

Weighted pseudo-values for partly unobserved group membership in paediatric stem cell transplantation studies

Statistical Methods in Medical Research

2022, Vol. 31(1) 76–86

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802211041756

journals.sagepub.com/home/smm


Martina Mittlböck^{1,*} , Ulrike Pötschger^{2,*} ,
and Harald Heinzl¹ 

Abstract

Generalised pseudo-values have been suggested to evaluate the impact of allogeneic stem cell transplantation on childhood leukaemia. The approach compares long-term survival of two cohorts defined by the availability or non-availability of suitable donors for stem cell transplantation. A patient's cohort membership becomes known only after completed donor search with or without an identified donor. If a patient suffers an event during donor search, stem cell transplantation will no longer be indicated. In such a case, donor search will be ceased and cohort membership will remain unknown. The generalised pseudo-values approach considers donor identification as binary time-dependent covariate and uses inverse-probability-of-censoring weighting to adjust for non-identified donors. The approach leads to time-consuming computations due to multiple redefinitions of the risk set for pseudo-value calculation and an explicit adjustment for waiting-time bias. Here, the problem is looked at from a different angle. By considering the probability that a donor would have been identified after ceasing of donor search, weights for common pseudo-values are defined. This leads to a faster alternative approach as only a single risk set is necessary. Extensive computer simulations show that both, the generalised and the new weighted pseudo-values approach, provide approximately unbiased estimates. Confidence interval coverage is satisfactory for typical clinical scenarios. In situations, where donor identification takes considerably longer than usual, the weighted pseudo-values approach is preferable. Both approaches complement each other as they have different potential in addressing further aspects of the underlying medical question.

Keywords

Cumulative hazard ratio, non-proportional hazards, selection bias, immortal time bias, generalised pseudo values, binary time-dependent covariate

1 Background

Childhood leukaemia studies that compare allogeneic stem cell transplantation (SCT) with conventional chemotherapy usually exhibit specific features that result in statistically challenging data analyses. After being considered eligible for SCT, a patient is included in the study, conventional chemotherapy is administered, and donor search is started in international donor registries if no suitable related donor has been identified. These registries contain pre-typed blood sample data of volunteers willing to become stem cell donors. If a potential donor is pre-selected and his/her willingness could be re-confirmed, then a new blood sample will be exactly typed. If a match is achieved, chemotherapy can be stopped

¹Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria

²Children's Cancer Research Institute, Vienna, Austria

*Both authors contributed equally.

Corresponding author:

Harald Heinzl, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria.
Email: harald.heinzl@meduniwien.ac.at

and SCT can be performed. Chemotherapy is the sole treatment for those patients, for whom no matched donor can be identified during the search period. According to their donor availability status, patients belong to either of two cohorts, that is, with and without available donor.

It may happen that a patient suffers death, relapse, or other primary disease-specific problems during the donor search period. Consequently, the indication for SCT will be lost, donor search will be ceased, and donor availability status will remain unknown for this patient. Note that donor availability status can, nonetheless, be considered as an external covariate as it exists beyond patient survival or disease progression.¹ After successful donor identification, SCT will be performed unless contraindications prevent it. Therefore, in order to avoid selection bias, it is crucial that statistical approaches consider donor identification rather than SCT applied.

Ideally, the so-called genetic randomisation approach would be used to evaluate SCT trials.²⁻⁴ This approach mimics the intention-to-treat principle of randomised trials by comparing the long-term survival of patients with available donor (SCT possible) and without available donor (SCT impossible). Unfortunately, missing donor availability status makes it often impossible to apply genetic randomisation and alternative approaches have to be considered.

The prevailing Cox model with a binary time-dependent covariate is hampered by non-proportional hazards. The benefits of an SCT are expected to fully unfold on the long run; on the short run, an increase in treatment-related events is commonly observed. As the Cox model considers relative instantaneous risks, it cannot be used to assess cumulative survival effects in a non-proportional hazard situation.^{5,6} Landmark analysis and dynamic prediction by landmarking are other popular approaches;⁷⁻¹⁰ however, their results can be affected by the arbitrarily chosen landmark time. Pötschger et al.¹¹ discuss all these and other approaches in more detail and suggest a new approach based on so-called generalised pseudo-values (GPVs) which considers long-term survival, mimics the intention-to-treat principle, yet overcomes the missing donor availability and the non-proportional hazards problem.

Unfortunately, the practical application of GPVs is computationally intensive and time-consuming. This is, among others, especially problematic during study design, when the required sample size or statistical power has to be assessed by means of computer simulation. Furthermore, suboptimal confidence interval coverage rates were observed in certain cases, and an ad-hoc parametric bootstrap procedure was developed to overcome this issue.¹¹ Meanwhile, a programming error in the original implementation of the procedure was discovered. The error has, by coincidence, resulted in reasonable confidence intervals for the cases considered¹¹; however, after error recovery, it became apparent that the intervals are in fact too wide leading to unacceptably high coverage rates. As no workable alternative solution is available, the ad-hoc parametric bootstrap procedure will not be pursued any further for the GPV approach.

Here, the new approach of weighted pseudo-values (WPVs) is proposed as alternative to GPVs; both estimating cohort survival probabilities at a pre-specified time point and the associated cumulative hazard ratio (CHR). WPVs are based on an individually adapted probability weighting of common pseudo-values and can be efficiently computed in a single step. In the following, GPVs and WPVs will be introduced in the section ‘Methods’. They will be compared by means of a simulation study in the section ‘Results’. Further issues and additional aspects will be summarised and discussed in the section ‘Discussion’.

2 Methods

In the following, the term survival is linked to the outcome-related events of interest, like death, or a composite endpoint that additionally includes relapse or other disease-related events. Furthermore, the term donor search-ceasing events comprise all events that permanently contraindicate SCT and thus lead to stop of donor search. Note that donor search-ceasing events do not have to be identical with outcome-related events. For example, if overall survival is studied, then death of patient will cease donor search as well as define the outcome of interest, whereas a relapse will only cease donor search. Since donor availability is an external covariate, it is independent of these donor search-ceasing events.

The survival probability at time t^* of patients with a donor available, $S_1(t^*)$, is to be compared with the survival probability of those without an available donor, $S_0(t^*)$. Here, t^* is a pre-specified time point used to indicate long-term survival or cure of the patients.

The maximum donor search time $t^{\text{search}} \leq t^*$ is used to determine the donor availability status (see Figure 1). That is, a donor is considered available/not available, if it can/cannot be identified up to t^{search} , respectively. If a patient without identified donor suffers a donor search-ceasing event (or is censored) before t^{search} , then the donor availability status will remain unknown for this patient.

Define the random variable W , $0 \leq W \leq t^{\text{search}}$, as waiting time in the initial state 0 until a donor is identified and the patient moves to state 1. Let $\lambda_{01}(t)$ and $f_{01}(t) = \lambda_{01}(t) \exp\left[-\int_0^t \lambda_{01}(v)dv\right]$ denote the hazard function and the density of the

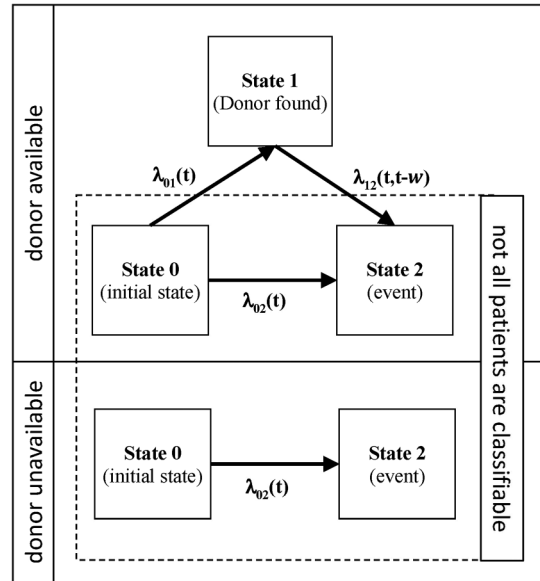


Figure 1. Stochastic processes of the two identical populations that only differ with regard to their donor availability status; $\lambda_{01}(t)$, $\lambda_{02}(t)$ and $\lambda_{12}(t, t - w)$ denote hazard functions, and w is the time of transition to state 1. In the donor available group, all patients leave state 0 either to state 1 or state 2 by no later than t^{search} .

distribution of W , respectively. Then the survival probability at t^* for patients with an available donor is

$$S_1(t^*) = \int_0^{t^*} f_{01}(w) S_1(t^*|w) dw,$$

where $S_1(t^*|w) = \exp\left[-\left\{\int_0^w \lambda_{02}(v)dv + \int_w^{t^*} \lambda_{12}(v, v - w)dv\right\}\right]$ denotes survival in a population with donor available at a fixed waiting time $W = w$. The hazard function $\lambda_{12}(t, t - w)$ describes the transition from state 1 to the event state 2; it can depend on the time t since the patient entered the study (in initial state 0) as well as the time $t - w$ since a donor was identified and the patient moved to state 1.

The survival probability at t^* for patients without donor available is

$$S_0(t^*) = \exp\left[-\left\{\int_0^{t^*} \lambda_{02}(v)dv\right\}\right]$$

where $\lambda_{02}(t)$ is the hazard function of transition from initial state 0 to the final state 2 (Figure 1).

Note that the hazard $\lambda_{02}(t)$ for the direct transition from the initial state 0 to the event state 2 is equal for both populations up to t^{search} ; thereafter there will be, by definition, no patients anymore in state 0 of the donor available population.

The cHR at time t^* for the comparison of patients with and without an available donor is $\text{cHR}(t^*) = \log(S_1(t^*)) / \log(S_0(t^*))$.

2.1 Common pseudo-values

Given n patients, the common pseudo-value at time t^* for the i th patient is

$$\hat{V}_i(t^*) = n\hat{S}(t^*) - (n - 1)\hat{S}^{-i}(t^*), \tag{1}$$

$i = 1, \dots, n$, where $\hat{S}(t^*)$ denotes the Kaplan-Meier estimate at t^* and $\hat{S}^{-i}(t^*)$ denotes the Kaplan-Meier estimate at t^* without the i th patient.¹²

As common pseudo-values are asymptotically conditionally unbiased given covariates, they can be used as a response variable in a regression model.^{13,14} However, i.i.d. observations and independent right censoring have to be assumed; and the risk set at time t^* has to be sufficiently large.

2.2 Generalised pseudo-values

Assume that for the first m out of n patients a transition to state 1 has been observed, and denote the observed transition times w_1, \dots, w_m . The GPV approach of Pötschger et al.¹¹ now defines $n + m$ GPVs, $\hat{\mathbf{V}} = (\hat{V}_{1,0}(t^*), \dots, \hat{V}_{n,0}(t^*), \hat{V}_{1,1}(t^*|w_1), \dots, \hat{V}_{m,1}(t^*|w_m))'$.

Here $\hat{V}_{i,0}(t^*) = n\hat{S}_0(t^*) - (n-1)\hat{S}_0^{-i}(t^*)$, $i = 1, \dots, n$, are common pseudo-values with $\hat{S}_0(t^*)$ denoting the Kaplan-Meier estimate at t^* for the risk of a direct transition from state 0 to state 2; where patients with an available donor are censored at their observed transition times w_1, \dots, w_m .

The remaining m GPVs are $\hat{V}_{i,1}(t^*|w_i) = \hat{S}_0(w_i) \hat{U}_{i,1}(t^*|w_i)$, with $\hat{U}_{i,1}(t^*|w_i) = n_i \hat{S}(t^*|T \geq w_i) - (n_i - 1) \hat{S}^{-i}(t^*|T \geq w_i)$, $i = 1, \dots, m$. $\hat{S}_0(w_i)$ has just been introduced above; and $\hat{S}(t^*|T \geq w_i)$ is the Kaplan-Meier estimate at t^* based on all n_i patients still at risk at w_i .

Finally, weights $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_{1,0}, \dots, \hat{\gamma}_{n,0}, \hat{\gamma}_{1,1}, \dots, \hat{\gamma}_{m,1})'$ have to be defined for $\hat{\mathbf{V}}$. Trivially, $\hat{\gamma}_{i,0} = 1$ for $i = 1, \dots, n$. In contrast, $\hat{\gamma}_{1,1}, \dots, \hat{\gamma}_{m,1}$ are needed to adjust for unobserved $0 \rightarrow 1$ transitions due to donor search-ceasing events and censoring. This is achieved by inverse probability weighting with $\hat{G}(w)$, the Kaplan-Meier estimate of all n patients where both, donor search-ceasing events and censoring, are considered as events; whereas observed $0 \rightarrow 1$ transitions are considered as censored at w_1, \dots, w_m . That is, $\hat{\gamma}_{i,1} = m \hat{G}^{-1}(w_i) / (\sum_{j=1}^m \hat{G}^{-1}(w_j))$, $i = 1, \dots, m$; note that $\sum_{i=1}^m \hat{\gamma}_{i,1} = m$.

Now, $\bar{V}_0 = \frac{1}{n} \sum_{i=1}^n \hat{\gamma}_{i,0} \hat{V}_{i,0}(t^*) = \frac{1}{n} \sum_{i=1}^n \hat{V}_{i,0}(t^*)$ and $\bar{V}_1 = \frac{1}{m} \sum_{i=1}^m \hat{\gamma}_{i,1} \hat{V}_{i,1}(t^*|w_i)$ are consistent estimators of $S_0(t^*)$ and $S_1(t^*)$, respectively.

GPVs can be used as response variable in a weighted generalised linear model with log-minus-log link function $g(y) = \log(-\log(y))$, a normal response probability distribution, and a robust sandwich variance estimator. Let $\mathbf{x} = (x_1, \dots, x_{n+m})'$ with $x_i = 0$ for $i = 1, \dots, n$ and $x_i = 1$ otherwise. With the linear predictor $\beta_0 + \beta_1 x_i$, the parameter estimates correspond to $\hat{\beta}_0 = g(\bar{V}_0)$ and $\hat{\beta}_1 = g(\bar{V}_1) - g(\bar{V}_0)$, so that $\exp(-\exp(\hat{\beta}_0))$ and $\exp(-\exp(\hat{\beta}_0 + \hat{\beta}_1))$ are the estimates for $S_0(t^*)$ and $S_1(t^*)$, respectively. Finally, $\exp(\hat{\beta}_1)$ provides a consistent estimate of the cHR at time t^* for the comparison of patients with and without an available donor.

2.3 Weighted pseudo-values

The novel WPV approach proportionally allocates a patient with uncertain group membership into the two groups with and without donor according to its probability that a donor might still be found until t^{search} . There are $n = n_U + m + n_C$ patients (Table 1); for notational convenience assume that the first n_U patients are those where it is known that a donor is unavailable (i.e. none has been identified until t^{search}), the next m patients are those with an identified donor; and the remaining n_C patients are those where donor search was ceased and group membership is unknown.

For the latter patients, let κ_i denote the probability that a donor will be identified until t^{search} , given that no donor has been found until t_i , the time when donor search was ceased. Now, $\kappa_i = \frac{S_D(t_i) - S_D(t^{\text{search}})}{S_D(t_i)}$, $i = 1, \dots, n_C$, where $S_D(t_i)$ is the probability that no donor has been found for the i th patient by t_i , so that, consequently, $S_D(t^{\text{search}})$ is the probability of belonging to the group without donor (additional details are provided in Section A of Supplemental Materials). A natural estimator of κ_i is provided by $\hat{\kappa}_i = \frac{\hat{S}_D(t_i) - \hat{S}_D(t^{\text{search}})}{\hat{S}_D(t_i)}$, $i = 1, \dots, n_C$; where $\hat{S}_D(t)$ is the Kaplan-Meier estimate for time to donor identification. Note that here all donor search-ceasing events as well as censoring are considered as censoring events, which are non-informative with regard to time to donor identification.

Consequently, the probability of no donor available is estimated by $1 - \hat{\kappa}_i$, $i = 1, \dots, n_C$. Note that if – for some reason – a donor is found after donor search has already been ceased, then it would be sensible to include this knowledge by setting $\hat{\kappa}_i = 1$ for the patient concerned.

Table 1. Weights of the WPV and the GPV approach.

Observed patients $n = n_U + m + n_C$	WPV weights for the estimation of		GPV weights for the estimation of	
	$S_0(t^*)$	$S_1(t^*)$	$S_0(t^*)$	$S_1(t^*)$
Without donor n_U	1	0	1	–
With donor m	0	1	1	$\hat{\gamma}_{i,1}$
Donor search ceased n_C	$(1 - \hat{\kappa}_i)$	$\hat{\kappa}_i$	1	–
Sum of weights per column	$n_U + \sum_{i=1}^{n_C} (1 - \hat{\kappa}_i)$	$m + \sum_{i=1}^{n_C} \hat{\kappa}_i$	n	$\sum_{i=1}^m \hat{\gamma}_{i,1}$
Total sum of weights	n		$n + m$	

Now, for all patients common pseudo-values $\hat{V}_i(t^*)$, $i = 1, \dots, n$, (see equation (1)) are calculated.

The WPV approach defines a vector with $n + n_C = n_U + m + 2n_C$ of these common pseudo-values $\hat{\mathbf{V}}_{\mathbf{D}} = (\hat{V}_1(t^*), \dots, \hat{V}_n(t^*), \hat{V}_{n_U+m+1}(t^*), \dots, \hat{V}_n(t^*))'$, where the pseudo-values of the patients of the n_C group, $\hat{V}_{n_U+m+1}(t^*), \dots, \hat{V}_n(t^*)$, occur twice. The corresponding weight vector is $\hat{\boldsymbol{\gamma}}_{\mathbf{D}} = (\hat{\gamma}_{1,D}, \dots, \hat{\gamma}_{n+n_C,D})'$ with $\hat{\gamma}_{i,D} = 1$ for $i = 1, \dots, n_U + m$, $\hat{\gamma}_{i,D} = 1 - \hat{\kappa}_{i-(n_U+m)}$ for $i = n_U + m + 1, \dots, n$, and $\hat{\gamma}_{i,D} = \hat{\kappa}_{i-n}$ for $i = n + 1, \dots, n + n_C$.

Let $\mathbf{x}_{\mathbf{D}} = (x_{1,D}, \dots, x_{n+n_C,D})'$ indicate group membership with $x_{i,D} = 0$ for $i = 1, \dots, n_U$ and $i = n_U + m + 1, \dots, n$ and $x_{i,D} = 1$ otherwise. Then $\bar{V}_{0,D} = \frac{1}{\sum_{i=1}^{n+n_C} (1-x_{i,D})\hat{\gamma}_{i,D}} \sum_{i=1}^{n+n_C} (1-x_{i,D})\hat{\gamma}_{i,D}\hat{V}_{i,D}(t^*)$ and $\bar{V}_{1,D} = \frac{1}{\sum_{i=1}^{n+n_C} x_{i,D}\hat{\gamma}_{i,D}} \sum_{i=1}^{n+n_C} x_{i,D}\hat{\gamma}_{i,D}\hat{V}_{i,D}(t^*)$ are consistent estimators of $S_0(t^*)$ and $S_1(t^*)$, respectively.^{13,15,16}

The vector $\hat{\mathbf{V}}_{\mathbf{D}}$ can be used as response variable in a weighted generalised linear model with log-minus-log link $g(y) = \log(-\log(y))$, a normal response probability distribution, and a robust sandwich variance estimator. Using the linear predictor $\beta_{0,D} + \beta_{1,D}x_{i,D}$, then, in analogy to the GPV approach, the parameter estimates correspond to $\hat{\beta}_{0,D} = g(\bar{V}_{0,D})$ and $\hat{\beta}_{1,D} = g(\bar{V}_{1,D}) - g(\bar{V}_{0,D})$, respectively, and $\exp(\hat{\beta}_{1,D})$ is a consistent estimator of the cHR at time t^* for the comparison of patients with and without an available donor.

An overview of the weights of both approaches and their contribution to the estimation of $S_0(t^*)$ and $S_1(t^*)$ can be found in Table 1. A SAS macro for the WPV approach is available via GitHub.^{17,18}

3 Results

3.1 Simulation studies

The two simulation studies of Pötschger et al.¹¹ are used to enable a straightforward comparison of the GPV with the WPV approach. In short, survival times were generated using the inversion method,¹⁹ which was extended by using a parametric mixture cure model²⁰ to allow for a plateau of the survival curve that represents cured patients. Uniform censoring times between 0 to 6 years and 0 to 11 years were superimposed in the simulation scenarios of simulation studies 1 and 2, respectively. For each scenario, sample sizes of 400 and 1000 were generated for 1000 simulation runs.

At first, the WPV approach is applied to the original simulated data sets of Pötschger et al.,¹¹ then the WPV results are compared to the original GPV results (Figures 2 to 5). Results of a considerably extended version of simulation study 2 are provided in the Supplemental Materials (Section D, Figures S2 to S13).

For the sake of convenience, the bias distributions (Figures 2, 4, and Supplemental Figures S2 to S7) are labelled $S_0(t^*)$, $S_1(t^*)$ and cHR(t^*), although they actually show the bias of the corresponding parameter estimates of the particular generalised linear models; that is, $\hat{\beta}_0$, $\hat{\beta}_0 + \hat{\beta}_1$ and $\hat{\beta}_1$ for the GPV, and $\hat{\beta}_{0,D}$, $\hat{\beta}_{0,D} + \hat{\beta}_{1,D}$ and $\hat{\beta}_{1,D}$ for the WPV approach, respectively.

The grey horizontal band in the confidence interval (CI) coverage plots (Figures 3, 5, and Supplemental Figures S8 to S13) marks the area between the 0.025 and the 0.975 quantile of a binomial distribution with parameters 1000 (number of trials, here a number of simulation runs) and 0.95 (success probability, here true confidence level $1 - \alpha$). That is, after 1000 simulation runs, the observed coverage rate of a CI with a true confidence level $1 - \alpha = 0.95$ is contained in this band with a probability of 0.95 ($= 0.975 - 0.025$). Hence, an observed coverage rate below or above this band may be interpreted as indication that the considered CI actually is too liberal or too conservative, respectively.

Both approaches, GPV and WPV, can be implemented using standard routines available in the majority of statistical software packages; the following numerical results have been obtained with SAS software.¹⁷

Simulation study 1

This simulation study was performed to examine bias and 95% CI coverage rate of the survival probabilities of the two approaches in an extreme setting with discrete waiting times at either 0.5, 1 or 3 years. No donors are available for 25% of the patients. Donors are available for the remaining 75% so that always 25% belong to each of the three waiting times of 0.5, 1 and 3 years, respectively. The outcome is evaluated at 5 years ($t^*=5$).

As donor search-ceasing events or censoring prevent donor identification, available donors are identified on average only with a probability of 0.88, 0.74 and 0.28 at 0.5, 1 and 3 years, respectively. Consequently, only 22%, 18% and 7% of all patients (m in Table 1) will have donors identified at 0.5, 1 and 3 years, respectively, and the remaining 53% will have no identified donor ($n_u + n_c$, see Table 1); remember that by definition there are 25% of patients in each of the four groups.

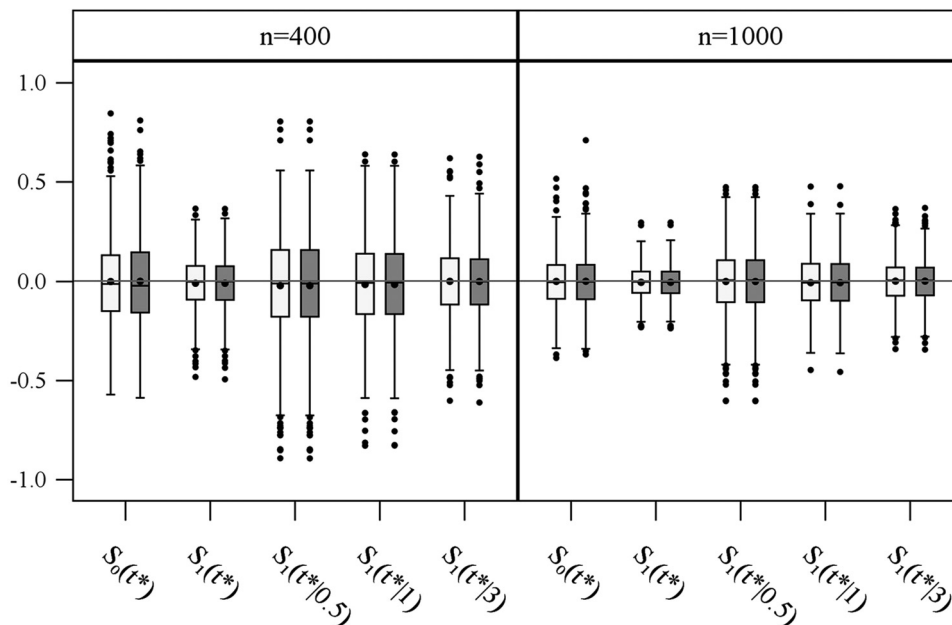


Figure 2. Bias distribution of log-minus-log survival probabilities estimated with the GPV (light grey) and WPV (dark grey) approach in simulation study I (1000 repetitions) for sample sizes of 400 and 1000. By definition, a donor can only be identified at 0.5, 1 or 3 years after start of donor search.
 GPV: generalised pseudo-value; WPV: weighted pseudo-value.

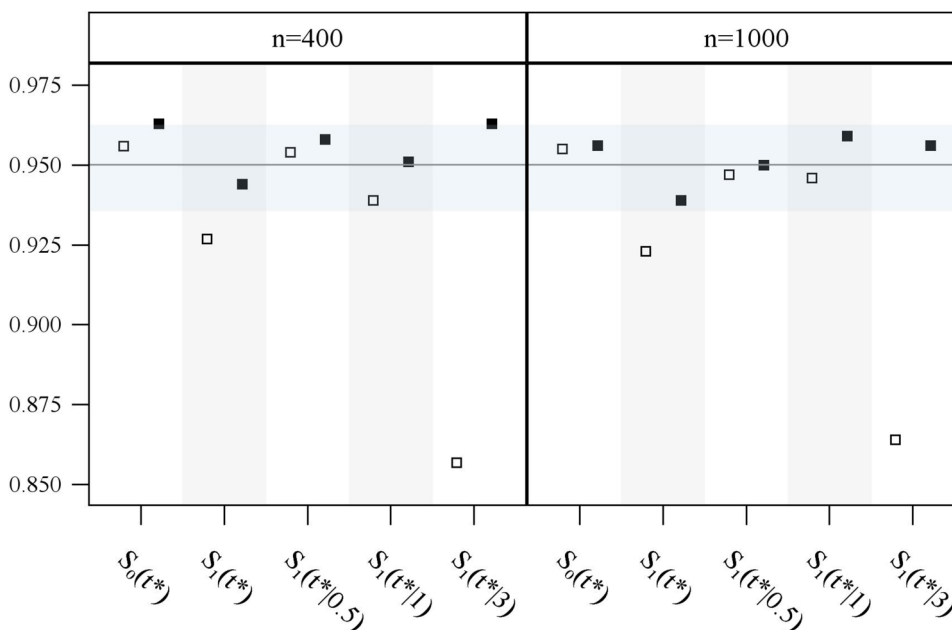


Figure 3. Observed coverage rates of 95% confidence intervals for survival probabilities estimated with the GPV (square) and WPV (square-filled) approach for simulation study I (1000 repetitions) and sample sizes of 400 and 1000. By definition, a donor can only be identified at 0.5, 1 or 3 years after start of donor search.
 GPV: generalised pseudo-value; WPV: weighted pseudo-value.

The GPV approach accounts for this missing observations at the three discrete donor identification times by weighting observed donor identifications to reconstruct the true waiting time distribution, leading to average estimated weights $\hat{\gamma}_{i,1}$ of 0.72, 0.86 and 2.26 for $w = 0.5, 1$ and 3 , respectively.

In contrary, the WPV approach reduces the number of patients in the group of no identified donor and enriches the donor groups with waiting times at 0.5, 1 and 3 years according to their probability that a donor might be identified beyond patients' observed survival time. That is, the underlying group frequencies of 25% for all four subgroups are approximately restored. This enrichment of the three donor available groups to the cost of the no identified donor group exemplifies how the WPV approach prevents immortal time bias.

Figure 2 shows that both approaches result in unbiased parameter estimates for the no donor group as well as the donor available groups; either when donor availability is considered in general, or when survival is separately estimated for the given waiting times at 0.5, 1 and 3 years (see Supplemental Materials, Section B). Figure 3 gives coverages of 95% CIs for $S_0(t^*)$, $S_1(t^*)$, $S_1(t^*|0.5)$, $S_1(t^*|1)$ and $S_1(t^*|3)$, respectively. The CI coverage rates for $S_1(t^*|3)$ of the GPV approach are much too low as donor identification occurs late at 3 years; this feature also shows through in the general case $S_1(t^*)$. The novel and computationally easier WPV approach shows very good CI coverage rates in all situations.

Simulation study 2

This simulation study is performed to assess unbiasedness of the parameter estimates and CI coverage of both approaches in four realistic settings from paediatric oncology and three theoretical settings.¹¹ The realistic situations represent the most common departures from proportional hazards: differences in long-term survival only (A), crossing survival curves (B, C), and a situation where only differences in short-term survival (D) are observed. The theoretical settings consider a proportional hazards model (E), a no-effect model (F), and a scenario with unrealistically long waiting times (G); see Figure S1 in Supplemental Materials, Section C. The proportions of available donors are 0.25 for scenario A, 0.4 for scenarios B–F and 0.45 for scenario G, respectively. The outcome is evaluated at 5 years ($t^*=5$), and the donor search period is set to 5 years ($t^{\text{search}} = 5$), too.

Figure 4 shows that both approaches result in unbiased parameter estimates for all seven simulation scenarios A–G and the sample sizes $n = 400$ and $n = 1000$, respectively.

Coverage rates of 95% CIs are provided in Figure 5. Note that the CI for $\text{cHR}(t^*)$ is the most important one as it corresponds to a statistical test of comparing both therapies at t^* . For both approaches, the 95% CI of $S_0(t^*)$ shows good coverage except for GPV in scenario G when $n = 1000$. This, however, seems to be a simulation artefact as the performance for $n = 400$ is adequate. WPV shows acceptable coverage in all situations for both, $S_1(t^*)$ and the cumulative hazard ratio $\text{cHR}(t^*)$; however, the coverage rates of GPV are only acceptable for scenarios A–F. Analogous to simulation study 1,

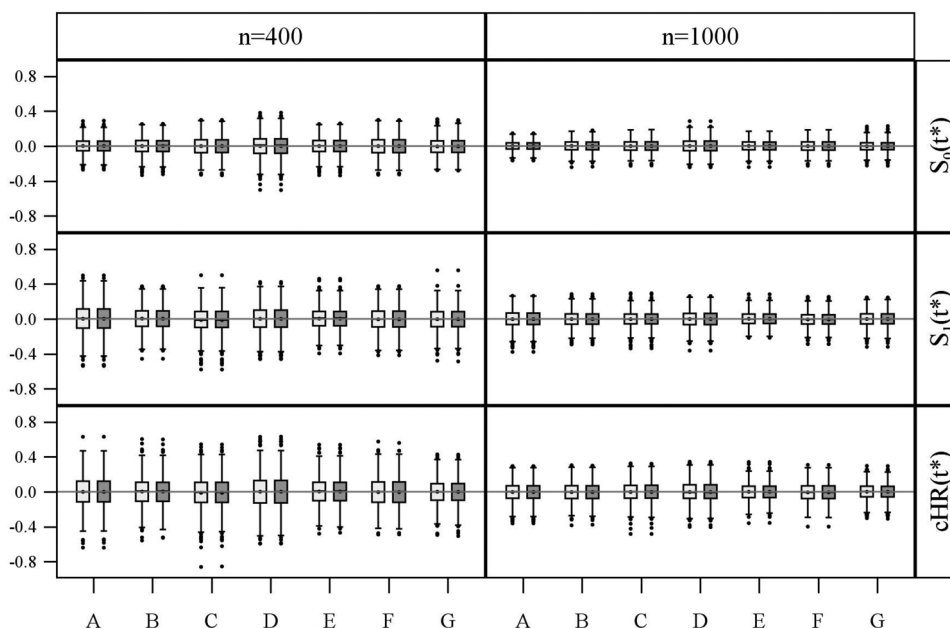


Figure 4. Bias distribution of log-minus-log survival probabilities and log-cumulative hazard ratios estimated with the GPV (light grey) and WPV (dark grey) approach in simulation study 2 (1000 repetitions) for sample sizes of 400 and 1000, for each of the seven scenarios A–G (see Figure S1 in Section C of Supplemental Materials).
GPV: generalised pseudo-value; WPV: weighted pseudo-value.

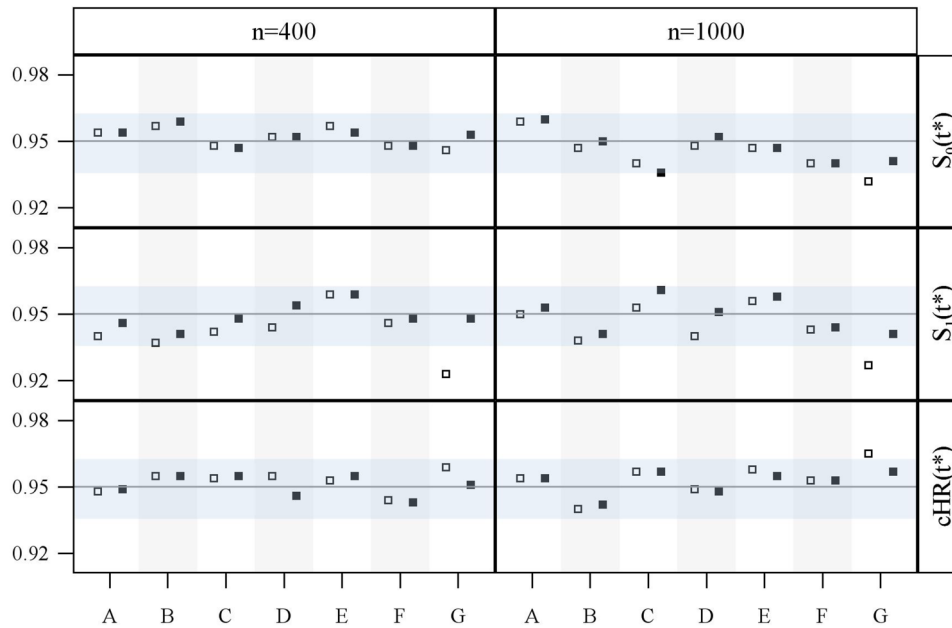


Figure 5. Observed coverage rates of 95% confidence intervals for survival probabilities and cumulative hazard ratios estimated with the GPV (square) and WPV (square-filled) approach for simulation study 2 (1000 repetitions) and sample sizes of 400 and 1000, for each of the seven scenarios A–G (see Figure S1 in Section C of Supplemental Materials). GPV: generalised pseudo-value; WPV: weighted pseudo-value.

GPV shows a too low coverage for $S_1(t^*)$ when donor identification takes considerably longer (scenario G); this is accompanied by a slightly conservative coverage for $cHR(t^*)$ when $n = 1000$.

Additional extensive simulation results can be found in Section D of Supplemental Materials, where the proportion of available donors and t^* have been varied. Again, the general unbiasedness of both approaches turns out (Supplemental Figures S2 to S7).

The CI coverage rates for $S_0(t^*)$ and $S_1(t^*)$ in Supplemental Figures S8 to S13 confirm the results of Figure 5; both approaches perform well in scenarios A–F with realistic waiting time distribution, irrespective of donor availability proportion and t^* -value. For scenario G, considerably low CI coverage rates for $S_1(t^*)$ emerge again for the GPV approach, whereas the WPV approach shows adequate CI coverage (Supplemental Figures S8 to S13).

Both approaches show adequate CI coverage rates for $cHR(t^*)$ in scenarios A–F. For the extreme scenario G, in particular, when $t^*=5$, too high coverage rates may occur in both approaches (Supplemental Figures S8 to S13).

In summary, the good performance in nearly all investigated settings shows that the WPV approach does not depend on the proportional hazards assumption and that it is able to properly address immortal time bias.

3.2 Real data example: SCT in paediatric leukaemia patients

Below, the WPV result is compared to the results of a Cox model and the GPV approach using real data from an international study in 542 children with newly diagnosed Philadelphia chromosome-positive (PH+) acute lymphoblastic leukaemia.^{11,21} The study compared SCT to conventional chemotherapy; the primary endpoint was disease-free survival (DFS). For 325 of 542 patients, a donor could be identified, so that they could switch from chemotherapy to SCT after a median waiting time of 5.1 months. The remaining 217 patients were treated with chemotherapy only. Here we assume that donor identification leads to SCT in a timely manner. The Cox model with a binary time-dependent covariate (indicating SCT after donor identification) gives a hazard ratio (HR) of 0.91 in favour of SCT (95% CI: 0.72–1.14; $p = 0.39$).

This result is questionable as it assumes a constant SCT effect on DFS. The problem is exemplified in Figure 6 where the instantaneous HR after SCT is flexibly plotted over time.^{22–24} As expected, SCT initially results in an increased hazard (compared to conventional chemotherapy) with a maximum at about 0.5 years and decreasing risk thereafter. At about 1 year instantaneous hazards in both groups are equal ($HR = 1$), nevertheless, cumulated risks until 1 year are still to the disadvantage of SCT. However, it can neither be seen from the (non-significant) constant nor from the (significant) time-flexible HRs (Figure 6) whether SCT has a positive benefit–risk balance after all. Hence, since long-term survival is the crucial and almost exclusive interest in paediatric oncology, here assessed at 5 years after patients have entered

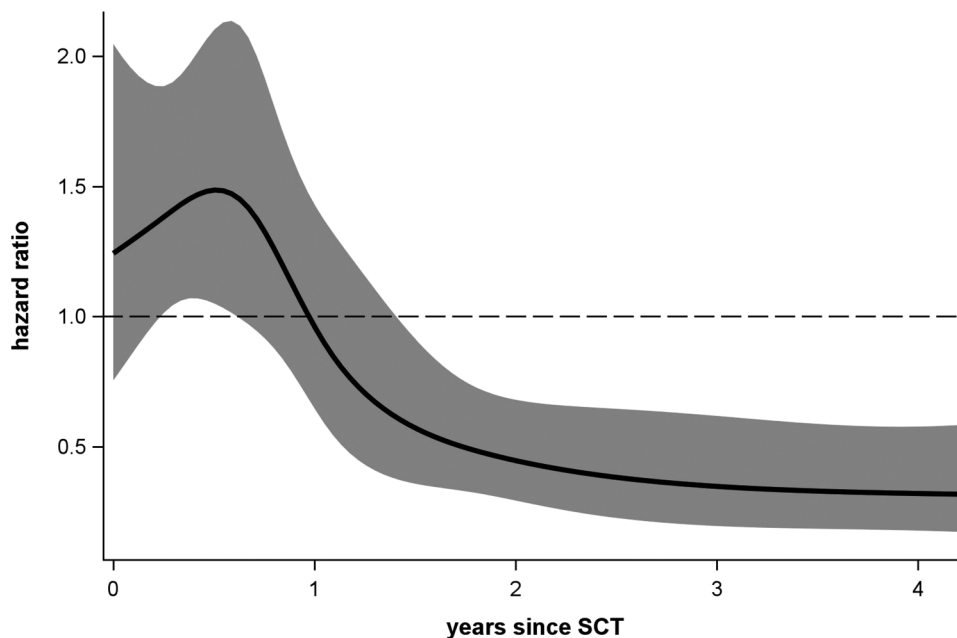


Figure 6. Instantaneous hazard ratio function and 95% pointwise confidence band for the PH+ acute lymphoblastic leukaemia dataset estimated by a restricted cubic spline function with 5 knots at 0.24, 0.68, 1.09, 1.86 and 4.2 years (corresponds to 5%, 27.5%, 50%, 72.5% and 95% of all outcome related events); a hazard ratio less than one favours stem cell transplantation (SCT).

the study, the Cox model with time-flexible hazard ratios will admittedly bring some additional insights, yet it cannot be considered the optimal statistical approach for treatment decisions.

The 5-year DFS estimates with and without donor are 41.7% (95% CI: 36.7%–46.5%) and 32.1% (25.4%–38.9%) for the GPV approach, and 41.7% (36.6%–46.7%) and 31.6% (24.6%–38.7%) for the WPV approach, respectively. The GPV approach yields a 5-year cHR of 0.77 favouring SCT (95% CI: 0.61–0.97; $p = 0.027$), the easier to calculate WPV approach gives a similar cHR of 0.76 (95% CI: 0.60–0.96; $p = 0.021$). Hence, it can be concluded that the long-term benefit of SCT outweighs its short-term risk (Figure 6).

For 325 patients, a donor is identified (m in Table 1). For 217 patients with no identified donor ($n_u + n_c$, see Table 1), it is known that $n_u = 45$ patients do not have a donor available. Now, with the WPV approach, it is straightforward to communicate that for the remaining $n_c = 172$ patients it is expected that about 38 of them belong to the donor available group, and about 134 do not have a suitable donor available. Overall, it is expected that 67% of all patients have a suitable donor registered in the databases.

4 Discussion

The choice between two treatment strategies with respect to long-term survival is not straightforward, when the feasibility of one of the strategies emerges over time, non-proportional hazards are present, and there is a risk for immortal time bias. A typical example where all these issues co-occur are childhood leukaemia studies comparing allogeneic SCT with conventional chemotherapy.

It is expected that SCT causes a short-term increase in the hazard for overall or disease-free survival and yields a long-term benefit thereafter^{3,25,26} (also see Figure 6). Due to patient-specific reasons, donor search may be ceased and an available donor cannot be identified anymore. Therefore, in order to allow a fair comparison between SCT and the competing treatment strategy, the latent binary covariate donor availability has to be considered in statistical analysis.

The WPV approach uses common pseudo-values and attaches appropriate weights to them to reconstruct the cohorts with and without donor available (Figure 1); that is, the weighted partitioning of observations with undefined group membership is the key feature of the method. It is important to note, that the weights depend on the elapsed time of donor search. The more time has been elapsed, the less likely is the identification of an available donor.

In contrast, the GPV approach generalises common pseudo-values in various sophisticated ways to account for immortal time bias and related issues, and weights them thereafter to reconstruct the distribution of the time to donor identification,

which is affected by donor search-ceasing events and censoring. Hence, the generalisation of the pseudo-values is the key feature of this method.

Both approaches are approximately unbiased and show reasonable CI coverage in scenarios with waiting time distributions that are typical for stem cell transplantation studies with non-related donors. However, confidence interval coverage, in particular of the GPV approach, may be negatively affected in situations where it takes much longer than usual to find suitable donors.

With regard to ease of software implementation and computational speed, WPV clearly outperforms GPV. The former approach requires the computation of common pseudo-values only, whereas the latter approach requires different rounds of computations with different starting times and patient risk sets. In the considered simulation scenarios of Section D of Supplemental Materials, the median WPV runtime was at least 40 times to more than 90 times faster than the corresponding median GPV runtime (see Figure S14 in Section E of Supplemental Materials). The WPV runtime advantage over GPV increases with the proportion of identified donors as more computationally time-consuming $\hat{U}_{i,1}(t^*|w_i)$ -values have to be calculated. The smaller runtime advantage in scenario G compared to scenarios A–F confirms that the proportion of identified rather than available donors is crucial (Supplemental Figure S14), as the considerably longer waiting times of scenario G lead to less identified donors and, consequently, less $\hat{U}_{i,1}(t^*|w_i)$ -calculations in the GPV approach.

When it comes to interpretation, both approaches provide the same primary message in terms of long-term survival estimation and group difference testing; however, both approaches may give different additional insights. Consider the real data example above (PH+ acute lymphoblastic leukaemia patients); here the WPV approach enables to directly conclude that an extremely fast donor search might have identified up to about 38 additional donors. Nonetheless, there are no donors available for around 33% of all patients; a number that can only be reduced by increasing the number of potential stem cell donors. In contrast, the GPV approach might be used to investigate other (artificial) distributions of the time to donor identification by adapting the inverse probability weights. By doing so, effects of accelerated or deferred donor identification on patients' survival could be directly studied.

Finally, consider the completely different setting of a diagnostic study with survival time outcome. To enable the computation of common diagnostic measures (like sensitivity, specificity) at a pre-specified time point, Antolini and Valsecchi²⁷ suggested to impute the disease status (case or control) of censored subjects according to the probability that the subject is a case or a control. The weights of this “direct estimation by imputation”—approach share the same underlying idea as the WPV-weights.

Conclusion

When donor availability status is unknown due to donor search-ceasing events or censoring, then both, the WPV and the GPV approach, are able to successfully deal with the time-dependent nature of donor identification and partially unobserved donor availability, thereby mimicking the genetic randomization approach for SCT trials in paediatric oncology.³ From a practical point of view, the WPV approach is computationally faster and easier to implement than the GPV approach. Previous SCT studies that have been analysed within the Cox model framework may be reanalysed with the new approach to reassess treatment effects and resulting treatment recommendations.

Acknowledgements

The authors thank the Ponte di Legno group for giving access to the data on Philadelphia chromosome-positive acute lymphoblastic leukaemia. They want to thank Christina Peters (St. Anna Children's Hospital, Vienna, Austria) and Susanne Karlsruher (Children's Cancer Research Institute, Vienna, Austria) for giving them an understanding of the modalities of unrelated donor search.

Availability of data

The authors are not allowed to make the PH+ acute lymphoblastic leukaemia dataset available.


Declaration of conflicting interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCID iDs

Martina Mittlböck  <https://orcid.org/0000-0003-1097-4047>

Ulrike Pötschger  <https://orcid.org/0000-0001-7024-7082>

Harald Heinzl  <https://orcid.org/0000-0003-4711-2379>

Supplemental material

Supplemental material for this article is available online.

References

1. Kalbfleisch JD and Prentice RL. *The statistical analysis of failure time data*. 2nd ed. New York: Wiley, 2002.
2. Balduzzi A, Valsecchi MG, Uderzo C, et al. Chemotherapy versus allogeneic transplantation for very-high-risk childhood acute lymphoblastic leukaemia in first complete remission: comparison by genetic randomisation in an international prospective study. *Lancet* 2005; **366**: 635–642.
3. Wheatley K and Gray R. Commentary: Mendelian randomization – an update on its use to evaluate allogeneic stem cell transplantation in leukaemia. *Int J Epidemiol* 2004; **33**: 15–17.
4. Gray R and Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. *Bone Marrow Transplant* 1991; **7**: 9–12.
5. Klein JP, Logan B, Harhoff M, et al. Analyzing survival curves at a fixed point in time. *Stat Med* 2007; **26**: 4505–4519.
6. Logan BR, Klein JP and Zhang MJ. Comparing treatments in the presence of crossing survival curves: an application to bone marrow transplantation. *Biometrics* 2008; **64**: 733–740.
7. Anderson JR, Cain KC and Gelber RD. Analysis of survival by tumor response. *J Clin Oncol* 1983; **1**: 710–719.
8. Simon R and Makuch RW. A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. *Stat Med* 1984; **3**: 35–44.
9. van Houwelingen HC and Putter H. *Dynamic prediction in clinical survival analysis*. Boca Raton: CRC Press, 2012.
10. Beyersmann J, Allignol A and Schumacher M. *Competing risks and multistate models with R*. New York: Springer, 2012.
11. Pötschger U, Heinzl H, Valsecchi MG, et al. Assessing the effect of a partly unobserved, exogenous, binary time-dependent covariate on survival probabilities using generalised pseudo-values. *BMC Med Res Methodol* 2018; **18**: 14.
12. Andersen PK and Pohar Perme M. Pseudo-observations in survival analysis. *Stat Methods Med Res* 2010; **19**: 71–99.
13. Graw F, Gerds TA and Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal* 2009; **15**: 241–255.
14. Andersen PK, Klein JP and Rosthøj S. Generalized linear models for correlated pseudo-observations with applications to multi-state models. *Biometrika* 2003; **90**: 15–27.
15. Overgaard M, Pamer ET and Pedersen J. Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations. *Ann Stat* 2017; **45**: 1988–2015.
16. Jacobsen M and Martinussen T. A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations. *Scand J Stat* 2016; **43**: 845–862.
17. SAS software. © 2016 by SAS Institute Inc., Cary, NC, USA, Version 9.4
18. Mittlböck M. WPV - a SAS macro for weighted pseudo-values. <https://github.com/MartinaMittlboeck/WPV>. 2021.
19. Bender R, Augustin T and Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med* 2005; **24**: 1713–1723.
20. Sposto R. Cure model analysis in cancer: an application to data from the children's cancer group. *Stat Med* 2002; **21**: 293–312.
21. Arico M, Schrappe M, Hunger SP, et al. Clinical outcome of children with newly diagnosed Philadelphia chromosome-positive acute lymphoblastic leukemia treated between 1995 and 2005. *J Clin Oncol* 2010; **28**: 4755–4761.
22. Harrell FE Jr. *Regression modeling strategies*. 2nd ed. Cham: Springer, 2015.
23. Heinzl H and Kaider A. Gaining more flexibility in Cox proportional hazards regression models with cubic spline functions. *Comput Methods Programs Biomed* 1997; **54**: 201–208.
24. Heinzl H, Kaider A and Zlabinger G. Assessing interactions of binary time-dependent covariates with time in Cox proportional hazards regression models using cubic spline functions. *Stat Med* 1996; **15**: 2589–2601.
25. Klein JP. Statistical challenges in comparing chemotherapy and bone marrow transplantation as a treatment for leukemia. In: Jewell NP (ed) *Lifetime data: models in reliability and survival analysis*. Dordrecht: Kluwer Academic Publishers, 1996, pp. 175–185.
26. Marubini E and Valsecchi MG. *Analysing survival data from clinical trials and observational studies*. Chichester: Wiley, 1995.
27. Antolini L and Valsecchi MG. Performance of binary markers for censored failure time outcome: nonparametric approach based on proportions. *Stat Med* 2012; **31**: 1113–1128.