\beta,\lambda,\rho) = \prod_{i=1}^{n}\{q(\mathbf{X}_i)^{\alpha}-1]f(Y_i|\mathbf{Z}_i)\}^{\Delta_i}\alpha^{-\Delta_i}S_{\text{pop}}(Y_i|\mathbf{X}_i,\mathbf{Z}_i)^{1-\alpha\Delta_i}.$$

Note that by assuming that $q(\mathbf{x}) = [1+\alpha \exp(\gamma^T \mathbf{x})]^{-1/\alpha}$, the two unifying models are equivalent, as explained by Peng & Xu (2012).

4.2.2. Biological perspective. The biological development of the promotion time cure model assumes that the survival time T is generated by the latent survival times $\tilde{T}_k$ such that $T = \min\{\tilde{T}_k, k = 1, \ldots, N\}$. Cooner et al. (2007) proposed widening this relationship and considered that a number r out of the N carcinogenic cells need to be activated in order to produce a failure time. The survival time is then defined as $T = \tilde{T}_{(r)}$, for $1 \leq r \leq N$, where $\tilde{T}_{(r)}$, $r = 1, \ldots, N$ represent the ordered latent activation times. For the promotion time cure model, they assumed that $\tilde{T}_k$ are i.i.d. random variables with distribution function $F(\cdot)$. The associated survival function for the whole population is given by

$$S_{\text{pop}}(t) = E_{(N,r)}[S(t|N,r)], \tag{20.}$$

where $E_{N,r}$ is the expectation with respect to the joint distribution of (N,r), and $S(t|N,r) = P(T > t|N,r)$ is given by

$$P(T > t|N,r) = I(N = 0) + \sum_{j=0}^{r-1} \binom{N}{j} F(t)^j S(t)^{N-j} I(N \geq r \geq 1),$$

with $I(\cdot)$ the indicator function, and $S(t) = 1 - F(t)$. The cure probability for this model is given by $\lim_{t\to\infty} S_{\text{pop}}(t) = P(N = 0)$. The variable r is considered as a threshold variable determining the survival time T. It can be considered as a constant, as a function of N, or as a random variable. Moreover, depending on its value, different activation schemes are possible.

When r is considered as random, a conditional distribution is specified for $r|N$. In order to model the survival time, Cooner et al. (2007) proposed decomposing the joint distribution of r and N in Expression 20 as the product of the conditional distribution of r given N and the marginal distribution of N. In such a case, a so-called hierarchical-activation scheme is obtained. They considered two types of conditional distributions for $r|N$: a mixture distribution where a positive mass on $\{1, N\}$ is attributed to $r|N$ with probability π and $1 - \pi$ respectively, and a binomial distribution for $r - 1|N$ with parameter $N - 1$ and π. For the marginal distribution of N, four main distributions were considered: Poisson, Bernoulli, binomial, and geometric.

In the particular case where $r = 1$, that is, when only one of the N tumor cells needs to be activated in order to produce a tumor, the survival time will be equal to the first latent activation time associated with the first cell that gives a tumor, that is, $T = \min\{\tilde{T}_k, k = 1, \ldots, N\}$. In such a case the survival function reduces to

$$\begin{aligned} S_{\text{pop}}(t) &= E_N[S(t|N,1)] \\ &= E_N[I(N = 0) + S(t)^N I(N \geq 1)] \\ &= E_N[S(t)^N]. \end{aligned} \tag{21.}$$

A so-called first-activation scheme is obtained. When it is assumed that $N \sim \text{Poisson}(\theta)$, $\theta > 0$, the model becomes the promotion time cure model. When a Bernoulli distribution with parameter $0 \leq \theta \leq 1$ is assumed for N, the model reduces to the mixture cure model. Other distributions, such as a binomial or a geometric one, were considered by Cooner et al. (2007), and these distributions give other types of cure models. Another interesting distribution for N that has been proposed in the literature is the negative binomial. Tournoud & Ecochard (2008), Rodrigues et al. (2009), and de Castro et al. (2009) considered this distribution. The survival function for the whole population in such a case is given by

$$S_{\text{pop}}(t) = [1 + \rho\theta F(t)]^{-1/\rho}, \ \rho \geq -1, \ \theta > 0,$$

where $\theta = E(N)$, $\rho = -1/N$, and $V(N) = \theta + \rho\theta^2$. Interestingly, this model embeds the mixture (when $\rho = -1$) and the promotion time cure models (when $\rho \to 0$). Moreover, it is equivalent to the model given in Expression 19 proposed by Yin & Ibrahim (2005) when $\rho = -\alpha$ for ρ in $[-1, 0]$.

At the other extreme, Cooner et al. (2007) considered the case where $r = N$, that is, all tumor cells N need to be activated in order to produce a tumor. In such a case, the survival time will be defined as the largest latent activation time associated with the last activated cell, that is, $T = \max\{\tilde{T}_k, k = 1, \ldots, N\}$. This class of models is referred to as the last-activation scheme. The survival function is given by

$$S_{\text{pop}}(t) = P(N = 0) + 1 - E_N\left[F(t)^N\right].$$

As for the first-activation scheme, they considered different distributions for N: Bernoulli, binomial, Poisson, and geometric. They obtained different types of cure models different from the mixture and the promotion time cure models. In summary, the proposal of Cooner et al. (2007) represents a general class of cure models that allows more flexibility to model survival data with a cure fraction than the mixture or the promotion time cure model.

In terms of modeling, Cooner et al. (2007) proposed a Weibull distribution for $S(\cdot)$ that depends on covariates. For the parameter θ, they considered the case where it is a function of covariates assuming the form $\theta(\mathbf{x}) = \exp(\gamma^T \mathbf{x})$ and also the case where it does not depend on covariates. In an attempt to make the model more flexible, Cooner et al. (2009) proposed later a piecewise exponential distribution for the latent time $\tilde{T}_k$. For the first-activation scheme, Tournoud & Ecochard (2008) assumed that $\theta(\mathbf{x}) = \exp(\gamma^T \mathbf{x})$ when N follows a negative binomial distribution, whereas de Castro et al. (2009) considered that $\theta(\mathbf{x}) = [q(\mathbf{x})^{-\alpha} - 1]/\alpha$, where $q(\mathbf{x}) = [\exp(\gamma^T \mathbf{x})]/[1 + \exp(\gamma^T \mathbf{x})]$ when $\alpha \neq 0$ and that $\theta(\mathbf{x}) = -\log[q(\mathbf{x})]$ when $\alpha = 0$. Furthermore, de Castro et al. (2009) assumed that $F(\cdot)$ has a parametric distribution (they proposed a Weibull or a piecewise exponential distribution) and did not consider covariates.

Regarding estimation approaches, Cooner et al. (2007) and de Castro et al. (2009) proposed performing Bayesian inference. Cooner et al. (2007) considered a marginalized likelihood function because the survival function given in Expression 20 depends on N and r, which are latent, and discussed the choice of the prior distributions with an emphasis on the parameter θ in order to guarantee the identifiability of the model. They proposed using a MCMC algorithm to sample from the posterior distribution. de Castro et al. (2009) considered an observed-data likelihood and as Yin & Ibrahim (2005), they assumed that the parameters are independent a priori and chose a discrete uniform prior for α. A MCMC algorithm is also used to sample from the posterior distribution.

Besides the Cooner et al. (2007) proposal, an even more general model was proposed by Kim et al. (2011). Based on the same biological approach, they considered that an individual experiences a failure if a certain number of the N carcinogenic cells greater than or equal to a threshold variable R is activated, where R may or may not be independent of N. In such a way, they relax the dependence assumption on N formulated for r by Cooner et al. (2007). The survival time is then defined as $T = \tilde{T}_{(R)}$, and the survival function is given by

$$S_{\text{pop}}(t) = P(N < R) + E_{(N,R)}\left[\sum_{j=0}^{r-1} \binom{N}{j} F(t)^j S(t)^{N-j} I(N \geq R)\right],$$

where the cured proportion is $P(N < R)$. R, which can be fixed or random, may be considered as the antibody level of the immune system, for example. If R is assumed dependent on N, the joint distribution of (N, R) can be specified as the product of the conditional distribution of R given N, and of the marginal distribution of N. In such a case, the model becomes the hierarchical-activation

scheme from Cooner et al. (2007). If $R = 1$, the model reduces to the first-activation scheme because only one cell needs to be activated in order to produce a tumor. If, additionally, N follows a Poisson distribution with parameter θ, the model is the promotion time cure model. Equivalently, if R is considered as random with a geometric distribution, and if N follows a Poisson distribution, the model reduces to the promotion time cure model. Kim et al. (2011) only considered a Poisson distribution for N. However if we assume that $R = 1$ and that $N \sim$ Bernoulli(p), the model becomes the mixture cure model. As previously, Kim et al. (2011) assumed that $\theta(\mathbf{x}) = \exp(\gamma^T \mathbf{x})$ and considered a piecewise exponential distribution for the latent time $\tilde{T}_k$. A Bayesian approach was proposed to estimate the model, where they assumed that the distribution of R is known and that R and N are independent.

4.3. Model Selection

Unifying models offer a flexible way to model survival data with a cure fraction. As explained in the previous section, there exist several possible models for each proposal. A question of interest is how to choose the most adequate model. Two directions have been investigated. For models proposed by Yin & Ibrahim (2005) and Taylor & Liu (2007), the transformation parameter α is usually considered as random, and it is estimated alongside the other parameters. The best model is directly determined by the estimation process. However, this approach supposes that one can estimate the parameter α with precision, which seems to be difficult because there is usually not enough information about the parameter in the data, as explained by Yin & Ibrahim (2005). This point is corroborated by Diao & Yin (2012), who proposed a model similar to Yin & Ibrahim (2005) with a frailty, and who mentioned the fact that when the sample size is small, the likelihood function is quite flat for α. Taylor & Liu (2007) performed a simulation study where they evaluated a fixed versus random α, and they drew the same conclusion: When the sample size is small, it is complicated to obtain a precise estimate for α. An alternative is to consider a grid of values for α and to perform model comparison as proposed by Diao & Yin (2012).

For biologically specified models, because there is no transformation parameter, model selection is also performed based on model comparison. The main idea is to define different distributions for random quantities, such as for N or R, and to select the best fit according to some criteria. Cooner et al. (2007) proposed fitting several models to the data with different activation schemes and different distributions for N and comparing them based on the posterior predictive L-measure proposed by Laud & Ibrahim (1995) and Gelfand & Ghosh (1998), a measure that rewards goodness-of-fit, assessed via posteriori predictive comparison, and at the same time penalizes for complexity. de Castro et al. (2009) proposed a similar approach and used the deviance information criterion (DIC) and the conditional predictive ordinate statistic to compare the models. Kim et al. (2011) fit different models with different distributions for R to the data and compared them based on the DIC and the logarithm of the pseudomarginal likelihood. Others, such as Tournoud & Ecochard (2008) proposed simply choosing the most appropriate model based on the scientific knowledge of the disease, the hazard structure (proportional or not, depending on the distribution assumed for N), and the variance structure from the distribution of N.

Even if these proposals are targeted to compare flexible cure rate models, one can implement them to compare the mixture cure model and the promotion time cure model. Peng & Xu (2012) went in this direction and proposed performing likelihood ratio and score tests from the model given in Expression 19 in order to evaluate the adequacy of the mixture cure model ($H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$) and of the promotion time cure model ($H_0 : \alpha = 0$ versus $H_1 : \alpha > 0$). The proposed tests perform well with large sample sizes, but they are sensitive to a misspecification of the baseline distribution function F.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The research was supported by the European Research Council (2016–2021, Horizon 2020/ERC grant agreement No. 694409) and by IAP research network grant no. P7/06 of the Belgian government (Belgian Science Policy). The authors thank Aurélie Bertrand and Anouar El Ghouch for their technical support, Vincent Bremhorst for his helpful comments, and Pierre Dupont for the data.

LITERATURE CITED

Beran R. 1981. *Nonparametric regression with randomly censored survival data*. Tech. Rep., Univ. Calif., Berkeley

Berkson J, Gage RP. 1952. Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.* 47:501–15

Bertrand A, Legrand C, Carroll R, de Meester C, Van Keilegom I. 2017a. Inference in a survival cure model with mismeasured covariates using a SIMEX approach. *Biometrika* 104:31–50

Bertrand A, Legrand C, Léonard D, Van Keilegom I. 2017b. Robustness of estimation methods in a survival cure model with mismeasured covariates. *Comput. Stat. Data Anal.* 113:3–18

Boag JW. 1949. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Stat. Soc. B* 11:15–53

Bremhorst V, Lambert P. 2016. Flexible estimation in cure survival models using Bayesian P-splines. *Comput. Stat. Data Anal.* 93:270–84

Breslow NE. 1974. Covariates analysis of censored survival data. *Biometrics* 30:89–99

Cai C, Zou Y, Peng P, Zhang J. 2012. `smcure`: fit semiparametric mixture cure models. *R package for estimating semiparametric PH and AFT mixture cure models*. **https://CRAN.R-project.org/package=smcure**

Carroll R, Ruppert D, Stefanski L, Crainiceanu C. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton, FL: Chapman and Hall/CRC. 2nd ed.

Chen MH, Ibrahim JG, Sinha D. 1999. A new Bayesian model for survival data with a surviving fraction. *J. Am. Stat. Assoc.* 94:909–19

Cook JR, Stefanski LA. 1994. Simulation-extrapolation in parametric measurement error models. *J. Am. Stat. Assoc.* 89:1314–28

Cooner F, Banerjee S, Carlin BP, Sinha D. 2007. Flexible cure rate modeling under latent activation schemes. *J. Am. Stat. Assoc.* 102:560–72

Cooner FW, Yu X, Banerjee S, Grambsch PL, McBean AM. 2009. Hierarchical dynamic time-to-event models for post-treatment preventive care data on breast cancer survivors. *Stat. Model.* 9:119–35

Corbière F, Commenges D, Taylor JMG, Joly P. 2009. A penalized likelihood approach for mixture cure models. *Stat. Med.* 28:510–24

Cox DR. 1972. Regression models and life-tables. *J. R. Stat. Soc. B* 34:187–220

de Castro M Cancho VG, Rodrigues J. 2009. A Bayesian long-term survival model parametrized in the cured fraction. *Biom. J.* 3:443–55

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. B* 39:1–38

Diao G, Yin G. 2012. A general transformation class of semiparametric cure rate frailty models. *Ann. Inst. Stat. Math.* 64:959–89

Dirick L, Claeskens G, Baesens B. 2015. An Akaike information criterion for multiple event mixture cure models. *Europ. J. Oper. Res.* 241:449–57

Eilers PHC, Marx BD. 1996. Flexible smoothing B-splines and penalties (with discussion). *Stat. Sci.* 1:89–121

Fan J, Li R. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* 96:1348–60

Farewell VT. 1977. A model for binary variable with time-censored observations. *Biometrika* 64:43–46
Farewell VT. 1982. The use of a mixture model for the analysis of survival data with long-term survivors. *Biometrics* 38:1041–46
Farewell VT, Long DL, Tom BDM, Yiu S, Su L. 2017. Two-part and related regression models for longitudinal data. *Annu. Rev. Stat. Appl.* 4:283–315
Gelfand AE, Ghosh SK. 1998. Model choice: a minimum posteriori predictive loss approach. *Biometrika* 85:1–11
Ghitany ME, Maller RA, Zhou S. 1994. Exponential mixture models with long-term survivors and covariates. *J. Multivar. Anal.* 49:218–41
Hanin L, Huang L-S. 2014. Identifiability of cure models revisited. *J. Multivar. Anal.* 130:261–74
Hsu WW, Todem D, Kim KM. 2016. A sup-score test for the cure fraction in mixture models for long-term survivors. *Biometrics* 76:1348–57
Ibrahim JG, Chen MH, Sinha D. 2001. Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* 57:383–88
Kalbfleisch JD, Prentice RL. 1973. Marginal likelihoods based on Cox's regression and life model. *Biometrika* 60:267–78
Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53:457–81
Kim S, Chen MH, Dey DK, Gamerman D. 2007 Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Anal.* 13:17–35
Kim S, Chen MH, Dey DK. 2011. A new threshold regression model for survival data with a cure fraction. *Lifetime Data Anal.* 17:101–22
Kuk AYC, Chen CH. 1992. A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 79:531–41
Lam KF, Fong DYT, Tang OY. 2005. Estimating the proportion of cured patients in censored sample. *Stat. Med.* 24:1865–79
Lang S, Brezger A. 2004. Bayesian P-splines. *J. Comput. Graph. Stat.* 13:183–212
Laud P, Ibrahim J. 1995. Predictive model selection. *J. R. Stat. Soc. B* 57:247–62
Li SC, Taylor JMG. 2002. A semi-parametric accelerated failure time cure model. *Stat. Med.* 21:3235–47
López-Cheda A, Cao R, Jácome MA, Van Keilegom I. 2017. Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Comp. Stat. Data Anal.* 105:144–65
Liu X, Peng Y, Tu D, Liang H. 2012. Variable selection in semiparametric cure models based on penalized likelihood, with application to breast cancer clinical trials. *Stat. Med.* 31:2882–91
Lu W. 2008. Maximum likelihood estimation in the proportional hazards cure model. *Ann. Inst. Stat. Math.* 60:545–74
Lu W. 2010. Efficient estimation for an accelerated failure time model with a cure fraction. *Stat. Sin.* 20:661–74
Lu W, Ying Z. 2004. On semiparametric transformation cure models. *Biometrika* 91:331–43
Ma Y, Yin G. 2008. Cure rate model with mismeasured covariates under transformation. *J. Am. Stat. Assoc.* 103:743–56
Maller RA, Zhou S. 1996. *Survival Analysis with Long Term Survivors*. New York: Wiley
Meeker WQ. 1987. Limited failure population life tests: Application to integrated circuit reliability. *Technometrics* 29:51–65
Mizoi MF, Bolfarine H, Pedroso-De-Lima AC. 2007. Cure rate model with measurement error. *Comm. Stat. Simul. Comput.* 36:185–96
Moeschberger ML, Klein JP. 1985. A comparison of several methods of estimating the survival function when there is extreme right censoring. *Biometrics* 41:253–59
Patilea V, Van Keilegom I. 2017. A general approach for cure models in survival analysis. arXiv:1701.03769 [math.ST]
Peng Y. 2003a. Estimating baseline distribution in proportional hazards cure models. *Comp. Stat. Data Anal.* 42:187–201
Peng Y. 2003b. Fitting semiparametric cure models. *Comput. Stat. Data Anal.* 41:481–90
Peng Y, Dear KBG. 2000. A nonparametric mixture model for cure rate estimation. *Biometrics* 56:237–43

Peng Y, Dear KBG, Denham JW. 1998. A generalized *F* mixture model for cure rate estimation. *Stat. Med.* 17:813–30
Peng Y, Taylor JMG. 2014. Cure models. In *Handbook of Survival Analysis*, ed. J Klein, H van Houwelingen, JG Ibrahim, TH Scheike, pp. 113–34. Boca Raton, FL: Chapman and Hall
Peng Y, Taylor JMG. 2017. Residual-based model diagnosis methods for mixture cure models. *Biometrics* 73:495–505
Peng Y, Xu J. 2012. An extended cure model and model selection. *Lifetime Data Anal.* 18:215–33
Portier F, El Ghouch A, Van Keilegom I. 2017. Efficiency and bootstrap in the promotion time cure model. *Bernoulli* 23:3437–68
Prentice RL. 1974. A log-gamma model and its maximum likelihood estimation. *Biometrika* 61:538–44
Ritov Y. 1990. Estimation in a linear regression model with censored data. *Ann. Stat.* 18:303–28
Rodrigues J, Canch VG, de Castro M, Louzada-Neto F. 2009. On the unification of long-term survival models. *Stat. Probab. Lett.* 79:753–59
Schmidt P, Witte AD. 1989. Predicting criminal recidivism using split population survival time models. *J. Econom.* 40:141–59
Scolas S, El Ghouch A, Legrand C, Oulhaj A. 2016. Variable selection in a flexible parametric mixture cure model with interval-censored data. *Stat. Med.* 35: 1210–25
Sy JP, Taylor JMG. 2000. Estimation of a Cox proportional hazards cure model. *Biometrics* 56:227–36
Taylor JMG. 1995. Semi-parametric estimation in failure time mixture models. *Biometrics* 51:899–907
Taylor JMG, Liu N. 2007. Statistical issues involved with extending standards models. In *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A Doksum*, ed. V Nair, pp. 299–311. Singapore: World Sci.
Tournoud M, Ecochard R. 2008. Promotion time models with time-changing exposure and heterogeneity: application to infectious diseases. *Biom. J.* 3:395–407
Tsodikov A. 1998a. A proportional hazards model taking account of long-term survivors. *Biometrics* 54:1508–16
Tsodikov A. 1998b. Asymptotic efficiency of proportional hazards model with cure. *Stat. Probab. Lett.* 39:237–44
Tsodikov A. 2001. Estimation of survival based on proportional hazards when cure is a possibility. *Math. Comput. Model.* 33:1227–36
Tsodikov A. 2002. Semi-parametric models of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Stat. Med.* 21:895–920
Tsodikov AD, Ibrahim JG, Yakovlev AY. 2003. Estimating cure rates from survival data. *J. Am. Stat. Assoc.* 98:1063–78
Wang L, Pang D, Liang H. 2012. Two-component mixture cure rate model with spline estimated nonparametric components. *Biometrics* 68:726–35
Wang Y, Klijn JGM, Sieuwerts AM, Look MP, Yang F, et al. 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365:671–79
Wei LJ. 1992. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat. Med.* 11:1871–79
Wileyto EP, Li Y, Chen J, Heitjan DF. 2013. Assessing the fit of parametric cure models. *Biostatistics* 14:340–50
Xu J, Peng Y. 2014. Nonparametric cure rate estimation with covariates. *Can. J. Stat.* 42:1–17
Yakovlev AY, Asselain B, Bardou VJ, Fouquet A, Hoang T, et al. 1993. A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. *Biom. Anal. Données Spatio-temporelles* 12:66–83
Yakovlev AY, Tsodikov AD, Asselain B. 1996. *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. Singapore: World Sci.
Yamaguchi K. 1992. Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in Japan. *J. Am. Stat. Assoc.* 87:284–92
Yin G, Ibrahim JG. 2005. Cure rate model: a unified approach. *Can. J. Stat.* 33:559–70
Zhao Y, Lee AH, Yau KKW, Burke V, McLachlan GJ. 2009. A score test for assessing the cured proportion in long-term survivor mixture model. *Stat. Med.* 28:3454–66
Zhang J, Peng Y. 2007. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Stat. Med.* 26:3157–71
Zeng D, Yin G, Ibrahim JG. 2006. Semiparametric transformation for survival data with a cure fraction. *J. Am. Stat. Assoc.* 101:670–84

New From Annual Reviews:

Annual Review of Criminology

criminol.annualreviews.org • Volume 1 • January 2018

ONLINE NOW!

Co-Editors: **Joan Petersilia**, *Stanford University* and **Robert J. Sampson**, *Harvard University*

The *Annual Review of Criminology* provides comprehensive reviews of significant developments in the multidisciplinary field of criminology, defined as the study of both the nature of criminal behavior and societal reactions to crime. International in scope, the journal examines variations in crime and punishment across time (e.g., why crime increases or decreases) and among individuals, communities, and societies (e.g., why certain individuals, groups, or nations are more likely than others to have high crime or victimization rates). The societal effects of crime and crime control, and why certain individuals or groups are more likely to be arrested, convicted, and sentenced to prison, will also be covered via topics relating to criminal justice agencies (e.g., police, courts, and corrections) and criminal law.

TABLE OF CONTENTS FOR VOLUME 1:

Annual Review of Statistics and its Application

Volume 5, 2018

Contents

Errata

An online log of corrections to *Annual Review of Statistics and Its Application* articles may be found at http://www.annualreviews.org/errata/statistics