TEST manuscript No.
(will be inserted by the editor)

Goodness-of-fit test for a parametric survival function with cure fraction

Candida Geerdens · Paul Janssen · Ingrid Van Keilegom

Received: date / Accepted: date

Abstract We consider the survival function for univariate right censored event time data, when a cure fraction is present. This means that the population consists of two parts: the cured or non-susceptible group, who will never experience the event of interest versus the non-cured or susceptible group, who will undergo the event of interest when followed up sufficiently long. When modeling the data, a parametric form is often imposed on the survival function of the susceptible group. In this paper, we construct a simple novel test to verify the aptness of the assumed parametric form. To this end, we contrast the parametric fit with the nonparametric fit based on a rescaled Kaplan-Meier estimator. The asymptotic distribution of the two estimators and of the test statistic are established. The latter depends on unknown parameters, hence a bootstrap procedure is applied to approximate the critical values of the test. An extensive simulation study reveals the good finite sample performance of the developed test. To illustrate the practical use, the test is also applied on two real-life data sets.

Keywords Bootstrap · Cramér-von Mises · Cure fraction · Kaplan-Meier · Parametric models · Weak convergence.

Mathematics Subject Classification (2010) MSC 62N01 · MSC 62N02 · MSC 62N03

Candida Geerdens
Censtat, Universiteit Hasselt, Agoralaan 1, B-3590 Diepenbeek, Belgium
E-mail: candida.geerdens@uhasselt.be

Paul Janssen
Censtat, Universiteit Hasselt, Agoralaan 1, B-3590 Diepenbeek, Belgium
E-mail: paul.janssen@uhasselt.be

Ingrid Van Keilegom
Orstat, KU Leuven, Naamsestraat 69 - bus 3500, B-3000 Leuven, Belgium
E-mail: ingrid.vankeilegom@kuleuven.be

1 Introduction

In classical survival analysis it is usually assumed that all subjects under study will eventually experience the event of interest. For instance, in an experiment that studies the lifetime of certain electronic or mechanical devices, it is clear that all devices will sooner or later fail. Likewise, when studying the survival time of a certain group of patients in which all causes of death are confounded, it is clear that all patients will eventually die. However, there are also many contexts in which subjects under study never experience the event of interest. Their survival time is considered to be infinite. A prominent example is the lifetime of cancer patients after treatment. Due to medical advances part of the patient population will not die of cancer. Other relevant examples stem from epidemiology (e.g. in case of a disease outbreak a fraction of the population will not get infected), economy (e.g. in an employment study part of the population may never get a job) and criminology (e.g. in a recidivism study a portion of the released population will not recommit a crime). In order to take this special feature into account, a stream of models and inferential procedures have been developed in survival analysis. The models are called cure models, and we refer to the book of Maller and Zhou (1996) for early references on this topic.

Let T be the survival time, a positive random variable with survival function $S(t) = P(T > t)$. If a certain proportion of infinite survival times exist, we can write this survival function as $S(t) = P(T > t|Y = 0)P(Y = 0) + P(T > t|Y = 1)P(Y = 1)$, where Y is a binary variable taking the value 0 if a subject is cured (meaning that $T = \infty$ in that case) and 1 if it is not. Using the notation $1 - \phi = P(Y = 0) = \lim_{t\to\infty} S(t)$ for the so-called cure rate, we can write

$$S(t) = 1 - \phi + \phi S_1(t), \tag{1}$$

where $S_1(t) = P(T > t|Y = 1)$ is the (proper) survival function of the uncured subjects. As such, we can see that our population is a mixture of two sub-populations, namely the non-susceptible or cured ones (with mixture probability $1 - \phi$) and the susceptible or uncured ones (with mixture probability ϕ). We will assume in this paper that the event times are subject to random right censoring. In the presence of right censoring, all infinite survival times will be censored, as well as some of the finite survival times, which makes it hard to distinguish the cured subjects from the susceptible ones. When a subject is uncensored we know for sure that it will be a non-cured one, but a censored subject can be both cured or uncured. This is a major challenge, not only for the estimation, but even for the identification of the model. See Section 2 for a more detailed discussion.

The advantage of writing the survival function $S(t)$ as in (1) is that it allows to separately study the cure rate $1-\phi$ and the survival function of the susceptibles $S_1(t)$. In this paper we focus on the latter. A parametric form for $S_1(t)$ is often imposed in practice. An attractive feature of parametric models is that usually nice expressions exist for relevant parameters (e.g. median time to event) and functions (e.g. hazard) (Klein and Moeschberger 1997). However, to ensure sound inference, one should verify whether or not the chosen parametric survival function fits the data well. Starting with Pettitt and Stephens (1976) and Koziol and Green (1976), ample (Cramér-von Mises and other) tests have been suggested in classical survival analysis (e.g., Koziol

(1980), Hollander and Peña (1992) and Kim (2017)). In the context of survival data with cure fraction, the problem of testing the parametric form of the survival function $S_1(t)$ has so far only been studied in an informal way, no formal test statistics have been developed. Our paper will fill this important gap. The existing literature is limited to Maller and Zhou (1996), who compare, in an informal way, the parametric fit for the population survival function with the nonparametric fit based on a Kaplan-Meier estimator (Kaplan and Meier 1958). More precisely, Maller and Zhou (1996) use the correlation as test statistic, where a value close to 1 indicates a good parametric choice. No asymptotic justification is given. In this paper, we propose a novel and simple test to assess the aptness of the assumed parametric form. To this end, for the survival function of the susceptible subpopulation, the parametric fit is contrasted with the nonparametric fit in an L_2 distance test statistic. Opposed to Maller and Zhou (1996), we study the asymptotic behavior of the test under the null hypothesis and under local alternatives.

This paper is the first paper to develop a formal test procedure for the parametric form of the survival function $S_1(t)$. We believe that this development can in a later stage be used for more sophisticated testing procedures. For instance, an important extension would be the development of test procedures for the parametric form of the function $S_1(t)$ when the survival function $S(t)$ in (1) depends on a set of covariates. In that case, there exist two main streams of models, the so-called mixture cure models and the promotion time cure models. We refer to Peng and Taylor (2014) and Amico and Van Keilegom (2018) for recent review papers on these two classes of models. The mixture cure model utilizes the same structure as in (1), and therefore the main ideas of our approach should be extendable to mixture cure models. A major challenge is however the way the covariates are handled in that case. In a goodness-of-fit problem it is important to keep the parsimonious (full) model as flexible as possible by imposing a minimum of constraints on the model. However, when $S_1(t)$ depends on covariates, a completely nonparametric model for $S_1(t)$ would lead to curse-of-dimensionality problems, and one might therefore have to use dimension reduction techniques in the parsimonious model. As this is a challenging problem which is moreover not related to the problem we want to address in this paper, we prefer not to deal with the regression case in this paper, and to leave this for further research.

The paper is organized as follows. In Section 2 we define our test statistic, together with the nonparametric estimators of ϕ, $S(t)$ and $S_1(t)$. Section 3 contains the asymptotic results, and the regularity conditions under which these results are valid. In Section 4 we investigate the finite sample behavior of the test for diverse parametric survival functions. We consider various practical settings: a low versus a high susceptibility fraction and a short versus a long plateau in the population survival curve. A bootstrap algorithm that facilitates the calculation of an approximate p-value is described. In Section 5 we select a parametric model for the bone marrow transplant data (BMT data, Klein and Moeschberger (1997)) and for the UMARU impact study data (UIS data, Hosmer et al. (2008)), two real-life data sets with different complementary cure features. Indeed, the BMT data exhibit a low susceptibility rate and a long plateau, whereas the UIS data show a high susceptibility rate and a short plateau. We compare our results to those based on the correlation test by Maller and

Zhou (1996). Section 6 summarizes the most important findings of this paper. Finally, the Appendix contains the proofs of the asymptotic results. The paper comes with extensive online Supplementary Material on common parametric survival functions, a schematic presentation of the simulation algorithm as well as additional simulation and real-life data results.

2 Test procedure

As mentioned in the Introduction, we suppose that the time to event is subject to random right censoring. Let C denote the random censoring time independent of T. The observed data are $\tilde{T} = \min(T, C)$ and $\delta = I(T \leq C)$. In this paper, we assume that T and C are continuous. For a survival function with cure fraction Maller and Zhou (1992) proposed the following nonparametric estimators for ϕ and $S_1(t)$:

$$\hat{\phi} = 1 - \hat{S}(\max(\tilde{T}_i | \delta_i = 1)) \text{ and } \hat{S}_1(t) = \frac{\hat{\phi} - 1 + \hat{S}(t)}{\hat{\phi}}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimator of $S(t)$ (Kaplan and Meier 1958):

$$\hat{S}(t) = \prod_{\substack{\tilde{T}_i \leq t \\ \delta_i = 1}} \Big(1 - \frac{1}{n - i + 1}\Big),$$

based on an independent and identically distributed sample $(\tilde{T}_i, \delta_i)$ $(i = 1, \ldots, n)$ with the same distribution as $(\tilde{T}, \delta)$. We suppose that $0 < \phi < 1$.

Note that $\hat{\phi}$, our estimator of ϕ, corresponds to the estimator in Maller and Zhou (1992). Indeed, with $\tilde{T}_{(n)} = \max(\tilde{T}_i)$, $\hat{S}(\tilde{T}_{(n)}) = \hat{S}(\max\{\tilde{T}_i : \delta_i = 1\})$. Further, note that if $\tilde{T}_{(n)} = \tilde{T}^1_{(n)}$ with $\tilde{T}^1_{(n)} = \max\{\tilde{T}_i : \delta_i = 1\}$, then $\hat{S}(\tilde{T}_{(n)}) = 0$. This suggests that there is no cured group of subjects. To avoid this situation, which is in conflict with the considered model, we need that $\tau_{F_1} \leq \tau_G$, where $F_1 = 1 - S_1$ is the distribution of the susceptibles, G is the distribution of the censoring variable C, and for a generic distribution function L, $\tau_L = \inf\{t : L(t) = 1\}$.

Also note that $\hat{S}_1(t) = 0$ for $t > \max\{\tilde{T}_i : \delta_i = 1\}$, which classifies the censored observations larger than $\max\{\tilde{T}_i : \delta_i = 1\}$ as cured and, by doing so, makes it possible to identify in a practical way the immune subjects (and therefore eliminates the identification problem).

Important parametric models with $\tau_{F_1} = \infty$ include the Weibull, Gompertz and lognormal. Parametric models with $\tau_{F_1} < \infty$ are the uniform distribution on $[0, \theta]$ and a truncated version of the Weibull, Gompertz or lognormal (Table 1 of the Supplementary Material). If τ_{F_1} is finite, it can be considered as known or unknown. The latter case is the most realistic one in practice, and implies that τ_{F_1} is an additional parameter in the model.

To ensure valid inference, it is important to verify the aptness of the chosen parametric form for $S_1(t)$. To this end, consider the hypotheses

$$\mathcal{H}_0 : S_1 \in \{S_{1,\theta} : \theta \in \Theta\} \quad \text{versus} \quad \mathcal{H}_1 : S_1 \notin \{S_{1,\theta} : \theta \in \Theta\}$$

where Θ is the parameter space of the (p-dimensional vector of) parameter(s) θ in the assumed parametric form for $S_1(t)$. As test statistic we define the Cramér-von Mises distance

$$\Lambda_n = \sum_{i=1}^{n} \left\{ \hat{S}_1(\tilde{T}_i) - S_{1,\hat{\theta}}(\tilde{T}_i) \right\}^2, \tag{2}$$

where $\hat{S}_1(t)$ is the nonparametric estimator of Maller and Zhou (1992) and $\hat{\theta}$ is the maximum likelihood estimator for θ under the null hypothesis, defined as

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} \log L(\theta, \hat{\phi})$$

where the log-likelihood is

$$\log L(\theta, \phi) = \sum_{i=1}^{n} \left[\delta_i \log \left\{ \phi f_{1,\theta}(\tilde{T}_i) \right\} + (1 - \delta_i) \log \left\{ 1 - \phi + \phi S_{1,\theta}(\tilde{T}_i) \right\} \right]$$

and $f_{1,\theta}(t) = (\partial/\partial t) F_{1,\theta}(t)$ and $F_{1,\theta}(t) = 1 - S_{1,\theta}(t)$, i.e. we maximize the log-likelihood with respect to θ for ϕ fixed at its nonparametrically estimated value $\hat{\phi}$. By doing so, we ensure valid comparison of both estimators for the survival function of the susceptible subpopulation, $\hat{S}_1(t)$ resp. $S_{1,\hat{\theta}}(t)$.

3 Asymptotic theory

Let θ_0 be the true value of θ under $\mathcal{H}_0$. It is the solution (in θ) of the p-dimensional system of score equations $M(\theta, \phi_0) = 0$, where ϕ_0 is the true value of ϕ, $M(\theta, \phi) = E[m(\tilde{T}, \delta, \theta, \phi)]$ is the expected value with respect to the true distribution $\phi_0 F_{1,\theta_0}(\cdot)$ of T and $G(\cdot)$ of C, and

$$m(t, \delta, \theta, \phi) = \begin{cases} \delta \frac{\frac{\partial}{\partial \theta} f_{1,\theta}(t)}{f_{1,\theta}(t)} + (1 - \delta) \frac{\phi \frac{\partial}{\partial \theta} S_{1,\theta}(t)}{1 - \phi + \phi S_{1,\theta}(t)} & \text{for } t \in A_\theta \\ 0 & \text{for } t \notin A_\theta, \end{cases}$$

whenever $f_{1,\theta}(t) \neq 0$, where

$$A_\theta = \left\{ t \in \mathbb{R}^+ : \tfrac{\partial}{\partial \theta} f_{1,\theta}(t) \text{ and } \tfrac{\partial}{\partial \theta} S_{1,\theta}(t) \text{ exist} \right\},$$

and $\frac{\partial}{\partial \theta}$ denotes the vector of partial derivatives with respect to the components of θ. Note that when $\theta = \tau_{F_1}$ is finite and unknown, the survival function $S_{1,\theta}(t)$ is possibly not differentiable in θ at $t = \theta$ (for fixed t). Fortunately, as will be described below, the latter does not disturb the asymptotic theory for our test statistic. We assume throughout the paper that at most a finite number of values of t do not belong to A_θ. As such, if the true distribution $\phi_0 F_{1,\theta_0}(\cdot)$ is continuous on $[0, \tau_{F_1}]$, the function $M(\theta, \phi)$ is well-defined. See also assumption (A1) below. In addition,

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta} \| M_n(\theta, \hat{\phi}) \|,$$

where $M_n(\theta,\phi) = n^{-1}\sum_{i=1}^n m(\tilde{T}_i,\delta_i,\theta,\phi)$, and $\|\cdot\|$ is the Euclidean norm. Note that $\hat{\theta}$, obtained as the solution of the empirical score equations, corresponds to the maximum likelihood estimator of θ in Section 2, provided none of the data points $\tilde{T}_i$ belongs to the set A_θ^c (which will be the case with probability 1 under assumption (A1)). Note that since $M_n(\theta,\hat{\phi})$ is not necessarily smooth in θ, a solution of the equation $M_n(\theta,\hat{\phi}) = 0$ does not necessarily exist, and so we minimize the norm $\|M_n(\theta,\hat{\phi})\|$ instead. Define $\Gamma(\theta,\phi) = (\partial/\partial\phi)M(\theta,\phi)$, let $\Delta(\theta,\phi)$ be the $p \times p$-matrix of partial derivatives of $M(\theta,\phi)$ with respect to the components of θ, and let $\Delta = \Delta(\theta_0,\phi_0)$.

Finally, $\hat{\phi}$ and ϕ can be written as

$$\hat{\phi} = 1 - \hat{S}(\tau_{F_1}-) \qquad \text{and} \qquad \phi = 1 - S(\tau_{F_1}-). \tag{3}$$

Indeed, the Kaplan-Meier estimator $\hat{S}$ does not change after the last uncensored observation, thus $\hat{S}(\max(\tilde{T}_i|\delta_i = 1)) = \hat{S}(\tau_{F_1}-)$. On the other hand, $\phi = P(Y = 1) = P(T < \infty) = P(T < \tau_{F_1}) = 1 - S(\tau_{F_1}-)$.

Denoting the distribution function of C resp. $\tilde{T}$ by G resp. H, we make the following assumption, which is essential for the identifiability of our model (see e.g. Maller and Zhou (1992)):

$$\tau_{F_1} \leq \tau_G \leq \infty.$$

Since $\tau_F = \infty$ and $\tau_G \leq \infty$, we have that $\tau_H = \tau_G$, and hence $\tau_{F_1} \leq \tau_H \leq \infty$.

The asymptotic results given below are valid under the following regularity conditions (for the definition of a Donsker class, we refer to Chapter 2.1 in Van der Vaart and Wellner (1996)):

(A1) The cardinality of the set $A_{\theta_0}^c$ is finite.
(A2) For all $\delta > 0$, there exists a $\epsilon > 0$ such that $\inf_{\|\theta-\theta_0\|>\delta} \|M(\theta,\phi_0)\| > \epsilon$.
(A3) The matrix $\Delta(\theta,\phi_0)$ exists for all θ and is continuous in θ at $\theta = \theta_0$. Moreover, $\Delta = \Delta(\theta_0,\phi_0)$ is non-singular.
(A4) $\tau_{F_1} \leq \tau_G \leq \infty$.
(A5) T and C are independent and continuous.
(A6) $\int_0^{\tau_{F_1}} \frac{dF(t)}{(1-G(t))^2} < \infty$.
(A7) The class $\{(t,\delta) \to m(t,\delta,\theta,\phi) : \theta \in \Theta, \phi \in [0,1]\}$ is Donsker.
(A8) $E\big[m^2(\tilde{T},\delta,\theta_0,\phi_0)\big] < \infty$.
(A9) $\sup_{t\in B} \big|\frac{\partial}{\partial\theta} S_{1,\theta_0}(t)\big| < \infty$, where $B = [0,\tau_H] \cap [0,\infty) \cap A_{\theta_0}$.

Note that assumption (A1) is required to make sure that the discontinuity points in the set A_{θ_0} do not disturb the asymptotic theory. Assumption (A2) is an identifiability assumption, that is common in the context of M- and Z-estimators (see e.g. Chen et al. (2003)). Note that a sufficient (but not necessary) condition for (A6) is that $\tau_{F_1} < \tau_G$. If $f_{1,\theta}$ and $S_{1,\theta}$ are twice continuously differentiable with respect to θ (which is the case for most parametric families defined on $[0,\infty)$), then assumption (A7) is easy to verify, using e.g. Corollary 2.7.2 in Van der Vaart and Wellner (1996). On the other hand, if τ_{F_1} is finite and depends on θ, then the verification of assumption (A7) will depend on the parametric family at hand. For instance, for the uniform density $f_{1,\theta}(t) = I(0 \leq t \leq \theta)/\theta$, the monotonicity of the indicator

function together with Theorem 2.7.5 in Van der Vaart and Wellner (1996) ensure the Donsker property. Finally, assumptions (A8) and (A9) are regularity conditions needed for some technical developments in the proofs.

Our first result gives a uniform independent and identically distributed expansion and the weak convergence of the process $n^{1/2}\big(\hat{S}(\cdot)-S(\cdot)\big)$ in the space $\ell^\infty([0,\tau_H]\cap[0,\infty))$ of bounded functions defined on $[0,\tau_H]\cap[0,\infty)$, endowed with the uniform norm. Note that $\hat{S}(t)=\hat{S}(\tau_{F_1})$ and $S(t)=S(\tau_{F_1})$ for $t\geq\tau_{F_1}$, and so it suffices to study the process $\hat{S}(\cdot)-S(\cdot)$ on the interval $[0,\tau_{F_1}]\cap[0,\infty)$. The result is an extension of Theorem 1 in Lo and Singh (1986), who restrict attention to independent and identically distributed expansions that are uniform on $[0,\tau]$ for any $\tau<\tau_H$.

Proposition 1 *Assume (A4)-(A6). Then,*

$$\hat{S}(t)-S(t)=-n^{-1}\sum_{i=1}^{n}\xi(\tilde{T}_i,\delta_i,t)+r_n(t),$$

where $\sup_{t\in[0,\tau_H]\cap[0,\infty)}|r_n(t)|=o_P(n^{-1/2})$*,*

$$\xi(y,\delta,t)=(1-F(t))\Big\{\frac{I(y\leq t)\delta}{1-H(y)}-\int_0^{y\wedge t}\frac{dH^1(z)}{(1-H(z))^2}\Big\},$$

$a\wedge b=\min(a,b)$*,* $H^1(z)=P(\tilde{T}\leq z,\delta=1)$ *and* $0/0$ *is defined as* 0*. Moreover, the process* $n^{1/2}\big(\hat{S}(\cdot)-S(\cdot)\big)$ *converges weakly in the space* $\ell^\infty([0,\tau_H]\cap[0,\infty))$ *to a zero-mean Gaussian process with covariance function* $(t_1,t_2)\to E[\xi(\tilde{T},\delta,t_1)\xi(\tilde{T},\delta,t_2)]$*.*

We continue with a result on the limiting distribution of $n^{1/2}(\hat{\theta}-\theta_0)$.

Proposition 2 *Assume (A1)-(A7). Then,* $\hat{\theta}-\theta_0\to 0$ *in probability,*

$$\begin{aligned}\hat{\theta}-\theta_0&=-\Delta^{-1}\Big\{M_n(\theta_0,\phi_0)+\Gamma(\theta_0,\phi_0)(\hat{\phi}-\phi_0)\Big\}+o_P(n^{-1/2})\\&=-\Delta^{-1}n^{-1}\sum_{i=1}^{n}\big\{m(\tilde{T}_i,\delta_i,\theta_0,\phi_0)+\Gamma(\theta_0,\phi_0)\xi(\tilde{T}_i,\delta_i,\tau_{F_1}-)\big\}+o_P(n^{-1/2}),\end{aligned}$$

and $n^{1/2}(\hat{\theta}-\theta_0)\to N_p(0,\Delta^{-1}V\Delta^{-1})$ *in distribution, where* $V=Var\big\{m(\tilde{T},\delta,\theta_0,\phi_0)+\Gamma(\theta_0,\phi_0)\xi(\tilde{T},\delta,\tau_{F_1}-)\big\}$*, and where* $\xi(\tilde{T}_i,\delta_i,\tau_{F_1}-)=\lim_{t\uparrow\tau_{F_1}}\xi(\tilde{T}_i,\delta_i,t)$*.*

We are now ready to state the weak convergence of the process $n^{1/2}\big(\hat{S}_1(\cdot)-S_{1,\hat{\theta}}(\cdot)\big)$ in the space $\ell^\infty(B)$, with B as in (A9). Note that

$$\hat{S}_1(t)-S_{1,\hat{\theta}}(t)=\hat{S}_1(t)-S_1(t)-\Big(\tfrac{\partial}{\partial\theta}S_{1,\theta_0}(t)\Big)^t\big(\hat{\theta}-\theta_0\big)+o_P(\|\hat{\theta}-\theta_0\|),$$

thanks to Young's form of Taylor's expansion (see e.g. Serfling (1980)). Hence, making use of Propositions 1 and 2 above for the first and second term of this decomposition, we will develop an independent and identically distributed expansion for $\hat{S}_1(\cdot)-S_{1,\hat{\theta}}(\cdot)$ on B.

Theorem 1 *Assume (A1)-(A9).*

(a) Under $\mathcal{H}_0$*,*

$$\hat{S}_1(t) - S_{1,\hat{\theta}}(t) = n^{-1} \sum_{i=1}^{n} \eta(\tilde{T}_i, \delta_i, t) + R_n(t),$$

where $\sup_{t \in B} |R_n(t)| = o_P(n^{-1/2})$*, and where*

$$\begin{aligned}\eta(\tilde{T}_i, \delta_i, t) = &-\frac{1}{\phi_0}\xi(\tilde{T}_i, \delta_i, t) \\ &+\Big[\frac{1 - S(t)}{\phi_0^2} + \Big(\frac{\partial}{\partial\theta} S_{1,\theta_0}(t)\Big)^t \Delta^{-1} \Gamma(\theta_0, \phi_0)\Big]\xi(\tilde{T}_i, \delta_i, \tau_{F_1}-) \\ &+\Big(\frac{\partial}{\partial\theta} S_{1,\theta_0}(t)\Big)^t \Delta^{-1}\, m(\tilde{T}_i, \delta_i, \theta_0, \phi_0)\end{aligned}$$

for $t \in A_{\theta_0}$*.*

(b) Under $\mathcal{H}_0$*, the process* $n^{1/2}\big(\hat{S}_1(\cdot) - S_{1,\hat{\theta}}(\cdot)\big)$ *converges weakly in* $\ell^\infty(B)$ *to a zero-mean Gaussian process* $W(\cdot)$ *with covariance function*

$$Cov\big(W(t_1), W(t_2)\big) = E\big[\eta(\tilde{T}, \delta, t_1)\eta(\tilde{T}, \delta, t_2)\big].$$

The limiting distribution of our test statistic Λ_n now follows:

Corollary 1 *Assume (A1)-(A9). Then, under* $\mathcal{H}_0$*,*

$$\Lambda_n \to \int W^2(t)\, dH(t) \quad \textit{in distribution.}$$

Next, we consider the asymptotic distribution of the test statistic Λ_n under local alternatives of the form

$$\mathcal{H}_{1n} : S_1 = S_{1,\theta_0} + n^{-1/2} c(\tilde{S}_1 - S_{1,\theta_0}), \tag{4}$$

where $\tilde{S}_1$ is a fixed survival function and c is a fixed constant. Note that under $\mathcal{H}_{1n}$ we can also write S_1 as $S_1 = (1 - n^{-1/2}c)S_{1,\theta_0} + n^{-1/2}c\tilde{S}_1$, which shows that S_1 is a survival function, provided that n is sufficiently large so that $0 \leq c \leq \sqrt{n}$. We then have the following result.

Corollary 2 *Assume (A1)-(A9). Then, under* $\mathcal{H}_{1n}$*,*

$$\Lambda_n \to \int [W(t) + cb(t)]^2\, dH(t),$$

in distribution, where

$$b(t) = \frac{\frac{\partial S_{1,\theta_0}}{\partial\theta}(t)}{\frac{\partial M}{\partial\theta}(\theta_0, \phi_0)} E_{\tilde{S}}[m(\tilde{T}, \delta, \theta_0, \phi_0)] + \tilde{S}_1(t) - S_{1,\theta_0}(t),$$

where $\tilde{S}(t) = \tilde{S}_1(t)\phi_0 + 1 - \phi_0$ *and where* $E_{\tilde{S}}[m(\tilde{T}, \delta, \theta_0, \phi_0)]$ *denotes the expected value assuming that the survival function of* T *equals* $\tilde{S}$*.*

Note that the asymptotic distribution of the test statistic, both under the null hypothesis and under local alternatives, depends on several unknown quantities. Moreover, convergence in distribution might be rather slow. We therefore propose a parametric bootstrap algorithm to obtain an approximate p-value for the test. Details are given below.

Step 1. Obtain, under $\mathcal{H}_0$, the estimators $\hat{\theta}$ and $\hat{\phi}$. Furthermore, calculate the test statistic Λ_n defined in (2).
Step 2. For $b = 1, \ldots, n_b$ create resamples in the following way :
Step 2.1. Generate Y_i^b where $Y_i^b \sim Be(\hat{\phi})$, $i = 1, \ldots, n$.
Step 2.2. Generate T_i^b from $S_{1,\hat{\theta}}(\cdot)$ if $Y_i^b = 1$, and set $T_i^b = \infty$ if $Y_i^b = 0$.
Step 2.3. Generate C_i^b from $1 - \hat{G}(\cdot)$, the Kaplan-Meier estimator of $1 - G(\cdot)$.
Step 2.4. Set $\tilde{T}_i^b = \min(T_i^b, C_i^b)$ and $\delta_i^b = I(T_i^b \leq C_i^b)$.
Step 2.5. Obtain the bootstrap value of the test statistic:

$$\Lambda_n^b = \sum_{i=1}^n \left\{ \hat{S}_1^b(\tilde{T}_i^b) - S_{1,\hat{\theta}^b}(\tilde{T}_i^b) \right\}^2$$

where $\hat{S}_1^b$ is the Kaplan-Meier estimator of S_1 and $\hat{\theta}^b$ is the estimator of θ, both based on the bootstrap data.
Step 3. Calculate the approximate p-value via $p_{\text{boot}} = \frac{1}{n_b} \sum_{b=1}^{n_b} I(\Lambda_n^b \geq \Lambda_n)$.

4 Simulation study

4.1 Non-local alternatives

Simulation setting. For the event times T_i of susceptible study items we consider four survival functions: Weibull, Gompertz, lognormal or uniform. Details on the parameter choices are listed in Table 1. The corresponding density, survival and hazard functions are visualized in Figure 1 of the Supplementary Material. Non-susceptible study items have an event at $T_i = \infty$. The observed data are given by $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ $(i = 1, \ldots, n)$, where the censoring times C_i $(i = 1, \ldots, n)$ are generated from a Weibull survival function with $\lambda = 0.05$ and $\rho = 3$, resp. $\lambda = 0.05$ and $\rho = 1.5$, leading to a short, resp. a long plateau in the tail of the Kaplan-Meier estimator of the survival function of the event times T_i $(i = 1, \ldots, n)$. The probability of being susceptible is chosen to be $\phi = 0.25$ or $\phi = 0.75$, representing a low, resp. a high susceptibility rate. The Supplementary Material contains a schematic presentation of the simulation algorithm. In each of the aforementioned scenarios, we assess the fit of a Weibull, a Gompertz, a lognormal and a uniform survival function to the susceptible subpopulation. The results are based on 1000 data sets with sample size $n = 150$ or $n = 500$, each supplemented with 500 bootstrap samples.

Simulation results. Before interpreting the estimation and the testing results, we point out that within a particular scenario the support of the censoring distribution (or the length of the follow-up period) can be linked to the censoring rate. For example, if $\phi = 0.25$ then we know that we have a censoring rate of at least 75%. For a long follow-up, it is more likely that the event of a susceptible study item will be observed within the study period than e.g. for a short follow-up, resulting in a censoring

rate closer to 75% for the former. The same holds true if $\phi = 0.75$, where we have a censoring rate of at least 25%.

The estimation accuracy is evaluated by the root mean squared error (RMSE) over all time points

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \{\tilde{S}_1(\tilde{T}_i) - S_1(\tilde{T}_i)\}^2}$$

where $\tilde{S}_1(t) \in \{S_{1,\hat{\theta}}(t), \hat{S}_1(t)\}$ is the parametric or the nonparametric counterpart of the true survival function $S_1(t)$. As such we define RMSE_p and RMSE_np. The average of the obtained values over the 1000 trials are summarized in Table 2 ($\phi = 0.75$, long plateau) and Table 3 ($\phi = 0.25$, short plateau), as well as Table 2 ($\phi = 0.75$, short plateau) and Table 3 ($\phi = 0.25$, long plateau) of the Supplementary Material. The results on the estimated parameter values are deferred to the Supplementary Material (Table 8 to Table 11). Note that ϕ is obtained nonparametrically, even if $S_1(t)$ is estimated parametrically. As expected, if the model is correct then RMSE_p is (on average) smaller that RMSE_np. Further, if the model is correct then parameters are (on average) estimated close to their true value. Estimation is somewhat more accurate for $\phi = 0.75$ and in the long plateau scenarios. In all scenarios, ϕ is (on average) estimated on target.

We evaluate the testing strategy through the empirical type I error and the empirical power attained at significance levels $\alpha = 0.05$ and $\alpha = 0.10$. The rejection rates are reported in Table 2 ($\phi = 0.75$, long plateau) and Table 3 ($\phi = 0.25$, short plateau), as well as in Table 2 ($\phi = 0.75$, short plateau) and Table 3 ($\phi = 0.25$, long plateau) of the Supplementary Material. It follows that the empirical type I error is usually (slightly) lower than the nominal level for the short plateau scenarios and a small sample size, but that it is close to the latter for the long plateau scenarios and a higher sample size. Similarly, the empirical power is higher for long plateau scenarios as well as for a higher susceptibility rate. Also, a larger sample size leads to a substantial increase in the empirical power. The results for the short plateau scenarios are somewhat less impressive than the ones for the long plateau scenarios. The latter is to be expected. A short plateau implies a more difficult detection of the cure fraction and therefore causes a more difficult fit and goodness-of-fit verification of the mixture cure model. Overall, the difference between a Weibull and a Gompertz survival function as well as between a uniform and a Gompertz survival function seems harder to detect, while the distinction between other survival functions appears to be easier. A comparison of the true survival function with its various estimated parametric counterparts illustrates and supports this statement, see Figure 4 to Figure 19

Table 1 Simulation settings for the event times of susceptible subjects, see also Table 1 of the Supplementary Material.

	λ	ρ	μ	γ	θ	mean	variance
Weibull	0.50	1.50				1.433	0.947
Gompertz	0.25			0.75		1.543	0.789
lognormal			0.50	0.25		1.868	0.991
uniform					3.25	1.625	0.880

Table 2 Simulation results - $\phi = 0.75$ and long plateau. Average censoring rate; average of the root mean squared error for the nonparametric estimator (RMSEnp) and the parametric estimator (RMSEp); rejection rates at 5% and at 10%; AIC preference. The results are based on 1000 trials and 500 bootstrap samples.

true	fit	n	cens.	RMSEnp	RMSEp	rej. rate 5%	rej. rate 10%	AIC pref.
Weibull	Weibull	150	0.318	0.420	0.320	0.042	0.099	
		500	0.318	0.415	0.317	0.051	0.092	
	Gompertz	150	0.318	0.419	0.451	0.433	0.561	0.872
		500	0.318	0.415	0.767	0.910	0.941	0.986
	lognormal	150	0.318	0.420	0.575	0.760	0.854	0.957
		500	0.318	0.413	0.926	1	1	1
	uniform	150	0.318	0.419	2.245	0.999	0.999	0.996
		500	0.318	0.413	5.039	1	1	1
Gompertz	Weibull	150	0.324	0.416	0.455	0.497	0.618	0.933
		500	0.324	0.415	0.678	0.986	0.992	0.999
	Gompertz	150	0.324	0.416	0.313	0.058	0.117	
		500	0.324	0.414	0.314	0.055	0.100	
	lognormal	150	0.324	0.417	0.961	0.998	1	1
		500	0.324	0.416	1.698	1	1	1
	uniform	150	0.324	0.416	1.006	0.791	0.863	0.851
		500	0.324	0.415	2.513	1	1	0.999
lognormal	Weibull	150	0.344	0.432	0.516	0.784	0.867	0.960
		500	0.344	0.433	0.807	1	1	1
	Gompertz	150	0.344	0.434	1.014	0.998	0.999	0.997
		500	0.344	0.430	1.896	1	1	1
	lognormal	150	0.344	0.434	0.330	0.059	0.113	
		500	0.344	0.432	0.330	0.054	0.097	
	uniform	150	0.344	0.434	2.151	1	1	1
		500	0.344	0.432	5.111	1	1	1
uniform	Weibull	150	0.330	0.419	0.594	0.926	0.968	1
		500	0.330	0.416	0.972	1	1	1
	Gompertz	150	0.330	0.419	0.451	0.605	0.731	0.996
		500	0.330	0.417	0.659	0.993	0.999	1
	lognormal	150	0.330	0.418	1.069	1	1	1
		500	0.330	0.413	1.902	1	1	1
	uniform	150	0.330	0.419	0.074	0.055	0.099	
		500	0.330	0.418	0.041	0.040	0.098	

in the Supplementary Material. The parameter estimates in these fitted models are obtained by taking the average over 1000 estimated values based on samples of size $n = 500$. It follows that the estimated Weibull and Gompertz survival functions are often close to each other and to the true, Weibull or Gompertz, survival function. The same holds true for a uniform versus a Gompertz survival function. The tables also show the AIC preference rate, i.e. the extent to which the correct parametric survival function is preferred over an incorrect parametric survival function as based on a AIC comparison. Here, AIC is defined as $-2\log(L) + 2n_{par}$, where $\log(L)$ is the value of the log-likelihood of the considered model and n_{par} is the number of estimated parameters. As mentioned before, ϕ is estimated non-parametrically and hence not a direct part of the AIC calculation. A smaller AIC value corresponds to a better parametric model. It follows that a lower AIC preference rate goes along with a smaller

Table 3 Simulation results - $\phi = 0.25$ and short plateau. Average censoring rate; average of the root mean squared error for the nonparametric estimator (RMSE_{np}) and the parametric estimator (RMSE_{p}); rejection rates at 5% and at 10%; AIC preference. The results are based on 1000 trials and 500 bootstrap samples.

true	fit	n	cens.	RMSE_{np}	RMSE_{p}	rej. rate 5%	rej. rate 10%	AIC pref.
Weibull	Weibull	150	0.803	1.138	0.921	0.032	0.078	
		500	0.803	1.302	1.029	0.029	0.085	
	Gompertz	150	0.803	1.137	1.002	0.047	0.111	0.651
		500	0.803	1.300	1.212	0.156	0.237	0.845
	lognormal	150	0.803	1.137	1.176	0.264	0.411	0.739
		500	0.803	1.299	1.993	0.590	0.726	0.913
	uniform	150	0.803	1.138	1.029	0.049	0.101	0.265
		500	0.803	1.301	2.119	0.569	0.714	0.877
Gompertz	Weibull	150	0.812	1.081	1.018	0.141	0.256	0.714
		500	0.812	1.078	1.405	0.446	0.610	0.863
	Gompertz	150	0.812	1.079	0.923	0.027	0.070	
		500	0.812	1.079	0.908	0.034	0.070	
	lognormal	150	0.812	1.081	1.733	0.566	0.715	0.866
		500	0.812	1.079	3.345	0.945	0.974	0.993
	uniform	150	0.812	1.081	0.843	0.045	0.095	0.089
		500	0.812	1.078	1.138	0.044	0.119	0.438
lognormal	Weibull	150	0.826	1.329	1.204	0.112	0.190	0.715
		500	0.826	1.523	1.550	0.401	0.540	0.891
	Gompertz	150	0.826	1.330	1.302	0.335	0.490	0.892
		500	0.826	1.514	1.682	0.905	0.954	0.996
	lognormal	150	0.826	1.332	1.004	0.038	0.090	
		500	0.826	1.520	1.083	0.033	0.080	
	uniform	150	0.826	1.330	1.606	0.216	0.325	0.964
		500	0.826	1.520	2.056	0.228	0.481	1
uniform	Weibull	150	0.820	1.082	1.193	0.207	0.376	0.944
		500	0.820	1.042	1.833	0.773	0.897	0.983
	Gompertz	150	0.820	1.084	0.987	0.053	0.140	0.932
		500	0.820	1.043	1.092	0.195	0.318	0.964
	lognormal	150	0.820	1.082	1.839	0.528	0.708	0.953
		500	0.820	1.043	3.475	0.980	0.994	0.998
	uniform	150	0.820	1.083	0.800	0.097	0.177	
		500	0.820	1.043	0.525	0.054	0.112	

empirical power of the proposed test, reflecting/confirming the difficulty to obtain a clear detection in some scenarios. Finally, in each simulation scenario the size of the discrepancy between RMSE_{p} and RMSE_{np}, the rejection rate as well as the AIC preference rate point towards a similar conclusion, i.e. a larger discrepancy in RMSE goes together with a higher rejection rate and a higher AIC preference rate, all indications of an incorrect parametric model choice.

To further illustrate the adequacy of our test, we compare its performance to that of the correlation test by Maller and Zhou (1996). The latter involves the estimated survival function with cure fraction and the Kaplan-Meier estimate for the population survival function, each evaluated at the observed event times. The null hypothesis states that the chosen parametric form is suitable, which implies that the correlation is nearly 1. Unfortunately, the asymptotic null distribution of the correlation test is

Table 4 Simulation results - comparison with the test of Maller and Zhou (1996). Rejection rates at 5% and at 10% for the test of Maller and Zhou (1996) and for the proposed test. The results are based on 1000 trials and 500 bootstrap samples.

setting	true	fit	n	Maller-Zhou test rej. rate 5%	Maller-Zhou test rej. rate 10%	Proposed test rej. rate 5%	Proposed test rej. rate 10%
long plateau $\phi = 0.75$	Weibull	Weibull	150	0.042	0.101	0.042	0.099
			500	0.045	0.089	0.051	0.092
		Gompertz	150	0.377	0.496	0.433	0.561
			500	0.831	0.895	0.910	0.941
		lognormal	150	0.707	0.800	0.760	0.854
			500	1	1	1	1
		uniform	150	0.999	1	0.999	0.999
			500	1	1	1	1
short plateau $\phi = 0.25$	Weibull	Weibull	150	0.037	0.079	0.032	0.078
			500	0.033	0.081	0.029	0.085
		Gompertz	150	0.069	0.146	0.047	0.111
			500	0.181	0.287	0.156	0.237
		lognormal	150	0.017	0.057	0.264	0.411
			500	0.054	0.188	0.590	0.726
		uniform	150	0.149	0.249	0.049	0.101
			500	0.791	0.866	0.569	0.714

unknown. Maller and Zhou (1996) therefore resorted to a simulation study, with specific assumptions on e.g. the censoring scheme. Although known here, the latter is often unknown for real life data. To overcome this issue, we opt to apply a bootstrap scheme similar to the one outlined in Section 3, except that we reject for small values of the test statistic instead of large values. As such, no extra assumptions on e.g. the censoring scheme are required. The results are summarized in Table 4. The table shows that both tests respect the nominal level quite well, although the scenario with a short plateau is clearly more difficult for both tests. Concerning the power, we see that in the scenario with a long plateau and $\phi = 0.75$, the proposed test has better power, whereas in the scenario with a short plateau and $\phi = 0.25$, the correlation test of Maller and Zhou (1996) behaves better, except for the fit with the log-normal distribution, where the test of Maller and Zhou (1996) has nearly no power. This suggests that the latter test could have difficulty to detect some alternatives, and in general might not be omnibus, while our newly proposed test is omnibus.

4.2 Local alternatives

Simulation setting. We focus on sample size $n = 500$. For the event times T_i of susceptible study items we consider two survival functions: a combination of Weibull (S_{1,θ_0}) and Gompertz ($\tilde{S}_1$) or a combination of Weibull (S_{1,θ_0}) and lognormal ($\tilde{S}_1$). The parameter choices are as listed in Table 1. The constant c in (4) equals $0, 5, 10, 15,$ 20 or $\sqrt{500}$, representing a transition from a Weibull survival function to a Gompertz, resp. a lognormal survival function. The latter is illustrated in Figure 2 of the Supplementary Material. Non-susceptible study items have an event at $T_i = \infty$. As before,

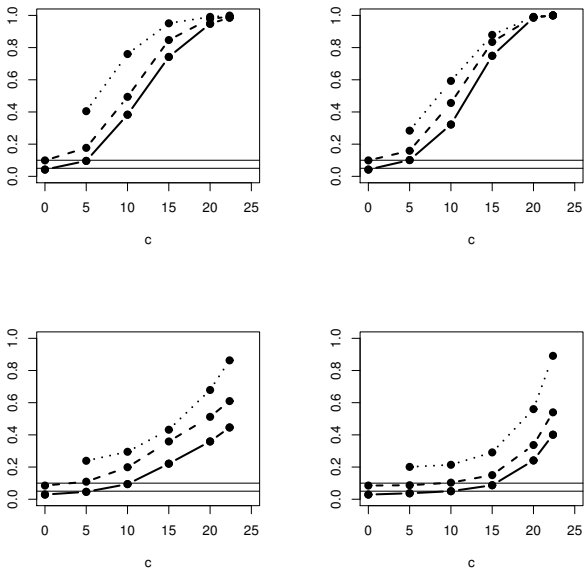

Fig. 1 Top panel: $\phi = 0.75$ and long plateau. Bottom panel: $\phi = 0.25$ and short plateau. Left panel: true survival function of susceptible subpopulation is a combination of Weibull and Gompertz with $c \in \{0, 5, 10, 15, 20, \sqrt{500}\}$, fitted survival function is Weibull. Right panel: true survival function of susceptible subpopulation is a combination of Weibull and lognormal with $c \in \{0, 5, 10, 15, 20, \sqrt{500}\}$, fitted survival function is Weibull. The solid line represents the 5% rejection rate, the dashed line is the 10% rejection rate and the dotted line gives the AIC preference rate. The results are based on 1000 trials and 500 bootstrap samples.

the censoring times C_i $(i = 1, \ldots, 500)$ are generated from a Weibull survival function with $\lambda = 0.05$ and $\rho = 3$ (short plateau), resp. $\lambda = 0.05$ and $\rho = 1.5$ (long plateau). The observed data are given by $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ $(i = 1, \ldots, 500)$. The probability of being susceptible is set to be $\phi = 0.25$ or $\phi = 0.75$. For each of the considered scenarios, we evaluate the fit of a Weibull survival function to the susceptible subpopulation. The results are based on 1000 data sets, each supplemented with 500 bootstrap samples to obtain an approximate p-value.

Simulation results. The estimation and testing results are listed in Table 4 to Table 7 of the Supplementary Material. Figure 1 and Figure 3 of the Supplementary Material visualize the testing results. Clearly the 5% and 10% rejection rates go up as the constant c increases from 0 to $\sqrt{500}$, i.e. as the true survival function moves from Weibull ($c = 0$) to Gompertz, resp. lognormal ($c = \sqrt{500}$), the test gains empirical power. As expected, the AIC preference rate also grows with increasing c.

5 Data application

We explore two real-life data sets. First, it is important to have evidence regarding the underlying assumption $\tau_{F_1} \leq \tau_G$. Maller and Zhou (1992) propose to reject $H_0 : \tau_{F_1} > \tau_G$ if the length of the interval $(2\tilde{T}^1_{(n)} - \tilde{T}_{(n)}, \tilde{T}^1_{(n)}]$ is too large, where $\tilde{T}^1_{(n)}$ is the largest uncensored failure time and $\tilde{T}_{(n)}$ is the largest failure time (censored or not). They further show that with

$$q_n(a) = \Big(1 - \frac{\text{number of uncensored observations} > a}{n}\Big)^n,$$

the estimated p-value is $q_n(v_n \vee 0)$, where v_n is the actual value of $V_n = 2\tilde{T}^1_{(n)} - \tilde{T}_{(n)}$.

Second, the presence of a cure fraction is established by applying the likelihood ratio test of Maller and Zhou (1996), which compares the fit of a model with no cure fraction to the fit of a model with cure fraction. The null hypothesis states that $\phi = 1$ (no cure fraction present) and the asymptotic null distribution of the likelihood ratio test corresponds to a $50-50$ mixture of a chi-square random variable with one degree of freedom and a point mass at 0. The test is accompanied by a visual inspection comparing both fits with the Kaplan-Meier estimate of the population survival function. For the data sets at hand we consider Weibull, Gompertz, lognormal and uniform as plausible parametric choices for the survival function of the entire population (model with no cure fraction) or the survival function of the susceptible group (model with cure fraction).

Third, the aptness of each parametric form is verified. Here, we apply the correlation test of Maller and Zhou (1996) as described in Section 4. The test comes along with a plot containing the corresponding pairs of estimated survival values. The closer these points are to the line with slope 1, or equivalently, the closer the correlation is to 1, the better the parametric fit. The choice of parametric form is further assessed via the test as presented in Section 2, supplemented with the suggested bootstrap scheme.

5.1 Dataset 1: bone marrow transplant

A transplant of bone marrow is a standard therapy for acute leukemia. Unfortunately, recovery is complex and not always successful. Here, we investigate the time to leukemia relapse (in years) based on data of 137 patients. The follow-up time is about 7 years, in which 42 patients (30.66%) experienced a relapse. The left panel of Figure 2 reveals a plateau in the Kaplan-Meier estimate of the population survival function for time to leukemia relapse, hence the presence of a cure fraction (patients who will never experience a relapse) is plausible. First, note that $q_n(v_n \vee 0) = 1.65 \times 10^{-22}$, i.e., the length of the plateau is clearly sufficient ($\tau_{F_1} \leq \tau_G$). Furthermore, the likelihood ratio test of Maller and Zhou (1996) (with actual values 32.48, 58.85, 23.80 and 47.92 for Weibull, uniform, lognormal and Gompertz and critical value equal to 2.71 for $\alpha = 0.05$) indicates that, regardless of the parametric form, a model with cure fraction is preferred (at 5%). Visual support can be found in Figure 20 of the Supplementary material. Indeed, for each parametric choice, the mixture model is closer to the Kaplan-Meier estimate of the population

survival function. The (nonparametrically) estimated probability of being susceptible is $\hat{\phi} = 0.375$. The BMT data are thus characterized by a long plateau in the survival plot and a rather low susceptibility rate. More details on the data can be found in Klein and Moeschberger (1997). Next, we assess the appropriateness of each of the parametric survival functions based on 1000 bootstrap samples. The results are summarized in Table 5 and Table 6. It follows that, for both the correlation test of Maller and Zhou (1996) as well as our test, the null hypothesis is only rejected for a uniform survival function (at 5%). The left panel of Figure 22 in the Supplementary material supports this conclusion, i.e. the dots corresponding to the uniform deviate a lot from the line with slope 1. Also, the AIC values of all parametric models are quite alike with a slight preference for lognormal. The right panel of Figure 2 visualizes the parametric estimates, together with the nonparametric counterpart. It seems that lognormal is closer to the nonparametric estimator for time values below 1.75, while Weibull and Gompertz provide a better fit for time values above 1.75. The uniform survival function clearly insuffices. The estimated parameter values are displayed in Table 12 of the Supplementary Material. To obtain standard errors 1000 bootstrap samples are generated.

Table 5 Results for bone marrow transplant (BMT) and UMARU impact study (UIS), based on the correlation test of Maller and Zhou (1996).

		Weibull	Gompertz	lognormal	uniform
BMT	Test statistic	0.995	0.995	0.996	0.974
	P-value	0.121	0.134	0.227	0.003
UIS	Test statistic	0.998	0.997	0.996	0.890
	P-value	0.057	0.004	0.002	0

5.2 Dataset 2: UMARU impact study

Drugs are addictive and hence many drug users experience a relapse. Here, we analyze the time to return to drug use (in years) after receiving therapy. The data contain information on the treatment duration and the site at which treatment took place.

Table 6 Results for bone marrow transplant (BMT) and UMARU impact study (UIS), based on the test in Section 2.

		Weibull	Gompertz	lognormal	uniform
BMT	Test statistic	0.189	0.158	0.171	2.012
	P-value	0.091	0.182	0.248	0.003
	AIC	201.852	203.352	200.048	207.172
UIS	Test statistic	0.106	0.253	0.423	18.748
	P-value	0.112	0.005	0.012	0
	AIC	269.780	280.758	281.358	465.094

Patients were randomly assigned to a short or a long program, either at site 1 or 2. Short versus long means 3 months resp. 6 months at site 1, while it corresponds to 6 months, resp. 12 months at site 2. Comparison of Kaplan-Meier curves via a log-rank test reveals a significant effect of treatment duration, but not of site ($\alpha = 5\%$). Based on this, we consider the data of patients receiving a short treatment, at either site 1 or 2. As such, we retain 289 patients of which 239 (82.70%) returned to drug use. The follow-up time is about 2 years. The left panel of Figure 2 reveals a plateau in the Kaplan-Meier estimate of the population survival function for time to reuse of drugs and thus suggests the presence of a cure fraction (patients who will never return to drug use). With $q_n(v_n \vee 0) = 2.46 \times 10^{-8}$ also here the condition $\tau_{F_1} \leq \tau_G$ is guaranteed. Moreover, the likelihood ratio test of Maller and Zhou (1996) (with actual values 54.17, 76.21, 9.56 and 45.52 for Weibull, uniform, lognormal and Gompertz and critical value equal to 2.71 for $\alpha = 0.05$) suggests that, regardless of the parametric form, a model with cure fraction is more appropriate (at 5%). Graphical support can be found in Figure 21 of the Supplementary material. Indeed, for each parametric choice, the mixture model is closer to the Kaplan-Meier estimate of the population survival function. The (nonparametrically) estimated probability of being susceptible is $\hat{\phi} = 0.835$. The UIS data thus show a short plateau in the survival plot and a rather high susceptibility rate. Full details on the data are given in Hosmer et al. (2008). Next, we assess the usefulness of each of the parametric survival functions based on 1000 bootstrap samples. The results are given in Table 5 and Table 6. It follows that, for both the correlation test of Maller and Zhou (1996) as well as our test, the null hypothesis is not rejected for Weibull (at 5%). The right panel of Figure 22 in the Supplementary material supports this conclusion, i.e. the dots corresponding to the Weibull seem to be close to the line with slope 1. Also, the AIC value is substantially smaller for Weibull than for Gompertz, lognormal and uniform, confirming the better fit of Weibull. The right panel of Figure 2 visualizes the parametric estimates, together with the nonparametric counterpart. The latter illustrates the deviating fit for lognormal and uniform. The estimated parameter values are given in Table 12 of the Supplementary Material. To obtain standard errors 1000 bootstrap samples are used.

6 Discussion

In this paper we proposed, for a mixture cure rate model, a test procedure to assess the parametric form imposed on the survival function of the susceptible subjects. We obtained asymptotic properties of the parametric and the nonparametric estimator for this survival function and established the asymptotic behavior of the proposed test statistic. A bootstrap algorithm facilitates the calculation of the p-value. The finite sample performance was evaluated in an extensive simulation study. The results indicate that the test procedure has reasonable to high power, especially when the cure proportion is small and when the plateau is long.

Acknowledgements This work was supported by the European Research Council [2016-2021, Horizon 2020/ERC grant No. 694409], by the Interuniversity Attraction Poles Programme [IAP-network P7/06] of the Belgian Science Policy Office, and the Research Foundation Flanders (FWO), Scientific Research

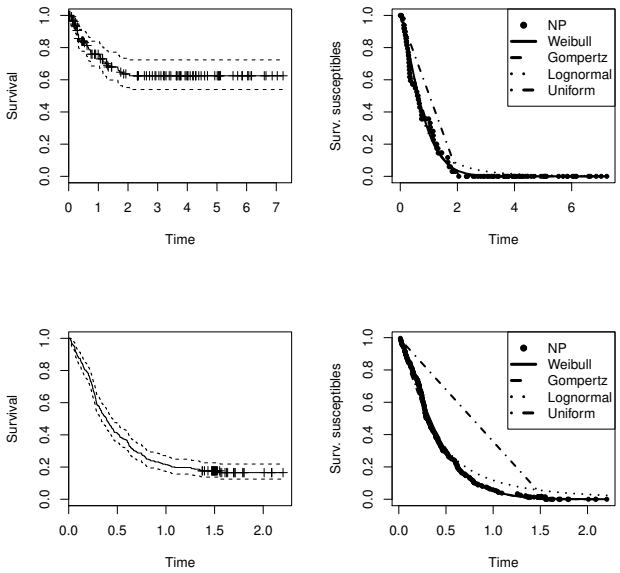

Fig. 2 Top panel: bone marrow transplant. Bottom panel: UMARU impact study. Left panel: Kaplan-Meier estimator of the population survival function. Right panel: estimators of the survival function for the susceptible subpopulation.

Community on 'Asymptotic Theory for Multidimensional Statistics' [W000817N]. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government - department EWI.

References

Amico M, Van Keilegom I (2018) Cure models in survival analysis. Annu Rev Stat Appl 5:311–342

Chen X, Linton O, Van Keilegom I (2003) Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71:1591–1608

Hollander M, Peña E (1992) A chi-squared goodness-of-fit test for randomly censored data. J Amer Statist Assoc 87:458–463

Hosmer D, Lemeshow S, May S (2008) Applied Survival Analysis, Regression Modeling of Time-to-Event Data. Wiley

Kaplan E, Meier P (1958) Nonparametric estimation from incomplete observations. J Amer Statist Assoc 53:457–481

Kim N (2017) Goodness-of-fit tests for randomly censored Weibull distributions with estimated parameters. Commun Statist Applic Meth 24:519–531

Klein J, Moeschberger M (1997) Survival Analysis, Techniques for Censored and Truncated Data. Springer

Koziol J (1980) Goodness-of-fit tests for randomly censored data. Biometrika 67:693–696

Koziol J, Green S (1976) A Cramér-von Mises statistic for randomly censored data. Biometrika 63:465–474

Lo SH, Singh K (1986) The product-limit estimator and the bootstrap: some asymptotic representations. Probab Theory Rel Fields 71:455–465

Maller R, Zhou S (1992) Estimating the proportion of immunes in a censored sample. Biometrika 79:731–739

Maller R, Zhou S (1996) Survival Analysis with Long Term Survivors. Wiley
Pardo-Fernández J, Van Keilegom I, González-Manteiga W (2007) Goodness-of-fit tests for parametric models in censored regression. Canad J Statist 35:249–264
Peng Y, Taylor JMG (2014) Cure models. Handbook of Survival Analysis, Handbooks of Modern Statistical Methods series, ed J Klein, H van Houwelingen, JG Ibrahim, TH Scheike, chapter 6 Boca Raton, FL : Chapman & Hall pp 113–134
Pettitt A, Stephens M (1976) Modified Cramér-von Mises statistics for censored data. Biometrika 63:291–298
Rao C (1965) Linear Statistical Inference and its Applications. Wiley
Sánchez-Sellero C, González-Manteiga W, Van Keilegom I (2005) Uniform representation of product-limit integrals with applications. Scand J Statist 32:563–581
Serfling R (1980) Approximation Theorems of Mathematical Statistics. Wiley
Van der Vaart A, Wellner J (1996) Weak Convergence and Empirical Processes. Springer

Appendix : Proofs of the asymptotic results

Proof (of Proposition 1) As noted just before the statement of this proposition, the process $n^{1/2}(\hat{S}(\cdot) - S(\cdot))$ is constant from τ_{F_1} on. Therefore, it suffices to study this process on $[0, \tau_{F_1}] \cap [0, \infty)$. To this end, we will use the results in Sánchez-Sellero et al. (2005). Although the latter paper considers the case where no cure fraction is present, it can be easily seen that the presence of a cure fraction does not alter the results, since inference is based on survival data available on $[0, \tau_H] \cap [0, \infty)$, so the shape of the survival function $S(t)$ for t larger than τ_H has no impact on the results. In the absence of covariates and when the data are not subject to truncation, Theorem 1 in the latter paper provides an independent and identically distributed representation for Kaplan-Meier integrals of the form $\int \varphi(s)\, d(\hat{S}(s) - S(s))$ uniformly over functions φ belonging to VC-subgraph classes of functions. Consider the family $\{\varphi_t(s) = I(s \leq t) : t \in [0, \tau_{F_1}] \cap [0, \infty)\}$. This is a VC-subgraph class, since the collection of all subgraphs of these functions is easily seen to be the set of all rectangles of the form $[0, t] \times [0, 1]$ with $0 \leq t \leq \tau_{F_1}$, and this is a VC-class of sets, since its VC-index equals 2, so finite. See Van der Vaart and Wellner (1996), page 141, for more details about VC-subgraph classes. Hence, condition $(\varphi 1)$ in Sánchez-Sellero et al. (2005) is satisfied. Condition $(\varphi 2)$ in the latter paper is easily seen to hold since the family $\varphi_t(s)$ is uniformly bounded by the envelope $I(s \leq \tau_{F_1})$ which satisfies the conditions in $(\varphi 2)$ thanks to assumption (A6) and because $\phi < 1$. Similarly, condition (Λ) in Sánchez-Sellero et al. (2005) is satisfied thanks to assumption (A6) and because $\phi < 1$. Since $\int \varphi_t(s)\, d(\hat{S}(s) - S(s)) = \hat{S}(t) - S(t)$, it follows from Theorem 1 in Sánchez-Sellero et al. (2005) that

$$\hat{S}(t) - S(t) = -n^{-1} \sum_{i=1}^{n} \xi(\tilde{T}_i, \delta_i, t) + r_n(t),$$

where $\sup_{t \in [0, \tau_{F_1}] \cap [0, \infty)} |r_n(t)| = o_P(n^{-1/2})$, where the formula of $\xi(\tilde{T}_i, \delta_i, t)$ is obtained after some straightforward algebraic calculations. In particular, this shows that the process $n^{1/2}(\hat{S}(\cdot) - S(\cdot))$ converges weakly in $\ell([0, \tau_{F_1}] \cap [0, \infty))$, since the class $\{(y, \delta) \to \xi(y, \delta, t) : t \in [0, \tau_{F_1}] \cap [0, \infty)\}$ is a Donsker class thanks to Lemma 1 below.

Proof (of Proposition 2) We prove the asymptotic normality of $\hat{\theta}$ by checking the conditions of Theorems 1 and 2 in Chen et al. (2003). Theorem 1 gives conditions under which $\hat{\theta}$ is weakly consistent, which is required for the asymptotic normality that is established in Theorem 2. Let us check the conditions of these two theorems, one by one. Condition (1.1) is satisfied by definition of the estimator $\hat{\theta}$, whereas condition (1.2) holds true thanks to assumption (A2). The continuity of $M(\theta, \phi)$ with respect to ϕ in $\phi = \phi_0$ stated in condition (1.3) is obviously satisfied. For condition (1.4) we know that $\hat{\phi} - \phi_0 = O_P(n^{-1/2}) = o_P(1)$, because of (3) and Proposition 1. Finally, condition (1.5) holds true thanks to (A7). This shows that $\hat{\theta} - \theta_0 \to 0$ in probability.

Next, we verify the conditions of Theorem 2 in Chen et al. (2003). Condition (2.1) is verified by the definition of $\hat{\theta}$. Condition (2.2) is satisfied thanks to assumption (A3). Next, for condition (2.3) note that

$$\begin{aligned}\Gamma(\theta, \phi) &= E\Big[(1-\delta)\frac{\frac{\partial}{\partial\theta}S_{1,\theta}(\tilde{T})(1-\phi+\phi S_{1,\theta}(\tilde{T})) - \phi\frac{\partial}{\partial\theta}S_{1,\theta}(\tilde{T})(-1+S_{1,\theta}(\tilde{T}))}{(1-\phi+\phi S_{1,\theta}(\tilde{T}))^2}\Big] \\ &= E\Big[(1-\delta)\frac{\frac{\partial}{\partial\theta}S_{1,\theta}(\tilde{T})}{(1-\phi+\phi S_{1,\theta}(\tilde{T}))^2}\Big].\end{aligned}$$

Hence, it is easily seen that (2.3)(i) is satisfied, whereas for (2.3)(ii) we have that

$$\|\Gamma(\theta, \phi_0) - \Gamma(\theta_0, \phi_0)\| \, \|\phi - \phi_0\| \leq \|\theta - \theta_0\| \, \|\phi - \phi_0\| = o(1)\delta_n$$

whenever $\|\phi - \phi_0\| \leq \delta_n$ and $\|\theta - \theta_0\| \leq \delta_n$ with $\delta_n \to 0$. Condition (2.4) follows from the fact that $\hat{\phi} - \phi_0 = O_P(n^{-1/2}) = o_P(n^{-1/4})$, whereas condition (2.5) follows from (A7). Finally, for (2.6) note that

$$\begin{aligned}&n^{1/2}\big\{M_n(\theta_0, \phi_0) + \Gamma(\theta_0, \phi_0)(\hat{\phi} - \phi_0)\big\} \\ &= n^{-1/2}\sum_{i=1}^{n}\big\{m(\tilde{T}_i, \delta_i, \theta_0, \phi_0) + \Gamma(\theta_0, \phi_0)\xi(\tilde{T}_i, \delta_i, \tau_{F_1}-)\big\} + o_P(n^{-1/2}),\end{aligned}$$

by Proposition 1, since $\hat{\phi} - \phi_0 = -(\hat{S}(\tau_{F_1}-) - S(\tau_{F_1}-))$, and this converges to a zero-mean normal distribution with variance-covariance matrix given by V. It now follows from the proof of Theorem 2 in Chen et al. (2003) that

$$n^{1/2}(\hat{\theta}-\theta_0) = -\Delta^{-1}n^{-1/2}\sum_{i=1}^{n}\big\{m(\tilde{T}_i, \delta_i, \theta_0, \phi_0)+\Gamma(\theta_0, \phi_0)\xi(\tilde{T}_i, \delta_i, \tau_{F_1}-)\big\}+o_P(n^{-1/2}),$$

which converges in distribution to a normal random variable with mean zero and variance-covariance matrix $\Delta^{-1}V\Delta^{-1}$.

Proof (of Theorem 1) (a) We decompose our process $n^{1/2}\big(\hat{S}_1(\cdot) - S_{1,\hat{\theta}}(\cdot)\big)$ under $\mathcal{H}_0$ as follows :

$$\hat{S}_1(t) - S_{1,\hat{\theta}}(t) = \hat{S}_1(t) - S_1(t) - \Big(\frac{\partial}{\partial\theta}S_{1,\theta_0}(t)\Big)^t\big(\hat{\theta} - \theta_0\big) + o_P(\|\hat{\theta} - \theta_0\|). \quad (5)$$

Note that

$$\begin{aligned}\hat{S}_1(t) - S_1(t) &= \frac{1}{\phi_0}\big(\hat{S}(t) - S(t)\big) + \frac{1-\hat{S}(t)}{\phi_0\hat{\phi}}(\hat{\phi}-\phi_0) \\ &= \frac{1}{\phi_0}\big(\hat{S}(t) - S(t)\big) + \frac{1-S(t)}{\phi_0^2}(\hat{\phi}-\phi_0) + o_P(n^{-1/2}), \qquad (6)\end{aligned}$$

uniformly in $t \in B$, provided $\hat{\phi} - \phi_0 = O_P(n^{-1/2})$ and $\sup_{t\in B}|\hat{S}(t) - S(t)| = O_P(n^{-1/2})$, which follows from Proposition 1 and the continuous mapping theorem. This combined with (5) and the linear representation of $\hat{\theta} - \theta_0$ given in Proposition 2 yields

$$\begin{aligned}\hat{S}_1(t) - S_{1,\hat{\theta}}(t) &= \frac{1}{\phi_0}\big(\hat{S}(t) - S(t)\big) \qquad (7)\\ &+\Big[\frac{1-S(t)}{\phi_0^2} + \Big(\frac{\partial}{\partial\theta}S_{1,\theta_0}(t)\Big)^t \Lambda^{-1}(\theta_0,\phi_0)\Gamma(\theta_0,\phi_0)\Big](\hat{\phi}-\phi_0) \\ &+\Big(\frac{\partial}{\partial\theta}S_{1,\theta_0}(t)\Big)^t \Lambda^{-1}(\theta_0,\phi_0)M_n(\theta_0,\phi_0) + o_P(n^{-1/2}).\end{aligned}$$

The result of part (a) now follows by combining (7) and Propositions 1 and 2.

(b) It suffices to show that the class $\{(y,\delta) \to \eta(y,\delta,t) : t \in B\}$ is Donsker. Since sums of Donsker class are again Donsker (see Example 2.10.7 in Van der Vaart and Wellner (1996)), we need to show that the classes corresponding to each of the three terms in the definition of $\eta(y,\delta,t)$ are Donsker. For the first term we refer to Lemma 1 below, and the fact that the function $\xi(y,\delta,t)$ is constant for $t \geq \tau_{F_1}$. The second and third term are both a product of a bounded function depending on t but not on y and δ (thanks to assumption (A9)), and a function that is independent of t and which has finite variance (thanks to assumptions (A6) and (A8)). Hence, it is easy to see that these classes are also Donsker.

Lemma 1 *Assume (A4)-(A6). Then, the class* $\{(y,\delta) \to \xi(y,\delta,t) : t \in [0,\tau_{F_1}] \cap [0,\infty)\}$ *is a* P*-Donsker class, where* P *is the joint probability measure of* $(\tilde{T},\delta)$.

Proof We consider the two terms of the function $\xi(y,\delta,t)$ separately, defining two subclasses of functions denoted by $\mathcal{F}_1$ and $\mathcal{F}_2$. For $\mathcal{F}_2$, note that $\int_0^{y\wedge t}\frac{dH^1(z)}{(1-H(z))^2}$ is an increasing and bounded function of t thanks to assumption (A6), and hence by Theorem 2.7.5 in Van der Vaart and Wellner (1996), the class $\{y \to \int_0^{y\wedge t}\frac{dH^1(z)}{(1-H(z))^2} : t \in [0,\tau_{F_1}] \cap [0,\infty)\}$ is Donsker. Multiplying these functions by $1-F(t)$ does not alter the Donsker property, since $1-F(t)$ is a deterministic and bounded function. Hence, $\mathcal{F}_2$ is Donsker. Next, for the class $\mathcal{F}_1$, note that

$$E\Big[\frac{I(\tilde{T}\leq\tau_{F_1})\delta}{(1-H(\tilde{T}))^2}\Big] < \infty,$$

thanks to assumption (A6). Hence, for a given $\varepsilon > 0$, we can divide the interval $[0,\tau_{F_1}] \cap [0,\infty)$ into subintervals $[t_j, t_{j+1}]$, $j = 1,\ldots,K\varepsilon^{-2}$, such that

$$E\Big[\frac{I(t_j\leq\tilde{T}\leq t_{j+1})\delta}{(1-H(\tilde{T}))^2}\Big] \leq \varepsilon^2$$

for each j. This shows that the bracketing number $N_{[]}(\varepsilon, \mathcal{F}_1, P)$ is bounded by $K\varepsilon^{-2}$. Moreover, the envelope function $I(\tilde{T} \leq \tau_{F_1})\delta(1-H(\tilde{T}))^{-1}$ has a weak second moment thanks to assumption (A6). Hence, the class $\mathcal{F}_1$ is also Donsker (see Theorem 2.5.6 in Van der Vaart and Wellner (1996)). Since sums of Donsker classes are again Donsker (see Example 2.10.7 in Van der Vaart and Wellner (1996)), the result follows.

Proof (of Corollary 1) The proof relies on the Helly-Bray Theorem (see e.g. p. 97 in Rao (1965)). For more details we refer e.g. to the proof of Corollary 4 in Pardo-Fernández et al. (2007), in which the convergence of a Cramér-von Mises statistic is shown that has the same structure as our statistic.

Proof (of Corollary 2) Under $\mathcal{H}_{1n}$ we consider the following decomposition:

$$\begin{aligned}\hat{S}_1 - \hat{S}_{1,\hat{\theta}} &= (\hat{S}_1 - S_1) + (S_1 - S_{1,\hat{\theta}}) \\ &= (\hat{S}_1 - S_1) + (S_{1,\theta_0} - S_{1,\tilde{\theta}_{0,n}}) + (S_{1,\tilde{\theta}_{0,n}} - S_{1,\hat{\theta}}) + n^{-1/2}c(\tilde{S}_1 - S_{1,\theta_0}) \\ &= T_1 + T_2 + T_3 + T_4,\end{aligned}$$

where $\tilde{\theta}_{0,n}$ maximizes $E_{\mathcal{H}_{1n}}(\log L(\theta, \phi_0))$ with respect to θ, and where $E_{\mathcal{H}_{1n}}$ denotes the expected value under $\mathcal{H}_{1n}$. Hence, $M_{\mathcal{H}_{1n}}(\tilde{\theta}_{0,n}, \phi_0) = 0$, where $M_{\mathcal{H}_{1n}}(\theta, \phi) = E_{\mathcal{H}_{1n}}[m(\tilde{T}, \delta, \theta, \phi)]$. We will now treat the four terms $T_1, \ldots, T_4$ separately.

The term T_1 can be decomposed as follows:

$$T_1(t) = \frac{\hat{\phi} - 1 + \hat{S}(t)}{\hat{\phi}} - \frac{\phi_0 - 1 + S(t)}{\phi_0} = \frac{\hat{S}(t) - S(t)}{\phi_0} + \frac{1 - S(t)}{\phi_0^2}(\hat{\phi} - \phi_0) + o_P(n^{-1/2}),$$

similarly as in the proof of Theorem 1. A linear expansion for the expression $\hat{S}(t) - S(t)$ can be derived in a similar way as in the proof of Proposition 1, except that the data $(\tilde{T}_i, \delta_i)$, $i = 1, \ldots, n$, are now a triangular array. This leads to

$$T_1(t) = n^{-1} \sum_{i=1}^{n} \Big[-\frac{1}{\phi_0} \xi_{\mathcal{H}_{1n}}(\tilde{T}_i, \delta_i, t) + \frac{1 - S(t)}{\phi_0^2} \xi_{\mathcal{H}_{1n}}(\tilde{T}_i, \delta_i, \tau_{F_1} -) \Big] + o_P(n^{-1/2}).$$

Next, for $T_2(t)$ note that $T_2(t) = -\frac{\partial S_{1,\theta_0}(t)}{\partial \theta}(\tilde{\theta}_{0,n} - \theta_0) + o_P(\|\tilde{\theta}_{0,n} - \theta_0\|)$. Note that $0 = M_{\mathcal{H}_{1n}}(\tilde{\theta}_{0,n}, \phi_0) = M_{\mathcal{H}_{1n}}(\theta_0, \phi_0) + (\tilde{\theta}_{0,n} - \theta_0)\frac{\partial M_{\mathcal{H}_{1n}}}{\partial \theta}(\theta_0, \phi_0) + o_P(\|\tilde{\theta}_{0,n} - \theta_0\|)$. Hence, using the notation $H^j_{\mathcal{H}_{1n}}(t) = P_{\mathcal{H}_{1n}}(\tilde{T} \leq t, \delta = j)$, $j = 0, 1$, where

$P_{\mathcal{H}_{1n}}$ denotes the probability under $\mathcal{H}_{1n}$, we have

$$\begin{aligned}
&\tilde{\theta}_{0,n} - \theta_0 \\
&= -\frac{1}{\frac{\partial M_{\mathcal{H}_{1n}}}{\partial \theta}(\theta_0, \phi_0)} \left[M_{\mathcal{H}_{1n}}(\theta_0, \phi_0) + o_P(\|\tilde{\theta}_{0,n} - \theta_0\|) \right] \\
&= -\frac{1}{\frac{\partial M_{\mathcal{H}_{1n}}}{\partial \theta}(\theta_0, \phi_0)} \Big\{ \int m(t, 1, \theta_0, \phi_0)\, dH^1_{\mathcal{H}_{1n}}(t) + \int m(t, 0, \theta_0, \phi_0)\, dH^0_{\mathcal{H}_{1n}}(t) \\
&\qquad + o_P(\|\tilde{\theta}_{0,n} - \theta_0\|) \Big\} \\
&= -\frac{1}{\frac{\partial M_{\mathcal{H}_{1n}}}{\partial \theta}(\theta_0, \phi_0)} \Big\{ -\int m(t, 1, \theta_0, \phi_0)(1 - G(t))\, dS(t) + \int m(t, 0, \theta_0, \phi_0) S(t)\, dG(t) \\
&\qquad + o_P(\|\tilde{\theta}_{0,n} - \theta_0\|) \Big\} \\
&= -\frac{1}{\frac{\partial M_{\mathcal{H}_{1n}}}{\partial \theta}(\theta_0, \phi_0)} \Big\{ E_S[m(\tilde{T}, \delta, \theta_0, \phi_0)] + o_P(\|\tilde{\theta}_{0,n} - \theta_0\|) \Big\}, \qquad (8)
\end{aligned}$$

where $E_S[m(\tilde{T}, \delta, \theta_0, \phi_0)]$ denotes the expected value assuming that the survival function of T equals S, and where

$$\begin{aligned}
S(t) &= S_1(t)\phi_0 + 1 - \phi_0 \\
&= S_{1,\theta_0}(t)\phi_0 + n^{-1/2} c \phi_0 (\tilde{S}_1(t) - S_{1,\theta_0}(t)) + 1 - \phi_0 \\
&= (1 - n^{-1/2} c)[S_{1,\theta_0}(t)\phi_0 + 1 - \phi_0] + n^{-1/2} c [\tilde{S}_1(t)\phi_0 + 1 - \phi_0],
\end{aligned}$$

which we abbreviate by $(1 - n^{-1/2}c) S_{\theta_0}(t) + n^{-1/2} c \tilde{S}(t)$. Noting that $E_{S_{\theta_0}}[m(\tilde{T}, \delta, \theta_0, \phi_0)] = 0$, (8) equals

$$-\frac{n^{-1/2} c}{\frac{\partial M}{\partial \theta}(\theta_0, \phi_0)} E_{\tilde{S}}[m(\tilde{T}, \delta, \theta_0, \phi_0)] + o_P(n^{-1/2}).$$

Let us now consider the term T_3. Following similar steps as in the proof of Proposition 2, it follows that

$$\begin{aligned}
T_3(t) &= -\frac{\partial S_{1,\theta_0}(t)}{\partial \theta}(\hat{\theta} - \tilde{\theta}_{0,n}) + o_P(\|\hat{\theta} - \tilde{\theta}_{0,n}\|) \\
&= \frac{\partial S_{1,\theta_0}(t)}{\partial \theta} \Delta_{\mathcal{H}_{1n}} n^{-1} \sum_{i=1}^{n} \Big\{ m(\tilde{T}_i, \delta_i, \tilde{\theta}_{0,n}, \phi_0) + \Gamma_{\mathcal{H}_{1n}}(\tilde{\theta}_{0,n}, \phi_0) \xi(\tilde{T}_i, \delta_i, \tau_{F_1} -) \Big\} \\
&\quad + o_P(n^{-1/2}),
\end{aligned}$$

where $\Delta_{\mathcal{H}_{1n}}$ is the $p \times p$ matrix of partial derivatives of $M_{\mathcal{H}_{1n}}(\theta, \phi)$ evaluated at $(\tilde{\theta}_{0,n}, \phi_0)$, and $\Gamma_{\mathcal{H}_{1n}}(\tilde{\theta}_{0,n}, \phi_0) = \frac{\partial}{\partial \phi} M_{\mathcal{H}_{1n}}(\tilde{\theta}_{0,n}, \phi_0)$.

Finally, the term T_4 is a bias term, which cannot be simplified further. This shows the result.