Electronic Journal of Statistics
Vol. 18 (2024) 3107–3169
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2249

A two-sample comparison of mean survival times of uncured subpopulations

Dennis Dobler[*1] **and Eni Musta**[*2]

[1] *Department of Mathematics, Vrije Universiteit Amsterdam*
e-mail: dobler@statistik.tu-dortmund.de

[2] *Korteweg-de Vries Institute for Mathematics, University of Amsterdam*
e-mail: e.musta@uva.nl

Abstract: Comparing the survival times among two groups is a common problem in time-to-event analysis, for example if one would like to understand whether one medical treatment is superior to another. In the standard survival analysis setting, there has been a lot of discussion on how to quantify such difference and what can be an intuitive, easily interpretable, summary estimand. In the presence of subjects that are immune to the event of interest ('cured'), we illustrate that it is not appropriate to just compare the overall survival functions. Instead, it is more informative to compare the cure fractions and the survival of the uncured subpopulations separately from each other. Our research is mainly driven by the question: if the cure fraction is similar for two available treatments, how else can we determine which is preferable? To this end, we estimate the mean survival times in the uncured fractions of both treatment groups and develop both permutation and asymptotic tests for inference. We first propose a nonparametric approach which is then extended to account for covariates by means of the semi-parametric logistic-Cox mixture cure model. The methods are illustrated through practical applications to breast cancer and leukemia data.

MSC2020 subject classifications: Primary 62N02, 62G09; secondary 62G20, 62F12.
Keywords and phrases: Asymptotic statistics, cure models, estimand, inference, logistic-Cox mixture model, random permutation, right-censoring.

Received August 2023.

Contents

[*] The authors contributed equally and are given in alphabetical order.

1. Introduction

In many applications, it is of interest to compare survival probabilities among two different samples, e.g., two treatment arms. One common approach is to test for the equality of the survival functions although it does not provide information on the size of the difference. Alternatively, as a graphical tool, one could plot the difference between the two estimated survival curves together with confidence bands. However, in practice it is preferred to have a summary measure of such difference. This facilitates the understanding and interpretation of study results even though it provides limited information since no single metric can capture the entire profile of the difference between two survival curves. The hazard ratio (HR) is commonly used to quantify this difference under the assumption that the ratio of the two hazard functions remains constant over time; for example, the proportionality of hazard rates is central to the famous semiparametric model by Cox [13]. However, such assumption is often not satisfied in practice and the use of the HR would be problematic.

An alternative approach is given by the restricted mean survival time (RMST), which is a popular estimand. The difference between restricted mean survival times for different groups has been advocated as a useful summary measure that offers clinically meaningful interpretation [54, 43, 44, 1, 63]. The RMST is defined as the expected lifetime truncated at a clinically relevant time point τ. The restriction to τ is used to accommodate the limited study duration, as a result of which the upper tail of the survival function cannot be estimated, unless one is willing to assume a specific parametric model for extrapolation beyond the range of the observed data. In observational studies, one can then adjust for imbalances in the baseline covariates between the two groups by using regression-based methods to estimate the RMST; see [11, 12, 1] and references therein. [57] developed a nonparametric estimation procedure for causal estimands in the presence of a cure fraction: the restricted average causal effect and the survival probability causal effect; more on cure models will follow below.

In two-sample problems, one popular class of hypothesis tests arises from the random permutation technique. That is, keeping the data fixed, a random

permutation of the sample group correspondence annihilates the differences between both (permuted) groups. If the two sample groups had the same data generating mechanism, i.e., if the samples were exchangeable, then the score of a test statistic can be deemed one particular realization of the permutation-based test statistic. This explains the property that permutation tests are exact for finite sample sizes in the special case of exchangeability. Recently, [22] and [58] investigated a random permutation method for inference on the difference in restricted mean (net) survival times. While their test is finitely exact under exchangeable data, [22] stated for the case of non-exchangeable data that "Further research to develop methods for constructing confidence intervals for RMST difference with a small sample data is warranted. It is quite challenging to construct an exact confidence interval for the difference in RMST." [15] continued in the direction of this remark and analyzed a studentized permutation version of the just-mentioned approach. Their resulting hypothesis test is exact under exchangeability and it even controls the type-I error probability asymptotically under non-exchangeability. Because of this additional feature of exactness under exchangeability, permutation tests also enjoy great popularity in survival analytic applications beyond the RMST: for instance, [4] and [14] researched permutation-based weighted log-rank tests.

In this paper, we will consider the not unusual case that some of the subjects are immune to the event of interest ('cured') instead of the classical survival problem. The challenge arises because, in absence of additional information, the cured subjects (for which the event never takes place) cannot be distinguished from the susceptible ones as a result of censoring. Cure rate models, which account for the presence of a cured subpopulation have become increasingly popular particularly for analyzing cancer survival and evaluating the curative effect of treatments. More in general, they have found applications in many other domains including fertility studies, credit scoring, demographic and social studies analyzing among other things time until marriage, time until rearrest of released prisoners, time until one starts smoking. For a complete review on cure models methodology and applications, we refer the reader to [2, 42, 27]. In presence of immune subjects, comparing survival between two samples becomes more complicated than in the standard setting since one can compare overall survivals, cure chances, and survival probabilities for the uncured subjects. Several papers have focused on testing for differences in the cure rates in a nonparametric setting. [25, 21, 48, 26] On the other hand, different methods have been proposed to test for equality of the survival functions among the uncured subpopulations. [29, 51, 64, 7, 5]

However, as in the standard survival analysis setting, testing for equality of the two distributions, i.e., with the aim to detect any difference in the two distributions, would not be sufficient for many practical applications. Thus, apart from comparing the cure probabilities, it would be meaningful to compare relevant estimands for the subpopulation of uncured subjects. To this end, we propose to analyze mean survival times of the uncured. This has the advantage of being an easy-to-interpret extension of the RMST in the present context, whereas we will not impose a time restriction. Hence, we will use the abbrevia-

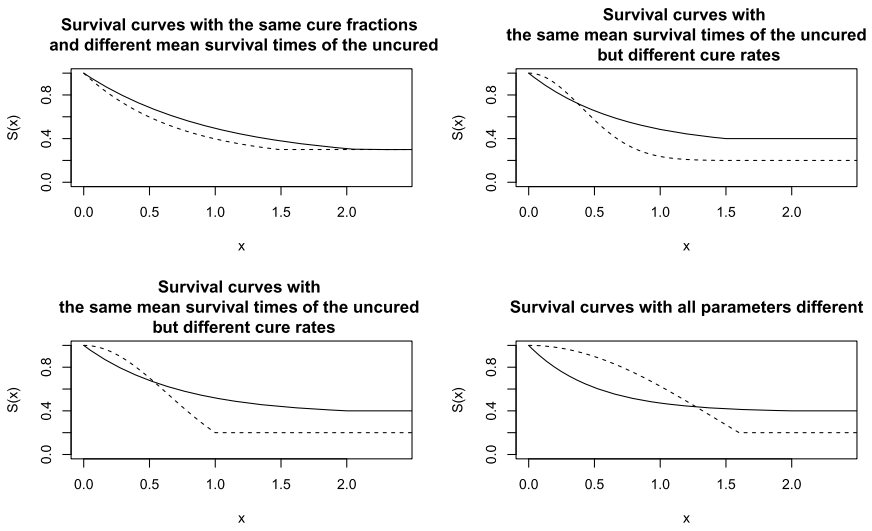

FIG 1. *Different constellations of survival curves in two-sample problems.*

tion MST for our purposes. To illustrate the different concepts in the context of survival models with cure fractions, see Figure 1. It illustrates different scenarios of different or equal cure fractions p, time points τ_0 when the plateau is reached, and mean survival times for the uncured patients. Recently, a nonparametric estimation procedure for the restricted average causal effect was developed in [57]. Differently from our approach, they did not condition on the subpopulation of the uncured individuals but instead consider the overall restricted mean survival time of the general population and focus on a propensity score adjustment for observational data. We argue that in the presence of a cure fraction, even in the context of a randomized clinical trial, it is more informative to compare separately the cure rates and the survival of the uncured. However, we discuss that conditioning on the uncured subpopulation makes our estimand difficult to interpret in a causal framework.

Measures other than the first moment could obviously also be used to summarize the survival curves of the uncured patients, e.g., the median or other quantiles of these proper survival functions, or other moments. In our opinion, however, the mean offers the easiest interpretation: how much (in absolute numbers) of the wholly available area of τ_0 is below the survival curve? The more, the better. This offers another means of comparing the usefulness of two (or more) treatments: which treatment prolongs the mean survival times most effectively, next to comparing the cure fractions?

The agenda for the present paper and the benefits of the proposed procedure are as follows:

- We will propose an estimand for comparing the lifetimes of the uncured subjects via mean survival times.
- This will allow for comparisons in two-sample problems.

- Restrictions of time will not be necessary.
- Only weak assumptions, e.g., for the sake of identifiability will be made.
- In particular, hazard rates are generally allowed to be discontinuous.
- Inference will be based on the random permutation method which gives rise to finitely exact hypothesis tests in the case of exchangeable samples and, otherwise, good small sample properties.

We start by considering a nonparametric model with a constant cure rate and then extend the method to a semiparametric mixture cure model that allows for expressions of mean survival times conditionally on covariates. For the latter, we will assume a logistic-Cox mixture cure model [50, 40] since it is the most widely used in practice. In particular, a logistic model is assumed for the cure probabilities and a Cox proportional hazards (PH) model for the survival times of the uncured subjects. Note, however, that the baseline hazards for the uncured subpopulations in the two groups are in general different, leading to nonproportional hazards. Despite this model choice, the estimation procedure and the results could be similarly extended to other semiparametric mixture cure models, e.g. [28, 62, 32]. Recently, the problem of estimating the conditional mean survival time for the uncured in a one-sample context has been considered in [10]. They assume a semiparametric proportional model for the mean (residual) life of the uncured and propose an estimation method via inverse-probability-of-censoring weighting and estimating equations. As a consequence, their approach also requires estimation of the censoring distribution.

In the nonparametric setting, it has been observed that the permutation approach improves upon the asymptotic method for small sample sizes, while maintaining good behavior asymptotically [22, 58, 15]. To the best of our knowledge, the permutation approach has not been used before for semiparametric models in a similar context of maximum likelihood-based estimators that we will pursue. One notable permutation-based inference approach in the survival literature concerns the weighted logrank test [4]; there, the semiparametric model arises from the form of the null hypothesis which is a cone or subspace of hazard derivatives.

Given also the complexity of our model, several challenges arise from both the computational and theoretical point of view. In order to obtain results on the asymptotic validity of the permutation approach, we first derive a general Donsker-type theorem for permutation based Z-estimators. Secondly, when fitting the model to the permuted sample, there is an issue of model misspecification. This leads to the convergence of the estimators to a minimizer of the Kullback–Leibler divergence. Thirdly, since the variance estimators need to be computed via a bootstrap procedure, the combination of bootstrap and permutation becomes computationally expensive. Hence, it is of interest to investigate whether the gain in accuracy is sufficient to compensate for the increased computation cost compared to the asymptotic approach.

This article is organized as follows. In Section 2, we focus on the simpler unconditional nonparametric approach describing the model, estimand, estimation procedure and the random permutation scheme. We present large sample

properties, which are crucial for inference, and also offer a toy example for a discussion about comparisons of survival times in the two-sample problem. In Section 3, we propose an estimator in a semi-parametric conditional model for the difference in conditional mean survival times of the uncured and derive its asymptotic distribution. The random permutation approach for inference is extended to this setting and its asymptotic validity is justified. Section 4 contains a description of an extensive simulation study as well as the numerical results. Data from studies of leukemia and breast cancer are illustrated and analyzed in the light of the proposed methods in Section 5. We conclude with a discussion in Section 6. All proofs are contained in Appendices A–C. The R code with an implementation of our methods is available in the GitHub repository `https://github.com/eni-musta/MST_uncured`.

2. Unconditional model

2.1. Model and notation

We consider i.i.d. survival times $T_{11}, \dots, T_{1n_1}$ and $T_{21}, \dots, T_{2n_2}$ from two independent groups ($i = 1, 2$) that consist of a mixture of cured and uncured individuals, meaning that a fraction of the study population in each group (the cured ones) would not experience the event of interest. We denote the event time of the cured individuals by ∞ and assume that, for the uncured ones, the event can happen on the interval $[0, \tau_{0,i}]$, $i = 1, 2$, respectively for each group. We do not assume the cure threshold $\tau_{0,i} < \infty$ to be known in advance but depending on the application at hand one might have some information about it; for example in oncology based on the medical knowledge $\tau_{0,i}$ is expected to be somewhere between 5 or 10 years depending on the cancer type.

For the remainder of this paper, we denote by $(\Omega, \mathcal{A}, \mathbb{P})$ the underlying probability space, $\mathbb{E}$ denotes expectation, $\xrightarrow{d}$ denotes convergence in distribution, $\overset{d}{=}$ denotes equality in distribution, and $\xrightarrow{p}$ denotes convergence in probability.

Let us denote by F_i and $S_i = 1 - F_i$, $i = 1, 2$, the improper cumulative distribution function and the improper survival function of the time-to-event variables in the two treatment groups. Let $F_{u,i}$ and $S_{u,i} = 1 - F_{u,i}$ be the proper cumulative distribution function and survival function for the uncured individuals in the two treatment groups. Let

$$p_i = \mathbb{P}(T_{i1} > \tau_{0,i}) = S_i(\tau_{0,i}) \in (0, 1)$$

denote the cure fractions in both groups. In particular, we have

$$F_i(t) = (1 - p_i)F_{u,i}(t), \qquad S_i(t) = p_i + (1 - p_i)S_{u,i}(t). \tag{1}$$

In the presence of right censoring, instead of the survival times, we observe the follow-up times $Y_{i1}, \dots, Y_{in_i}$ and the censoring indicators $\Delta_{i1}, \dots, \Delta_{in_i}$, where $Y_{ij} = \min\{T_{ij}, C_{ij}\}$, $\Delta_{ij} = \mathbb{1}_{\{T_{ij} \le C_{ij}\}}$ and C_{ij} are the censoring times. We assume that censoring is independent of the survival times and has bounded

support $[0, \tau_i]$ in each group. In particular, because of the finite censoring times, all the cured individuals will be observed as censored. In order to be able to identify the cure fraction, we need $\tau_{0,i} \leq \tau_i$ and F_i continuous at τ_i in case $\tau_{0,i} = \tau_i$, which is known as the sufficient follow-up assumption [34]. The idea is that, since the cure status is not observed and $F_{u,i}(\cdot)$ is left unspecified, if $\tau_i < \tau_{0,i}$ then the events $\{T_{i1} \in (\tau_i, \tau_{0,i}]\}$ and $\{T_{i1} = \infty\}$ would be indistinguishable. As a result, the cure rate could not be identified. A statistical test for this assumption is proposed in [34] but its practical behavior is not very satisfactory given the unstable behaviour of the Kaplan–Meier estimator in the tail region. In practice, a long plateau of the Kaplan–Meier estimator, containing many censored observations, is considered to be an indication of sufficient follow-up.

Comparison of overall survival

We first illustrate why comparing overall survival functions is not appropriate in the presence of a cure fraction. The difference in overall survival combines together the difference in cure fractions and in the survival times of the uncured in a way that it is difficult to interpret. For example, if group one has a higher cure fraction but lower survival times for the uncured, the overall survival functions might cross and the difference between them would be a weighted combination of the two effects:

$$S_1(t) - S_2(t) = (p_1 - p_2)\{1 - S_{u,2}(t)\} + (1 - p_1)\{S_{u,1}(t) - S_{u,2}(t)\}.$$

On the other hand, if the two groups have the same cure fraction p, then

$$S_1(t) - S_2(t) = (1 - p)\{S_{u,1}(t) - S_{u,2}(t)\}$$

This means that, particularly for a large cure fraction, the observed difference in overall survival functions is much smaller than the actual difference of the survival functions for the uncured.

Using a one number summary of the difference in overall survival is even more problematic in the presence of a cure fraction. First, the proportional hazard assumption is clearly violated on the level of the whole population and, as a result, the hazard ratio cannot be used. For the Mann–Whitney effect, what counts are the chances of having longer survival times for one group compared to the other, but the actual difference between these times does not matter. So one cannot distinguish between having a larger cure fraction or just slightly longer survival.

Consider for example the following hypothetical scenario: patients receiving treatment A have a 20% chance of being cured while with treatment B there is no cure chance; a random person receiving treatment B lives several months longer compared to an uncured patient who received treatment A. Let T_1 and T_2 represent the random lifetimes of patients receiving treatments A and B, respectively. According to the above description, the Mann–Whitney effect would then be

$$\mathbb{P}(T_1 > T_2) + \frac{1}{2}\mathbb{P}(T_1 = T_2) = 0.2 < 0.5,$$

leading to the conclusion that treatment B should be preferred. This is counter-intuitive because, given the small difference in survival times of the uncured, in practice one would probably prefer the treatment that offers some chance of getting cured. On the other hand, if one would use the restricted mean survival time as an estimand, the actual survival times matter. However, because of the restriction to a specified point τ (duration of the study), there would still be no distinction between the cured individuals and those who survive more than τ:

$$RMST_{\tau}^{(i)} = \mathbb{E}[\min(T_i, \tau)] = (1 - p_i)\mathbb{E}[\min(T_i, \tau)|T_i < \tau] + \tau p_i, \quad i = 1, 2.$$

To illustrate this, consider the following example: patients receiving treatment A have 20% chance of being cured, while with treatment B there is no cure chance; a random person receiving treatment B lives on average 60 months, while an uncured patient who received treatment A lives on average 24 months. Let us assume that $\tau_{0,1} = \tau_{0,2} = 120$ months. If the duration of the study was also 120 months (sufficient follow-up), we would obtain

$$RMST_{120}^{(1)} - RMST_{120}^{(2)} = 0.8 \cdot \mathbb{E}[T_1|T_1 < 120] + 120 \cdot 0.2 - \mathbb{E}[T_2] = -16.8,$$

leading to the conclusion that treatment B should be preferred. However, if the study had continued for longer, e.g., 240 months, we would obtain

$$RMST_{240}^{(1)} - RMST_{240}^{(2)} = 0.8 \cdot \mathbb{E}[T_1|T_1 < 240] + 240 \cdot 0.2 - \mathbb{E}[T_2] = 7.2,$$

suggesting that treatment A is better, which contradicts the previous conclusion.

For these reasons, we think that in the presence of a cure fraction, it is more informative to compare separately the cure fractions and the survival functions of the uncured. In practice, one can then make a personalized decision by choosing to put more weight to one component compared to the other, based on the life expectancy if uncured and the risks one is willing to take. For example, for children there is an essential difference between cure and 10 year survival, while such difference might be less significant for elderly patients.

Mean survival time of the uncured

The problem of comparing cure fractions has already been considered in the literature. Here, we focus on comparing the survival times of the uncured. In particular, we propose the mean survival time as a summary estimand.

We are interested in the difference of mean survival times of the uncured individuals among the two groups,

$$MST_{u,1} - MST_{u,2} = \mathbb{E}[T_{11} \mid T_{11} < \infty] - \mathbb{E}[T_{21} \mid T_{21} < \infty].$$

In combination with the cure fractions, such mean survival times provide useful summaries of the improper survival curves. Using the relations in (1), we obtain the following expression for the mean survival times:

$$MST_{u,i} = \int_0^{\tau_{0,i}} S_{u,i}(s)\,\mathrm{d}s = \int_0^{\tau_{0,i}} \frac{S_i(s) - p_i}{1 - p_i}\,\mathrm{d}s, \qquad i = 1, 2.$$

2.2. Estimation, asymptotics, and random permutation

Estimators and their large sample properties

Estimation of the cure rate and of the nonparametric survival function in this setting has been considered in [33, 34] and is based on the Kaplan–Meier (KM) estimator. In particular, we can estimate S_i and p_i by

$$\hat{S}_i(t) = \prod_{s\in(0,t]} \left(1 - \frac{\mathrm{d}N_i(s)}{R_i(s)}\right) \quad \text{and} \quad \hat{p}_i = \hat{S}_i(Y_{i,(m_i)}),$$

respectively, where $Y_{i,(m_i)}$ is the largest observed event time in group i; here, m_i is the total number of observed events in group i and the subscript (m_i) indicates the m_i-th order statistic of the observable event times. Furthermore, in the display above, the counting processes $N_i(s) = \sum_{j=1}^{n_i} \mathbb{1}_{\{Y_{ij}\le s, \Delta_{ij}=1\}}$ count the number of observed events up to time s and $R_i(s) = \sum_{j=1}^{n_i} \mathbb{1}_{\{Y_{ij}\ge s\}}$ count the numbers of individuals at risk at time s. By a plug-in method, we estimate the mean survival time of the uncured by

$$\widehat{MST}_{u,i} = \int_0^{Y_{i,(m_i)}} \frac{\hat{S}_i(s) - \hat{p}_i}{1 - \hat{p}_i}\,\mathrm{d}s$$

Let $\hat{m} = \widehat{MST}_{u,1} - \widehat{MST}_{u,2}$ be an estimator of $m = MST_{u,1} - MST_{u,2}$. Using the asymptotic properties of the KM estimator, we obtain the following result for the mean survival times, for which we define

$$v_i(t) = \int_0^t \frac{\mathrm{d}F_i(s)}{S_i(s)\{1 - H_i(s-)\}}, \tag{2}$$

$H_i(t) = \mathbb{P}(Y_{i1} \le t) = \{1 - F_i(t)\}\{1 - G_i(t)\}$, and $G_i(t) = \mathbb{P}(C_{i1} \le t)$ denotes the distribution of the censoring times.

Theorem 2.1. *For $i = 1, 2$, assume $p_i \in (0, 1)$ and that one of the following conditions holds:*

a) $\tau_{0,i} < \tau_i$
b) $\tau_{0,i} = \tau_i$, F_i *is continuous and*

$$\int_0^{\tau_i} \frac{\mathrm{d}F_i(t)}{1 - G_i(t-)} < \infty; \tag{3}$$

c) $\tau_{0,i} = \tau_i$, F_i *is continuous at* τ_i, $\lim_{t\uparrow\tau_i}\{F_i(\tau_i) - F_i(t)\}^2 v_i(t) = 0$, *and*

$$\lim_{t\uparrow\tau_i} \int_t^{\tau_i} \frac{\mathbb{1}_{\{0\le G_i(s-)<1\}} S_i(s)}{\{1 - G_i(s-)\}S_i(s-)}\,\mathrm{d}F_i(s) = 0.$$

Then the random variable $\sqrt{n_i}(\widehat{MST}_{u,i} - MST_{u,i})$ is asymptotically normally distributed, i.e.

$$\sqrt{n_i}(\widehat{MST}_{u,i} - MST_{u,i}) \xrightarrow{d} N_i \sim \mathcal{N}(0, \sigma_i^2)$$

as $n \to \infty$. The limit variance σ_i^2 is defined in (10) in the appendix.

The extra technical conditions in b) and c) of the previous theorem are the conditions needed to obtain weak convergence of the normalized Kaplan–Meier estimator to a Gaussian process [20, 61]. Note that in the particular case $\tau_{0,i} < \tau_i$, the conditions unrelated to the continuity of F_i are automatically satisfied, leading to no extra requirements for case a). Continuity of F_i is nowhere needed in case a).

Remark 1. Theorem 2.1 still holds in the case of $p_i = 0$, i.e., in the absence of a cure fraction. In this case, the estimand reduces to the common (unrestricted) mean survival time. However, we have not made this explicit in the statement of the theorem because we generally assume the presence of a cure fraction which should, in practice, be confirmed by medical expert knowledge.

From the independence assumption between the two groups and Theorem 2.1, we obtain the following result for which we define $a_n = \sqrt{n_1 n_2/(n_1+n_2)}$.

Corollary 2.1.1. *Assume that $n_1/(n_1+n_2) \to \kappa \in (0,1)$ as $\min(n_1, n_2) \to \infty$. Under any of the two conditions in Theorem 1, we have that $a_n(\hat{m} - m)$ is asymptotically normally distributed with mean zero and variance*

$$\sigma^2 = (1-\kappa)\sigma_1^2 + \kappa\sigma_2^2.$$

The canonical plug-in estimator of σ^2, say

$$\hat{\sigma}^2 = \frac{n_2}{n_1+n_2}\hat{\sigma}_1^2 + \frac{n_1}{n_1+n_2}\hat{\sigma}_2^2 \tag{4}$$

is obviously consistent. Hence, the combination of Theorem 2.1 and Corollary 2.1.1 could be used to justify inference methods for $MST_{u,1} - MST_{u,2}$ based on the asymptotic normal approximation. However, such inference procedures can usually be made more reliable by means of resampling methods.

Inference via random permutation across samples

We propose random permutation to construct inference methods for m. To introduce this procedure, let $\pi = (\pi_1, \ldots, \pi_{n_1+n_2})$ be any permutation of $(1, 2, \ldots, n_1+n_2)$. When applied to the pooled sample, say, $(Y_1, \Delta_1), \ldots, (Y_{n_1+n_2}, \Delta_{n_1+n_2})$, this permutation leads to the permuted samples $(Y_{\pi_1}, \Delta_{\pi_1}), \ldots, (Y_{\pi_{n_1}}, \Delta_{\pi_{n_1}})$ and $(Y_{\pi_{n_1+1}}, \Delta_{\pi_{n_1+1}}), \ldots, (Y_{\pi_{n_1+n_2}}, \Delta_{\pi_{n_1+n_2}})$.

In the special case of exchangeability, i.e. $H_0^{exch.}: (Y_{1j}, \Delta_{1j}) \stackrel{d}{=} (Y_{2j}, \Delta_{2j})$, $\hat{m} = \widehat{MST}_{u,1} - \widehat{MST}_{u,2}$ would have the same distribution as $\hat{m}^\pi = \widehat{MST}^\pi_{u,1} - \widehat{MST}^\pi_{u,2}$. Here, $\widehat{MST}^\pi_{u,i}$ are the estimators of the mean survival times, just based on the i-th permuted sample. So, under a sharp null hypothesis $H_0^{exch.}$ of exchangeability, a test for the equality of mean survival times would reject the null hypothesis if $\hat{m}$ belongs to the $\alpha \times 100\%$ most extreme values of $\hat{m}^\pi$ across all $(n_1+n_2)!$ permutations.

However, under the weak null hypothesis of equal mean survival times, $H_0: m = 0$, the samples are in general not exchangeable. As a consequence, the

asymptotic variances of $\hat{m}$ and $\hat{m}^\pi$ cannot be assumed equal and hence they must be studentized. We will thus focus on $\hat{m}^\pi/\hat{\sigma}^\pi$ as the permutation version of $\hat{m}/\hat{\sigma}$, where

$$\hat{\sigma}^{\pi 2} = \frac{n_2}{n_1+n_2}\hat{\sigma}_1^{\pi 2} + \frac{n_1}{n_1+n_2}\hat{\sigma}_2^{\pi 2}, \tag{5}$$

and $\sigma_i^{\pi 2}$ is the plug-in variance estimator based on the i-th permuted sample. Consequently, our aim is to compare $\hat{m}/\hat{\sigma}$ to the conditional distribution of $\hat{m}^\pi/\hat{\sigma}^\pi$ given the data to reach a test conclusion.

Because it is computationally infeasible to realize $\hat{m}^\pi/\hat{\sigma}^\pi$ for all $(n_1+n_2)!$ permutations, we will realize a relatively large number B of random permutations π and approximate the conditional distribution by the collection of the realized $\hat{m}^\pi/\hat{\sigma}^\pi$ of size B.

In the following, we will discuss the asymptotic behaviour of $\hat{m}^\pi/\hat{\sigma}^\pi$ to justify the validity of the resulting inference procedures. From now on, we understand the weak convergence of conditional distributions in probability as the convergence of these distributions to another with respect to the bounded Lipschitz metric in probability; see e.g. Theorem 1.12.4 in [55].

Theorem 2.2. *Assume that $\tau_{0,i} < \tau_i$ and $p_i \in (0,1)$ for $i = 1,2$. Then, as $n_1, n_2 \to \infty$ with $n_1/(n_1+n_2) \to \kappa \in (0,1)$, the conditional distribution of $a_n\hat{m}^\pi$ given the data converges weakly in probability to the zero-mean normal distribution with variance given in (11) in the appendix.*

Under continuity assumptions similar to those in Theorem 2.1 and additional assumptions on the censoring distributions, we conjecture that a similar weak convergence result holds for the case of $\tau_{0,i} = \tau_i$. This could potentially be shown by extending the results of [16] to the random permutation method instead of the classical bootstrap. Deriving such results, however, is beyond the scope of the present paper.

The structure of the asymptotic variance in the previous theorem motivates a canonical permutation-type variance estimator $\hat{\sigma}^{\pi 2}$, that is, the plug-in estimator based on the pooled sample. Due to the obvious consistency of this estimator, we arrive at the following main result on the permuted studentized mean survival time:

Corollary 2.2.1. *Assume that $\tau_{0,i} < \tau_i$ and $p_i \in (0,1)$ and F_i is continuous for $i = 1,2$. Then, as $n_1, n_2 \to \infty$ with $0 < \liminf n_1/(n_1+n_2) \leq \limsup n_1/(n_1+n_2) < 1$, the (conditional) distributions of $a_n\hat{m}^\pi/\hat{\sigma}^\pi$ (given the data) and $a_n(\hat{m}-m)/\hat{\sigma}$ converge weakly in probability to the same limit distribution which is standard normal.*

We conclude this section with a remark on the inference procedures deduced from the random permutation approach.

Remark 2. Corollary 2.2.1 gives rise to asymptotically exact 1- and 2-sided tests for the null hypotheses $H_0^{(1)} : m \leq 0$, $H_0^{(2)} : m \geq 0$, or $H_0^{(3)} : m = 0$ against the respective complementary alternative hypotheses: comparing $\hat{m}/\hat{\sigma}$ with data-dependent critical value(s) obtained from the collection of realized

$\hat{m}^\pi/\hat{\sigma}^\pi$ (for fixed data) allows for controlling the chosen significance level $\alpha \in (0,1)$ as $n_1, n_2 \to \infty$ under the assumptions made above. A similar remark holds true for more general null hypotheses in which m is compared to some hypothetical value $m_0 \in \mathbb{R}$. In addition, as a well-known property of permutation tests based on studentized test statistics, the just-mentioned tests are exact in the special case of exchangeability between both sample groups for which $m_0 = 0$ is automatically fulfilled. Similarly, by inverting hypothesis tests into confidence intervals, asymptotically exact confidence intervals for m can be constructed; see the Section 4.1 for details.

3. Conditional model

3.1. Model and notation

Assume now that in addition to the follow-up times $Y_{i1}, \dots, Y_{in_i}$ and the censoring indicators $\Delta_{i1}, \dots, \Delta_{in_i}$, we also observe two covariate vectors $X_{ij} \in \mathbb{R}^p$ and $Z_{ij} \in \mathbb{R}^q$, $i = 1, 2$, $j = 1, \dots, n_i$, representing the variables that affect the probability of being susceptible (p_i; incidence) and the survival of the uncured ($S_{u,i}$; latency), respectively. In this way, we allow for these two components of the model to be affected by different variables. However, we do not exclude situations in which the two vectors X and Z are exactly the same or share some components.

Using the framework of mixture cure models, the relations in (1) now hold conditionally on the covariates. In particular, we have that the survival function of T_{i1} given X_{i1} and Z_{i1} is given by

$$S_i(t|x,z) = \mathbb{P}(T_{i1} > t | X_{i1} = x, Z_{i1} = z) = p_i(x) + (1 - p_i(x)) S_{u,i}(t|z), \qquad (6)$$

where $S_{u,i}(t|z) = \mathbb{P}(T_{i1} > t | Z_{i1} = z, T_{i1} < \infty)$ denotes the conditional survival function of the susceptibles, which is independent of X given Z, and $p_i(x) = \mathbb{P}(T_{i1} = \infty | X_{i1} = x)$ denotes the conditional cure probability in group i, which is independent of Z given X. Instead of independent censoring, now we assume that censoring is independent of the survival times conditionally on the covariates: $T \perp C \mid (X, Z)$.

Among various modeling approaches for the incidence and the latency, the most common choice in practice is a parametric model, such as logistic regression, for the incidence and a semiparametric model, such as Cox proportional hazards, for the latency [27, 60, 49, 59]. The popularity of such choice is primarily due to simplicity and interpretability, particularly when dealing with multiple covariates. We focus on this type of models and assume that

$$1 - p_i(x) = \phi(\gamma_i^T x),$$

where $\phi : \mathbb{R} \to [0,1]$ is a known function, $\gamma_i \in \mathbb{R}^{p+1}$ and γ_i^T denotes the transpose of the vector γ_i. Here, the first component of x is taken to be equal

to one and the first component of γ_i corresponds to the intercept. In particular, for the logistic model, we have

$$\phi(u) = \frac{e^u}{1+e^u}. \tag{7}$$

One can in principle allow also for a different function ϕ in the two groups but for simplicity we assume that to be the same. For the latency, we assume a semiparametric model $S_{u,i}(t|z) = S_{u,i}(t|z;\beta_i,\Lambda_i)$ depending on a finite-dimensional parameter $\beta_i \in \mathbb{R}^q$, and a function Λ_i. For example, for the Cox proportional hazards model, we have

$$S_{u,i}(t|z) = \exp\{-\Lambda_i(t)\exp(\beta_i^T z)\}, \tag{8}$$

where Λ_i is the baseline cumulative hazard in group i.

One challenge with mixture cure models is model identifiability, i.e., ensuring that different parameter values lead to different distributions of the observed variables. General identifiability conditions for semiparametric mixture cure models were derived by [38]. In the particular case of the logistic-Cox model the conditions are:

(I1) for all x, $0 < \phi(\gamma_i^T x) < 1$,
(I2) the function $S_{u,i}$ has support $[0;\tau_{0,i}]$ for some $\tau_{0,i} < \infty$,
(I3) $P(C_{i1} > \tau_{0,i}|X_{i1};Z_{i1}) > 0$ for almost all X_{i1} and $Z_{i,1}$,
(I4) the matrices $Var(X_{i1})$ and $Var(Z_{i1})$ are positive definite.

Condition I3 corresponds again to the assumption of sufficient follow-up. In practice, this can be evaluated based on the plateau of the Kaplan–Meier estimator and the expert (medical) knowledge.

In the presence of covariate information, we are now interested in the difference of mean survival times of the uncured individuals among the two groups conditional on the covariate values:

$$MST_{u,1,z} - MST_{u,2,z} = \mathbb{E}[T_{11} \mid T_{11} < \infty, Z_{11} = z] - \mathbb{E}[T_{21} \mid T_{21} < \infty, Z_{21} = z].$$

Note that we use only the covariate Z because that affects the survival of the uncured individuals. In combination with the conditional cure probabilities $p_i(x)$, such conditional mean survival times provide useful summaries of the conditional survival curves.

3.2. Estimation, asymptotics, and random permutation

The conditional mean survival time can be written as

$$MST_{u,i,z} = \int_0^{\tau_{0,i}} S_{u,i}(s|z)\,\mathrm{d}s \qquad i = 1,2.$$

This leads to the following estimator:

$$\widehat{MST}_{u,i,z} = \int_0^{Y_{i,(m_i)}} \hat{S}_{u,i}(s|z)\,\mathrm{d}s,$$

where $\hat{S}_{u,i}(\cdot|z)$ is an estimate of the conditional survival function for the uncured and $Y_{i,(m_i)}$ is the largest observed event time in group i.

Next we focus on the logistic-Cox mixture model, given by (7)–(8), and consider the plug-in estimate

$$\hat{S}_{u,i}(t|z) = \exp\left(-\hat{\Lambda}_i(t)e^{\hat{\beta}_i^T z}\right),$$

where $\hat{\Lambda}_i$, $\hat{\beta}_i$ are the maximum likelihood estimates of Λ_i and β_i respectively. Maximum likelihood estimation in the logistic-Cox model was initially proposed by [50, 40] and is carried out via the EM algorithm. The procedure is implemented in the R package `smcure` [9]. In practice, the survival $\hat{S}_{u,i}(t|z)$ is forced to be equal to zero beyond the last event $Y_{i,(m_i)}$, meaning that the observations in the plateau are considered as cured. This is known as the zero-tail constraint as suggested in [50, 52] and is reasonable under the assumption of sufficient follow-up: $\tau_{0,i} < \tau_i$, which follows from (I3).

The asymptotic properties of the maximum likelihood estimates $\hat{\Lambda}_i$, $\hat{\beta}_i$, $\hat{\gamma}_i$ were derived in [31] under the following assumptions:

(A1) The function $\Lambda_i(t)$ is strictly increasing, continuously differentiable on $[0, \tau_{0,i})$ and $\Lambda_i(\tau_{0,i}) := \lim_{t\to\tau_{0,i}} \Lambda_i(t) < \infty$.
(A2) γ_i, β_i lie in the interiors of compact sets and the covariate vectors Z_{ij} and X_{ij} have compact support: there exist $K_i > 0$ such that: $\mathbb{P}(\|Z_{ij}\| < K_i \text{ and } \|X_{ij}\| < K_i) = 1$.
(A3) There exists a constant $\epsilon > 0$ such that $\mathbb{P}(T_{i1} = \tau_{0,i} \mid T_{i1} < \infty, Z_{i1}) > \epsilon$ with probability one.
(A4) $\mathbb{P}(Y_{i1} \geq t \mid Z_{i1}, X_{i1})$ is continuous in $t \leq \tau_{0,i}$.

Assumptions (A1),(A3) are formulated slightly in a different way in [31] but, given the identifiability constraints (I1)–(I4), they reduce to the ones stated above. In particular, assuming that the survival distribution for the uncured has a positive mass at the end point of the support (A3) is a technical condition needed to guarantee that Λ_i stays bounded on $[0, \tau_{0,i}]$ while ensuring the identifiability of the model. In the Cox model without cure fraction, one does not encounter this problem because the support of the event times is larger than the follow-up of the study. Even though (A3) might seem not realistic, one can think of such assumption being satisfied with a very small ϵ. In such case it is unlikely to observe events at $\tau_{0,i}$ as we see in real-life scenarios. If instead of the maximum likelihood estimation, one considers estimation via presmoothing as proposed in [37], this condition can be avoided at the price of additional technicalities. This is because the conditional probability of $\{T = \infty\}$ is identified beforehand by means of a nonparametric smooth estimator. As a result, when estimating Λ_i in the second step, one could restrict to a smaller interval $[0, \tau^*] \subset [0, \tau_{0,i}]$; see the discussion in Section 5.1 of [37].

Let $\hat{m}_z = \widehat{MST}_{u,1,z} - \widehat{MST}_{u,2,z}$ be an estimator of $m_z = MST_{u,1,z} - MST_{u,2,z}$. Using the large sample properties of the estimators $\hat{\Lambda}_i$ and $\hat{\beta}_i$ from [31], we first derive the limit distribution of the process $\sqrt{n_i}\{\hat{S}_{u,i}(\cdot|z) - S_{u,i}(\cdot|z)\}$ and then obtain the following result for the conditional mean survival time.

Theorem 3.1. *Assume that the identifiability conditions (I1)–(I4) and the assumptions (A1)–(A4) are satisfied. Then, for any $z \in \mathcal{Z}$, the random variable $\sqrt{n_i}(\hat{E}_{i,z} - E_{i,z})$ is asymptotically normally distributed, i.e.,*

$$\sqrt{n_i}(\widehat{MST}_{u,i,z} - MST_{u,i,z}) \xrightarrow{d} N_i \sim \mathcal{N}(0, \sigma^2_{i,z})$$

as $n_i \to \infty$. The limit variance $\sigma^2_{i,z}$ is defined in (18) in the Appendix B.

Remark 3. Based on the proof of Theorem 3.1, we cannot guarantee that the convergence result still holds under the absence of a cure fraction, i.e., $p_i \equiv 0$, $i = 1, 2$.

From the independence assumption between the two groups and Theorem 3.1, we obtain the following result:

Corollary 3.1.1. *Assume that $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$ as $\min(n_1, n_2) \to \infty$. Then $a_n(\hat{m}_z - m_z)$ is asymptotically normally distributed with mean zero and variance*

$$\sigma^2_z = (1 - \kappa)\sigma^2_{1,z} + \kappa\sigma^2_{2,z}.$$

We restrict for simplicity to the logistic-Cox model and the maximum likelihood estimation method but the previous results can be generalized to other estimation methods or other semiparametric mixture cure models. For example, if the presmoothing approach introduced in [37] is used instead of the MLE, then the asymptotic properties could be derived in the same way using Theorem 4 in [37]. We also note that because of their complicated expressions, the variances of the estimators in the semiparametric mixture cure model are estimated via a bootstrap procedure [9]. As a result, we will also use the bootstrap to estimate σ_z.

As in Section 2, we would like to investigate whether the asymptotic inference can be made more reliable by means of a permutation approach. Again $\pi = (\pi_1, \ldots, \pi_{n_1+n_2})$ denotes any permutation of $(1, 2, \ldots, n_1 + n_2)$. Write $(Y_j, \Delta_j, X_j, Z_j)$, $j = 1, \ldots, n_1 + n_2$ for the pooled sample which consists of the data points of the first group ($j \le n_1$) and those of the second group ($j > n_1$). Applying π to the pooled sample leads to the permuted samples

$$(Y_{\pi_1}, \Delta_{\pi_1}, X_{\pi_1}, Z_{\pi_1}), \ldots, (Y_{\pi_{n_1}}, \Delta_{\pi_{n_1}}, X_{\pi_{n_1}}, Z_{\pi_{n_1}}) \quad \text{and}$$
$$(Y_{\pi_{n_1+1}}, \Delta_{\pi_{n_1+1}}, X_{\pi_{n_1+1}}, Z_{\pi_{n_1+1}}), \ldots, (Y_{\pi_{n_1+n_2}}, \Delta_{\pi_{n_1+n_2}}, X_{\pi_{n_1+n_2}}, Z_{\pi_{n_1+n_2}}).$$

In the special case of exchangeability, for any z, $\hat{m}_z$ would have the same distribution as $\hat{m}^\pi_z = \widehat{MST}^\pi_{u,1,z} - \widehat{MST}^\pi_{u,2,z}$. Here, $\widehat{MST}^\pi_{u,i,z}$ are the estimators of the conditional mean survival times, just based on the i-th permuted sample.

Since the samples are in general not exchangeable and, as a consequence, the asymptotic variances of $\hat{m}_z$ and $\hat{m}^\pi_z$ cannot be assumed equal, we use their studentized version. We will thus focus on $\hat{m}^\pi_z/\hat{\sigma}^\pi_z$ as the permutation version of $\hat{m}_z/\hat{\sigma}_z$, where $\hat{\sigma}^{\pi 2}_z$ is the estimated variance of $\hat{m}^\pi_z$, estimated via bootstrap.

Because it is computationally infeasible to realize $\hat{m}_z^\pi/\hat{\sigma}_z^\pi$ for all $(n_1+n_2)!$ permutations, we will realize a relatively large number B of random permutations π and approximate the conditional distribution by the collection of the realized $\hat{m}_z^\pi/\hat{\sigma}_z^\pi$ of size B. In the following, we will discuss the asymptotic behaviour of $\hat{m}_z^\pi/\hat{\sigma}_z^\pi$ to justify the validity of the resulting inference procedures.

One challenge that arises in this setting is that, since we are assuming a semiparametric model, the permuted samples will in general not follow the same model. Hence, when we fit the logistic-Cox model to obtain the estimates in the permuted samples, the model is misspecified. Hence, we first show in a series of lemmas in Appendix B that the maximum likelihood estimators converge to the parameters of a logistic-Cox likelihood that minimize the Kullback-Leibler (KL) divergence from the true distribution of the pooled data; denote this by $\bar{\mathbb{P}}$. We can indeed argue that such a minimizer exists and we assume that it is unique. In case of non-uniqueness, we expect that the results can be extended and, depending on the starting point of the algorithm, the estimates would converge to one of such minimizers. However, such extension is beyond the scope of the current paper. In practice, we observed that the EM algorithm converges and the limit was stable with respect to the initial point, which might indicate that the minimizer was indeed unique.

Secondly, to obtain the asymptotic distribution of the permuted estimators, we first obtain a general Donsker-type theorem for permutation based Z-estimators (see Appendix C). That result holds in a great generality, so it would also apply to countless other two-sample problems. Thus, it is of interest of its own. But let us first return to the main result about the permuted estimators in the present context. To this end, we define the KL divergence of a probability measure $\mathbb{P}_1$ from another one, $\mathbb{P}_2$:

$$\mathrm{KL}(\mathbb{P}_1|\mathbb{P}_2) = \int \log\Big(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_2}\Big)\mathrm{d}\mathbb{P}_1;$$

here, $\mathbb{P}_1$ is assumed absolutely continuous with respect to $\mathbb{P}_2$, such that the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_2}$ exists.

Theorem 3.2. *Assume that the identifiability conditions (I1)–(I4) and (A1)–(A4) hold. Assume also that the minimizer of* $\operatorname{argmin}\mathrm{KL}(\bar{\mathbb{P}}|\,\cdot\,)$ *over the model space of logistic-Cox models is unique; also see* (19) *in Appendix B. Then, for any* $z \in \mathcal{Z}$, *as* $\min(n_1,n_2)\to\infty$ *with* $n_1/(n_1+n_2)\to\kappa\in(0,1)$, *the conditional distribution of* $a_n\hat{m}_z^\pi$ *given the data converges weekly in probability to the zero-mean normal distribution with variance* $\sigma_z^{\pi 2}$ *given in* (27) *in Appendix B.*

Corollary 3.2.1. *Under the assumptions of the previous theorem, for any* $z\in\mathcal{Z}$, *as* $\min(n_1,n_2)\to\infty$ *with* $0<\liminf n_1/(n_1+n_2)\le\limsup n_1/(n_1+n_2)<1$, *the (conditional) distributions of* $a_n\hat{m}_z^\pi/\hat{\sigma}_z^\pi$ *(given the data) and* $a_n(\hat{m}_z-m_z)/\hat{\sigma}_z$ *converge weakly in probability to the same limit distribution which is standard normal.*

4. Simulation study

4.1. Unconditional model

In this section, we study the finite sample performance of both the asymptotic and the permutation approach for the nonparametric model described in Section 2. We construct confidence intervals for m and test the one sided hypothesis $H_0 : m \leq 0$ versus $H_1 : m > 0$. In order to cover a wide range of scenarios, we consider nine settings as described below. Settings NP1-NP4 correspond to having two samples with the same mean survival time for the uncured ($m = 0$) but possibly different distributions, while settings NP5-NP9 correspond to having two samples with different mean survival times for the uncured ($m \neq 0$) with different magnitudes and signs for m. The cure and censoring rates also vary across the settings. In all of the following settings, the survival times for the uncured are truncated at $\tau_{0,i}$ equal to the 99% quantile of their distribution in order to satisfy the assumption of compact support. The censoring times are generated independently from an exponential distribution with parameter $\lambda_{C,i}$ and are truncated at $\tau_i = \tau_{0,i} + 2$. The truncation of the censoring times is done only to reflect the bounded follow-up but τ_i does not play any role apart from the fact that $\tau_i > \tau_{0,i}$. Note also that the reported censoring rate includes the cured subjects, which are always observed as censored, hence it is larger than the cure rate.

Simulation settings

Setting NP1. The two samples are exchangeable ($m = 0$): cure rate 40%, Weibull distribution for the uncured with shape and scale parameters 0.75 and 1.5 respectively, censoring rate around 50% ($\lambda_{C,i} = 0.3$, $i = 1, 2$).

Setting NP2. The uncured have the same Weibull distribution in both samples ($m = 0$) with shape and scale parameters 0.75 and 1.5 respectively. The cure rate is 20% in sample 1 and 60% in sample 2, the censoring rate is around 30% and 70% in sample 1 and 2 respectively ($\lambda_{C,1} = 0.25$, $\lambda_{C,2} = 0.5$).

Setting NP3. The uncured have the same Weibull distribution in both samples ($m = 0$) with shape and scale parameters 0.75 and 1.5 respectively. The cure rate is samples 1 and 2 is 60% and 20% respectively; censoring rate is around 65% and 25% respectively ($\lambda_{C,1} = 0.3$, $\lambda_{C,2} = 0.1$).

Setting NP4. The uncured in sample 1 follow a Weibull distribution with shape and scale parameters 0.75 and 1 respectively, while the uncured in sample 2 follow a Gompertz distribution with scale parameter 1 and shape parameter 0.327. The parameters are chosen such that the two groups have the same mean survival time (so $m = 0$). The cure rate is 40% in both samples; censoring rate is around 50% in both samples ($\lambda_{C,1} = 0.2$, $\lambda_{C,2} = 0.15$).

Setting NP5. The uncured in the two samples have different Gompertz distributions with the same scale parameter 1 and shape parameters 0.1 and 0.5

respectively. The difference of mean survival times is $m = 1.09$. For this choice of parameters, the supports of the event times in the two samples are $[0, 3.8]$ and $[0, 2.3]$ respectively. The cure rate is 40% in both samples; censoring rate is around 65% in sample 1 and 45% in sample 2 ($\lambda_{C,1} = 0.3$, $\lambda_{C,2} = 0.1$).

Setting NP6. This is the same as Setting NP5 but the two groups are exchanged, i.e. $m = -1.09$.

Setting NP7. The uncured in the two samples have different Gompertz distributions with the same scale parameter 1 and shape parameters 0.08 and 0.1 respectively. The difference of mean survival times is $m = 0.18$ For this choice of parameters, the supports of the event times in the two samples are $[0, 4.1]$ and $[0, 3.9]$ respectively. The cure rate is 60% in sample 1 and 20% in sample 2; censoring rate is around 70% in sample 1 and 40% in sample 2 ($\lambda_{C,1} = 0.2$, $\lambda_{C,2} = 0.15$).

Setting NP8. The survival distributions of the uncured are as in Setting NP7, i.e., $m = 0.18$. The cure rate in samples 1 and 2 is 30% and 20% respectively; censoring rate is around 44% in sample 1 and 34% in sample 2 ($\lambda_{C,1} = 0.1$, $\lambda_{C,2} = 0.1$).

Setting NP9. The event times of the uncured in the sample 1 follow a Gompertz distribution with scale parameter 1 and shape parameter 0.08, while in sample 2 they follow a Weibull distribution with shape and scale parameters 2 and 0.28 respectively. Both distributions have support $[0, 4.1]$ but different mean survival times ($m = 0.52$). The cure rate is 40% in both samples; censoring rate is around 50% in both samples ($\lambda_{C,1} = 0.1$, $\lambda_{C,2} = 0.1$).

Simulation results

First, considering different sample sizes $n_1 = 2n_2 \in \{50, 200\}$ or $n_1 = n_2 \in \{100, 200\}$, 95% confidence intervals for m are constructed based on both the asymptotic approximation and the permutation approach:

$$I_n = [\hat{m} \mp q_{1-\alpha/2}\hat{\sigma}/a_n], \qquad I_n^\pi = [\hat{m} - q^\pi_{1-\alpha/2}\hat{\sigma}, \hat{m} - q^\pi_{\alpha/2}\hat{\sigma}],$$

where $\hat{\sigma}$ is given in (4), $q_{1-\alpha}$ denotes the $100(1-\alpha)$%-quantile of the standard normal distribution, $\alpha = 0.05$ and $q^\pi_{1-\alpha}$ denotes the $100(1-\alpha)$%-quantile of the conditional distribution of $\hat{m}^\pi/\hat{\sigma}^\pi$ for the permutation approach. Note that $q^\pi_{1-\alpha}$ is a random quantile that depends on the available data, with $a_n q^\pi_{1-\alpha} \xrightarrow{p} q_{1-\alpha}$ as $n \to \infty$. We take $B = 500$ random permutations, which seemed to be sufficient since increasing B to 1,000 did not have much effect in the results. Average length and coverage probabilities over 1,000 repetitions are reported in Table 1. The coverage rates closest to 95% among both types of confidence intervals is printed in bold-type.

We observe that the confidence intervals based on the permutation approach are in general slightly wider and have better coverage, particularly for small

TABLE 1

Coverage probabilities (CP) in % and average length (L) of 95% confidence intervals using the asymptotic approach (M1) and the permutation approach (M2) for different sample sizes. Next to the setting name we report the true m in parentheses and below the true standard deviations of the survival times in each sample (sd_1, sd_2).

Sett.		$n_1 = 2n_2 = 50$		$n_1 = n_2 = 100$		$n_1 = 2n_2 = 200$		$n_1 = n_2 = 200$	
		M1	M2	M1	M2	M1	M2	M1	M2
NP1 (0)	L	1.34	1.15	0.83	0.88	0.74	0.80	0.62	0.64
(0.86, 0.86)	CP	88.4	**93.1**	93.0	**94.3**	93.3	**95.4**	94.1	**94.2**
NP2 (0)	L	1.04	1.22	0.91	1.00	0.84	0.92	0.73	0.78
(0.86, 0.86)	CP	76.0	**87.2**	86.6	**90.0**	83.5	**87.6**	89.4	**90.6**
NP3 (0)	L	1.18	1.37	0.82	0.87	0.66	0.71	0.60	0.62
(0.86, 0.86)	CP	89.8	**93.5**	88.8	**90.1**	**93.3**	92.9	92.0	**92.2**
NP4 (0)	L	1.44	1.64	1.08	1.12	0.87	0.90	0.83	0.84
(1.47, 1.33)	CP	86.3	**87.2**	84.2	**86.5**	**89.5**	88.9	89.2	**89.3**
NP5 (1.09)	L	1.02	1.22	0.67	0.69	0.53	0.54	0.48	0.48
(0.91, 0.57)	CP	93.1	**95.9**	93.6	**94.6**	**95.0**	94.8	**95.2**	95.4
NP6 (−1.09)	L	1.18	1.34	0.67	0.69	0.64	0.66	0.48	0.48
(0.57, 0.91)	CP	86.4	**89.2**	93.6	**94.7**	**93.3**	**93.3**	**95.0**	94.7
NP7 (0.18)	L	1.31	1.52	0.83	0.85	0.67	0.69	0.60	0.60
(0.95, 0.91)	CP	92.3	**95.7**	94.3	**94.5**	**93.3**	92.9	**93.7**	93.6
NP8 (0.18)	L	1.07	1.15	0.64	0.65	0.55	0.55	0.45	0.45
(0.95, 0.91)	CP	94.1	**95.6**	**93.3**	**93.3**	**94.8**	94.7	**93.9**	93.5
NP9 (0.52)	L	1.18	1.30	0.71	0.72	0.61	0.61	0.50	0.50
(0.95, 0.86)	CP	92.5	**94.1**	94.2	**94.8**	94.2	**94.6**	94.4	**94.6**

sample sizes. As the sample sizes increase, the two approaches give more comparable results. For some settings, much larger sample sizes are needed to have coverage close to the nominal level but, for most of them, coverage is close to 95%. When the sample sizes are the same, settings NP2 and NP3 are almost the same, with setting NP3 having less censoring, leading to shorter confidence intervals and better coverage. When $n_1 = 2n_2$, setting NP2 is more difficult because the smaller sample has a very large cure and censoring rate, leading to worse coverage probabilities. Similarly, when the sample sizes are the same, setting NP5 and NP6 are the same, leading to same length confidence intervals and approximately same coverage (due to sampling variation). When $n_1 = 2n_2$, setting NP6 is more difficult because it has higher censoring rate in the smaller sample. As a result, we observe longer confidence intervals and worse coverage probabilities. Setting NP8 is similar to setting NP7 but the first sample has lower cure rate, leading to shorter confidence intervals. When the two samples are not comparable in terms of cure and censoring rate, increasing the sample size of the sample in which it is easier to estimate $MST_{u,i}$, does not usually lead to better coverage (compare settings NP2 and NP3, NP5 and NP6). On the other hand, increasing the sample size of the sample in which estimation of $MST_{u,i}$ is more difficult usually leads to better coverage. In Table 1, we also report the standard deviations of the survival times of the uncured for each sample. We notice that in general the variances are comparable across settings and do not show any particular relation with the performance of the methods, except setting NP4 which has higher variance and lower coverage.

TABLE 2

Coverage probabilities (CP) in % and average length (L) of 95% confidence intervals for m using the asymptotic approach (M1) and the permutation approach (M2) for different sample sizes.

Sett.		$n_1, 2n_2 = 600$		$n_1, 2n_2 = 1200$		$n_1, 2n_2 = 4000$		$n_1, 2n_2 = 10000$	
		M1	M2	M1	M2	M1	M2	M1	M2
NP2	L	0.57	0.60	0.44	0.45	0.26	0.26	0.12	0.12
$(m = 0)$	CP	84.4	**87.6**	88.4	**89.9**	**92.1**	92.0	**94.5**	94.2

		$n_1, n_2 = 500$		$n_1, n_2 = 1000$		$n_1, n_2 = 2000$	
		M1	M2	M1	M2	M1	M2
NP3	L	0.40	0.40	0.28	0.28	0.20	0.20
$(m = 0)$	CP	**93.8**	93.6	**92.5**	91.9	**94.8**	94.4
NP4	L	0.55	0.56	0.39	0.40	0.28	0.28
$(m = 0)$	CP	**92.3**	**92.3**	**94.2**	**94.2**	94.1	**94.5**

We further investigate settings NP2, NP3, NP4 which exhibit the worst performance in terms of coverage probabilities. In setting NP2, as estimation in the second sample is more difficult (because of higher cure and censoring rates), the performance of both the asymptotic and permutation approach in worse when the size of sample 1 is larger than the size of sample 2. In setting NP3, estimation of the first sample is more challenging and we observed that the coverage is better when the size of sample 1 is larger. In setting NP4, both samples have the same censoring and cure rate but the coverage seems to be worse when the sample sizes are the same. Results for larger sample sizes under the most difficult scenarios for each of these three settings are reported in Table 2. They show that, as expected, the coverage probabilities for both approaches converge to the nominal level.

In addition, we selected 2 of the settings (setting NP4 and NP9) and further investigated the effect of the censoring and cure rates for sample sizes $100-100$ and $200-100$. First, we keep the cure rate fixed at 40% (moderate) and consider 3 censoring levels: 45% (low), 50% (moderate) and 60% (high). Secondly, we vary the cure rate: 20% (low), 40% (moderate) and 60% (high), while maintaining the same moderate censoring level equal to the cure rate plus 10%. Average length and coverage probabilities over 1,000 repetitions are reported in Table 3. As expected, we observe that, as the censoring or cure rate increases, the length of the confidence intervals increases. In setting NP4 the coverage deteriorates significantly for a high censoring rate, while in setting NP9 the coverage remains stable and close to the nominal value throughout all scenarios.

Next, we consider a one-sided hypothesis test for $H_0 : m \leq 0$ versus $H_1 : m > 0$ at level 5%. The rejection rates for the test are reported in Table 4. Looking at settings NP1-NP4 and NP6 for which H_0 is true (with $m = 0$ for settings NP1-NP4), we observe that most of the time the rejection rate is lower or close to 5% for both methods. Setting NP2 is again the most problematic one with rejection rate higher than the level of the test. This might be because in setting NP2 the second sample has a high cure (and censoring) rate, which might lead to underestimation of the mean survival times for the uncured in sample 2 and

TABLE 3
Coverage probabilities (CP) in % and average length (L) of 95% confidence intervals using the asymptotic approach (M1) and the permutation approach (M2) for different sample sizes, censoring and cure rates.

			censoring rate					
			low		moderate		high	
Sett.	n_1/n_2		M1	M2	M1	M2	M1	M2
NP4	100/100	L	0.87	0.89	1.08	1.12	1.12	1.23
		CP	92.4	**92.6**	84.0	**86.5**	68.7	**72.2**
($m = 0$)	200/100	L	0.67	0.71	0.87	0.90	1.06	1.17
		CP	**93.8**	92.9	**89.5**	88.9	76.8	**77.3**
NP9	100/100	L	0.67	0.68	0.71	0.72	0.79	0.82
		CP	**94.1**	93.9	94.2	**94.8**	92.9	**94.3**
($m = 0.52$)	200/100	L	0.58	0.58	0.61	0.61	0.69	0.71
		CP	**94.5**	94.2	94.2	**94.6**	**94.7**	94.6
			cure rate					
			low		moderate		high	
Sett.	n_1/n_2		M1	M2	M1	M2	M1	M2
NP4	100/100	L	0.85	0.92	1.08	1.12	1.21	1.21
		CP	90.6	**91.4**	84.0	**86.5**	88.1	**89.0**
($m = 0$)	200/100	L	0.66	0.72	0.87	0.90	1.20	1.31
		CP	**93.0**	**93.0**	**89.5**	88.9	80.6	**80.7**
NP9	100/100	L	0.59	0.60	0.71	0.72	0.92	0.96
		CP	**94.3**	**94.3**	94.2	**94.8**	94.4	**95.5**
($m = 0.52$)	200/100	L	0.51	0.51	0.61	0.61	0.80	0.82
		CP	95.8	**95.6**	94.2	**94.6**	93.2	**93.8**

as a result an overestimation of m. For settings NP2, NP3, and NP4 we also considered larger sample sizes. The results are reported in Table 5. In particular, we observe that the rejection rates in setting NP2 decrease and approaches the significance level as the sample size increases. In terms of power, as m or the sample size increase, the power increases. Both methods are comparable, with the permutation approach usually leading to slightly lower rejection rate under both hypothesis. Again, in settings NP4 and NP9 we also investigate the effect of the censoring and cure rate as above. Results are given in Table 6. As the censoring or cure rate increases, the power of the test decreases, while the rejection rate in setting NP4 (H_0 is true) remains below the 5% level throughout all scenarios.

Finally, to acknowledge that the case of unbalanced sample sizes where the smaller sample meets the higher censoring rate, we would like to point to [15]. For very small sample sizes, their studentized permutation test about the RMST also exhibited the worst control of the type-I error rate in this challenging context; see Table 1 therein, and also Tables S.1 and S.2 in the supplementary material accompanying that paper. Of note, their proposed permutation test is still quite accurate with a size not exceeding 6.8% even in the most challenging setting.

To give an idea of the computational cost of both approaches we provide in Table 7 the average running times over 10 replications for some of the settings and sample sizes. The computations are done on an Intel Core i7 CPU laptop. We observe that the permutation approach is around 500 times more computa-

TABLE 4

Rejection rate in % for testing the hypothesis $H_0 : m \leq 0$ versus $H_1 : m > 0$ at level 5% using the asymptotic approach (M1) and the permutation approach (M2) for different sample sizes. Settings under H_0 are indicated by NP* *($m = 0$) and* NP** *($m < 0$).*

Sett.	m	$n_1 = 2n_2 = 50$		$n_1 = n_2 = 100$		$n_1 = 2n_2 = 200$		$n_1 = n_2 = 200$	
		M1	M2	M1	M2	M1	M2	M1	M2
NP1*	0	5.6	5.0	6.9	4.9	5.9	5.4	6.2	5.5
NP2*	0	20.9	19.0	17.5	14.0	18.0	17.2	14.3	12.1
NP3*	0	0.9	0.8	2.3	1.7	0.6	0.5	2.4	2.2
NP4*	0	0.6	0.2	2.3	0.6	0.3	0.2	1.9	0.7
NP5	1.09	96.5	94.1	100	100	100	100	100	100
NP6**	−1.09	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NP7	0.18	11.7	9.5	21.3	21.0	23.0	22.1	30.5	30.3
NP8	0.18	15.1	13.0	30.3	29.3	33.3	33.0	45.0	45.0
NP9	0.52	55.4	50.0	87.4	86.4	94.8	94.4	97.3	98.7

TABLE 5

Rejection rates in % for testing the hypothesis $H_0 : m \leq 0$ versus $H_1 : m > 0$ at level 5% using the asymptotic approach (M1) and the permutation approach (M2) for different sample sizes. For all settings $m = 0$.

Sett.	$n_1 = 2n_2 = 600$		$n_1 = 2n_2 = 1{,}200$		$n_1 = 2n_2 = 4{,}000$		$n_1 = 2n_2 = 10{,}000$	
	M1	M2	M1	M2	M1	M2	M1	M2
NP2*	18.8	15.0	15.5	13.2	7.7	10.1	3.4	7.0

	$n_1 = n_2 = 500$		$n_1 = n_2 = 1{,}000$		$n_1 = n_2 = 2{,}000$	
	M1	M2	M1	M2	M1	M2
NP3*	2.8	2.7	3.6	3.8	4.1	4.2
NP4*	2.0	1.7	2.3	1.9	3.4	3.2

tionally expensive than the asymptotic one because essentially the estimation procedure has to be repeated for the 500 permutation samples.

4.2. Conditional model

In this section, we investigate the practical performance of the permutation approach and of the asymptotic method when comparing conditional mean survival times for the uncured subpopulations as described in Section 3. We consider two samples of size 200 and 100, respectively, from logistic-Cox mixture cure models. Note that, in practice, semiparametric cure models are usually not used for sample sizes much smaller than these because of their complexity (more parameters need to be estimated compared to standard Cox model for example) and the need to observe a long plateau with a considerable amount of censored observations (as a confirmation of the sufficiently long follow-up assumption). Since the permutation approach is computationally intensive and asymptotically we expect the behavior of the two methods to be more similar, we also did not consider larger sample sizes. Instead, we focus on three different scenarios as described below by varying the distributions of the uncured subjects, the cure proportions and the censoring rates. For simplicity, we also consider the same covariates in the incidence and latency components, i.e., $X = Z$.

TABLE 6

Rejection rate in % for testing the hypothesis $H_1 : m > 0$ at level 5% using the asymptotic approach (M1) and the permutation approach (M2) for different sample sizes, censoring and cure rates.

			censoring rate					
			low		moderate		high	
Sett.	m	n_1/n_2	M1	M2	M1	M2	M1	M2
NP4*	0	100/100	3.0	1.6	2.3	0.6	1.7	0.2
		200/100	2.9	1.0	0.3	0.2	2.4	0.0
NP9	0.52	100/100	91.9	91.6	87.4	86.4	79.7	79.1
		200/100	95.9	95.0	94.8	94.4	89.0	87.5
			cure rate					
			low		moderate		high	
Sett.	m	n_1/n_2	M1	M2	M1	M2	M1	M2
NP4*	0	100/100	2.8	1.0	2.3	0.6	2.4	0.8
		200/100	3.6	1.1	0.3	0.2	1.9	0.1
NP9	0.52	100/100	95.7	95.5	87.4	86.4	73.1	72.0
		200/100	99.2	99.0	94.8	94.4	79.5	77.6

TABLE 7

Average running time (in seconds) over 10 replications of the asymptotic and permutation method.

	$n_1 = n_2 = 100$		$n_1 = n_2 = 200$	
Sett.	asymp.	perm.	asymp.	perm.
NP1*	0.52	279.13	0.94	461.76
NP2*	0.65	333.00	0.97	465.88
NP8	1.17	576.14	2.57	1203.77

Simulation settings

Setting SP1. Both samples are generated from the logistic-Cox mixture cure model with Weibull baseline distribution with shape parameter 0.75 and scale parameters 1.5 and 2, respectively. The survival times of the uncured subjects are truncated at $\tau_{0,i}$, $i = 1, 2$ equal to the 99% quantile of the corresponding baseline Weibull distribution in order to have finite supports. We consider two independent covariates Z_1 and Z_2, which affect both the cure probability and the survival of the uncured. In the first sample, $Z_1 \sim N(0, 1)$, $Z_2 \sim Bern(0.4)$ while in the second sample $Z_1 \sim N(1, 1)$, $Z_2 \sim Bernoulli(0.6)$. The regression coefficients are $\gamma_1 = (0, 0.5, 0.8)$, $\beta_1 = (0.3, 0.5)$, $\gamma_2 = (0.1, 1, 0.6)$, $\beta_2 = (0.3 + \log(0.75), 0.5)$. This corresponds to having around 43% and 24% cured subjects in each sample. The censoring times are generated independently of the other variables from exponential distributions with parameters 0.4 and 0.2, respectively. They are truncated at $\tau_i = \tau_{0,i} + 2$, $i = 1, 2$ to reflect the limited length of studies in practice. The censoring rate in sample 1 is 52%, while in sample 2 it is 28%. In both cases, we have around 15% of the observations in the plateau. In this setting, the covariate distributions, the cure and censoring rates, and the survival distributions of the uncured are different among the two samples. However, depending on the covariate values, the conditional mean survival times of the uncured can be the same, i.e., $m_z = 0$. We consider

a range of possible covariate values, see Table 8, including some extreme and unlikely values in order to get more different mean survival times between the two groups.

Setting SP2. Both samples are generated from logistic-Cox mixture cure models with Gompertz baseline distribution with shape parameter 1 and rate parameters 0.1, 0.3, respectively. The survival times of the uncured subjects are truncated at $\tau_{0,i}$, $i = 1, 2$, equal to the 99% quantiles of the corresponding baseline distributions in order to have finite supports. We consider two independent covariates, $Z_1 \sim N(0, 1)$ and $Z_2 \sim Unif(-1, 1)$ with the same distribution in both samples. The regression coefficients are $\gamma_1 = \gamma_2 = (0.8, -1, 1)$, $\beta_1 = (-0.6, 0.5)$, $\gamma_2 = (0.1, 1, 0.6)$, $\beta_2 = (-0.05, 0.4)$. This corresponds to having around 35% cured subjects in each sample. The censoring times are generated independently of the other variables from exponential distributions with parameters, 0.1 and 0.2, respectively. They are truncated at $\tau_i = \tau_{0,i} + 2$, $i = 1, 2$, to reflect the limited length of studies in practice. The censoring rate in sample 1 is 46%, while in sample 2 it is 48%. In both cases, we have around 20% of the observations in the plateau. In this setting, the covariate distributions, the cure rates, and the censoring rates are the same for both samples. The survival distributions of the uncured are different but again, for certain values of the covariates, the conditional mean survival times of the uncured are the same. We consider different covariate values as in Table 8. In particular, z_8 is a very extreme and unlikely value but it was considered in order to have a case with larger negative value for m_z.

Setting SP3. Both samples are generated from the same distribution as for sample 1 in setting SP1. This means that the two samples are exchangeable and for any covariate value we have $m_z = 0$.

Setting SP4. In this setting the data is not generated under the logistic-Cox model to investigate robustness to model misspecification. The link function for the incidence model is given by

$$\phi(s) = \frac{1}{2}[1 + \tanh(0.2 + s^3)],$$

which is shown in Figure 2 together with the logistic function for comparison. The survival times for the uncured are generated from the accelerated failure time model

$$S_{u,i}(t|z) = \exp\{-\Lambda_i(t \exp(\beta_i^T z))\}$$

with Weibull baseline distribution with shape parameter 0.75 and scale parameters 1.5 and 2.5 respectively. The survival times of the uncured subjects are truncated at $\tau_{0,i}(z)$, $i = 1, 2$, equal to the 99% quantile of the corresponding conditional distributions. We consider two independent covariates Z_1 and Z_2 as in setting SP1 which affect both the cure probabilities and the survival of the uncured. The regression coefficients are $\gamma_1 = (-0.2, 0.5, 0.8)$, $\gamma_2 = (-0.05, 1, 0.2)$, $\beta_1 = (0.3, 0.1)$, $\beta_2 = (-0.1, 0.1)$. This corresponds to having around 38% and 18% cured observations in each sample. The censoring times are generated inde-

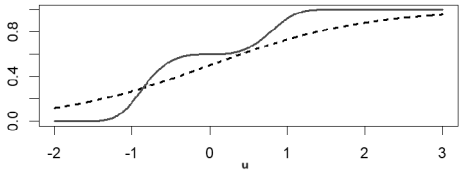

FIG 2. *The link function for the incidence model in setting SP4 (solid line) and the logistic function (dashed line).*

TABLE 8
Covariate values z and corresponding difference in conditional mean survival times for the uncured m_z for Settings SP1, SP2 and SP4.

Sett. SP1	$z_1 = (0,1)$ $m_1 = 0.11$	$z_2 = (-1,0)$ $m_2 = 0.5$	$z_3 = (1,0)$ $m_3 = 0$	$z_4 = (1,1)$ $m_4 = 0$
	$z_5 = (2,1)$ $m_5 = -0.07$	$z_6 = (4,0)$ $m_6 = -0.3$	$z_7 = (-4,0)$ $m_7 = 1.68$	$z_8 = (-3,1)$ $m_8 = 0.83$
Sett. SP2	$z_1 = (-2,0)$ $m_1 = 0$	$z_2 = (-1.85,0.8)$ $m_2 = 0$	$z_3 = (-2.16,-0.8)$ $m_3 = 0$	$z_4 = (0,0)$ $m_4 = 0.79$
	$z_5 = (1,0.5)$ $m_5 = 1.16$	$z_6 = (-1,-0.5)$ $m_6 = 0.43$	$z_7 = (-3,0.5)$ $m_7 = -0.31$	$z_8 = (-6,0)$ $m_8 = -0.82$
	$z_9 = (2,0)$ $m_9 = 1.65$	$z_{10} = (-1.5,0)$ $m_{10} = 0.18$	$z_{11} = (-2.5,0)$ $m_{11} = -0.16$	
Sett. SP4	$z_1 = (0,1)$ $m_1 = 0.36$	$z_2 = (-1,0)$ $m_2 = 0.14$	$z_3 = (1,0)$ $m_3 = 0.53$	$z_4 = (-1.94,1)$ $m_4 = 0$
	$z_5 = (-4,1)$ $m_5 = -0.24$	$z_6 = (-1.85,0)$ $m_6 = 0$	$z_7 = (3,0)$ $m_7 = 0.71$	$z_8 = (-2.5,1)$ $m_8 = -0.08$

pendently of the other variables from exponential distributions with parameters 0.4 and 0.2, respectively. They are truncated at $\tau_i = 10$, which is larger than $\tau_{0,i}(z)$ for the observed values of z to guarantee sufficient follow-up. The censoring rate in sample 1 is 50%, while in sample 2 it is 23%. In both cases, we have around 12% of the observations in the plateau. We consider different covariate values as in Table 8.

Simulation results

For each setting and covariate value, we construct $1-\alpha = 95\%$ confidence intervals for m_z based on the asymptotic and the permutation approach

$$I_z^* = [\hat{m}_z \mp q_{1-\alpha/2}\hat{\sigma}_z/a_n], \qquad I_z^\pi = [\hat{m}_z - q^\pi_{1-\alpha/2}, \hat{m}_z - q^\pi_{\alpha/2}\hat{\sigma}_z]$$

where $\hat{m}_z$ is computed as in Section 3.2, $\hat{\sigma}_z^2$ is the variance of $a_n\hat{m}_z$ estimated via the bootstrap, $q_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution and $q^\pi_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the conditional distribution of $\hat{m}^\pi_z/\hat{\sigma}^\pi_z$. Due to the computational cost, we use 100 bootstrap samples

TABLE 9

Coverage probabilities (CP) in % and length (L) of 95% confidence intervals using asymptotic approach (M1) and the permutation approach (M2) for different choices of z in Setting SP1. Next to the covariate we report m_z in parentheses and the true standard deviations of the survival times of the uncured conditional on z for each sample.

	z_1 (−0.11) (0.48,0.33)		z_2 (0.5) (1.13,0.59)		z_3 (0) (0.61,0.58)		z_4 (0) (0.32,0.32)	
	M1	M2	M1	M2	M1	M2	M1	M2
L	0.498	0.476	1.294	1.385	0.703	0.716	0.372	0.353
CP	96.5	**95.3**	**90.6**	89.8	94.5	**95.5**	97.4	**96.3**
	z_5 (−0.07) (0.22,0.32)		z_6 (−0.3) (0.73,0.56)		z_7 (1.68) (1.70,0.62)		z_8 (0.83) (1.22,0.34)	
	M1	M2	M1	M2	M1	M2	M1	M2
L	0.429	0.393	1.168	1.012	2.606	3.119	1.846	1.925
CP	98.3	**96.4**	**95.6**	91.4	68.6	**72.5**	**85.9**	85.6

TABLE 10

Coverage probabilities (CP) in % and length (L) of 95% confidence intervals using the asymptotic approach (M1) and the permutation approach (M2) for different choices of z in Setting SP2. Next to the covariate we report m_z in parentheses and the true standard deviations of the survival times of the uncured conditional on z for each sample.

	z_1 (0) (0.67,0.67)		z_2 (0) (0.60,0.60)		z_3 (0) (0.74,0.73)		z_4 (0.79) (0.91,0.68)	
	M1	M2	M1	M2	M1	M2	M1	M2
L	0.970	0.952	1.003	0.985	1.308	1.291	0.641	0.64
CP	**93.9**	93.3	**92.9**	92.8	**94.1**	93.4	**97.0**	**97.0**
	z_5 (1.16) (0.95,0.65)		z_6 (0.43) (0.85,0.71)		z_7 (−0.31) (0.48,0.61)		z_8 (−0.82) (0.19,0.62)	
	M1	M2	M1	M2	M1	M2	M1	M2
L	1.038	1.060	0.851	0.845	1.311	1.263	2.097	1.669
CP	**97.1**	97.3	**94.3**	94.2	**92.7**	91.9	**89.9**	81.5
	z_9 (1.65) (0.94,0.70)		z_{10} (0.18) (0.73,0.67)		z_{11} (−0.16) (0.60,0.66)			
	M1	M2	M1	M2	M1	M2		
L	1.526	1.571	0.785	0.775	1.158	1.130		
CP	98.1	**97.5**	**94.0**	93.3	**93.2**	93.1		

combined with 500 random permutation samples. This procedure was repeated 1,000 times. The lengths and coverage probabilities of the confidence intervals are given in Tables 9–12. The coverage rates closest to 95% among both types of confidence intervals is printed in bold-type. For the exchangeable Setting SP3, since the permutation confidence intervals are exact, we only provide the results of the asymptotic approach.

We observe that the coverage of both confidence intervals is very low for some covariate values. That happens mainly when m_z is large in absolute value (either positive or negative depending on the setting). This seems to be because, for certain covariate values, the errors that we make in the estimation of the coefficients and baseline survival get amplified when computing the survival function

TABLE 11

Coverage probabilities (CP) in % and length (L) of 95% confidence intervals using the asymptotic method for different choices of z in Setting SP3. The true value of m_z is 0 for all z. In parentheses we report the true standard deviations of the survival times of the uncured conditional on z (same for both samples).

	z_1 (0.48)	z_2 (1.13)	z_3 (0.61)	z_4 (0.32)	z_5 (0.22)	z_6 (0.73)	z_7 (1.70)	z_8 (1.22)
L	0.707	0.795	0.974	0.587	0.680	1.338	3.512	2.726
CP	94.7	91.0	95.3	97.0	98.6	99.0	83.5	90.1

TABLE 12

Coverage probabilities (CP) in % and length (L) of 95% confidence intervals using asymptotic approach (M1) and the permutation approach (M2) for different choices of z in Setting SP4. Next to the covariate we report m_z in parentheses and the true standard deviations of the survival times of the uncured conditional on z for each sample.

	z_1 (0.36) (0.78,0.39)		z_2 (0.14) (1.16,0.39)		z_3 (0.53) (0.64,0.49)		z_4 (0) (1.39,0.32)	
	M1	M2	M1	M2	M1	M2	M1	M2
L	0.759	0.764	1.313	1.348	0.751	0.745	1.748	1.787
CP	**89.9**	87.3	75.1	**83.5**	**40.4**	32.8	57.1	**64.2**
	z_5 (−0.24) (2.58,0.26)		z_6 (0) (1.5,0.35)		z_7 (0.71) (0.35,0.59)		z_8 (−0.08) (1.65,0.30)	
	M1	M2	M1	M2	M1	M2	M1	M2
L	2.786	3.109	1.779	1.861	1.102	1.005	2.062	2.153
CP	41.9	**51.7**	56.3	**66.2**	**15.5**	11.5	50.7	**58.1**

conditional on z, resulting in a biased estimate for m_z. Much larger sample sizes would be needed to get a good estimate of m_z for such covariate values z. In the other cases, the coverage probabilities are close to the nominal value and the two methods are comparable. For some z, the permutation approach does slightly better than the asymptotic one, but vice versa for other choices of z. For settings SP1 and SP3 we notice that the performance of both methods is worse when there is large variability of the survival times of the uncured in each sample, for example for covariate values z_2, z_7, z_8. However, this does not seem to be the case in setting SP2. For setting SP4 with model misspecification we observe that the coverage probabilities are in general very low for both methods. This poor performance seems to be mainly related to the misspecified Cox model for the latency. For example, for the worse case with covariate value z_7, the mean survival time in the first (second) sample is underestimated (overestimated) since $\beta_1^T z_7 > 0$ ($\beta_2^T z_7 > 0$), leading to large underestimation of m_z. Hence, one needs to be very cautious to model misspecification. For the real data application, we have first performed a goodness of fit test for the model assumption.

Next, we consider testing the hypothesis $H_0 : m_z = 0$ against $H_1 : m_z \neq 0$ at level $\alpha = 5\%$. Again the variance of $\hat{m}_z$ is estimated via the bootstrap with 100 bootstrap samples and the quantiles of $\hat{m}_z/\hat{\sigma}_z$ are estimated via 500 permutation samples. The procedure was repeated 1,000 times. In Tables 13–15, we report

TABLE 13

Rejection rates of H_0 in % for the asymptotic approach (M1) and the permutation approach (M2) and different choices of z in Setting SP1. Value of m_z is reported in parentheses.

	z_1 (−0.11)		z_2 (0.5)		z_3 (0)		z_4 (0)	
	M1	M2	M1	M2	M1	M2	M1	M2
Rejection rate	14.1	11.8	22.9	15.2	5.5	4.5	2.6	3.7
	z_5 (−0.07)		z_6 (−0.3)		z_7 (1.68)		z_8 (0.83)	
	M1	M2	M1	M2	M1	M2	M1	M2
Rejection rate	6.9	13.6	8.5	17.8	33.6	19.5	26.7	22.1

TABLE 14

Rejection rates of H_0 in % for the asymptotic approach (M1) and the permutation approach (M2) and different choices of z in Setting SP2. Value of m_z is reported in parentheses.

	z_1 (0)		z_2 (0)		z_3 (0)		z_4 (0.79)	
	M1	M2	M1	M2	M1	M2	M1	M2
Rejection rate	6.1	6.7	7.1	7.2	5.1	6.6	99.9	100.0
	z_5 (1.16)		z_6 (0.43)		z_7 (−0.31)		z_8 (−0.82)	
	M1	M2	M1	M2	M1	M2	M1	M2
Rejection rate	99.5	99.3	48.3	49.3	15.4	19.9	30.6	52.7
	z_9 (1.65)		z_{10} (0.18)		z_{11} (−0.16)			
	M1	M2	M1	M2	M1	M2		
Rejection rate	99.2	99.0	14.1	15.1	9.7	11.8		

TABLE 15

Rejection rates of H_0 in % for the asymptotic method and different choices of z in Setting SP3. H_0 is true for all choices of z.

	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
Rejection rate	5.3	9.0	4.7	3.0	1.4	1.0	16.5	9.9

the percentages of the cases in which H_0 was rejected.

In Settings SP1 and SP2, the levels of the test seem to be close to the nominal level and the power is larger when $|m_z|$ is larger, even though it does not only depend on $|m_z|$ but also on the sign of m_z (deviations in conditional mean survival times might be easier to detect in one direction compared to the other). In Setting SP1, the asymptotic method has more power when $m > 0$ (m_1, m_2, m_7, m_8), while the permutation approach has more power when $m < 0$ (m_5, m_6). In Setting SP2, we observe that the results for both methods are comparable when $m > 0$ but the permutation approach has more power when $m < 0$. For the exchangeable setting, the level of the asymptotic test is larger than the nominal value for some of the covariate values (the ones for which the coverage probabilities were anti-conservative; see above). For setting SP4, the level of the test is very far from the nominal one for both the asymptotic and permutation approach, which shows as before that model misspecification can have serious consequences.

TABLE 16

Rejection rates of H_0 in % for the asymptotic approach (M1) and the permutation approach (M2) and different choices of z in Setting SP4. Value of m_z is reported in parentheses.

	z_1 (0.36)		z_2 (0.14)		z_3 (0.53)		z_4 (0)	
	M1	M2	M1	M2	M1	M2	M1	M2
Rejection rate	37.2	29.4	42.5	35.0	7.0	5.6	42.9	35.8
	z_5 (−0.24)		z_6 (0)		z_7 (0.71)		z_8 (−0.08)	
	M1	M2	M1	M2	M1	M2	M1	M2
Rejection rate	44.3	33.4	43.7	33.8	7.8	12.7	42.3	34.2

TABLE 17

Average running time (in seconds) over 5 replications of the asymptotic and permutation method.

	$n_1 = n_2 = 100$	
Sett.	asymp.	perm.
SP1	33.57	11662.04
SP2	16.83	9803.54

To compare the computational cost of both approaches we provide in Table 17 the average running times over 5 replications for settings 1 and 2 and sample size $n_1 = n_2 = 100$. The computations are done on an Intel Core i7 CPU laptop. Overall, we conclude that, unless the two samples are exchangeable, there is no clear advantage of using the permutation approach to justify its much higher computational cost. This is different from what is observed previously in the literature and it might be related to the fact that the logistic-Cox model is misspecified in the permutation samples. Computationally, the EM algorithm still converges and is stable with respect to the initial estimates. This suggest that the problem should not be about the existence of a *unique* maximizer of the likelihood for the misspecified model.

5. Application

In this section, we consider two real data applications of the developed methods. The first one is a leukemia study with relatively small sample size for which the nonparametric approach is appropriate, while the second is an observational study of breast cancer for which the conditional semiparametric approach is more appropriate.

5.1. Leukemia study

We consider a real data set from research on leukemia [24]; the study ran from March 1982 to May 1987. 91 patients were treated with high-dose chemoradiotherapy, followed by a bone marrow transplant. $n_1 = 46$ patients received allogeneic marrow from a matched donor and $n_2 = 45$ patients without a matched donor received autologous marrow, i.e. their own. They were followed for 1.4 to 5 years. For other details such as additional patient characteristics and the

frequency of the graft-versus-host disease among allogeneically transplanted patients we refer to the original study [24].

In our analysis, we are going to re-analyze relapse-free survival. The data sets are available in the monograph [42]. They contain the (potentially right-censored) times to relapse or death (in days), together with the censoring status. It was argued in both [42] and [24] that a cure model is appropriate for the data. The possibility of cure of leukemia patients after a bone marrow transplant has been also argued in the medical literature around the time of this study, e.g. [53, 47]. The authors of [42] first fit parametric accelerated failure time mixture cure models (see Section 2.6) and did not find a significant difference in either cure rates or survival times for the uncured. Additionally, in Section 3.6 of [42] they fit semiparametric logistic-Cox and logistic-accelerated failure time (AFT) models. Under the logistic-Cox model they did find a borderline significant effect for the survival time distribution (latency) of the uncured between both treatment groups (p-value of 0.054). In particular, they concluded that the autologous group has significantly higher hazard than the allogeneic group with an estimated hazard ratio of 1.88, p-value 0.04, and 95% confidence interval (1.03,3.45). On the other hand, with the semiparametric logistic-AFT model, no statistically significant difference is detected between the two groups (p-values of 0.42 for the incidence and 0.19 for the latency). However, the authors surmised that these non-significant results are due to a bad fit of the logistic-AFT model, as the logistic-Cox model found a borderline significant effect of the treatment group.

Let us briefly summarize the data sets. The censoring percentages amounted to 28% and 20% in the allogeneic and autologous groups, respectively, i.e. 13 and 9 patients in absolute numbers. The percentages of data points in plateaus were 15% and 16% and the estimated cure fractions $\hat{p}_i = \hat{S}_i(\tau_{0,i})$ were 26% and 19%, respectively. The test of [25] for the equality of cure fractions resulted in a non-significant p-value of 0.453.

Figure 3 shows an illustration of the Kaplan–Meier curves. It shows that the curves are crossing and very close to each other in during the first weeks. After that, the curve for the autologous group clearly stays below the one for the allogeneic group. However, that discrepancy melts down as time progresses, hence the non-significant p-value for the equality of cure fractions.

On the other hand, the difference in estimated mean survival times for the uncured patients amounts to $\hat{m} = 129$. The 95% confidence intervals for those mean survival differences were [3, 255] days (asymptotic) and [1, 255] days (permutation). The two-sided hypothesis tests for equal mean survival times of the uncured resulted in the p-values 0.045 (asymptotic) and 0.046 (permutation), i.e., just significant at the significance level $\alpha = 5\%$. The random permutation-based inference methods have been run with 5,000 iterations. The asymptotic and the permutation-based inference methods thus agreed on their outcomes, despite the rather small sample sizes. In view of the rather low censoring rates and the simulation results for moderate censoring settings presented in Section 4.1, we deem all applied inference methods reliable.

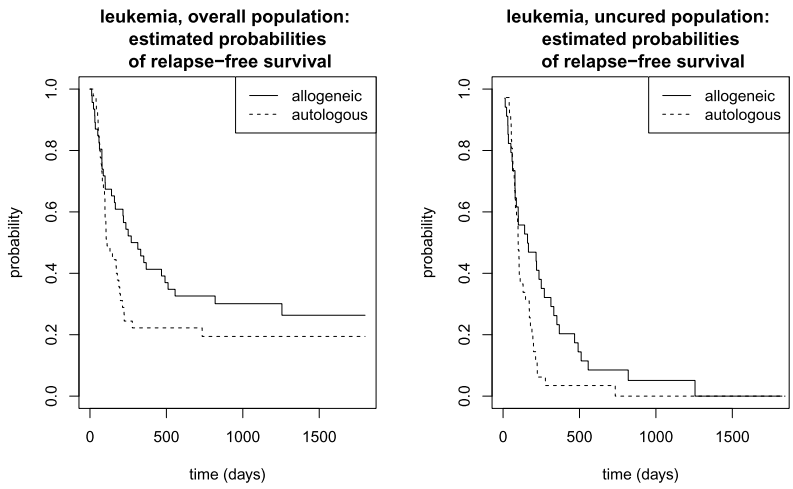

FIG 3. *Kaplan–Meier curves for relapse-free survival of the leukemia data.*

5.2. Breast cancer study

In this section, we analyze a data set about breast cancer that is freely available. In Appendix D, we provide the R code for accessing the data set. The data come from an observational study that included 286 lymph-node-negative breast cancer patients collected between 1980 and 1995. Thereof, 209 patients were estrogen-receptor-positive (ER+) and 77 were ER-negative (ER-). These two will later form the subgroups to be analyzed in a two-sample inference problem. As additional covariates, we consider the patients' age (ranging from 26 to 83 with a median of 52 years) and a tumour size score which is an integer number between 1 and 4. We refer to [56] for a more complete description of the study and other specifics of the dataset. Additionally, [3] compared this dataset in the light of two competing models and corresponding statistical methods: a logistic-Cox cure model versus a Single-Index/Cox model.

We, on the other hand, do not model the ER-status semiparametrically but nonparametrically by means of two subgroups. Our aim is to conduct a regression analysis to investigate differences in disease progression expectations for ER+/- patients while taking the covariates tumour size (ordinal) and age into account. The outcomes of this two-sample analysis could be used to justify why the two groups should not be pooled, and how or how not to model the ER-status semiparametrically. These questions are relevant if one wishes to make predictions, e.g., for the remaining expected lifetime of a patient.

From a technical point of view, we consider the composite endpoint of relapse-free survival (measured in months), which here means that deaths and the occurrence of distant metastases are combined into one event of interest. We excluded those 8 patients from our analysis who had a tumour size exceeding 2. This results in two samples of sizes $n_1 = 203$ (ER+) and $n_2 = 75$ (ER-).

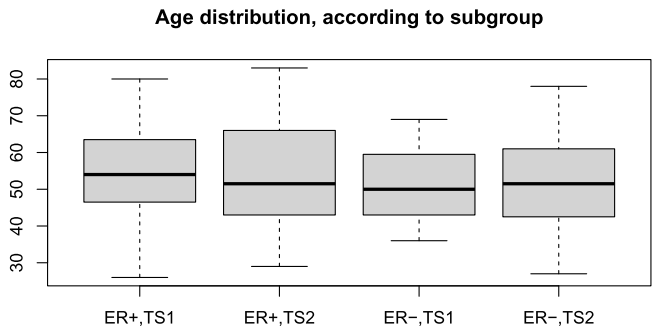

FIG 4. *Boxplots summarizing the age distributions in the four relevant subgroups. TS is short for tumour size.*

Let us briefly summarize the data: in the ER+ subgroup, about 55% have a tumour size score of 1, as opposed to 47% in the ER- subgroup. The age distributions in the four subgroups (ER+/-, tumour size score 1/2) are generally similar (rather symmetric, no outliers), although the patients with ER- and smaller tumour sizes exhibit a smaller dispersion in age; see Figure 4. For both groups, the latest uncensored events were observed after 80 and 48 months, respectively. The censoring rates amount to 62% and 64%, respectively, and the majority of censorings occurred in the plateau. Thus, there is sufficient follow-up. Apart from the plateau of the Kaplan–Meier estimator, the possibility of cure for breast cancer is recognized even from a medical perspective, see for example [8, 6, 23]. It is also mentioned in the initial study of the data [56] that around 60–70% of patients with lymph-node-negative breast cancer are cured.

Figure 5 shows the nonparametric Kaplan–Meier estimates for relapse-free survival for the subsamples of ER+ and ER- patients. These two curves are crossing twice: once, but insignificantly, soon after the time origin, and once again after the last observed event in the ER- subgroup. These crossings underline that the classical proportional hazards model [13] might not be appropriate for a combined modeling of all these data within a single, extended logistic-Cox cure model: such a model would contradict crossing Kaplan–Meier curves as seen in Figure 5 (after rescaling both curves to exhibit the same cure rate).

Our present approach is to compare the outcomes of two independently fitted logistic-Cox models for both sample groups ER+/-. We have first checked the proportional hazards assumption for the latency parts of the model by means of the test proposed in [41]. The resulting p-values are 1 and 0.96, respectively, for the covariates age and tumour size in sample 1; 1 and 0.81, respectively, for the covariates age and tumour size in sample 2. Thus, there is no reason to reject the proportional hazards assumption. We would also like to point out that we do not rely on the proportional hazards assumption for ER status, which is nonparametrically modeled in terms of two separate subgroups. We also test the logistic model assumption for the incidence component using the goodness-of-fit test proposed in [35]. The resulting p-values, estimated via bootstrap, were 0.982

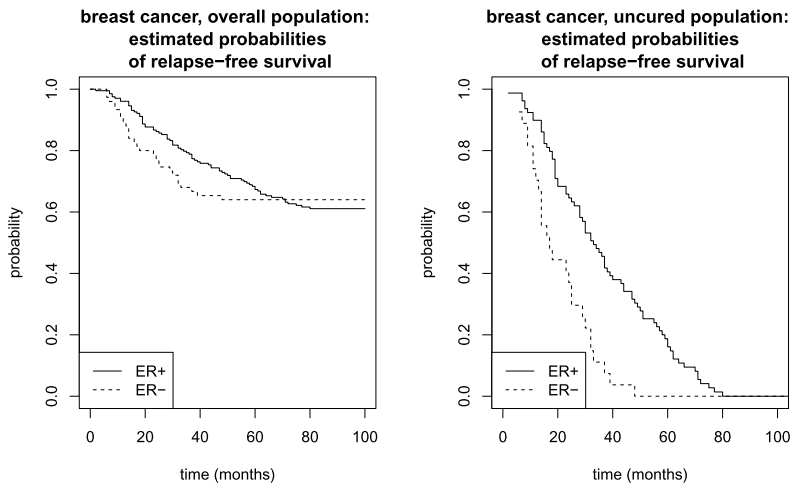

FIG 5. *Nonparametric Kaplan–Meier curves for the estrogen receptor positive (ER+) and negative (ER-) subgroups (time in months) of the breast cancer data.*

TABLE 18

Point estimates, standard errors, and p-values of the parametric model components rounded to three decimal places; values obtained from the `smcure` *package in R.*

ER	γ_0	γ_1	γ_2	β_1	β_2
+	−0.838(0.438)	−0.018(0.013)	0.272(0.296)	−0.014(0.011)	0.482(0.260)
	$p = 0.056$	$p = 0.158$	$p = 0.359$	$p = 0.206$	$p = 0.063$
-	0.303(0.754)	0.011(0.021)	−0.568(0.505)	0.010(0.026)	0.437(0.505)
	$p = 0.688$	$p = 0.591$	$p = 0.261$	$p = 0.708$	$p = 0.387$

for sample 1 and 0.988 for sample 2, showing no evidence against the logistic model assumption.

Thus continuing with the two independently fitted logistic-Cox models, Table 18 contains all point estimates of the parametric model components. Note that none of the covariates was found to have a significant influence on any of the two models. From the point estimates, we also see that for ER+ patients age seems protective for both, incidence ($\gamma_1 < 0$) and latency ($\beta_1 < 0$), and generally harmful for ER- ($\gamma_1, \beta_1 > 0$). Also, a bigger tumour size is generally harmful for ER+ patients ($\gamma_2, \beta_2 > 0$) but, for ER- patients, it has a protective influence on incidence ($\gamma_2 < 0$) although a harmful influence on the latency ($\beta_2 > 0$). Testing whether there is a significant difference of any of these parameters from 0 can be easily achieved by studentizing the parameter point estimates and comparing the results with quantiles from the standard normal distribution or those from a corresponding permutation version of the studentized parameter estimates (see Table 18).

Just like for the nonparametric test, the results from the semiparametric model were unambiguous (see Table 19): for all considered covariate combinations, uncured patients with ER+ have, at level $\alpha = 5\%$, a significantly larger

TABLE 19

p-values for two-sided testing of equal mean survival times, $H_0 : m_z = 0$, and corresponding 95%-confidence intervals for the differences (rounded to full months), for different covariate combinations. TS is short for tumour size and ±10 y refers to the mean age 53.9 years plus or minus ten years.

		point estimate	hypothesis test	confidence interval	
$X = Z$	method	$\hat{m}_z$	p-value	lower	upper
TS1, −10 y	asymptotic	15	0.019	2	27
	permutation	15	0.020	2	27
TS1, mean age	asymptotic	18	< 0.001	9	27
	permutation	18	< 0.001	9	27
TS1, +10 y	asymptotic	22	< 0.001	13	31
	permutation	22	< 0.001	13	31
TS2, −10 y	asymptotic	11	0.014	2	19
	permutation	11	0.032	0	19
TS2, mean age	asymptotic	14	0.001	6	23
	permutation	14	0.008	4	22
TS2, +10 y	asymptotic	17	0.002	6	28
	permutation	17	0.012	5	28

expected mean survival time than uncured patients with ER-. This difference seems to grow with progressing age, especially for patients with tumour size 1. This is in line with the parameter estimates related to age and latency. The influence of the tumour size on the mean survival difference is not that obvious. All in all, both the asymptotic and the permutation-based method resulted in very similar outcomes.

6. Discussion

In this article, we considered a two-sample comparison of survival data in the presence of a cure fraction. In such situations, instead of just looking at the overall survival function, it is more informative to compare the cured fractions and the survival of the uncured subpopulations. We propose the use of the mean survival time as estimand, i.e., a summary measure of the survival curve for the uncured subjects since it is easy to interpret. The mean survival time of an uncured subpopulation is closely related to the restricted mean survival time (RMST) which is popular in the survival literature without a cure fraction.

We introduced a nonparametric estimator of the mean survival time for the uncured in absence of covariates and a semiparametric estimator for the conditional MST_u that allows to adjust for potential confounders. Asymptotic and permutation-based approaches were developed for inference on the difference between the MST_u. In the nonparametric setting, based on our simulation results, both methods were quite reliable, with the permutation approach being recommended particularly for small sample sizes. However, more caution is required when applying the methods in the presence of a high censoring or cure rate, for which larger sample sizes are needed to obtain reliable results. In the semiparametric setting, we encountered several theoretical and computational challenges related to the permutation approach in a semi-parametric model.

Moreover, based on our simulation study, we did not observe a clear advantage of the permutation method over the asymptotic one, contrarily to existing findings in the literature for the nonparametric setting.

The MST_u is useful in assessing whether there is a difference in the survival times of the uncured among the two groups. However, we would like to point out that, even within a randomized controlled trial, MST_u does not necessarily have a causal interpretation as a direct effect of the treatment on survival of the uncured. This is because of the conditioning on being uncured; the uncured subpopulations between the two treatment arms might, in general, fail to be comparable. In contrast, the approach for causal estimands in the presence of a cure fraction in [57] does not refer specifically to the uncured subpopulation.

The following situation exemplifies the circumstances of our approach: it might be that treatment A is more beneficial in curing patients compared to treatment B and those who do not get cured with treatment A are the patients with worse condition. As a result, the survival of the uncured for treatment A might be worse compared to treatment B, but that does not mean that treatment A shortens the survival of the uncured. However, the fact the MST_u does not have an interpretation as a direct causal effect on the uncured is not a problem when the goal is to choose which treatment should be preferred. Randomized clinical trial data allow us to understand the causal effect of the treatment on the joint distribution of $\mathbb{1}_{T=\infty}$ and $T\mathbb{1}_{T<\infty}$. One can then in practice define a utility function, for example $w_1\mathbb{1}_{T=\infty} + w_2 T\mathbb{1}_{T<\infty}$ for certain weights w_1, w_2 that represent whether the curative effect or the life prolonging one is more important. Maximizing the expected utility reduces to choosing the treatment that maximizes $w_1\mathbb{P}(T=\infty) + w_2\mathbb{P}(T<\infty)MST_u$ (similarly also conditionally on a given covariate). On the other hand, if one is interested only in the causal effect of the treatment on the uncured subpopulation, determining the relevant quantity is not straightforward and extra caution is required since we are conditioning on a post-treatment variable, which cannot be directly intervened upon. If all the variables that can affect both the cure status and the survival time of the uncured were observed, one can condition on those to estimate the effect of treatment on the uncured as in our conditional semiparametric approach.

We would also like to point out that, instead of comparing the MST_u's over some time horizon $[0, \tau_0]$, trivial adjustments of the methods and proofs would allow for comparing the mean residual survival times as an alternative estimand. Similarly as in [10], these are defined as $MRST_{u,t} = \mathbb{E}[T - t \mid t < T < \infty]$. Moreover, here we considered the standard cure model setting where it is not possible to identify the cured subjects. Another interesting extension of this work would be to allow for partially observed cure status, as done for example in [45] in the context of nonparametric latency estimation.

Furthermore, as one referee pointed out, it is straightforward to extend the present theory to estimands of the form

$$\int_0^{\tau_{0,i}} S_{u,i}(s)\phi(s)ds \tag{9}$$

for a bounded and continuous weight function ϕ. For example, the choice $\phi(s) =$

$2s$ would lead to the second moments of the survival times of the uncured patients; higher moments could be expressed as well. Such estimands seem particularly important whenever the focus is not (only) on the mean but (also) on other aspects of a population. Tests for homoskedasticity, i.e., equal variances across both treatment groups, could be an interesting application, after some more theoretical preparations; note that the variances are not of the form (9) due to the subtraction of the squared mean. This could be considered in future research.

In the semiparametric setting, we employed a logistic-Cox mixture cure model because it is the most widely used semiparametric cure model. However, the logistic model for the incidence could be replaced by any other parametric model that might be more suitable in specific applications. Also, the Cox model for the latency component could for example be replaced by an accelerated failure time model; this would have the advantage that covariate-dependent terminal times $\tau_{0,i}$ of the disease occurrence could be modeled. In such cases, both the estimators and the theory would need some adjustments but we expect that the same challenges would arise and the practical performance would be similar to the current model. Another possibility would be to allow for covariates without imposing a specific model on the latency. For example, [39] focused on estimating the cure rate, while leaving the distribution of the uncured unspecified, and they obtained a nonparametric estimator for the conditional survival function of the uncured. Another nonparametric estimator was proposed in [30], relying on the Beran estimator for the conditional survival function. It would be of interest to further extend our method to these nonparametric settings, which are more robust towards model misspecification but require choice of tuning parameters and suffer from the 'course-of-dimensionality' in presence of multiple covariates.

Apart from the permutation approach for inference, one could also consider a pooled bootstrap approach; cf. Section 3.7.2 of [55]. However, even with the pooled bootstrap we would encounter the same theoretical and computational problems and we expect that the practical behavior would be similar to the permutation approach. Also, one would lose the benefit of the finite sample exactness of the permutation-based inference procedures under exchangeability.

Finally, to avoid the model misspecification issue for the permutation approach in the semiparametric setting, one could estimate a mixed model instead of a logistic-Cox model in the permutation samples, i.e., fitting the correct model of the pooled data, which is a linear combination of two logistic-Cox models. Afterwards, we could estimate the MST as usual. Then one would still need to develop the asymptotic results theoretically for the resulting estimator; these asymptotics will not be the same as in the standard logistic-Cox model. In practice, we do not expect this approach to behave better since more parameters need to be estimated for each permuted sample.

Appendix A: Proofs for Section 2

Proof of Theorem 2.1. Since the KM estimator remains constant after the last observed event time, we have $\hat{p}_i = \hat{S}_i(\tau_{0,i})$ and

$$\widehat{MST}_{u,i} = \int_0^{\tau_{0,i}} \frac{\hat{S}_i(s) - \hat{S}_i(\tau_{0,i})}{1 - \hat{S}_i(\tau_{0,i})} \, \mathrm{d}s.$$

Hence $\widehat{MST}_{u,i} - MST_{u,i} = \psi(\hat{S}_i) - \psi(S_i)$, where

$$\psi : \tilde{D}[0, \tau_{0,i}] \to \mathbb{R} \qquad \psi(\theta) = \int_0^{\tau_{0,i}} \frac{\theta(s) - \theta(\tau_{0,i})}{1 - \theta(\tau_{0,i})} \, \mathrm{d}s$$

and

$$\tilde{D}[0, \tau_{0,i}] = \{f \in D[0, \tau_{0,i}] \mid \sup_{t \in [0, \tau_{0,i}]} |f(t)| < 1\},$$

where $D[0, \tau_{0,i}]$ is the space of càdlàg functions on the interval $[0, \tau_{0,i}]$, equipped with the Skorohod topology. We will derive the asymptotic distribution using the functional delta method based on standard limit results for the KM estimator.

First, we consider for simplicity the case of a continuous distribution F_i and assume that either condition a) or b) of the theorem holds. From results of [19, 61], we have that the stochastic process $\sqrt{n_i}\{\hat{S}_i(t) - S_i(t)\}$ converges weakly in $D[0, \tau_{0,i}]$ to the process $S_i(t) \cdot B(v_i(t))$ where B is a standard Brownian motion and v_i is defined as in (2). Returning to the delta method, the function ψ is Hadamard-differentiable tangentially to $C[0, \tau_{0,i}]$ with derivative $\mathrm{d}\psi_\theta$ given by

$$\mathrm{d}\psi_\theta \cdot h = h(\tau_{0,i}) \int_0^{\tau_{0,i}} \frac{\theta(s) - \theta(\tau_{0,i})}{\{1 - \theta(\tau_{0,i})\}^2} \, \mathrm{d}s + \int_0^{\tau_{0,i}} \frac{h(s) - h(\tau_{0,i})}{1 - \theta(\tau_{0,i})} \, \mathrm{d}s.$$

Here $C[0, \tau_{0,i}] \subset D[0, \tau_{0,i}]$ is the space of continuous functions on the interval $[0, \tau_{0,i}]$, equipped with the topology induced by the supremum norm. By Theorem 3.9.4 in [55], we conclude that $\sqrt{n_i}\{\widehat{MST}_{u,i} - MST_{u,i}\}$ converges weakly to

$$\begin{aligned} N_i &= S_i(\tau_{0,i}) B(v_i(\tau_{0,i})) \int_0^{\tau_{0,i}} \frac{S_i(s) - S_i(\tau_{0,i})}{\{1 - S_i(\tau_{0,i})\}^2} \, \mathrm{d}s \\ &\quad + \int_0^{\tau_{0,i}} \frac{S_i(s) B(v_i(s)) - S_i(\tau_{0,i}) B(v(\tau_{0,i}))}{1 - S_i(\tau_{0,i})} \, \mathrm{d}s \\ &= p_i B(v_i(\tau_{0,i})) \int_0^{\tau_{0,i}} \frac{S_i(s) - p_i}{(1 - p_i)^2} \, \mathrm{d}u + \int_0^{\tau_{0,i}} \frac{S_i(s) B(v(s)) - p_i B(v_i(\tau_{0,i}))}{1 - p_i} \, \mathrm{d}s \end{aligned}$$

The variable N_i is normally distributed with mean zero and variance

$$\begin{aligned} \sigma_i^2 &= \int_0^{\tau_{0,i}} \int_0^{\tau_{0,i}} \frac{S_i(s) S_i(t)}{(1 - p_i)^2} v_i(s \wedge t) \, \mathrm{d}s \, \mathrm{d}t + \frac{p_i^2}{(1 - p_i)^2} (MST_{u,i} - \tau_{0,i})^2 v_i(\tau_{0,i}) \\ &\quad + 2 \frac{p_i}{(1 - p_i)^2} (MST_{u,i} - \tau_{0,i}) \int_0^{\tau_{0,i}} S_i(s) v_i(s) \, \mathrm{d}s. \end{aligned} \tag{10}$$

If F_i is not continuous and either condition a) or c) of the theorem is satisfied, we only have weak convergence in $D[0,\tau_i]$ of the stopped process

$$\begin{aligned}\hat{L}_i(t) &= \sqrt{n_i}\frac{S_i(t)}{S_i(t\wedge Y_{i,(n)})}\left\{\hat{S}_i(t\wedge Y_{i,(n)}) - S_i(t\wedge Y_{i,(n)})\right\}\\ &= \sqrt{n_i}\left\{\frac{S_i(t)\hat{S}_i(t\wedge Y_{i,(n)})}{S_i(t\wedge Y_{i,(n)})} - S_i(t)\right\}\end{aligned}$$

where $Y_{i,(n)}$ denotes the largest observation in group i (see for example Theorem 3.14 in [34]). If $\tau_{0,i} < \tau_i$, then $Y_{i,(n)} > \tau_{0,i}$ with probability converging to one, which leads to the uniform convergence of $\sqrt{n_i}\{\hat{S}_i(t) - S_i(t)\}$ on $D[0,\tau_{0,i}]$ as in part a). Otherwise, if $\tau_{0,i} = \tau_i$, applying the Delta method as before, we would obtain the limit distribution of

$$\sqrt{n_i}\left\{\psi(\hat{Q}_i) - \psi(S_i)\right\}, \qquad Q(t) = \frac{S_i(t)\hat{S}_i(t\wedge Y_{i,(m_i)})}{S_i(t\wedge Y_{i,(m_i)})}$$

It then remains to deal with the difference $\sqrt{n_i}\left\{\psi(\hat{S}_i) - \psi(\hat{Q}_i)\right\}$ and show that it converges to zero. For this one can use Lemma 3 in [61], which does not require continuity of F_i. □

Proof of Theorem 2.2. We begin by analyzing the asymptotic behaviour of

$$(W_1^\pi, W_2^\pi) = \sqrt{\frac{n_1 n_2}{n_1+n_2}}(\hat{S}_1^\pi - \hat{S}, \hat{S}_2^\pi - \hat{S})$$

as a random element of $(D[0,\tau_0])^2$, where $\tau_0 = \max(\tau_{0,1},\tau_{0,2})$ and $\hat{S}$ denotes the Kaplan–Meier estimator based on the pooled sample. We refer to Lemma 2 in the online supplementary material to [17] for the result that, as $\min(n_1,n_2)\to\infty$, the conditional distribution of (W_1^π, W_2^π) converges weakly on $(D[0,\tau_0])^2$ in probability to the distribution of the Gaussian process

$$((1-\kappa)S(\cdot)B(v(\cdot)), -\kappa S(\cdot)B(v(\cdot))).$$

Here, B again denotes a standard Brownian motion,

$$S(t) = \exp\Big(-\int_0^t \frac{\kappa(1-G_1(u-))\mathrm{d}F_1(u) + (1-\kappa)(1-G_2(u-))\mathrm{d}F_2(u)}{\kappa(1-G_1(u-))S_1(u-) + (1-\kappa)(1-G_2(u-))S_2(u-)}\Big)$$

is the limit of the pooled Kaplan–Meier estimator, and

$$v(t) = \int_0^t \frac{\kappa(1-G_1(u-))\mathrm{d}F_1(u) + (1-\kappa)(1-G_2(u-))\mathrm{d}F_2(u)}{\{\kappa(1-G_1(u-))S_1(u-) + (1-\kappa)(1-G_2(u-))S_2(u-)\}^2}.$$

Now, write $\widehat{MST}_{u,i}^\pi = \psi(\hat{S}_i^\pi)$ and $\widehat{MST}_u = \psi(\hat{S})$ for the estimated mean survival time based on the pooled sample.

We pointed out the Hadamard-differentiability of ψ in the proof of Theorem 2.1 above. This functional is even uniformly Hadamard-differentiable on suitable subspaces, as we will point out in a forthcoming paper. Thus, a simple extension of Theorem 3.9.11 in [55] to the random permutation case applies and it follows that, conditionally on the data,

$$\sqrt{n_1 n_2/(n_1+n_2)}(\widehat{MST}_{u,1}^{\pi} - \widehat{MST}_u, \widehat{MST}_{u,2}^{\pi} - \widehat{MST}_u)$$

converges weakly to the following two-dimensional Gaussian random vector, say, (N_1, N_2) in probability:

$$\begin{aligned}
&\Big[S(\tau_0)B(v(\tau_0))\int_0^{\tau_0}\frac{S(s)-S(\tau_0)}{\{1-S(\tau_0)\}^2}\,\mathrm{d}s \\
&\quad+\int_0^{\tau_0}\frac{S(s)B(v(s))-S(\tau_0)B(v(\tau_0))}{1-S(\tau_0)}\,\mathrm{d}s\Big]\cdot(1-\kappa,-\kappa) \\
&=\Big[pB(v(\tau_0))\int_0^{\tau_0}\frac{S(s)-p}{\{1-p\}^2}\,\mathrm{d}s \\
&\quad+\int_0^{\tau_0}\frac{S(s)B(v(s))-pB(v(\tau_0))}{1-p}\,\mathrm{d}s\Big]\cdot(1-\kappa,-\kappa),
\end{aligned}$$

where $p = S(\tau_0)$. Finally, we take the difference of both entries of the pair to conclude that the weak limit of the conditional distribution of

$$\sqrt{n_1 n_2/(n_1+n_2)}(\widehat{MST}_{u,1}^{\pi} - \widehat{MST}_{u,2}^{\pi})$$

is normal with mean zero and variance

$$\begin{aligned}
\sigma^{\pi 2} = &\int_0^{\tau_0}\int_0^{\tau_0}\frac{S(s)S(t)}{(1-p)^2}v(s\wedge t)\,\mathrm{d}s\,\mathrm{d}t+\frac{p^2}{(1-p)^2}(MST_u-\tau_0)^2v(\tau_0) \\
&+2\frac{p}{(1-p)^2}(MST_u-\tau_0)\int_0^{\tau_0}S(s)v(s)\,\mathrm{d}s.
\end{aligned}\tag{11}$$

As is apparent from a comparison of (10) and (11), both limit variances coincide up to sample size-related factors if both samples are exchangeable. □

Appendix B: Proofs for Section 3

Proof of Theorem 3.1. Let $\theta_i = (\gamma_i, \beta_i)$ denote the vector of parameters of the semiparametric mixture cure model. Consider the space $\mathcal{H}_m = \{h = (h_1, h_2) \in BV[0, \tau_{0,i}] \times \mathbb{R}^{p+q} : \|h_1\|_v + \|h_2\|_1 \le m\}$, where $m < \infty$ and $\|h_1\|_v$ is the absolute value of $h_1(0)$ plus the total variation of h_1 on the interval $[0, \tau_{0,i}]$. From Theorem 3 in [31] we have that the process

$$\langle n_i^{1/2}(\hat{\Lambda}_i - \Lambda_i), n_i^{1/2}(\hat{\theta}_i - \theta_i)\rangle(h) = n_i^{1/2}\int_0^{\tau_{0,i}} h_1(s)\,\mathrm{d}(\hat{\Lambda}_i - \Lambda_i)(s) + n_i^{1/2}h_2^T(\hat{\theta}_i - \theta_i)\tag{12}$$

indexed by $h \in \mathcal{H}_m$ converges weakly in $l^\infty(\mathcal{H}_m)$ to a tight Gaussian process G_i in $l^\infty(\mathcal{H}_m)$ with mean zero and covariance process

$$Cov(G_i(h), G_i(h^*)) = \int_0^{\tau_{0,i}} h_1(s)\sigma^{-1}_{(1),i}(h^*)(s)\mathrm{d}\Lambda_i(s) + h_2^T\sigma^{-1}_{(2),i}(h^*)$$

where

$$\sigma_{(1),i}(h)(s) = \mathbb{E}\left[\mathbb{1}_{\{Y_{i1}\geq s\}}V_i(s;\theta_i,\Lambda_i)(h)g_i(s;\theta_i,\Lambda_i)e^{\beta_i^T Z_{i1}}\right]$$
$$-\mathbb{E}\left[\int_s^{\tau_{0,i}} \mathbb{1}_{\{Y_{i1}\geq u\}}V_i(s;\theta_i,\Lambda_i)(h)g_i(u;\theta_i,\Lambda_i)\{1-g_i(u;\theta_i,\Lambda_i)\}e^{2\beta_i^T Z_{i1}}\mathrm{d}\Lambda_i(u)\right], \tag{13}$$

$$\sigma_{(2),i}(h) = \mathbb{E}\left[\int_0^{\tau_{0,i}} \mathbb{1}_{\{Y_{i1}\geq s\}}W_i(s;\theta_i,\Lambda_i)V_i(s;\theta_i,\Lambda_i)(h)g_i(s;\theta_i,\Lambda_i)e^{\beta_i^T Z_{i1}}\mathrm{d}\Lambda_i(s)\right] \tag{14}$$

and

$$g_i(s;\theta,\Lambda) = \frac{\phi(\gamma^T X_{i1})\exp\left(-\Lambda(s)\exp\left(\beta^T Z_{i1}\right)\right)}{1-\phi(\gamma^T X_{i1})+\phi(\gamma^T X_{i1})\exp\left(-\Lambda(s)\exp\left(\beta^T Z_{i1}\right)\right)},$$
$$V_i(s;\theta_i,\Lambda_i)(h) = h_1(s) - \{1-g_i(s;\theta_i,\Lambda_i)\}e^{\beta_i^T Z_{i1}} \times \int_0^s h_1(u)\mathrm{d}\Lambda_i(u) + h_2^T W_i(s;\theta_i,\Lambda_i)$$
$$W_i(s;\theta_i,\Lambda_i) = \left(\{1-g_i(s;\theta_i,\Lambda_i)\}X_{i1}^T, \left[1-\{1-g_i(s;\theta_i,\Lambda_i)\}e^{\beta_i^T Z_{i1}}\Lambda_i(s)\right]Z_{i1}^T\right)^T.$$

Using this result we first obtain weak convergence of the process $\sqrt{n}\{\hat{S}_{u,i}(t|z) - S_{u,i}(t|z)\}$ for fixed $z \in \mathcal{Z}$. By definition and a series of Taylor expansions we have

$$\begin{aligned}
&\sqrt{n_i}\{\hat{S}_{u,i}(t|z) - S_{u,i}(t|z)\}\\
&= \sqrt{n_i}\left\{\exp\left(-\hat{\Lambda}_i(t)e^{\hat{\beta}_i^T z}\right) - \exp\left(-\Lambda_i(t)e^{\beta_i^T z}\right)\right\}\\
&= -\sqrt{n_i}\left\{\hat{\Lambda}_i(t)e^{\hat{\beta}_i^T z} - \Lambda_i(t)e^{\beta_i^T z}\right\}\exp\left(-\Lambda_i(t)e^{\beta_i^T z}\right) + R_1\\
&= -\sqrt{n_i}\left\{\hat{\Lambda}_i(t) - \Lambda_i(t)\right\}e^{\beta_i^T z}\exp\left(-\Lambda_i(t)e^{\beta_i^T z}\right)\\
&\quad - \sqrt{n_i}\left\{e^{\hat{\beta}_i^T z} - e^{\beta_i^T z}\right\}\Lambda_i(t)\exp\left(-\Lambda_i(t)e^{\beta_i^T z}\right) + R_1 + R_2\\
&= -\sqrt{n_i}\left\{\hat{\Lambda}_i(t) - \Lambda_i(t)\right\}e^{\beta_i^T z}\exp\left(-\Lambda_i(t)e^{\beta_i^T z}\right)\\
&\quad - \sqrt{n_i}\left\{\hat{\beta}_i - \beta_i\right\}^T ze^{\beta_i^T z}\Lambda_i(t)\exp\left(-\Lambda_i(t)e^{\beta_i^T z}\right) + R_1 + R_2 + R_3
\end{aligned}$$

where the remainder terms R_1, R_2, R_3 converge uniformly to zero in probability because of the boundedness of Λ_i, β_i and Theorems 2-3 in [31]. Note that, as in assumption (A1), we denote by $S_{u,i}(\tau_{0,i}|z)$ the left limit $S_{u,i}(\tau_{0,i} - |z)$ so that

$S_{u,i}(\cdot|z)$ is a continuous function, bounded from below away from zero. This modification does not influence the value of $MST_{u,i,z}$.
Consider the functions $h \in \mathcal{H}_m$ of the form

$$\begin{aligned} h_{t,z} &= (h_{1;t,z}, h_{2;t,z}) \\ &= \left(\mathbb{1}_{[0,t]}(\cdot) e^{\beta_i^T z} \exp\left(-\Lambda_i(t) e^{\beta_i^T z}\right), \left(\mathbf{0}_p, z e^{\beta_i^T z} \Lambda_i(t) \exp\left(-\Lambda_i(t) e^{\beta_i^T z}\right) \right) \right), \end{aligned} \tag{15}$$

where $\mathbf{0}_p$ denotes a zero vector in $\mathbb{R}^p$ since we are not interested in the γ component. For an appropriate choice of m and any $t \in [0, \tau_{0,i}]$, such functions belong to $\mathcal{H}_m$ because of assumptions (A1)–(A2). For these functions we have

$$\sqrt{n_i}\{\hat{S}_{u,i}(t|z) - S_{u,i}(t|z)\} = \langle n_i^{1/2}(\hat{\Lambda}_i - \Lambda_i), n^{1/2}(\hat{\theta}_i - \theta_i) \rangle (h_{t,z}) + o_P(1)$$

from which we conclude that the process $\sqrt{n}\{\hat{S}_{u,i}(t|z) - S_{u,i}(t|z)\}$ converges weakly in $D[0, \tau_{0,i}]$ to a mean zero Gaussian process $G^*_{i,z}$ with covariance

$$\begin{aligned} \rho_{i,z}(t, t^*) &= Cov(G^*_{i,z}(t), G^*_{i,z}(t^*)) \\ &= \int_0^{\tau_{0,i}} h_{1;t,z}(s) \sigma^{-1}_{(1),i}(h_{t^*,z})(s) \mathrm{d}\Lambda_i(s) + h^T_{2;t,z} \sigma^{-1}_{(2),i}(h_{t^*,z}), \end{aligned} \tag{16}$$

where $h_{t,z}$ is as in (15) and $\sigma_{(1),i}, \sigma_{(2),i}$ as in (13), (14). Next we use the delta method to obtain the asymptotic distribution of the conditional mean survival time. We have

$$\begin{aligned} \sqrt{n_i}(\widehat{MST}_{u,i,z} - MST_{u,i,z}) = \sqrt{n_i} \int_0^{\tau_{0,i}} \{\hat{S}_{u,i}(s|z) - S_{u,i}(s|z)\} \mathrm{d}s \\ - \sqrt{n_i}\{\tau_{0,i} - Y_{i,(m_i)}\} \hat{S}_{u,i}(\tau_{0,i}). \end{aligned} \tag{17}$$

The first term in the previous equation is equal to $\sqrt{n_i}\{\psi(\hat{S}_{u,i}) - \psi(S_{u,i})\}$ where

$$\psi : D[0, \tau_{0,i}] \to \mathbb{R} \qquad \psi(\xi) = \int_0^{\tau_{0,i}} \xi(s)\, \mathrm{d}s$$

The function ψ is Hadamard-differentiable with derivative

$$\mathrm{d}\psi_\xi \cdot h = \int_0^{\tau_{0,i}} h(s) \mathrm{d}s.$$

By Theorem 3.9.4 in [55] it follows that $\sqrt{n_i}\{\psi(\hat{S}_{u,i}) - \psi(S_{u,i})\}$ converges weakly to

$$N_i = \int_0^{\tau_{0,i}} G^*_{i,z}(s) \mathrm{d}s$$

which is normal distributed with mean zero and variance

$$\sigma^2_{i,z} = \int_0^{\tau_{0,i}} \int_0^{\tau_{0,i}} \rho_{i,z}(s,t)\, \mathrm{d}s\, \mathrm{d}t, \tag{18}$$

where $\rho_{i,z}$ is defined in (16). Next we show that the second term on the right hand side of (17) converges to zero in probability, from which we can conclude that the asymptotic distribution of $\sqrt{n_i}(\widehat{MST}_{u,i,z} - MST_{u,i,z})$ is determined by that of the first term. Since $\hat{S}_{u,i}(\tau_{0,i})$ converges to $S_{u,i}(\tau_{0,i})$, it is sufficient to show that $\sqrt{n_i}\{\tau_{0,i} - Y_{i,(m_i)}\} = o_P(1)$. This result have been proved in Lemma 5 of [30] under slightly different assumptions. Hence, for completeness we prove it again for our setting below. For any $\delta > 0$ we have

$$
\begin{aligned}
&\mathbb{P}(\sqrt{n_i}\{\tau_{0,i} - Y_{i,(m_i)}\} > \delta) \\
&= \mathbb{P}\left(Y_{i,(m_i)} < \tau_{0,i} - \frac{\delta}{\sqrt{n_i}}\right) \\
&= \mathbb{P}\left(\Delta_{i1} Y_{i1} < \tau_{0,i} - \frac{\delta}{\sqrt{n_i}}\right)^{n_i} \\
&= \left[1 - \mathbb{P}\left(\Delta_{i1} Y_{i1} \geq \tau_{0,i} - \frac{\delta}{\sqrt{n_i}}\right)\right]^{n_i} \\
&= \left[1 - \int_{\mathcal{Z}\times\mathcal{X}} \int_{\tau_{0,i} - \frac{\delta}{\sqrt{n_i}}}^{\tau_{0,i}} \mathbb{P}\left(C_{i1} \geq t | x, z\right) \mathrm{d}F_{T_{i1}|X_{i1},Z_{i1}}(t|x,z) \, \mathrm{d}F_{(Z_{i1},X_{i1})}(z,x)\right]^{n_i} \\
&\leq \left[1 - \int_{\mathcal{Z}\times\mathcal{X}} \mathbb{P}\left(C_{i1} > \tau_0 | x, z\right) \phi(\gamma_i^T x) \mathbb{P}(T_{i1} = \tau_{0,i} | T_{i1} < \infty, z) \, \mathrm{d}F_{(Z_{i1},X_{i1})}(z,x)\right]^{n_i}.
\end{aligned}
$$

Because of assumptions (I1),(I3) and (A3), for some $K > 0$ we have

$$
\begin{aligned}
&\mathbb{P}(\sqrt{n_i}\{\tau_{0,i} - Y_{i,(m_i)}\} > \delta) \\
&\leq [1 - K\epsilon]^{n_i} \to 0.
\end{aligned}
$$

This concludes the proof. □

In order to prove our main Theorem 3.2, we first need some preliminary results, which are provided in the following lemmas. In what follows, we assume that conditions (I1)–(I4) and (A1)–(A4) are satisfied.

Denote by $\hat{\Lambda}^{\pi}_{n_i,i}$ and $\hat{\theta}^{\pi}_{n_i,i}$ the estimators of Λ and $\theta = (\gamma, \beta)$ obtained by fitting a logistic-Cox model to the i-th permuted sample, which will be denoted for notational convenience as $(\Delta^{\pi}_{i1}, Y^{\pi}_{i1}, X^{\pi}_{i1}, Z^{\pi}_{i1}), \ldots, (\Delta^{\pi}_{in_i}, Y^{\pi}_{in_i}, X^{\pi}_{in_i}, Z^{\pi}_{in_i})$. Let $\bar{\Lambda}_{n_1+n_2}$ and $\bar{\theta}_{n_1+n_2}$ denote the estimators of Λ and θ based on the pooled sample $(\Delta_1, Y_1, X_1, Z_1), \ldots, (\Delta_{n_1+n_2}, Y_{n_1+n_2}, X_{n_1+n_2}, Z_{n_1+n_2})$. Note that the true distribution of the pooled sample is $\bar{\mathbb{P}} = \kappa\mathbb{P}_1 + (1-\kappa)\mathbb{P}_2$, where $\mathbb{P}_i$ denotes the distribution of the ith sample. In particular, $\mathbb{P}$ does not correspond to a logistic-Cox model. Let $Q_{\Lambda,\theta}$ be the corresponding distribution of a logistic-Cox model with parameters (Λ, θ) and corresponding log-likelihood

$$
\begin{aligned}
l(\delta, y, x, z; \Lambda, \theta) = \delta &\left\{\log \phi(\gamma^T x) + \log f_u(y|z; \Lambda, \beta)\right\} \\
&+ (1-\delta) \log\left\{1 - \phi(\gamma^T x) + \phi(\gamma^T x) S_u(y|z; \Lambda, \beta)\right\}
\end{aligned}
$$

By assumption (I2), the event times on the pooled sample happen on $[0, \bar{\tau}_0]$, where $\bar{\tau}_0 = \max\{\tau_{0,1}, \tau_{0,2}\}$. Hence $S_u(t|z; \Lambda, \beta) = 0$ for $t \geq \bar{\tau}_0$ which corresponds

to Λ being defined on $[0, \bar{\tau}_0)$. In addition, $\bar{\mathbb{P}}(\Delta = 1, Y = \bar{\tau}_0) > 0$ by assumptions (A3) and (I3). Hence, we can restrict on distributions $Q_{\Lambda,\theta}$ that have a positive mass at $\bar{\tau}_0$, meaning that $\lim_{t\to\bar{\tau}_0} \Lambda(t) < +\infty$ and we can denote the limit by $\Lambda(\bar{\tau}_0)$. To reflect the existence of the jump in the likelihood, for the terms with $\Delta = 1$ and $Y = \bar{\tau}$ we have $f_u(\bar{\tau}|z;\Lambda,\beta) = S_u(\bar{\tau} - |z;\Lambda,\beta) = \exp(-\Lambda(\bar{\tau}_0)e^{\beta^T z})$, instead of the usual expression $f_u(t|z;\Lambda,\beta) = \lambda_u(t|z;\beta)S_u(\bar{\tau} - |z;\Lambda,\beta)$. Here λ_u denotes the hazard function corresponding to S_u and the baseline hazard function corresponding to Λ will be denoted by λ. Define

$$(\bar{\Lambda}, \bar{\theta}) = \underset{\Lambda,\theta}{\operatorname{argmax}}\, \mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta, Y, X, Z; \Lambda, \theta)] = \underset{\Lambda,\theta}{\operatorname{argmin}}\, \mathrm{KL}(\bar{\mathbb{P}}|Q_{\Lambda,\theta}) \tag{19}$$

where $\mathrm{KL}(\cdot|\cdot)$ denotes the Kullback-Leibler divergence between two distributions.

Lemma B.1. *The argmax defined in* (19) *exists.*

Proof. We show that the argmax can be restricted on a bounded set, from which the existence follows because of continuity. In the three steps below we deal consequently with β, Λ and γ.

Step 1. First we show that for any $K > 0$ there exists $\bar{c} > 0$ such that for any $c \geq \bar{c}$ we have $\inf_{\tilde{\beta}\in S^{q-1}} \bar{\mathbb{P}}(c\,|\tilde{\beta}^T Z| > K) > 0$, where S^{q-1} is the unit circle in $\mathbb{R}^q$. Suppose by contradiction that there exists K such that for any c we have $\inf_{\tilde{\beta}\in S^{q-1}} \bar{\mathbb{P}}(c\,|\tilde{\beta}^T Z| > K) = 0$. Note that the infimum is actually a minimum because S^{q-1} is compact and the function is continuous. Hence, it means that for any c there exists $\tilde{\beta} \in S^{q-1}$ for which $\bar{\mathbb{P}}(c\,|\tilde{\beta}^T Z| \leq K) = 1$. Equivalently, for any $\epsilon > 0$, there exists $\tilde{\beta} \in S^{q-1}$ for which $\bar{\mathbb{P}}(|\tilde{\beta}^T Z| \leq \epsilon) = 1$. The closed subsets of S^{q-1} defined by $B_m = \{\tilde{\beta} \in S^{q-1} \,\big|\, \bar{\mathbb{P}}(|\tilde{\beta}^T Z| \leq \frac{1}{m}) = 1\}$ are non-empty for all m and $B_m \downarrow B = \cap_m B_m$. B cannot be empty because then $(B_m^c)_m$ form an open covering of the compact S^{q-1} and there would exists a finite sub-covering, which is impossible since all B_m are non-empty. It follows that B is not empty, which is equivalent to saying that there exists $\tilde{\beta} \in S^{q-1}$ for which $\bar{\mathbb{P}}(|\tilde{\beta}^T Z| = 0) = 1$. This contradicts the assumption that $Var(Z)$ has full rank.
Next, let $\eta = \frac{1}{2}\Lambda(\bar{\tau}_0)\inf_z \bar{\mathbb{P}}(Y = \bar{\tau}_0|\Delta = 1, Z = z)$ and choose K such that $x \leq \eta e^x$ for all $x \geq K$. Let $\beta = c\tilde{\beta}$ with $\tilde{\beta} \in S^{q-1}$, $c > \bar{c}$. We will show that, as c increases, the expectation in (19) becomes arbitrarily small. For fixed γ and Λ, we can write

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta, Y, X, Z; \Lambda, \theta)] \\
&= \mathbb{E}_{\bar{\mathbb{P}}}[\Delta\beta^T Z \mathbb{1}_{\{Y<\bar{\tau}_0\}} - \Delta\Lambda(Y)e^{\beta^T Z}] + R_1 \\
&= \mathbb{E}_{\bar{\mathbb{P}}}\left[\left\{\Delta c\tilde{\beta}^T Z \mathbb{1}_{\{Y<\bar{\tau}_0\}} - \Delta\Lambda(Y)e^{c\tilde{\beta}^T Z}\right\} \mathbb{1}_{\{\tilde{\beta}^T Z>0\}} \mathbb{1}_{\{c\,|\tilde{\beta}^T Z|>K\}}\right] \\
&+ \mathbb{E}_{\bar{\mathbb{P}}}\left[\left\{\Delta c\tilde{\beta}^T Z \mathbb{1}_{\{Y<\bar{\tau}_0\}} - \Delta\Lambda(Y)e^{c\tilde{\beta}^T Z}\right\} \mathbb{1}_{\{\tilde{\beta}^T Z<0\}} \mathbb{1}_{\{c\,|\tilde{\beta}^T Z|>K\}}\right] + R_2,
\end{aligned}$$

where R_1 and R_2 denote terms that are bounded in absolute value. Using $\mathbb{E}_{\bar{\mathbb{P}}}[\Lambda(Y)|Z, \Delta = 1] > \Lambda(\bar{\tau}_0)\inf_z \bar{\mathbb{P}}(Y = \bar{\tau}_0|\Delta = 1, Z = z) = 2\eta > 0$, we ob-

tain the following bound

$$\begin{aligned}
\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)] &\leq \mathbb{E}_{\bar{\mathbb{P}}}\left[\left\{\Delta c\tilde{\beta}^T Z - 2\Delta\eta e^{c\tilde{\beta}^T Z}\right\}\mathbb{1}_{\{\tilde{\beta}^T Z>0\}}\mathbb{1}_{\{c\,|\tilde{\beta}^T Z|>K\}}\right] \\
&\quad + c\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta\tilde{\beta}^T Z\mathbb{1}_{\{Y<\bar{\tau}_0\}}\mathbb{1}_{\{\tilde{\beta}^T Z<0\}}\mathbb{1}_{\{c\,|\tilde{\beta}^T Z|>K\}}\right] + R_2 \\
&\leq -\eta\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta e^{c\tilde{\beta}^T Z}\mathbb{1}_{\{\tilde{\beta}^T Z>0\}}\mathbb{1}_{\{c\,|\tilde{\beta}^T Z|>K\}}\right] \\
&\quad - c\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta|\tilde{\beta}^T Z|\mathbb{1}_{\{Y<\bar{\tau}_0\}}\mathbb{1}_{\{\tilde{\beta}^T Z<0\}}\mathbb{1}_{\{c\,|\tilde{\beta}^T Z|>K\}}\right] + R_2 \\
&\leq -\eta\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta e^{c\tilde{\beta}^T Z}\mathbb{1}_{\{\tilde{\beta}^T Z>0\}}\mathbb{1}_{\{\bar{c}\,|\tilde{\beta}^T Z|>K\}}\right] \\
&\quad - c\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta|\tilde{\beta}^T Z|\mathbb{1}_{\{Y<\bar{\tau}_0\}}\mathbb{1}_{\{\tilde{\beta}^T Z<0\}}\mathbb{1}_{\{\bar{c}\,|\tilde{\beta}^T Z|>K\}}\right] + R_2
\end{aligned}$$

This further leads to

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)] \\
&\leq -\eta e^{cK/\bar{c}}\bar{\mathbb{P}}\left(\Delta=1,\tilde{\beta}^T Z>0,\bar{c}\,|\tilde{\beta}^T Z|>K\right) \\
&- c\frac{K}{\bar{c}}\bar{\mathbb{P}}\left(\Delta=1,Y<\bar{\tau}_0,\tilde{\beta}^T Z<0,\bar{c}\,|\tilde{\beta}^T Z|>K\right)+R_2 \\
&\leq -c\frac{K}{\bar{c}}\bar{\mathbb{P}}\left(\Delta=1,Y<\bar{\tau}_0,\bar{c}\,|\tilde{\beta}^T Z|>K\right)+R_2 \\
&\leq -c\frac{K}{\bar{c}}\inf_z\bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0|Z=z)\inf_{\tilde{\beta}\in S^{q-1}}\bar{\mathbb{P}}\left(\bar{c}\,|\tilde{\beta}^T Z|>K\right)+R_2
\end{aligned}$$

Since both infimums are strictly positive, $\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)]$ can be made arbitrarily small for c sufficiently large (and how large c should be does not depend on $\tilde{\beta}$). Hence, we can restrict the argmax on a bounded set for β.

Step 2. Next we show that, there exists $M>0$ such that it suffices to search for the maximizer among Λ that are bounded by M. Let Λ be such that $\Lambda(\bar{\tau}_0)>M$. We can construct $\tilde{\Lambda}(t)=c\Lambda(t)$ with $c=M/\Lambda(\bar{\tau}_0)\in(0,1)$. We have $\tilde{\Lambda}(\bar{\tau}_0)=M$ and $\tilde{\lambda}=c\lambda$. We show that

$$\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)]<\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\tilde{\Lambda},\theta)].$$

Indeed we have

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)]-\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\tilde{\Lambda},\theta)] \\
&=\mathbb{E}_{\bar{\mathbb{P}}}\Big[-\Delta\log c\mathbb{1}_{\{Y<\bar{\tau}_0\}}-\Delta\Lambda(Y)e^{\beta^T Z}+\Delta c\Lambda(Y)e^{\beta^T Z} \\
&\qquad +(1-\Delta)\log\frac{1-\phi(\gamma^T X)+\phi(\gamma^T X)S_u(Y|Z;\Lambda,\beta)}{1-\phi(\gamma^T X)+\phi(\gamma^T X)S_u(Y|Z;\tilde{\Lambda},\beta)}\Big]
\end{aligned}$$

Since $S_u(Y|Z;\tilde{\Lambda},\beta)>S_u(Y|Z;\Lambda,\beta)$ the ratio is smaller than 1 and as a result

the $(1-\Delta)$ term in the expectation is negative. Hence

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)] - \mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\tilde{\Lambda},\theta)] \\
&< \mathbb{E}_{\bar{\mathbb{P}}}\left[-\Delta\log c\mathbb{1}_{\{Y<\bar{\tau}_0\}} - \Delta\Lambda(Y)e^{\beta^T Z} + \Delta c\Lambda(Y)e^{\beta^T Z}\right] \\
&= -\bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0)\log c - (1-c)\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta\Lambda(Y)e^{\beta^T Z}\right] \\
&\leq -\bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0)\log c - (1-c)\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta\Lambda(Y)e^{\beta^T Z}\mathbb{1}_{\{Y=\bar{\tau}_0\}}\right] \\
&= -\bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0)\log c - (1-c)\Lambda(\bar{\tau}_0)\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta e^{\beta^T Z}\mathbb{1}_{\{Y=\bar{\tau}_0\}}\right].
\end{aligned}$$

Since we are restricting β on a compact and Z is assumed to have bounded support, there exist $c_2>0$ such that $e^{\beta^T Z}>c_2$ a.s.. It follows that

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)] - \mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\tilde{\Lambda},\theta)] \\
&< -\bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0)\log c - (1-c)\Lambda(\bar{\tau}_0)c_2\bar{\mathbb{P}}(\Delta=1,Y=\bar{\tau}_0) \\
&= \bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0)(\log\Lambda(\bar{\tau}_0)-\log M) - (\Lambda(\bar{\tau}_0)-M)c_2\bar{\mathbb{P}}(\Delta=1,Y=\bar{\tau}_0) \\
&< \bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0)(\Lambda(\bar{\tau}_0)-M)\frac{1}{M} - (\Lambda(\bar{\tau}_0)-M)c_2\bar{\mathbb{P}}(\Delta=1,Y=\bar{\tau}_0) \\
&= (\Lambda(\bar{\tau}_0)-M)\left\{\frac{1}{M}\bar{\mathbb{P}}(\Delta=1,Y<\bar{\tau}_0) - c_2\bar{\mathbb{P}}(\Delta=1,Y=\bar{\tau}_0)\right\} < 0
\end{aligned}$$

for large enough M since $\bar{\mathbb{P}}(\Delta=1,Y=\bar{\tau})>0$ by assumption. Hence we conclude that, there exists M such that it is sufficient to search for the maximizer among Λ's bounded by M.

Step 3. We can also restrict the argmax on a bounded set for γ because as $\|\gamma\|\to\infty$, for fixed values of β and Λ, the expectation converges to $-\infty$. Indeed we have

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathbb{P}}}[l(\Delta,Y,X,Z;\Lambda,\theta)] \\
&= \mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta\log\phi(\gamma^T X)\mathbb{1}_{\{\gamma^T X>0\}}\right] + \mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta\log\phi(\gamma^T X)\mathbb{1}_{\{\gamma^T X\leq 0\}}\right] + R_3,
\end{aligned}$$

where R_3 denotes terms bounded in absolute value. The first term is bounded and using the same reasoning as with β it can be shown that the second term converges to $-\infty$.

We conclude that we can restrict the argmax on a bounded set, from which the existance of the argmax follows as the criteria is continuous with respect to the parameters. □

In what follows, we assume that the maximizer $(\bar{\Lambda},\bar{\theta})$ is unique. It will also be useful to characterize it as the solution of the score equation defined similarly to [31]. As in the proof of Theorem 3.1, consider $\mathcal{H}_m = \{h=(h_1,h_2)\in BV[0,\bar{\tau}_0]\times\mathbb{R}^{p+q} : \|h_1\|_v + \|h_2\|_1 \leq m\}$, where $m<\infty$ and $\|h_1\|_v$ is the absolute value of $h_1(0)$ plus the total variation of h_1 on the interval $[0,\bar{\tau}_0]$. Define the functions

$$\begin{aligned}
\psi_{(\Lambda,\theta),h}(\delta,y,x,z) = \delta\left[h_1(y)+h_{21}^T x + h_{22}^T z\right] - \left\{\phi(\gamma^T x) - (1-\delta)g(y,\Lambda,\theta)\right\}h_{21}^T x \\
- \left\{\delta+(1-\delta)g(y,\Lambda,\theta)\right\}\left\{e^{\beta^T z}\int_0^y h_1(s)\mathrm{d}\Lambda(s) + e^{\beta^T z}\Lambda(y)h_{22}^T z\right\},
\end{aligned}$$

where

$$g(t,\Lambda,\theta)=\frac{\phi(\gamma^T x)\exp\left(-\Lambda(t)\exp\left(\beta^T z\right)\right)}{1-\phi(\gamma^T x)+\phi(\gamma^T x)\exp\left(-\Lambda(t)\exp\left(\beta^T z\right)\right)}.$$

We will denote by $\mathbb{P}^{\pi}_{n_i,i}\psi_{(\Lambda,\theta),h}$ the score function for the i-th permuted sample

$$\begin{aligned}\mathbb{P}^{\pi}_{n_i,i}\psi_{(\Lambda,\theta),h}=&\frac{1}{n_i}\sum_{j=1}^{n_i}\Delta^{\pi}_{ij}\left[h_1(Y^{\pi}_j)+h_{21}^T X^{\pi}_{ij}+h_{22}^T Z^{\pi}_{ij}\right]\\&-\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{\phi(\gamma^T X^{\pi}_{ij})-(1-\Delta^{\pi}_{ij})g^{\pi}_{ij}(Y^{\pi}_{ij},\Lambda,\theta)\right\}h_{21}^T X^{\pi}_{ij}\\&-\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{\Delta^{\pi}_{ij}+(1-\Delta^{\pi}_{ij})g^{\pi}_{ij}(Y^{\pi}_{ij},\Lambda,\theta)\right\}\\&\qquad\times\left\{e^{\beta^T Z^{\pi}_{ij}}\int_0^{Y^{\pi}_{ij}}h_1(s)\mathrm{d}\Lambda(s)+e^{\beta^T Z^{\pi}_{ij}}\Lambda(Y^{\pi}_{ij})h_{22}^T Z^{\pi}_{ij}\right\},\end{aligned}$$

where $h=(h_1,h_2)=(h_1,h_{21},h_{22})\in\mathcal{H}_m$ and

$$g^{\pi}_{ij}(t,\Lambda,\theta)=\frac{\phi(\gamma^T X^{\pi}_{ij})\exp\left(-\Lambda(t)\exp\left(\beta^T Z^{\pi}_{ij}\right)\right)}{1-\phi(\gamma^T X^{\pi}_{ij})+\phi(\gamma^T X^{\pi}_{ij})\exp\left(-\Lambda(t)\exp\left(\beta^T Z^{\pi}_{ij}\right)\right)}.$$

Similarly, $\mathbb{P}_{n_1+n_2}\psi_{(\Lambda,\theta),h}$ and $\bar{\mathbb{P}}\psi_{(\Lambda,\theta),h}$ are defined using the empirical distribution of the pooled sample or the true distribution of the pooled sample $\bar{\mathbb{P}}=\kappa\mathbb{P}_1+(1-\kappa)\mathbb{P}_2$ respectively. By definition $(\bar{\Lambda},\bar{\theta})$ is the solution of $\bar{\mathbb{P}}\psi_{(\Lambda,\theta),h}\overset{!}{=}0$.

Lemma B.2. *Assume that the maximizer $(\bar{\Lambda},\bar{\theta})$ defined in (19) is unique. The pooled maximum likelihood estimator $(\bar{\Lambda}_{n_1+n_2},\bar{\theta}_{n_1+n_2})$ is a (weakly) consistent estimator of $(\bar{\Lambda},\bar{\theta})$.*

Proof. We will pursue similar ideas as [31] in the proofs of his Lemma 2 and Theorem 2. Comparing the arguments in [31] that lead to the maximum likelihood estimator in Display (12) of that paper, it is evident that the pooled estimator $\bar{\Lambda}_{n_1+n_2}$ must exhibit a similar structure. In particular,

$$\begin{aligned}&\bar{\Lambda}_{n_1+n_2}(t)\\&=\int_0^t\frac{dN(s)}{\sum_{j=1}^{n_1+n_2}R_j(s)\exp(\bar{\beta}^T_{n_1+n_2}Z_j)\{\Delta_j+(1-\Delta_j)\bar{g}_j(Y_j;\bar{\theta}_{n_1+n_2},\bar{\Lambda}_{n_1+n_2})\}},\end{aligned}$$

where $N=N_1+N_2$ is the pooled counting process, $R_j(s)=\mathbb{1}_{\{Y_j\geq s\}}$ denotes the at-risk process of the j-th pooled individual and

$$\bar{g}_i(t;\Lambda,\theta)=\frac{\phi(\gamma^T X_i)\exp(-\Lambda(t)\exp(\beta^T Z_i))}{1-\phi(\gamma^T X_i)+\phi(\gamma^T X_i)\exp(-\Lambda(t)\exp(\beta^T Z_i))}.$$

Similarly, we define

$$\tilde{\Lambda}_{n_1+n_2}(t)=\int_0^t\frac{dN(s)}{\sum_{j=1}^{n_1+n_2}R_j(s)\exp(\bar{\beta}^T Z_j)\{\Delta_j+(1-\Delta_j)\bar{g}_j(Y_j;\bar{\theta},\bar{\Lambda})\}}.$$

We have already noticed that $\sup_{n_1,n_2} \tilde{\Lambda}_{n_1+n_2}(\tau) < \infty$ a.s.; also, following the lines of Lemma 2 (ii) in [31], there exists a non-negative and integrable function $\eta : [0, \bar{\tau}_0] \to (0, \infty)$ bounded away from 0 such that, for each $\omega \in \Omega$, there exists a subsequence $(n_{1,k_1}(\omega)+n_{2,k_2}(\omega))$ such that $\sup_{t\in(0,\bar{\tau}_0]} |\frac{\mathrm{d}\bar{\Lambda}_{n_{1,k_1}+n_{2,k_2}}}{\mathrm{d}\tilde{\Lambda}_{n_{1,k_1}+n_{2,k_2}}}(t)-\eta(t)| \to 0$. For notational convenience, we will write $\bar{n}_k = n_{1,k_1} + n_{2,k_2}$ from now on. Arguing for a fixed ω and along subsequences had similarly been done by [36] and [46] based on Helly's theorem.

To prove the desired consistency of the pooled estimators, we will show that the difference of the log-likelihoods, say $\bar{\ell}_{\bar{n}_k}(\bar{\Lambda}_{\bar{n}_k}, \bar{\theta}_{\bar{n}_k})$ and $\bar{\ell}_{\bar{n}_k}(\tilde{\Lambda}_{\bar{n}_k}, \bar{\theta})$ converges to zero. Clearly,

$$
\begin{aligned}
0 \le\ & \bar{\ell}_{\bar{n}_k}(\bar{\Lambda}_{\bar{n}_k}, \bar{\theta}_{\bar{n}_k}) - \bar{\ell}_{\bar{n}_k}(\tilde{\Lambda}_{\bar{n}_k}, \bar{\theta}) \\
=& \frac{1}{\bar{n}_k} \sum_{i=1}^{\bar{n}_k} \Big[\Delta_i \log \frac{\bar{g}_i(Y_i; \bar{\Lambda}_{\bar{n}_k}, \bar{\theta}_{\bar{n}_k})}{\bar{g}_i(Y_i, \tilde{\Lambda}_{\bar{n}_k}, \bar{\theta})} + \Delta_i \log \frac{\Delta\bar{\Lambda}_{\bar{n}_k}(Y_i)}{\Delta\tilde{\Lambda}_{\bar{n}_k}(Y_i)} + \Delta_i(\bar{\beta}_{n_k} - \bar{\beta})^T Z_i \\
& \qquad + \log \frac{\bar{S}_i(Y_i; \bar{\Lambda}_{\bar{n}_k}, \bar{\theta}_{\bar{n}_k})}{\bar{S}_i(Y_i; \tilde{\Lambda}_{\bar{n}_k}, \bar{\theta})} \Big] \\
=& \frac{1}{\bar{n}_k} \sum_{i=1}^{\bar{n}_k} \Big[\Delta_i \log \frac{\phi(\bar{\gamma}_{n_k}^T X_i)}{\phi(\bar{\gamma}^T X_i)} - \Delta_i \bar{\Lambda}_{n_k}(Y_i) \exp(\bar{\beta}_{n_k}^T X_i) + \Delta_i \tilde{\Lambda}_{n_k}(Y_i) \exp(\bar{\beta}^T X_i) \\
& \qquad + \Delta_i \log \frac{\Delta\bar{\Lambda}_{\bar{n}_k}(Y_i)}{\Delta\tilde{\Lambda}_{\bar{n}_k}(Y_i)} + \Delta_i(\bar{\beta}_{n_k} - \bar{\beta})^T Z_i + (1 - \Delta_i) \log \frac{\bar{S}_i(Y_i; \bar{\Lambda}_{\bar{n}_k}, \bar{\theta}_{\bar{n}_k})}{\bar{S}_i(Y_i; \tilde{\Lambda}_{\bar{n}_k}, \bar{\theta})} \Big]
\end{aligned}
$$

where and $\bar{S}_i(t; \Lambda, \theta) = 1 - \phi(\gamma^T X_i) + \phi(\gamma^T X_i) \exp(-\Lambda(t) \exp(\beta^T Z_i))$.

The space of bounded, increasing functions with discontinuities only at $\tau_{0,1}$ and $\tau_{0,2}$ is separable with respect to the supremum norm. Also, Euclidean spaces are separable. Denote by $(\Lambda_l, \theta_l)_{l\in\mathbb{N}}$ a countable subset that is dense in the product of the just-described spaces. For each $l \in \mathbb{N}$, by the strong law of large numbers,

$$
\begin{aligned}
\frac{1}{\bar{n}_k} \sum_{i=1}^{\bar{n}_k} \Big[& \Delta_i \log \phi(\gamma_l^T X_i) - \Delta_i \Lambda_l(t) \exp(\beta_l^T Z_i) + \Delta_i \log \mathrm{d}\Lambda_l(Y_i) + \Delta_i \beta_l^T Z_i \\
& + (1 - \Delta_i) \log \bar{S}_i(Y_i; \Lambda_l, \theta_l) \Big]
\end{aligned}
$$

converges a.s. to its expectation, i.e., for all $\omega \in \Omega_l$ with $\bar{\mathbb{P}}(\Omega_l) = 1$. From now on, we restrict ω to be in the intersection $\bigcap_{l\in\mathbb{N}} \Omega_l$ which also has probability 1. Consequently, also due to the continuity of the likelihoods in Λ and θ,

$$
\begin{aligned}
0 \le\ & \bar{\ell}_{\bar{n}_k}(\bar{\Lambda}_{\bar{n}_k}, \bar{\theta}_{\bar{n}_k}) - \bar{\ell}_{\bar{n}_k}(\tilde{\Lambda}_{\bar{n}_k}, \bar{\theta}) \\
=\ & \mathbb{E}_{\bar{\mathbb{P}}} \Big[\Delta_i \log \frac{\phi(\gamma^{*T} X_i)}{\phi(\bar{\gamma}^T X_i)} - \{\Lambda^*(t) \exp(\beta^{*T} Z_i) - \bar{\Lambda}(t) \exp(\bar{\beta}^T Z_i)\} + \Delta_i \log \frac{\mathrm{d}\Lambda^*(Y_i)}{\mathrm{d}\bar{\Lambda}(Y_i)} \\
& \qquad + \Delta_i(\beta^* - \bar{\beta})^T Z_i + \log \frac{\bar{S}_i(Y_i; \Lambda^*, \theta^*)}{\bar{S}_i(Y_i; \bar{\Lambda}, \bar{\theta})} \Big] + o(1)
\end{aligned}
$$

where the expectation is taken with respect to X, Y, Z and Λ^*, β^*, γ^* are fixed (depending on ω). For a.e. ω, the conditional expectation in the previous display represents a negative KL-divergence of the logistic-Cox model specified by $(\bar{\Lambda}, \bar{\theta})$ from the model specified by $(\Lambda^*(\omega), \theta^*(\omega))$. As a consequence, it must be 0, i.e., $\bar{\ell}(\Lambda^*, \theta^*) = \bar{\ell}(\bar{\Lambda}, \bar{\theta})$ $\bar{\mathbb{P}}$-a.e.. We use this fact to identify all model components, one by one; every equality below is to be understood $\bar{\mathbb{P}}$-a.s..

We first consider $\Delta = 0$ and $Y \geq \bar{\tau}_0$, for which $\bar{S}(Y; \Lambda^*, \theta^*) = \phi(\gamma^{*T} X)$ and $\bar{S}(Y; \bar{\Lambda}, \bar{\theta}) = \phi(\bar{\gamma}^T X)$. From this we can identify $\gamma^* = \bar{\gamma}$ a.s. for the logistic model. Next, for $\Delta = 0$ and $Y < \bar{\tau}_0$, we obtain $\bar{S}(Y; \Lambda^*, \theta^*) = \bar{S}(Y; \bar{\Lambda}, \bar{\theta})$, hence

$$\exp(-\Lambda^*(Y) \exp(\beta^{*T} Z)) = \exp(-\bar{\Lambda}(Y) \exp(\bar{\beta}^T Z))$$

Upon inserting different combinations of Y and Z, we conclude that $\beta^* = \bar{\beta}$ and $\Lambda^* = \bar{\Lambda}$ a.s.. □

The following lemma establishes the consistency of randomly permuted Z-estimators. Since the proof does not make use of the specific underlying model structure, it is clear that this result holds more generally, i.e., also beyond logistic-Cox cure models.

Lemma B.3. *Assume the maximizer* $(\bar{\Lambda}, \bar{\theta})$ *defined in* (19) *is unique. The permutation estimators* $(\hat{\Lambda}^\pi_{n_1,1}, \hat{\theta}^\pi_{n_1,1})$ *and* $(\hat{\Lambda}^\pi_{n_2,2}, \hat{\theta}^\pi_{n_2,2})$ *converge in probability to* $(\bar{\Lambda}, \bar{\theta})$.

Proof. First, we would like to point out that conditional convergence in probability (given a σ-algebra) is equivalent to the unconditional convergence in probability; a variant of Fact 1 in the Supporting Information of [18] similarly holds for the present setting. That is why we do not distinguish between conditional and unconditional consistency.

To prove the consistency of the permuted estimators, we are going to employ the permutation version of the score equations, i.e., $\mathbb{P}^\pi_{n_i,i} \psi_{(\Lambda,\theta),h} \overset{!}{=} 0$ for all indexing h, $i = 1, 2$. So far, we know by definition that $\mathbb{P}^\pi_{n_i,i} \psi_{(\hat{\Lambda}^\pi_{n_i,i}, \hat{\theta}^\pi_{n_i,i}),h} = 0$ and that $\mathbb{P}_{n_1+n_2} \psi_{(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2}),h} = 0$.

Also, since $n_1 \mathbb{P}^\pi_{n_1,1} \psi_{(\Lambda,\theta),h} + n_2 \mathbb{P}^\pi_{n_2,2} \psi_{(\Lambda,\theta),h} = (n_1 + n_2) \mathbb{P}_{n_1+n_2} \psi_{(\Lambda,\theta),h}$, we have that $n_1 \mathbb{P}^\pi_{n_1,1} \psi_{(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2}),h} = -n_2 \mathbb{P}^\pi_{n_2,2} \psi_{(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2}),h}$. Thus, because both of these permuted expressions are connected, we will only focus on the index $i = 1$ from now on.

Furthermore, upon integrating out all permutations, it is easy to see that the (conditional) expectation is $\mathbb{E}[\mathbb{P}^\pi_{n_1,1} \psi_{(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2}),h} \mid Y_i, \Delta_i, X_i, Z_i : i = 1, \ldots, n_1 + n_2] = 0$. Additionally, straightforward and standard algebra for permuted linear statistics for the conditional variance leads to

$$Var[\mathbb{P}^\pi_{n_1,1} \psi_{(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2}),h} \mid Y_i, \Delta_i, X_i, Z_i : i = 1, \ldots, n_1+n_2] = O_p((n_1+n_2)^{-1}).$$

Consequently, Chebychev's inequality (applied to the conditional distribution) verifies that the permutation-based score equations evaluated at the point

$(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2})$ all converge to 0 in probability. Similar convergences in probability (not necessarily to zero) also hold for other evaluation points.

Hence, the pooled estimator is asymptotically a solution to the permutation-based score equations. Now, since $(\hat{\Lambda}^{\pi}_{n_1,1}, \hat{\theta}^{\pi}_{n_1,1})$ is another (finite sample) solution and the "true" solution $(\bar{\Lambda}, \bar{\theta})$ is assumed to be unique, the permuted estimator must approach the pooled estimator in probability as the sample size goes in infinity. Anything else would contradict the continuity of the map $(\Lambda, \theta) \mapsto \bar{\mathbb{E}}_{\bar{\mathbb{P}}}(\bar{\ell}(\Lambda, \theta))$. □

Lemma B.4. *Assume the maximizer $(\bar{\Lambda}, \bar{\theta})$ defined in (19) is unique. Conditionally on the observations, the process*

$$\langle n_i^{1/2}(\hat{\Lambda}^{\pi}_{n_i,i} - \bar{\Lambda}_{n_1+n_2}), n_i^{1/2}(\hat{\theta}^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2})\rangle, \qquad i = 1, 2$$

defined as in (12) and indexed by $h \in \mathcal{H}_m$ converges weakly in $l^{\infty}(\mathcal{H}_m)$ to a tight Gaussian process G_i^* in $l^{\infty}(\mathcal{H}_m)$, in outer probability.

Proof. We will apply Theorem C.1 from Appendix C. We need to show that the sample specific estimators $(\hat{\Lambda}^{\pi}_{n_i,i}, \hat{\theta}^{\pi}_{n_i,i})$ satisfy the conditions of Theorem 3.3.1 in [55]. Consistency of the estimators was shown in Lemmas B.2 and B.3 above. Verification of the other conditions can be done as in [31]. We omit the other details here since the proof goes along the same lines as points a)-c) below.

a) To verify condition (28) of Theorem C.1, it suffices to show that for any sequence $\epsilon_{n_i} \to 0$,

$$\sup_{\substack{\|\Lambda - \bar{\Lambda}_{n_1+n_2}\|_{\infty} \le \epsilon_{n_i}, \\ \|\beta - \bar{\beta}_{n_1+n_2}\| \le \epsilon_{n_i}, \\ \|\gamma - \bar{\gamma}_{n_1+n_2}\| \le \epsilon_{n_i}}} \frac{\left|(\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{(\Lambda,\theta),h} - (\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2}),h}\right|}{n_i^{-1/2} \vee \|\beta - \bar{\beta}_{n_1+n_2}\| \vee \|\gamma - \bar{\gamma}_{n_1+n_2}\| \vee \|\Lambda - \bar{\Lambda}_{n_1+n_2}\|_{\infty}} \tag{20}$$

converges to zero in probability given the data. For simplicity, we can write

$$\begin{aligned}
&(\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{(\Lambda,\theta),h} - (\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{(\bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2}),h} \\
&= \sum_{j=1}^{6} (\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2}) a_{j,h}
\end{aligned}$$

where

$$\begin{aligned}
a_{1,h}(\delta, y, x, z) &= -h_{21}^T x \left\{\phi(\gamma^T x) - \phi(\bar{\gamma}_{n_1+n_2}^T x)\right\} \\
a_{2,h}(\delta, y, x, z) &= (1-\delta) h_{21}^T x \left\{g(y, \Lambda, \theta) - g(y, \bar{\Lambda}_{n_1+n_2}, \bar{\theta}_{n_1+n_2})\right\} \\
a_{3,h}(\delta, y, x, z) &= \delta \left\{ e^{\beta^T z} \int_0^y h_1(s) \mathrm{d}\Lambda(s) - e^{\bar{\beta}_{n_1+n_2}^T z} \int_0^y h_1(s) \mathrm{d}\bar{\Lambda}_{n_1+n_2}(s) \right\} \\
a_{4,h}(\delta, y, x, z) &= \delta h_{22}^T z \left\{ e^{\beta^T z} \Lambda(y) - e^{\bar{\beta}_{n_1+n_2}^T z} \bar{\Lambda}_{n_1+n_2}(y) \right\}
\end{aligned}$$

$$
\begin{aligned}
a_{5,h}(\delta,y,x,z) &= (1-\delta)\Big\{g(y,\Lambda,\theta)e^{\beta^T z}\int_0^y h_1(s)\mathrm{d}\Lambda(s) \\
&\qquad - g(y,\bar{\Lambda}_{n_1+n_2},\bar{\theta}_{n_1+n_2})e^{\bar{\beta}_{n_1+n_2}^T z}\int_0^y h_1(s)\mathrm{d}\bar{\Lambda}_{n_1+n_2}(s)\Big\} \\
a_{6,h}(\delta,y,x,z) &= (1-\delta)h_{22}^T z\Big\{g(y,\Lambda,\theta)e^{\beta^T z}\Lambda(y) \\
&\qquad - g(y,\bar{\Lambda}_{n_1+n_2},\bar{\theta}_{n_1+n_2})e^{\bar{\beta}_{n_1+n_2}^T z}\bar{\Lambda}_{n_1+n_2}(y)\Big\}.
\end{aligned} \tag{21}
$$

Next we consider the third term. The other terms can be handled similarly. First note that, by Lemma B.2, $\bar{\Lambda}_{n_1+n_2}$ and $\bar{\theta}_{n_1+n_2}$ are consistent estimates of $\bar{\Lambda}$ and $\bar{\theta}$, and as a result they are bounded on a set of probability converging to one. From a Taylor expansion we have

$$
\begin{aligned}
&(\mathbb{P}_{n_i,i}^{\pi} - \mathbb{P}_{n_1+n_2})a_{3,h} \\
&= (\beta - \bar{\beta}_{n_1+n_2})^T \int z\delta e^{\bar{\beta}_{n_1+n_2}^T z}\int_0^y h_1(s)\mathrm{d}\Lambda(s)\mathrm{d}(\mathbb{P}_{n_i,i}^{\pi} - \mathbb{P}_{n_1+n_2})(\delta,y,x,z) \\
&\quad + \int \delta e^{\bar{\beta}_{n_1+n_2}^T z}\int_0^y h_1(s)\mathrm{d}(\Lambda - \bar{\Lambda}_{n_1+n_2})(s)\mathrm{d}(\mathbb{P}_{n_i,i}^{\pi} - \mathbb{P}_{n_1+n_2})(\delta,y,x,z) \\
&\quad + o_P^*\left(\int \delta e^{\bar{\beta}_{n_1+n_2}^T z}\int_0^y h_1(s)\mathrm{d}(\Lambda - \bar{\Lambda}_{n_1+n_2})(s)\mathrm{d}(\mathbb{P}_{n_i,i}^{\pi} - \mathbb{P}_{n_1+n_2})(\delta,y,x,z)\right) \\
&\quad + o_P^*(\|\beta - \bar{\beta}_{n_1+n_2}\|)
\end{aligned}
$$

For the first term, since the class of functions that we are integrating is Donsker and uniformly bounded, by Theorem 3.7.2 in [55] it follows that, conditionally on the observations

$$
\sup_{\|\Lambda - \bar{\Lambda}_{n_1+n_2}\|_\infty \le \epsilon_{n_i}} \int z\delta e^{\bar{\beta}^T z}\int_0^y h_1(s)\mathrm{d}\Lambda(s)\mathrm{d}(\mathbb{P}_{n_i,i}^{\pi} - \mathbb{P}_{n_1+n_2})(\delta,y,x,z) = o_P^*(1). \tag{22}
$$

The second term can be rewritten as

$$
\int_0^{\bar{\tau}_0} D_n(s)h_1(s)\mathrm{d}(\Lambda - \bar{\Lambda}_{n_1+n_2})(s)
$$

where

$$
D_n(s) = \int \delta \mathbb{1}_{\{y>s\}} e^{\bar{\beta}^T z}\mathrm{d}(\mathbb{P}_{n_i,i}^{\pi} - \mathbb{P}_{n_1+n_2})(\delta,y,x,z).
$$

By integration by parts and the chain rule we have

$$
\begin{aligned}
&\int_0^{\bar{\tau}_0} D_n(s)h_1(s)\mathrm{d}(\Lambda - \bar{\Lambda}_{n_1+n_2})(s) \\
&= D_n(\bar{\tau}_0)h_1(\bar{\tau}_0)(\Lambda - \bar{\Lambda}_{n_1+n_2})(\bar{\tau}_0) - \int_0^{\bar{\tau}_0}(\Lambda - \bar{\Lambda}_{n_1+n_2})\mathrm{d}\left[D_n(s)h_1(s)\right] \\
&= D_n(\bar{\tau}_0)h_1(\bar{\tau}_0)(\Lambda - \bar{\Lambda}_{n_1+n_2})(\bar{\tau}_0) - \int_0^{\bar{\tau}_0}(\Lambda - \bar{\Lambda}_{n_1+n_2})(s)D_n(s)\mathrm{d}h_1(s)
\end{aligned}
$$

$$+\int \delta(\Lambda-\bar{\Lambda})(y)h_1(y)e^{\bar{\beta}^T z}\mathrm{d}(\mathbb{P}^{\pi}_{n_i,i}-\mathbb{P}_{n_1+n_2})(\delta,y,x,z)$$
$$+\int \delta(\bar{\Lambda}-\bar{\Lambda}_{n_1+n_2})(y)h_1(y)e^{\bar{\beta}^T z}\mathrm{d}(\mathbb{P}^{\pi}_{n_i,i}-\mathbb{P}_{n_1+n_2})(\delta,y,x,z)$$

Again, by Theorem 3.7.2 in [55], it follows that, conditionally on the observations $D_n = o_P^*(1)$. Since h_1 is bounded, it follows

$$\sup_{\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty\le\epsilon_{n_i}} \frac{\left|D_n(\bar{\tau}_0)h_1(\bar{\tau}_0)(\Lambda-\bar{\Lambda}_{n_1+n_2})(\bar{\tau}_0)\right|}{\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty} = o_P^*(1).$$

In addition we also have that $\sup_{s\in[0,\tau_0]}|D_n(s)| = o_P^*(1)$ conditionally on the observations and since h_1 is of bounded variation

$$\sup_{\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty\le\epsilon_{n_i}} \frac{\left|\int_0^{\bar{\tau}_0}(\Lambda-\bar{\Lambda}_{n_1+n_2})(s)D_n(s)\mathrm{d}h_1(s)\right|}{\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty}$$
$$\le \sup_{t\in[0,\bar{\tau}_0]}|D_n(s)|\int_0^{\bar{\tau}_0}|\mathrm{d}h(s)| = o_P^*(1).$$

Since $\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty \le \epsilon_{n_i}$ implies $\|\Lambda-\bar{\Lambda}\|_\infty \le \tilde{\epsilon}_{n_i}$ for some $\tilde{\epsilon}_{n_i}\to 0$, the class $\{g_\Lambda(y,\delta,z) = \delta(\Lambda-\bar{\Lambda})(y)h_1(y)e^{\bar{\beta}^T z} : \|\Lambda-\bar{\Lambda}\|_\infty \le \tilde{\epsilon}_{n_i}\}$ is a Donsker class (product of bounded variation functions, uniformly bounded) and

$$\mathbb{E}_{\bar{\mathbb{P}}}\left[\Delta(\Lambda-\bar{\Lambda})(Y)^2h_1(Y)^2e^{2\bar{\beta}^T Z}\right] = O(\tilde{\epsilon}_n^2) = o(1),$$

we have that, conditionally on the data,

$$\sup_{\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty\le\epsilon_{n_i}}\sqrt{n}\int\delta(\Lambda-\bar{\Lambda})(y)h_1(y)e^{\bar{\beta}^T z}\mathrm{d}(\mathbb{P}^{\pi}_{n_i,i}-\mathbb{P}_{n_1+n_2})(\delta,y,x,z) = o_P^*(1).$$

Finally, since $\|\bar{\Lambda}_{n_1+n_2}-\bar{\Lambda}\|_\infty\to 0$ a.s., by Proposition A.5.3 in [55] it follows that, conditionally on the data,

$$\int\delta(\bar{\Lambda}-\bar{\Lambda}_{n_1+n_2})(y)h_1(y)e^{\bar{\beta}^T z}\mathrm{d}(\mathbb{P}^{\pi}_{n_i,i}-\mathbb{P}_{n_1+n_2})(\delta,y,x,z) = o_P^*(1).$$

Combining all the results we obtain that

$$\sup_{\substack{\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty\le\epsilon_{n_i},\\ \|\beta-\bar{\beta}_{n_1+n_2}\|\le\epsilon_{n_i},\\ \|\gamma-\bar{\gamma}_{n_1+n_2}\|\le\epsilon_{n_i}}} \frac{\left|(\mathbb{P}^{\pi}_{n_i,i}-\mathbb{P}_{n_1+n_2})a_{3,h}\right|}{n_i^{-1/2}\vee\|\beta-\bar{\beta}_{n_1+n_2}\|\vee\|\gamma-\bar{\gamma}_{n_1+n_2}\|\vee\|\Lambda-\bar{\Lambda}_{n_1+n_2}\|_\infty}$$

converges to zero in probability, given the data. The terms related to the other a_j can be treated similarly.

b) Next we check condition (29). From (20) it follows in particular that, conditionally on the data,

$$\sqrt{n_i}(\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{(\bar{\Lambda},\bar{\theta}),h} - \sqrt{n_i}(\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{(\bar{\Lambda}_{n_1+n_2},\bar{\theta}_{n_1+n_2}),h} = o^*_P(1)$$

almost surely. Hence, it is sufficient to show that

$$\sqrt{n_i}(\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{(\bar{\Lambda},\bar{\theta}),h} \rightsquigarrow Z_1 \tag{23}$$

on $l^\infty(\mathcal{H}_m)$ in outer probability, where Z_1 is a tight random element (actually a Gaussian process). This follows from Theorem 3.7.1. in [55] since the class of functions $\{\psi_{(\bar{\Lambda},\bar{\theta}),h} : h \in \mathcal{H}_m\}$ is Donsker and bounded. This is already shown in step 1 of the proof of Theorem 3 in [31] (the class of functions is the same, just evaluated at a different point $(\bar{\Lambda},\bar{\theta})$).

c) Since the functions $\psi_{(\Lambda,\theta),h}$, $h \in \mathcal{H}_m$ are the same as in [31], it can be proved in the same way that $\bar{\mathbb{P}}\psi_{(\bar{\Lambda},\bar{\theta}),h}$ is Fréchet-differentiable at $(\bar{\Lambda},\bar{\theta})$ and the derivative is given by

$$(\bar{\mathbb{P}}\dot{\psi}_{(\bar{\Lambda},\bar{\theta})})((\Lambda,\theta) - (\bar{\Lambda},\bar{\theta}))(h) = \int_0^{\bar{\tau}_0} \bar{\sigma}_{(1)}(h)\,\mathrm{d}(\Lambda - \bar{\Lambda})(t) + (\theta - \bar{\theta})^T \bar{\sigma}_{(2)}(h)$$

where $\bar{\sigma}_{(1)}$, $\bar{\sigma}_{(2)}$ are defined as in (13)–(14) respectively, with $\mathbb{E}_{\mathbb{P}_i}$ replaced by $\mathbb{E}_{\bar{\mathbb{P}}}$ and evaluated at $(\bar{\Lambda},\bar{\theta})$ instead of (Λ_i,θ_i). Uniform Fréchet-differentiability at $\bar{\theta}$ is implied by the Fréchet-differentiability on a neighborhood of $\bar{\theta}$ and the (operator) norm-continuity of $(\Lambda,\theta) \mapsto \bar{\mathbb{P}}\dot{\psi}_{(\Lambda,\theta)}$. Based on the boundedness of all terms and because $h \in \mathcal{H}_m$, it is clear from the expressions in [31] that this continuity holds. Also the proof that the derivative is continuously invertible remains the same as in [31].

d) For condition (30) we have

$$\begin{aligned}
&(\mathbb{P}_{n_1+n_2}\psi_{\theta,h} - \mathbb{P}_{n_1+n_2}\psi_{\tilde{\theta},h}) - (\bar{\mathbb{P}}\psi_{\theta,h} - \bar{\mathbb{P}}\psi_{\tilde{\theta},h}) \\
&= (\mathbb{P}_{n_1+n_2} - \bar{\mathbb{P}})\psi_{\theta,h} - (\mathbb{P}_{n_1+n_2} - \bar{\mathbb{P}})\psi_{\tilde{\theta},h} = \sum_{j=1}^{6}(\mathbb{P}_{n_1+n_2} - \bar{\mathbb{P}})\bar{a}_{j,h}
\end{aligned}$$

where $\bar{a}_{j,h}$ are defined as in (21) replacing $(\bar{\Lambda}_{n_1+n_2},\bar{\theta}_{n_1+n_2})$ by $(\tilde{\Lambda},\tilde{\theta})$. We can deal with this similarly to what we did to show (20). The difference is that now (Λ,θ) and $(\tilde{\Lambda},\tilde{\theta})$ are fixed and we consider the class of functions with respect to $h \in \mathcal{H}_m$. For example, for the term corresponding to $\bar{a}_{3,h}$ we have

$$\begin{aligned}
&(\mathbb{P}_{n_1+n_2} - \bar{\mathbb{P}})\bar{a}_{3,h} \\
&= (\beta - \tilde{\beta})^T \int z\delta e^{\tilde{\beta}^T z} \int_0^y h_1(s)\mathrm{d}\tilde{\Lambda}(s)\mathrm{d}(\mathbb{P}_{n_1+n_2} - \bar{\mathbb{P}})(\delta,y,x,z) \\
&\quad + \int \delta e^{\tilde{\beta}^T z} \int_0^y h_1(s)\mathrm{d}(\Lambda - \tilde{\Lambda})(s)\mathrm{d}(\mathbb{P}_{n_1+n_2} - \bar{\mathbb{P}})(\delta,y,x,z) \\
&\quad + o^*_P\left(\int \delta e^{\tilde{\beta}^T z} \int_0^y h_1(s)\mathrm{d}(\Lambda - \tilde{\Lambda})(s)\mathrm{d}(\mathbb{P}_{n_1+n_2} - \bar{\mathbb{P}})(\delta,y,x,z)\right) \\
&\quad + o^*_P(\|\beta - \tilde{\beta}\|).
\end{aligned} \tag{24}$$

Note also that

$$\mathbb{P}_{n_1+n_2}-\bar{\mathbb{P}}=\frac{n_1}{n_1+n_2}(\mathbb{P}_{n_1}-\mathbb{P}_1)+\frac{n_2}{n_1+n_2}(\mathbb{P}_{n_2}-\mathbb{P}_2)+\left(\frac{n_1}{n_1+n_2}-\kappa\right)(\mathbb{P}_1-\mathbb{P}_2)$$

and $n_1/(n_1+n_2)\to\kappa$. Then the integral in the first term of (24) converges to zero since the class of functions that we are integrating is Donsker and uniformly bounded. The second term can be rewritten as

$$\int_0^{\bar{\tau}_0}\bar{D}_{n_1+n_2}(s)h_1(s)\mathrm{d}(\Lambda-\tilde{\Lambda})(s)$$

where

$$\bar{D}_{n_1+n_2}(s)=\int\delta\mathbb{1}_{\{y>s\}}e^{\tilde{\beta}^Tz}\mathrm{d}(\mathbb{P}_{n_1+n_2}-\bar{\mathbb{P}})(\delta,y,x,z).$$

By integration by parts and the chain rule we again have

$$\begin{aligned}&\int_0^{\bar{\tau}_0}\bar{D}_{n_1+n_2}(s)h_1(s)\mathrm{d}(\Lambda-\tilde{\Lambda})(s)\\&=\bar{D}_{n_1+n_2}(\bar{\tau}_0)h_1(\bar{\tau}_0)(\Lambda-\tilde{\Lambda})(\bar{\tau}_0)-\int_0^{\bar{\tau}_0}(\Lambda-\tilde{\Lambda})(s)\bar{D}_{n_1+n_2}(s)\mathrm{d}h_1(s)\\&\quad+\int\delta(\Lambda-\tilde{\Lambda})(y)h_1(y)e^{\tilde{\beta}^Tz}\mathrm{d}(\mathbb{P}_{n_1+n_2}-\bar{\mathbb{P}})(\delta,y,x,z).\end{aligned}$$

By the Glivenko-Cantelli theorem, applied for both $(\mathbb{P}_{n_i}-\mathbb{P}_i)$, $i=1,2$, and the boundedness of the integrand we have $\sup_{s\in[0,\bar{\tau}_0]}\bar{D}_{n_1+n_2}(s)=o_P(1)$. Since $h_1\in\mathcal{H}_m$ are uniformly bounded, it follows that

$$\sup_{h_1\in\mathcal{H}_m}\frac{|\bar{D}_{n_1+n_2}(\bar{\tau}_0)h_1(\bar{\tau}_0)(\Lambda-\tilde{\Lambda})(\bar{\tau}_0)|}{\|\Lambda-\tilde{\Lambda}\|_\infty}=o_P(1).$$

In addition, since h_1 are functions of bounded variation and uniformly bounded norm,

$$\begin{aligned}&\sup_{h_1\in\mathcal{H}_m}\frac{|\int_0^{\bar{\tau}_0}(\Lambda-\tilde{\Lambda})(s)\bar{D}_{n_1+n_2}(s)\mathrm{d}h_1(s)|}{\|\Lambda-\tilde{\Lambda}\|_\infty}\\&\le\sup_{s\in[0,\bar{\tau}_0]}|\bar{D}_{n_1+n_2}(s)|\int_0^{\bar{\tau}_0}|\mathrm{d}h_1(s)|=o_P(1).\end{aligned}$$

Finally, from Theorem 2.11.23 in [55] it follows that

$$\sup_{h_1\in\mathcal{H}_m}\frac{\sqrt{n_i}\int\delta(\Lambda-\tilde{\Lambda})(y)h_1(y)e^{\tilde{\beta}^Tz}\mathrm{d}(\mathbb{P}_{n_i}-\mathbb{P}_i)(\delta,y,x,z)}{\|\Lambda-\tilde{\Lambda}\|_\infty}$$

is bounded in probability. Combining all the results we obtain

$$\sup_{h_1\in\mathcal{H}_m}\left\|(\mathbb{P}_{n_1+n_2}-\bar{\mathbb{P}})a_{3,h}\right\|=o_P\left(\|\theta-\tilde{\theta}\|\vee\|\Lambda-\tilde{\Lambda}\|_\infty\right).$$

The other terms can be handled similarly obtaining

$$\|(\mathbb{P}_{n_1+n_2}\psi_{\theta,h}-\mathbb{P}_{n_1+n_2}\psi_{\tilde{\theta},h})-(\bar{\mathbb{P}}\psi_{\theta,h}-\bar{\mathbb{P}}\psi_{\tilde{\theta},h})\|=o_P\left(\|\theta-\tilde{\theta}\|\vee\|\Lambda-\tilde{\Lambda}\|_\infty\right).$$

□

Proof of Theorem 3.2. We proceed similarly to the proof of Theorem 3.1. From Lemma B.4, it follows that the process

$$\langle n_i^{1/2}(\hat{\Lambda}^{\pi}_{n_i,i} - \bar{\Lambda}_{n_1+n_2}), n_i^{1/2}(\hat{\theta}^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2})\rangle_{i=1,2}(h)$$

converges to a Gaussian process $G^* = (G_1^*, G_2^*)$ with $G_2^* = -\sqrt{\kappa/(1-\kappa)}G_1^*$. We start by deriving the weak convergence of the process $\sqrt{a_n}(\hat{S}^{\pi}_{n_1,1} - \bar{S}_{n_1+n_2}, \hat{S}^{\pi}_{n_2,2} - \bar{S}_{n_1+n_2})$, where $a_n = \sqrt{n_1 n_2/(n_1+n_2)}$. By a series of Taylor expansions we can write

$$\begin{aligned}
&\sqrt{a_n}\{\hat{S}^{\pi}_{n_1,1}(t|z) - \bar{S}_{n_1+n_2}(t|z)\} \\
&= \sqrt{a_n}\left\{\exp\left(-\hat{\Lambda}^{\pi}_{n_1,1}(t)e^{\hat{\beta}^{\pi T}_{n_1,1}z}\right) - \exp\left(-\bar{\Lambda}_{n_1+n_2}(t)e^{\bar{\beta}^T_{n_1+n_2}z}\right)\right\} \\
&= -\sqrt{a_n}\left\{\hat{\Lambda}^{\pi}_{n_1,1}(t)e^{\hat{\beta}^{\pi T}_{n_1,1}z} - \bar{\Lambda}_{n_1+n_2}(t)e^{\bar{\beta}^T_{n_1+n_2}z}\right\}\exp\left(-\bar{\Lambda}_{n_1+n_2}(t)e^{\bar{\beta}^T_{n_1+n_2}z}\right) + R_1 \\
&= -\sqrt{a_n}\left\{\hat{\Lambda}^{\pi}_{n_1,1}(t) - \bar{\Lambda}_{n_1+n_2}(t)\right\}e^{\bar{\beta}^T_{n_1+n_2}z}\exp\left(-\bar{\Lambda}_{n_1+n_2}(t)e^{\bar{\beta}^T_{n_1+n_2}z}\right) \\
&\quad - \sqrt{a_n}\left\{e^{\hat{\beta}^{\pi T}_{n_1,1}z} - e^{\bar{\beta}^T_{n_1+n_2}z}\right\}\bar{\Lambda}_{n_1+n_2}(t)\exp\left(-\bar{\Lambda}_{n_1+n_2}(t)e^{\bar{\beta}^T_{n_1+n_2}z}\right) + R_1 + R_2 \\
&= -\sqrt{a_n}\left\{\hat{\Lambda}^{\pi}_{n_1,1}(t) - \bar{\Lambda}_{n_1+n_2}(t)\right\}e^{\bar{\beta}^T z}\exp\left(-\bar{\Lambda}(t)e^{\bar{\beta}^T z}\right) \\
&\quad - \sqrt{a_n}\left\{\hat{\beta}^{\pi}_{n_1,1} - \bar{\beta}_{n_1+n_2}\right\}^T ze^{\bar{\beta}^T z}\bar{\Lambda}(t)\exp\left(-\bar{\Lambda}(t)e^{\bar{\beta}^T z}\right) + R_1 + R_2 + R_3
\end{aligned}$$

where the remainder terms converge to zero in probability. Considering functions $h \in \mathcal{H}_m$ of the form

$$\begin{aligned}
h_{t,z} &= (h_{1;t,z}, h_{2;t,z}) \\
&= \left(\mathbb{1}_{[0,t]}(\cdot)e^{\bar{\beta}^T z}\exp\left(-\bar{\Lambda}(t)e^{\bar{\beta}^T z}\right), \left(\mathbf{0}_p, ze^{\bar{\beta}^T z}\bar{\Lambda}(t)\exp\left(-\bar{\Lambda}(t)e^{\bar{\beta}^T z}\right)\right)\right),
\end{aligned}$$

where $\mathbf{0}_p$ denotes a zero vector in $\mathbb{R}^p$, we have

$$\begin{aligned}
&\sqrt{a_n}\{\hat{S}^{\pi}_{n_1,1}(t|z) - \bar{S}_{n_1+n_2}(t|z)\} \\
&= \sqrt{1-\kappa}\langle n_1^{1/2}(\hat{\Lambda}^{\pi}_{n_1,1} - \bar{\Lambda}_{n_1+n_2}), n_1^{1/2}(\hat{\theta}^{\pi}_{n_1,1} - \bar{\theta}_{n_1+n_2})\rangle(h_{t,z}) + o_{P^*}(1)
\end{aligned}$$

from which we conclude that, given the data, the process $\sqrt{a_n}\{\hat{S}^{\pi}_{n_1,1}(t|z) - \bar{S}_{n_1+n_2}(t|z)\}$ converges weakly in $D[0,\bar{\tau}_0]$ to a mean zero Gaussian process $\bar{G}_{1,z}$ with covariance

$$\begin{aligned}
\rho_{1,z}(s,t) &= Cov(\bar{G}_{1,z}(s), \bar{G}_{1,z}(t)) \\
&= (1-\kappa)\left\{\int_0^{\tau_{0,1}} h_{1;s,z}(u)\bar{\sigma}^{-1}_{(1)}(h_{t,z})(u)\mathrm{d}\bar{\Lambda}(u) + h^T_{2;s,z}\bar{\sigma}^{-1}_{(2)}(h_{t,z})\right\},
\end{aligned} \tag{25}$$

where $\bar{\sigma}_{(1)}$, $\bar{\sigma}_{(2)}$ are defined as in (13)–(14) respectively, with $\mathbb{E}_{\mathbb{P}_i}$ replaced by $\mathbb{E}_{\bar{\mathbb{P}}}$ and evaluated at $(\bar{\Lambda}, \bar{\theta})$ instead of (Λ_i, θ_i) Defining $\overline{MST}_{u,z}$ as the conditional

expected lifetime of the uncured in the pooled sample, we have

$$\begin{aligned}&\sqrt{a_n}(\widehat{MST}^{\pi}_{u,1,z} - \overline{MST}_{u,z})\\ &= \sqrt{a_n}\int_0^{\bar\tau_0}\{\hat S^{\pi}_{n_1,1}(t|z) - \bar S_{n_1+n_2}(t|z)\}\mathrm{d}t\\ &\quad - \sqrt{a_n}\{\bar\tau_0 - Y^{\pi}_{1,(m_1)}\}\hat S^{\pi}_{n_1,1}(\bar\tau_0) - \sqrt{a_n}\{\bar\tau_0 - Y_{(m)}\}\bar S_{n_1+n_2}(\bar\tau_0),\end{aligned} \tag{26}$$

where $Y^{\pi}_{1,(m_1)}$ and $Y_{(m)}$ denote the largest uncensored observation in the first permuted sample and the pooled sample respectively. As in the proof of Theorem 3.1, the second term and third term in the right hand side of the previous equation can be shown to converge to zero in probability. Considering the map

$$\psi : D[0,\bar\tau_0] \to \mathbb{R} \qquad \psi(\xi) = \int_0^{\bar\tau_0}\xi(s)\,\mathrm{d}s$$

which is Hadamard-differentiable, it follows that, given the data, $\sqrt{a_n}(\widehat{MST}^{\pi}_{u,1,z} - \overline{MST}_{u,z}, \widehat{MST}^{\pi}_{u,2,z} - \overline{MST}_{u,z})$ converges weakly to a two dimensional Gaussian random vector

$$(N_1,N_2) = \left(\int_0^{\bar\tau_0}\bar G_{1,z}(s)\,\mathrm{d}s, \int_0^{\bar\tau_0}\bar G_{2,z}(s)\,\mathrm{d}s\right) = \left(N_1, -\frac{\kappa}{1-\kappa}N_1\right).$$

Taking the difference of both entries of the pair, we conclude that, given the data, $\sqrt{a_n}(\widehat{MST}^{\pi}_{u,1,z} - \widehat{MST}^{\pi}_{u,2,z})$ converges weakly in probability to a mean-zero Gaussian random variable with variance

$$\sigma_z^{\pi 2} = \frac{1}{(1-\kappa)^2}\int_0^{\bar\tau_0}\int_0^{\bar\tau_0}\rho_{1,z}(s,t)\,\mathrm{d}s\,\mathrm{d}t, \tag{27}$$

where $\rho_{1,z}$ is defined in (25). This concludes the proof. □

Appendix C: Permutation of *Z*-estimators in a two-sample set-up

In this appendix, we discuss the asymptotic properties of randomly permuted Z-estimators. For this, we consider a two independent samples set-up with n_1 and n_2 i.i.d. random vectors $W_{11},\ldots,W_{1n_1} \sim \mathbb{P}_1$ and $W_{21},\ldots,W_{2n_2} \sim \mathbb{P}_2$, respectively. Let Θ be a subset of a Banach space, $\Psi_{n_1,1}, \Psi_{n_2,2} : \Theta \to \mathbb{L}$ be random maps, and $\Psi : \Theta \to \mathbb{L}$ be a deterministic map. Solutions (or approximate solutions) $\hat\theta_{n_i,i}$ to the equations $\Psi_{n_i,i}(\theta) \overset{!}{=} 0$ will be called *Z-estimators*. Due to the i.i.d. set-up, we assume the structure $\Psi_{n_i,i}(\theta)h = \mathbb{P}_{n_i,i}\psi_{\theta,h}$, for given measurable functions $\psi_{\theta,h}$ indexed by Θ and $h \in \mathcal{H}$ for some index set $\mathcal{H}$, where $\mathbb{P}_{n_i,i}$ denotes the i-th empirical process. Thus, we understand the equation system in the space $\mathbb{L} = \ell^\infty(\mathcal{H})$.

For the random permutation approach, we randomly re-assign the n_1+n_2 observations of the pooled sample $(W_{11},\ldots,W_{1n_1},W_{21},\ldots,W_{2n_2}) =: (W_1,\ldots,$

$W_{n_1+n_2}$) to the groups 1 and 2 without changing the original sample sizes. For a random permutation π of the numbers $1,\dots,n_1+n_2$, the permuted samples can be expressed as $W_{\pi(1)},\dots,W_{\pi(n_1)}$ and $W_{\pi(n_1+1)},\dots,W_{\pi(n_1+n_2)}$. For notational convenience, we denote the permuted samples by $W^\pi_{i1},\dots,W^\pi_{in_i}$, for sample group $i=1,2$, and the corresponding i-th permutation empirical process by $\mathbb{P}^\pi_{n_i,i}$. Let $\Psi^\pi_{n_i,i}(\theta)h = \mathbb{P}^\pi_{n_i,i}\psi_{\theta,h} \stackrel{!}{=} 0$ for all $h\in\mathcal{H}$ be the estimating equation corresponding to $\Psi_{n_i,i}(\theta)\stackrel{!}{=}0$, just based on the i-th permuted sample. We denote the (approximate) solution to the i-th permuted estimating equation by $\theta^\pi_{n_i,i}$. For future uses, let $\mathbb{G}^\pi_{n_i,i} = \sqrt{n_i}(\mathbb{P}^\pi_{n_i,i} - \mathbb{P}_{n_1+n_2})$ be the i-th normalized permutation empirical process, where $\mathbb{P}_{n_1+n_2}$ denotes the empirical process of the pooled sample. The centering at $\mathbb{P}_{n_1+n_2}$ seems reasonable, as this has an interpretation as a conditional expectation:

$$E(\mathbb{P}^\pi_{n_i,i}\psi_{\theta,h} \mid W_{ij} : i=1,2; j=1,\dots,n_i) = \mathbb{P}_{n_1+n_2}\psi_{\theta,h}.$$

Let $\bar\theta_{n_1+n_2}$ be the (approximate) solution to $\mathbb{P}_{n_1+n_2}\psi_{\theta,h}\stackrel{!}{=}0$ for all $h\in\mathcal{H}$.

The following theorem represents a version of Theorem 3.3.1 of [55] for the random permutation-based estimators.

Theorem C.1. *Assume that* $\frac{n_1}{n_1+n_2}\to\lambda\in(0,1)$ *as* $n_1+n_2\to\infty$, *define* $\bar{\mathbb{P}} = \lambda\mathbb{P}_1 + (1-\lambda)\mathbb{P}_2$, *and assume that Theorem 3.3.1 holds for each sample-specific Z-estimator* $\hat\theta_{n_1,1}$ *and* $\hat\theta_{n_2,2}$. *Let the criterion functions* $\psi_{\cdot,h}$ *be such that*

$$\|\mathbb{G}^\pi_{n_i,i}(\psi_{\theta^\pi_{n_i,i},h} - \psi_{\bar\theta_{n_1+n_2},h})\|_{\mathcal{H}} = o^*_P(1+\sqrt{n_i}\|\theta^\pi_{n_i,i} - \bar\theta_{n_1+n_2}\|). \tag{28}$$

Conditionally on $W_{11},W_{21},W_{12},W_{22},\dots$, *assume that*

$$(\sqrt{n_i}(\mathbb{P}^\pi_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{\bar\theta_{n_1+n_2},h})_{i=1}^2 \rightsquigarrow (Z_1,Z_2) \tag{29}$$

on $(\ell^\infty(\mathcal{H}))^2$ *in outer probability, where* (Z_1,Z_2) *is a tight random element.*

We assume that $\theta\mapsto\bar{\mathbb{P}}\psi_{\theta,h}$ *is uniformly Fréchet-differentiable in* $\ell^\infty(\mathcal{H})$ *at* $\bar\theta$ *with a continuously invertible derivative* $\bar{\mathbb{P}}\dot\psi_{\bar\theta,h}$, *and that, for any* $\theta,\tilde\theta$,

$$\|(\mathbb{P}_{n_1+n_2}\psi_{\theta,h} - \mathbb{P}_{n_1+n_2}\psi_{\tilde\theta,h}) - (\bar{\mathbb{P}}\psi_{\theta,h} - \bar{\mathbb{P}}\psi_{\tilde\theta,h})\|_{\mathcal{H}} = o^*_P(\|\theta-\tilde\theta\|) \tag{30}$$

as $n_1+n_2\to\infty$.

If $\theta^\pi_{n_i,i}$ *and* $\bar\theta_{n_1+n_2}$ *satisfy* $\|\mathbb{P}^\pi_{n_i,i}\psi_{\theta^\pi_{n_i,i},h}\|_{\mathcal{H}} = o^*_P(n^{-1/2})$, $i=1,2$, *and* $\|\mathbb{P}_{n_1+n_2}\psi_{\bar\theta_{n_1+n_2},h}\|_{\mathcal{H}} = o^*_P(n^{-1/2})$, *respectively, and if all three estimators converge in outer probability to* $\bar\theta$, *then*

$$\sqrt{n_i}(\bar{\mathbb{P}}\dot\psi_{\bar\theta,h})(\theta^\pi_{n_i,i} - \bar\theta_{n_1+n_2}) = -\sqrt{n_i}(\mathbb{P}^\pi_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{\bar\theta_{n_1+n_2},h} + o^*_P(1),\ i=1,2 \tag{31}$$

$$\rightsquigarrow -(Z_1,Z_2) \tag{32}$$

as $n_1 + n_2 \to \infty$ *conditionally on* $W_{11}, W_{21}, W_{12}, W_{22}, \ldots$, *in outer probability. Finally,*

$$(\sqrt{n_1}(\theta^{\pi}_{n_1,1} - \bar{\theta}_{n_1+n_2}), \sqrt{n_2}(\theta^{\pi}_{n_2,2} - \bar{\theta}_{n_1+n_2})) \rightsquigarrow -((\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h})^{-1}Z_1, (\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h})^{-1}Z_2)$$

conditionally on $W_{11}, W_{21}, W_{12}, W_{22}, \ldots$, *in outer probability.*

Note that, due to the equality $\mathbb{P}^{\pi}_{n_2,2} - \mathbb{P}_{n_1+n_2} = -\frac{n_1}{n_2}(\mathbb{P}^{\pi}_{n_1,1} - \mathbb{P}_{n_1+n_2})$, Z_1 and Z_2 are perfectly negatively linearly correlated: $Z_2 = -\sqrt{\frac{\lambda}{1-\lambda}}Z_1$.

Proof. The essential steps of this proof are similar to those in the proof of Theorem 3.3.1 of [55]. But for the sake of completeness, we shall present the whole proof.

The assumed consistencies of the pooled and the permuted estimators and then assumption (28) entail that

$$\begin{aligned}
&\sqrt{n_i}(\mathbb{P}_{n_1+n_2}\psi_{\theta^{\pi}_{n_i,i},h} - \mathbb{P}_{n_1+n_2}\psi_{\bar{\theta}_{n_1+n_2},h}) \\
&= \sqrt{n_i}(\mathbb{P}_{n_1+n_2}\psi_{\theta^{\pi}_{n_i,i},h} - \mathbb{P}^{\pi}_{n_i,i}\psi_{\theta^{\pi}_{n_i,i},h}) + o^*_P(1) \\
&= -\mathbb{G}^{\pi}_{n_i,i}\psi_{\theta^{\pi}_{n_i,i},h} + o^*_P(1) \\
&= -\sqrt{n_i}(\mathbb{P}^{\pi}_{n_i,i} - \mathbb{P}_{n_1+n_2})\psi_{\bar{\theta}_{n_1+n_2},h} + o^*_P(1 + \sqrt{n_i}\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\|).
\end{aligned} \tag{33}$$

Approximation (30) implies that the norm of the left-hand side of (33) equals

$$\sqrt{n_i}(\|\bar{\mathbb{P}}\psi_{\theta^{\pi}_{n_i,i},h} - \bar{\mathbb{P}}\psi_{\bar{\theta}_{n_1+n_2},h}\|_{\mathcal{H}} + o^*_P(\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\|)). \tag{34}$$

Additionally, the uniform Fréchet-differentiability of $\theta \mapsto \bar{\mathbb{P}}\psi_{\theta,h}$ at $\bar{\theta}$ and the continuous invertibility of the derivative $\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h}$ respectively imply that

$$\|\bar{\mathbb{P}}\psi_{\theta^{\pi}_{n_i,i},h} - \bar{\mathbb{P}}\psi_{\bar{\theta}_{n_1+n_2},h}\|_{\mathcal{H}} = \|(\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h})(\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2})\|_{\mathcal{H}} + o(\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\|) \tag{35}$$

and that the right-hand side in the previous display is bounded below by

$$c\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\| + o(\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\|)$$

for some positive constant c. Combine this with (34) and (33) to see that

$$\sqrt{n_i}\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\|(c + o^*_P(1)) \le O^*_P(1) + o^*_P(1 + \sqrt{n_i}\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\|),$$

conditionally on $W_{11}, W_{21}, W_{12}, W_{22}, \ldots$ in probability. Thus, in the same manner, $\theta^{\pi}_{n_i,i}$ is $\sqrt{n_i}$-consistent for $\bar{\theta}_{n_1+n_2}$ in norm.

Next, apply the approximation in (30) to the left-hand side of (33) and use the uniform Fréchet-differentiability of $\theta \mapsto \bar{\mathbb{P}}\psi_{\theta,h}$ at $\bar{\theta}$ to find that (33) equals

$$\sqrt{n_i}(\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h})(\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}) + o^*_P(\sqrt{n_i}\|\theta^{\pi}_{n_i,i} - \bar{\theta}_{n_1+n_2}\|).$$

Since $o_P^*(\sqrt{n_i}\|\theta^\pi_{n_i,i} - \bar{\theta}_{n_1+n_2}\|)$ and also $o_P^*(1 + \sqrt{n_i}\|\theta^\pi_{n_i,i} - \bar{\theta}_{n_1+n_2}\|)$ in the right-hand side of (33) are both $o_P^*(1)$, the assertion given in (31) follows from assumption (29).

The continuity of $(\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h})^{-1}$ together with the continuous mapping theorem and the just established conditional weak convergence (31) in outer probability imply the corresponding conditional weak convergence

$$(\sqrt{n_1}(\theta^\pi_{n_1,1} - \bar{\theta}_{n_1+n_2}), \sqrt{n_2}(\theta^\pi_{n_2,2} - \bar{\theta}_{n_1+n_2})) \rightsquigarrow -((\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h})^{-1}Z_1, (\bar{\mathbb{P}}\dot{\psi}_{\bar{\theta},h})^{-1}Z_2)$$

in outer probability. □

Appendix D: R code for accessing the breast cancer data set

```
if (!require("BiocManager", quietly = TRUE))
install.packages("BiocManager")
BiocManager::install("curatedBreastData")
library(curatedBreastData)

# read the data, study id 2034
data(curatedBreastDataExprSetList)
D=curatedBreastDataExprSetList$study_2034_GPL96_all

# Relapse-free survival
Data=data.frame(Y=D$RFS_months_or_MIN_months_of_RFS,status=1-D$RFS,
age=D$age,ER=D$ER_preTrt,sizeTum=D$tumor_stage_preTrt)
```

Acknowledgments

The authors would like to thank Sarah Friedrich for useful comments regarding the breast cancer data set, Merle Munko for helpful comments about uniform Hadamard differentiability and Joris Mooij for insightful comments regarding the causal interpretation and the decision theoretic perspective on choosing between two treatments. Furthermore, Dennis Dobler was supported by his new affiliations, i.e., Department of Statistics at TU Dortmund University and Research Center Trustworthy Data Science and Security of University Alliance Ruhr. Also, they would like to thank two anonymous referees and an associate editor whose comments and suggestions helped to improve the manuscript considerably.

References

[1] Ambrogi, F., Iacobelli, S. and Andersen, P. K. (2022). Analyzing differences between restricted mean survival time curves using pseudo-values. *BMC Medical Research Methodology* **22** 1–12.

[2] Amico, M. and Van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and its Application* **5** 311–342. MR3774750
[3] Amico, M., Van Keilegom, I. and Legrand, C. (2019). The single-index/Cox mixture cure model. *Biometrics* **75** 452–462. MR3999169
[4] Brendel, M., Janssen, A., Mayer, C.-D. and Pauly, M. (2014). Weighted logrank permutation tests for randomly right censored life science data. *Scandinavian Journal of Statistics* **41** 742–761. MR3249426
[5] Broët, P., De Rycke, Y., Tubert-Bitter, P., Lellouch, J., Asselain, B. and Moreau, T. (2001). A semiparametric approach for the two-sample comparison of survival times with long-term survivors. *Biometrics* **57** 844–852. MR1859819
[6] Broët, P., Kuznetsov, V. A., Bergh, J., Liu, E. T. and Miller, L. D. (2006). Identifying gene expression changes in breast cancer that distinguish early and late relapse among uncured patients. *Bioinformatics* **22** 1477–1485.
[7] Broet, P., Tubert-Bitter, P., De Rycke, Y. and Moreau, T. (2003). A score test for establishing non-inferiority with respect to short-term survival in two-sample comparisons with identical proportions of long-term survivors. *Statistics in Medicine* **22** 931–940.
[8] Brown, M., Tsodikov, A., Bauer, K. R., Parise, C. A. and Caggiano, V. (2008). The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: The California Cancer Registry, 1999–2004. *Cancer* **112** 737–747.
[9] Cai, C., Zou, Y., Peng, Y. and Zhang, J. (2012). smcure: An R-Package for estimating semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine* **108** 1255–1260.
[10] Chen, C.-M., Chen, H.-J. and Peng, Y. (2023). Mean residual life cure models for right-censored data with and without length-biased sampling. *Biometrical Journal* **65** 2100368. MR4603524
[11] Conner, S. C., Sullivan, L. M., Benjamin, E. J., LaValley, M. P., Galea, S. and Trinquart, L. (2019). Adjusted restricted mean survival times in observational studies. *Statistics in Medicine* **38** 3832–3860. MR3999249
[12] Conner, S. C. and Trinquart, L. (2021). Estimation and modeling of the restricted mean time lost in the presence of competing risks. *Statistics in Medicine* **40** 2177–2196. MR4239901
[13] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34** 187–202. MR0341758
[14] Ditzhaus, M. and Friedrich, S. (2020). More powerful logrank permutation tests for two-sample survival data. *Journal of Statistical Computation and Simulation* **90** 2209–2227. MR4124368
[15] Ditzhaus, M., Yu, M. and Xu, J. (2023). Studentized permutation method for comparing two restricted mean survival times with small sample from randomized trials. *Statistics in Medicine* **42** 2226–2240. MR4594709
[16] Dobler, D. (2019). Bootstrapping the Kaplan–Meier estimator on the

whole line. *Annals of the Institute of Statistical Mathematics* **71** 213–246. MR3898432

[17] Dobler, D. and Pauly, M. (2018). Bootstrap-and permutation-based inference for the Mann–Whitney effect for right-censored and tied data. *Test* **27** 639–658. MR3858451

[18] Dobler, D., Pauly, M. and Scheike, T. H. (2019). Confidence bands for multiplicative hazards models: Flexible resampling approaches. *Biometrics* **75** 906–916. MR4012096

[19] Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *The Annals of Statistics* **11** 49–58. MR0684862

[20] Gill, R. D. (1980). Censoring and stochastic integrals. *Statistica Neerlandica* **34** 124–124. MR0596815

[21] Halpern, J., Wm, B. and Jun, B. (1987). Cure rate models: power of the logrank and generalized Wilcoxon tests. *Statistics in Medicine* **6** 483–489.

[22] Horiguchi, M. and Uno, H. (2020). On permutation tests for comparing restricted mean survival time with small sample from randomized trials. *Statistics in Medicine* **39** 2655–2670. MR4133131

[23] Hortobagyi, G. (2003). The curability of breast cancer: present and future. *European Journal of Cancer Supplements* **1** 24–34.

[24] Kersey, J. H., Weisdorf, D., Nesbit, M. E., LeBien, T. W., Woods, W. G., McGlave, P. B., Kim, T., Vallera, D. A., Goldman, A. I., Bostrom, B. et al. (1987). Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia. *New England Journal of Medicine* **317** 461–467.

[25] Klein, J. P., Logan, B., Harhoff, M. and Andersen, P. K. (2007). Analyzing survival curves at a fixed point in time. *Statistics in Medicine* **26** 4505–4519. MR2410057

[26] Laska, E. M. and Meisner, M. J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics* **48** 1223–1234.

[27] Legrand, C. and Bertrand, A. (2019). Cure models in oncology clinical trials. *Textb Clin Trials Oncol Stat Perspect* **1** 465–492.

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

[29] Li, Y. and Feng, J. (2005). A nonparametric comparison of conditional distributions with nonnegligible cure fractions. *Lifetime Data Analysis* **11** 367–387. MR2186651

[30] López-Cheda, A., Jácome, M. A. and Cao, R. (2017). Nonparametric latency estimation for mixture cure models. *Test* **26** 353–376. MR3650531

[31] Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics* **60** 545–574. MR2434411

[32] Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika* **91** 331–343. MR2081305

[33] Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika* **79** 731–739. MR1209474

[34] Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley New York. MR1453117
[35] Müller, U. U. and Van Keilegom, I. (2019). Goodness-of-fit tests for the cure rate in a mixture cure model. *Biometrika* **106** 211–227. MR3912392
[36] Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics* **22** 712–731. MR1292537
[37] Musta, E., Patilea, V. and Van Keilegom, I. (2022). A presmoothing approach for estimation in mixture cure models. *Bernoulli* **28** 2689–2715. MR4474559
[38] Parsa, M. and Van Keilegom, I. (2023). Accelerated failure time vs Cox proportional hazards mixture cure models: David vs Goliath? *Statistical Papers* **64(3)** 835–855.
[39] Patilea, V. and Van Keilegom, I. (2020). A general approach for cure models in survival analysis. *The Annals of Statistics* **48** 2323–2346. MR4134797
[40] Peng, Y. and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56** 237–243.
[41] Peng, Y. and Taylor, J. M. (2017). Residual-based model diagnosis methods for mixture cure models. *Biometrics* **73** 495–505. MR3665966
[42] Peng, Y. and Yu, B. (2021). *Cure Models: Methods, Applications, and Implementation*. Chapman and Hall/CRC. MR4596124
[43] Royston, P. and Parmar, M. K. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine* **30** 2409–2421. MR2829141
[44] Royston, P. and Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology* **13** 1–15.
[45] Safari, W. C., López-de Ullibarri, I. and Jácome, M. A. (2023). Latency function estimation under the mixture cure model when the cure status is available. *Lifetime Data Analysis* **29** 608–627. MR4599636
[46] Scharfstein, D. O., Tsiatis, A. A. and Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis* **4** 355–391.
[47] Socié, G., Stone, J. V., Wingard, J. R., Weisdorf, D., Henslee-Downey, P. J., Bredeson, C., Cahn, J.-Y., Passweg, J. R., Rowlings, P. A., Schouten, H. C. et al. (1999). Long-term survival and late deaths after allogeneic bone marrow transplantation. *New England Journal of Medicine* **341** 14–21.
[48] Sposto, R., Sather, H. N. and Baker, S. A. (1992). A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics* **48** 87–99.
[49] Stringer, S., Denys, D., Kahn, R. S. and Derks, E. M. (2016). What

cure models can teach us about genome-wide survival analysis. *Behavior Genetics* **46** 269–280.

[50] Sy, J. P. and Taylor, J. M. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56** 227–236. MR1767631

[51] Tamura, R. N., Faries, D. E. and Feng, J. (2000). Comparing time to onset of response in antidepressant clinical trials using the cure model and the Cramer–von Mises test. *Statistics in Medicine* **19** 2169–2184.

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

[53] Thomas, E. D., Flournoy, N., Buckner, C. D., Clift, R. A., Fefer, A., Neiman, P. E. and Storb, R. (1977). Cure of leukemia by marrow transplantation. *Leukemia Research* **1** 67–70.

[54] Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., Schrag, D., Takeuchi, M., Uyama, Y., Zhao, L. et al. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* **32** 2380.

[55] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer-Verlag, New York With applications to statistics. MR1385671

[56] Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J. et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet* **365** 671–679.

[57] Wang, Z., Wang, C. and Wang, X. (2023). Estimating causal effects in observational studies for survival data with a cure fraction using propensity score adjustment. *Biometrical Journal* **65** 2100357. MR4677807

[58] Wolski, A., Graffeo, N., Giorgi, R. and working survival group, C. (2020). A permutation test based on the restricted mean survival time for comparison of net survival distributions in non-proportional excess hazard settings. *Statistical Methods in Medical Research* **29** 1612–1623. MR4106960

[59] Wycinka, E. and Jurkiewicz, T. (2017). Mixture cure models in prediction of time to default: comparison with logit and Cox models. In *Contemporary Trends and Challenges in Finance. Springer Proceedings in Business and Economics* 221–231. Springer, Cham.

[60] Yilmaz, Y. E., Lawless, J. F., Andrulis, I. L. and Bull, S. B. (2013). Insights from mixture cure modeling of molecular markers for prognosis in breast cancer. *Journal of Clinical Oncology* **31** 2047–2054.

[61] Ying, Z. (1989). A note on the asymptotic properties of the product-limit estimator on the whole line. *Statistics & Probability Letters* **7** 311–314. MR0980706

[62] Zhang, J. and Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine* **26** 3157–3171. MR2380509

[63] Zhao, L., Claggett, B., Tian, L., Uno, H., Pfeffer, M. A.,

Solomon, S. D., Trippa, L. and Wei, L. (2016). On the restricted mean survival time curve in survival analysis. *Biometrics* **72** 215–221. MR3500590

[64] Zhao, X. and Zhou, X. (2010). Empirical receiver operating characteristic curve for two-sample comparison with cure fractions. *Lifetime Data Analysis* **16** 316–332. MR2657893