

ipcwswitch: An R package for inverse probability of censoring weighting with an application to switches in clinical trials

Nathalie Graffeo, Aurélien Latouche, Christophe Le Tourneau, Sylvie Chevret

▶ **To cite this version:**

Nathalie Graffeo, Aurélien Latouche, Christophe Le Tourneau, Sylvie Chevret. ipcwswitch: An R package for inverse probability of censoring weighting with an application to switches in clinical trials. Computers in Biology and Medicine, 2019, 111, pp.103339. 10.1016/j.compbiomed.2019.103339 . hal-02349298

HAL Id: hal-02349298
https://cnam.hal.science/hal-02349298

Submitted on 20 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0010482519302082
Manuscript_e3585eadf7af9bb931f9b6c112b3f8b0

`ipcwswitch`: An `R` Package for Inverse Probability of Censoring Weighting with an Application to Switches in Clinical Trials

Nathalie Grafféo[a,b,c,*], Aurélien Latouche[d,e], Christophe Le Tourneau[e,f,g], Sylvie Chevret[a,b,h]

[a] *INSERMU1153, Statistic and Epidemiologic Research Center Sorbonne Paris Cité (CRESS), ECSTRA Team, Saint-Louis Hospital, 75010 Paris, France*
[b] *Paris Diderot University, Paris 7, SPC, Paris, France*
[c] *Aix Marseille Univ, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Marseille, France*
[d] *EA 4629, Conservatoire national des arts et métiers (Cnam), Paris, France*
[e] *INSERM U900, Institut Curie, Saint Cloud, France*
[f] *Department of Drug Development and Innovation, Institut Curie, Paris & Saint-Cloud, France*
[g] *Versailles-Saint-Quentin-en-Yvelines University, Montigny-le-Bretonneux, France*
[h] *SBIM, Saint-Louis Hospital, APHP, Paris, France*

[*] Corresponding author. SBIM, Hôpital Saint Louis, 1 Avenue Claude Vellefaux, 75010 Paris, France

Email addresses: `n_graffeo@yahoo.fr` (Nathalie Grafféo), `aurelien.latouche@curie.fr` (Aurélien Latouche), `christophe.letourneau@curie.fr` (Christophe Le Tourneau), `sylvie.chevret@paris7.jussieu.fr` (Sylvie Chevret)

Preprint submitted to Computers in Biology and Medicine *May 30, 2019*

© 2019 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

Abstract:

In randomized clinical trials (RCT), the analysis is based on the intent-to-treat principle to avoid any selection bias in the constitution of groups. However, estimates of overall survival can be biased when significant crossover occurs because the separation of randomized groups is lost. To handle these switches, the inverse probability of censoring weighting (IPCW) method has been proposed; however, it is still poorly used in RCT, notably because of its complex implementation. In particular, for time-to-event outcomes, it can be difficult to format data, especially when time-dependent covariates have to be managed, with different measurement times between patients. This paper aims to present the `R` package `ipcwswitch` with some guidance for the analysis of the treatment effect on survival in a hypothetical setting where all patients would have continued to take the randomization treatment. After a brief recall of the key principles of the IPCW method, each step of the implementation is described using a toy example. The guidelines are illustrated in a case study that aimed at evaluating the benefit of therapy based on tumour molecular profiling for advanced cancers, SHIVA01.

Keywords:

inverse probability weighting, `R`, randomized clinical trial, survival analysis, treatment switch

1. Introduction

Estimating the beneficial effect of a treatment is the primary goal of randomized clinical trials (RCT), based on intent-to-treat (ITT) analysis. In case of two parallel arms, the patients are assigned to the experimental or control arm according to randomization. However, treatment decisions may change over time, notably to better fit the evolution of patient status. This is almost systematically the case in oncology trials, where patients are allowed switching to the other arm in case of disease progression, defining a second line treatment. Although such crossovers do not affect progression free survival (PFS) endpoints, estimates of overall survival (OS) using ITT analysis can be biased when significant crossover occurs because the separation of randomized groups is lost [1, 2].

To handle these switches, different statistical methods have been investigated. The easiest approaches, consisting in excluding switchers or censoring at switch (in a "per-protocol analysis"), or including treatment as a time-varying covariate to assess the effect of treatment actually received by the patients, appeared subject to selection bias [3], and should not be considered [4]. Two-stage methods, using accelerated failure time models, have been developed to provide randomization-based efficacy estimators [5, 6, 3]. In these models, the patient's observed event time is related to his counterfactual event time, that which would have been observed for the patient if he had not received any treatment [7]. However, it involves fitting parametric models to the data. To estimate the causal treatment effect when selective switches or dropouts may induce dependent censoring, other solutions involving counterfactual outcomes such as the inverse probability of censoring

weighting (IPCW) have been proposed [8]. Its use in clinical trials has been recommended, under the condition that appropriate data and sufficient sample size are available [9, 10, 11, 12]. It represents an adjustment method that is likely to produce lower bias than the ITT analysis, provided the switching proportion is not extreme [4].

However, the use of the IPCW method along with marginal structural models (MSM) to deal with treatment changes in the RCT setting remains uncommon, notably when dealing with a time-to-event outcome. This could be explained at least partially because of the difficulty in its implementation given it is not straightforward to apply these methods directly. Moreover, the lack of available software for a long time may have also limited their use and routine implementation to handle time-varying weights [9]. Recently, several authors gave some practical guidelines to facilitate the use of the IPCW method. For example, Dodd et al. [12] provided a `Stata` code to implement the weights and to fit the model. Willems et al. [11] proposed a "user friendly" implementation of IPCW in `R` [13]. Notably, using the `survSplit` function from the `survival` package [14] on a toy example, the authors provided a nice four steps algorithm, which can be easily reproduced. However, with regards to time-dependent covariates, the authors referred to the `unfold` function from the `RcmdrPlugin.survival` package [15]. Using such a function could be not straightforward, especially when dealing with different observation times of measurement for the patients. Indeed, data processing for time-dependent covariates using the `survSplit` function could be tricky since it requires the additional building of a function that generates the time-dependent covariates from the output of `survSplit` (see, e.g., page

19 of Thomas and Reyes [16]). Then, up to our knowledge, there is no available straightforward `R` function dealing with time-dependent covariates when times of measurements are not the same.

In this article, we propose the `R` package `ipcwswitch` to estimate the treatment effect on OS in the RCT setting, taking into account the actual treatment switches over time. The package is available from the Comprehensive `R` Archive Network at `https://CRAN.R-project.org/package=ipcwswitch`. We present a step-by-step tutorial on how the IPCW modeling can be used in this framework, i.e., when informative non-compliance to the treatment allocated by randomization may occur. Note that our aim is not to assess the performances of the IPCW method. For this, we refer the reader to the works of Latimer et al. (see e.g.,Latimer et al. [4]). By contrast, our aim is to provide tools for statisticians to facilitate the use of this method in the specific setting of RCT with a right-censored end point and time-dependent covariates with different times of measurement across patients.

This paper is structured as follows. In Section 2 we give the key principles to IPCW. We describe a step-by-step overview of the IPCW implementation in Section 3. We illustrate the use of the package `ipcwswitch` in a case study, in Section 4. Some concluding remarks are given in Section 5.

2. Key principles of IPCW

2.1. Context

Let consider a randomized clinical trial where n patients are randomly allocated to one of two parallel arms, with a primary endpoint being right censored such as overall survival. Let consider that patients are allowed to switch

in case of a disease progression. The first step of the IPCW method consists in censoring patients who interrupt their initially assigned treatment. Note that we will refer to this censoring as the "treatment censoring", whereas the "censoring" term will only be used with reference to the usual right censoring, that is, when the patient drops out of the study before death. However, the treatment censoring mechanism may depend on the baseline and time-dependent factors that have lead to this change of treatment. Since these factors are likely to have some prognostic value, they are associated both with the endpoint and the treatment censoring mechanism. The IPCW method introduces weights to correct for this dependent censoring.

2.2. Weights derivation

The second step of the IPCW method is thus to use inverse probability of censoring weights. At a given time t, the weights are defined as the inverse of the probability of having remained on the randomized treatment, that is, of being uncensored by treatment, until time t given still on randomized treatment before time t and given the observed values of the measured baseline and time-dependent confounders at t. Here, the confounders refer to factors that are a common cause of the treatment exposure and the treatment censoring.

More precisely, let T_i^C be the time to treatment censoring for patient i. The patient's weights up to the ordered treatment censoring times t_k, $(1 \leq k \leq j)$, are defined as:

$$w_i(t_k) = \prod_{l=1}^{k} \frac{1}{P\left(T_i^C > t_l \;\mid\; T_i^C > t_{l-1}, \bar{\mathbf{X}}_{t_l} = \bar{x}_{t_l,i}, \mathbf{V} = v_i\right)}, \qquad (1)$$

where $t_0 = 0$, $\bar{x}_{t_l,i}$ denotes the observed history of the time-varying confounders up to t_l, and v_i the observed time-fixed covariates. Unstable weights can occur when the denominator is small, that is when some strata defined by the conditioned part of the above mentioned probabilities are small. Then, as suggested by Robins [17], one can use stabilized weights to obtain narrower distributions. Specifically, the weight denominators do not change whereas the numerators are obtained in the same way as denominators but adjusting only for baseline covariates:

$$sw_i(t_k) = \prod_{l=1}^{k} \frac{P\left(T_i^C > t_l \;\mid\; T_i^C > t_{l-1}, \mathbf{V} = v_i\right)}{P\left(T_i^C > t_l \;\mid\; T_i^C > t_{l-1}, \bar{\mathbf{X}}_{t_l} = \bar{x}_{t_l,i}, \mathbf{V} = v_i\right)}. \qquad (2)$$

Because underlying mechanisms of treatment censoring can differ between the two arms, the weights have to be computed in each arm, separately.

Let $\lambda_C(t|\mathbf{Z_i}(t)) = \lambda_{0C}(t)exp(\boldsymbol{\beta}_C{}^t\mathbf{Z}_i(t))$ denote the hazard of treatment discontinuation or switch from a Cox proportional hazards model conditional on covariates $\mathbf{Z}$. In each arm, the probabilities of treatment censoring can be estimated at time t using the Kaplan-Meier product limit estimator [18] as follows:

$$\prod_{\{j:t_j<t,\delta_{C,j}=1\}} \left[1 - \hat{\Lambda}_{0C}(t_j)exp(\hat{\boldsymbol{\beta}}_C{}^t\mathbf{Z_i}(t_j))\right],$$

where, for patient j, t_j is the minimum time between the death time and the treatment censoring time, and $\delta_{C,j}$ is the treatment censoring indicator. Individual weights are then easily derived from equation (1) or (2).

2.3. Cox marginal structural model

The IPCW method requires the use of a marginal structural model (MSM) to describe the relationship between the treatment arm and the primary end-

point, that is overall survival. A Cox MSM is a Cox model that estimates marginal effects that would have been observed in the absence of switch or discontinuation of treatment. More specifically, assuming that all confounders have been observed, applying these weights to the Cox partial likelihood estimators creates a pseudo–population that would have been studied if the patients had complied with their assigned treatment arm [19, 8, 18, 9]. Thus, as given by Robins [8], the partial likelihood score becomes:

$$U(\beta) = \sum_{\{j:\delta_j=1\}} \hat{w}_j(t_j) \times \left[A_j - \frac{\sum_{i=1}^{n} Y_i(t_j)\hat{w}_i(t_j)A_i exp(\hat{\beta}A_i)}{\sum_{i=1}^{n} Y_i(t_j)\hat{w}_i(t_j)exp(\hat{\beta}A_i)} \right],$$

where δ_j and Y_j are the death and the at-risk indicator of patient j, respectively, and A_j is the binary covariate indicating his(her) treatment arm (where 1 denotes the experimental treatment and 0 otherwise).

Under correct specification of the model, solving the IPCW score equation provides a consistent and asymptotically normal estimator of the parameter β [8]. Moreover, using stabilized weights also provides consistent estimates of β [17]. In other words, the "hypothetical" causal effect of the experimental treatment on the overall survival is obtained using these IPCW weights in a Cox marginal structural model. Specifically, the estimand represents the effect in a hypothetical setting where all patients would have continued to take the randomization treatment.

3. Overview of the IPCW implementation

We present here some guidance to format and analyse some RCT data when assessing the benefit of any experimental treatment on overall survival using the weights from the IPCW method in a Cox MSM as described above.

Let consider a very simple dataset mimicking RCT data of 3 patients with a single time-fixed covariate and a single time-dependent covariate that can be measured up to 3 times:

```
R> toydata
id     randt     lastdt status age ps1 ps2 ps3         dt2         dt3 arm    swtrtdt
1 2018-01-12 2018-03-02      1  20   0   0   0  2018-02-02  2018-03-01   A 2018-03-01
2 2017-11-04 2017-12-15      1  50   1  NA   2        <NA>  2017-12-12   B       <NA>
3 2017-05-20 2018-01-04      0  40   0   0   1  2017-08-02  2018-01-02   A       <NA>
```

This Toy Example Data is given in wide format. It contains the following variables:

- `id` is the patient's identifier,
- `randt` is the date of the randomization visit,
- `lastdt` is the date of latest news,
- `status` is equal to 1 if the patient dies on `lastdt` (and 0 otherwise),
- `age` is the patient's age,
- `ps1`, `ps2`, `ps3` are the values of a covariate measured at dates `randt`, `dt2`, `dt3`, respectively
- `arm` is the patient's randomized arm,
- `swtrtdt` is the date when the patient initiates the other arm treatment (NA if does not happen).

Step 0: Handling time-dependent covariates

As explained in Section 2, two time-dependent Cox models, one for treatment censoring and the Cox MSM, are specified. As a result, values of the covariates are to be computed for all at-risk patients at each failure time (due to the treatment censoring or to the death) in the dataset. We used a counting process method with multiple records for each individual, where each record corresponds to a time interval during which all covariates remain constant. For further analyses, the required variables are the patient's `id`, the times of death `lastdt` and the treatment censoring times that correspond to the times of treatment switch `swtrtdt`, the status indicator `status`, the fixed covariates `age`, the time-dependent covariate `ps1, ps2, ps3` along with their times of change `randt, dt2, dt3`, and the time of treatment interruption `swtrtdt`.

Step 1: Data formatting – wide vs. long format

Before computing weights using a time-dependent Cox model with time-dependent confounders and time-dependent treatment, the estimation of the outcome models requires some tricky data formatting. The wide format refers to a dataset with one row per subject, while the long format refers to a dataset with possibly several rows per subject. When data are provided in the wide format, some data processing are needed to build separated rows according to the observed changes in the time-varying confounders, the observed changes in the administered treatment (i.e., treatment censoring time), and the end of the follow-up (i.e., death or lost of follow-up). Functions are available to convert the wide format to the long format, such as the `survSplit` function from the `survival` package [14]. Nevertheless, to our knowledge, there is no

available straightforward R function dealing with time-dependent covariates when patients do not share the same dates of measurements. Thus, we developed the specific R package ipcwswitch.

First, the timesToKeep function is run on the toydata as follows:

```
R> kept.t  <- timesTokeep(toydata, id = "id",
+    tstart = "randt", tstop = "lastdt",
+    mes.cov = list(c("ps1", "ps2", "ps3")),
+    time.cov = list(c("randt", "dt2", "dt3")))
```

For example, for patient id=3, we obtain:

```
R> kept.t[[1]][[3]]
[1] "2017-05-20" "2018-01-02" "2018-01-04"
```

Then, the long format is obtained by the wideToLongTDC function:

```
R> toy.long <- wideToLongTDC(toydata, id = "id",
+    tstart = "randt", tstop = "lastdt",
+    event = "status",
+    bas.cov = c("age", "arm", "swtrtdt"),
+    mes.cov = list(TDconf=c("ps1", "ps2", "ps3")),
+    time.cov = list(c("randt", "dt2", "dt3")),
+    times = kept.t[[1]])
```

We obtain the following dataframe, which represents the Toy Example Data in long format. The variable TDconf aggregates the repeated values ps1, ps2, and ps3 in a single variable:

```
  id     tstart      tstop event age arm    swtrtdt TDconf
1  1 2018-01-12 2018-03-02     1  20   A 2018-03-01      0
2  2 2017-11-04 2017-12-12     0  50   B       <NA>      1
```

```
3  2 2017-12-12 2017-12-15     1  50   B       <NA>      2
4  3 2017-05-20 2018-01-02     0  40   A       <NA>      0
5  3 2018-01-02 2018-01-04     0  40   A       <NA>      1
```

Step 2: Data formatting to compute the IPCW weights

Then, another data formatting is required to allow the computation of the inverse probability of censoring weights. We have to define the exposure times, that is when each patient has been treated according to the randomization or not.

First, among the only patients who have been treated at randomization, those who switched are to be censored at the time of switch. Note that from this point, we work with duration times that are obtained subtracting the starting date of the follow-up from the date we are interested in.

This is illustrated in the toy example, where patients are censored when initiating the other arm treatment, that is, at time `swtrtdt`; a treatment censoring indicator `cens` is thus added to the previous dataset to indicate such a switch. We thus developed the `cens.ipw` function that should be used as follows:

```
R> toy.long2 <- cens.ipw(toy.long, id = "id", tstart = "tstart", tstop = "tstop",
+    event = "event", arm = "arm",
+    realtrt = FALSE, censTime ="swtrtdt")
```

resulting in the following formatted data:

```
  id tstart tstop event age arm swtrtdt TDconf cens
1  1      0    48     0  20   A      48      0    1
2  2      0    38     0  50   B      NA      1    0
3  2     38    41     1  50   B      NA      2    0
```

```
4  3      0   227     0  40   A      NA      0    0
5  3    227   229     0  40   A      NA      1    0
```

Then, only times related to either the treatment switches or the occurrences of death were retained in the dataset. Nevertheless, weights need to be computed at all the observed event or treatment censoring times in the sample [20]. Thus, each patient's follow-up must be split according to all these dates.

Considering the toy example, all the treatment censoring and death times are collected. Because the weights have to be computed in each arm, the treatment censoring times are gathered in each arm separately, whereas the death times are gathered in both arms, as follows:

```
R> rep.times1 <- unique(c(toy.long2$tstop[toy.long2$cens==1
+                            & toy.long2$arm == "A"],
+                        toy.long2$tstop[toy.long2$event==1]))
R> rep.times2 <- unique(c(toy.long2$tstop[toy.long2$cens==1
+                            & toy.long2$arm == "B"],
+                        toy.long2$tstop[toy.long2$event==1]))
```

We finally obtained the terminal dataframe using the `replicRows` function:

```
R> toy.rep <- replicRows(toy.long2, tstart = "tstart", tstop = "tstop",
+    event = "event", cens = "cens",
+    times1 = rep.times1, times2 = rep.times2, arm = "arm")
  id age arm swtrtdt TDconf cens tstart tstop event
1  1  20   A      48      0    0      0    41     0
2  1  20   A      48      0    1     41    48     0
3  2  50   B      NA      1    0      0    38     0
4  2  50   B      NA      2    0     38    41     1
5  3  40   A      NA      0    0      0    41     0
6  3  40   A      NA      0    0     41    48     0
```

```
7  3  40   A      NA      0    0     48   227     0
8  3  40   A      NA      1    0    227   229     0
```

Step 3: Computation of stabilized weights

Once data have been adequately formatted, weights are to be estimated in each arm as explained in Section 2. To this end, the `survfit` function from the `survival` package can be fitted to the result from the `coxph` function for treatment censoring adjusted on baseline and time-dependent confounders. This is the standard approach when events and censoring occur simultaneously given we considered that the event comes first. More precisely, in case of ties occurring at a given time t, the treatment censoring at t does not contribute to the likelihood at the current event time but contributes to the likelihood at the following event time. Last, because stabilized weights produce smaller confidence intervals and increase the statistical efficiency [20], we choose to present the single output of stabilized weights. From our toy example, the dataframe with stabilized truncated weights is obtained by:

```
R > toy.rep <- ipcw(data = toy.rep, id = "id", tstart = randt, tstop = lastdt,
+   cens = cens, arm = "arm", bas.cov = c("age"),
+   conf = c("TDconf"), trunc = 0.01, type='kaplan-meier')
```

Additionally, truncating the weights can help when dealing with extreme weights. For example, one can try different levels of truncation, such as truncating the left tail at the 1st percentile and the right tail at the 99th percentile. Hence, plotting those truncated stabilized weights can be useful to determine the cut-off that provides both a mean close to 1 and a small range. To this end, it is possible to use the `ipwplot` function from the `ipw` package [20].

The plots of un-truncated and truncated weights are obtained by:

```
# un-truncated stabilized weights
ipwplot(weights = toy.rep$weights,
    timevar = toy.rep$randt,
    binwidth = 50,
    xlab = "Time since enrolment (days)",
    ylab = "Logarithm of the stabilized\n un-truncated weights"))
# truncated stabilized weights
ipwplot(weights = toy.rep$weights.trunc,
    timevar = toy.rep$randt,
    binwidth = 50,
    xlab = "Time since enrolment (days)",
    ylab = "Logarithm of the truncated stabilized weights\n (trunc = 0.01)"))
```

Please, note that the arguments of the `ipcw` function are explained in Appendix, Section Appendix A, whereas those of the `ipwplot` function are given in the accompanying paper of the `ipw` package [20]. Since the `ipcw` function allows the calculation of un-truncated weights, the user has the possibility to compare the weights before and after truncation and to better fit the model by including different potential confounding factors. Also, giving less importance to some patients by truncating the weights can lead to biased estimates. Thus, a choice has to be made between the size of the bias and the strength of the instability of the weights. In addition, let us mention that the `ipwtm` function in the `ipw` package [20] can be used to assess the IPCW from adequately formatted data. However, there are two conceptual differences between the `ipw::ipwtm` and the `ipcwswitch::ipcw` functions. First, in the `ipw` package, the authors chose not to follow the above-mentioned "standard approach", that is they consider that the event does not come first in case

of ties regarding event and treatment-censoring. Secondly, their procedure is based on the assessment of the survival probabilities from the estimated baseline hazard, whereas we directly estimate these probabilities using the Kaplan-Meier estimator. Specifically, using the `ipcwswitch::ipcw` with the option `type='efron'` allows the user to find the results obtained using the `ipw::ipwtm` function, when considering the corresponding intervals.

Step 4: Marginal estimates of the causal treatment effect on the pseudo-population

The weights being estimated, they should be incorporated in a marginal Cox model of the hazard of death. They are obtained running the `coxph` function of the `survival` package, specifying `weights` in the arguments of the function.

In our example, the Cox MSM using the truncated stabilized weights is fitted as follows:

```
fit.stab.w <- coxph(Surv(randt, lastdt, event) ~ age + cluster(id),
    data = toy.rep, weights = toy.rep$weights.trunc)
```

Note that, in the `coxph` function, one has to add the option `cluster` to use the robust variance in order to handle data correlation.

4. Case study

4.1. Design of the trial

The motivating example was the first randomized clinical trial that aimed at comparing molecularly targeted therapy based on tumour molecular profiling (MTA) versus conventional therapy (CT) for advanced cancer (SHIVA01) [21].

A switch to the other arm was scheduled to be proposed at disease progression for patients in both treatment groups. The primary endpoint of the study was progression-free survival, defined as the time from randomization to death from any cause or progression. Overall, 100 patients were randomly allocated to the MTA arm and 97 patients to the CT arm. The primary analysis of the trial was based on the intent-to-treat population [21], ignoring the time-dependent treatment decisions that may occur after disease progression.

At the time of terminal analysis, there were 167 disease progressions (83 in the MTA arm and 84 in the CT one) and 27 deaths before progression (17 and 10, respectively). Among the 167 disease progressions, 95 patients switched to the other arm. More precisely, in the MTA arm, 25 patients switched to the CT arm (14 of whom died, 56%), while 70 patients in the CT switched to the MTA arm (with 42 subsequent deaths, 60%) (Figure 1). Since treatment switches were motivated by disease progression, we used overall survival as the main endpoint in this case study. The proposed methodology will account for dependent censoring, providing a causal estimate of the treatment effect that would have been obtained if no switches had been observed.

4.2. Illustration of the IPCW implementation on the case study

We now illustrate the IPCW implementation on the case study described above, following the algorithm introduced in Section 3. Note that we used anonymized data for reproducibility purpose, since both code and data are available in the `ipcwswitch` package.

For each observation, the follow-up started on the date of the randomization visit, that is before patients were administered the study treatment. Applying the IPCW method, patients who never initiated their assigned

treatment had to be excluded. Therefore, while there were 197 patients in the ITT analysis, only 193 patients were analysed in the IPCW method described below.

Baseline time-fixed covariates were considered, namely age at randomization, gender, number of previous lines of treatment, the dichotomized Royal Marsden Hospital score (0 or 1 *vs.* 2 or 3) and the altered molecular pathway (distinguishing 3 pathways, namely hormone receptors pathway, PI3K/ AKT/mTOR pathway, and RAF/MEK pathway). Time-varying confounders, measured at each visit scheduled every two months up to study completion, were the Eastern Cooperative Oncology Group (ECOG) performance status, the presence of concomitant treatments and the need of platelet transfusions. All confounders were selected with the physicians. To deal with missing values, the last observation carried forward method was used to impute values after available measurements.

First, a long format dataset was obtained. As reported in Figure 1, 68 patients from the CT arm and 25 patients from the MTA arm, who initiated the other-arm treatment, defined treatment censoring observations. Then, the weights were computed, taking into account the potential confounders. We chose the stabilized weights with a cut-off truncation at 0.05. Indeed, there were those combining a log(mean) close to 0 and a small range (see Figure 2).

Finally, a Cox marginal structural model was fitted on the entire follow-up. It provided an estimated causal hazard ratio of 1.30 (95%CI = $[0.81, 2.08]$). It could be contrasted to those estimates obtained in the ITT population (1.19, 95%CI = $[0.84, 1.68]$), or a Cox model adjusted on both baseline

and time-dependent covariates (1.035, 95%CI = $[0.74, 1.44]$. Anyway, results were not statistically significant at the 5% level. Note that, using the `ipw::ipwtm` function, we obtained an estimated causal hazard ratio of 1.30 (95%CI = $[0.81, 2.07]$). The difference is rather low, although there were 29 event times that were also treatment censoring times in the studied data.

4.3. Discussion of the case study

This work was motivated and illustrated on data from the SHIVA01 trial. These anonymized data were used as illustrative of our package and not to report new findings of the SHIVA01 trial. Nevertheless, one should not avoid the checking of the underlying assumptions of the marginal structural model, that is exchangeability, consistency, positivity and correctness of the weight-generating and survival models [17]. The low and "moderate", as termed by Latimer, switching proportions observed in each arm (25% and 70%, respectively) legitimized the use of the IPCW method [4]. Also, to avoid instability in estimates, we used truncated weights, which is a common practice when data from some levels of the confounders are rare, leading to extreme IPCW weights. The most demanding assumption is the absence of unmeasured confounders to achieve unbiased estimates of treatment effect. Since this exchangeability assumption cannot be tested, it is necessary to collect all the baseline covariates and the time-dependent confounders impacting both the probability of being "treatment-censored" and the survival outcome, to insure most of the confounding information could be taken into account.

Some authors suggested to fit a time-dependent Cox proportional hazards model for treatment censoring to check the presence of dependent censoring, that is the dependence between treatment censoring times and the covariates

suspected of being confounders [18, 11]. Among the three chosen potential confounders, only the need for platelet transfusions was statistically significant (results not shown). Given the fact that using only this covariate as a time-dependent confounder did not markedly change the estimated causal hazard ratio, we kept as time-dependent confounders those three indicated by the physicians. Also, in the SHIVA01 trial, toxic effects were reported according to their association with the received treatment. However, regarding only the toxic effects related to the assigned treatment, there were not enough patients in the MTA arm to be included in the analysis. Therefore, the results should be interpreted with caution. Conversely, one can deal with a high number of potential confounding factors. Some methods to dimension reduction could be considered [18, 22, 12]. We did not use such a method in our application because the set of potential confounders that were identified by clinicians was small.

Recently, some authors showed that using the IPCW method on a post-hoc analysis of the phase 3 PREVAIL study resulted in a greater effect of the experimental treatment on overall survival than reported in the ITT analysis [23]. However, causal methods, such as IPCW, remain slightly used in clinical trials setting [24]. This could be explained in several ways. First it requires many data transformations and algorithms, that may have limited its easy use; this justified our own work. Moreover, as already mentioned, the IPCW is rather sensitive to small sample sizes and extreme switching proportions [5]. It further needs not all patients from the control arm to have switched unless the method is prone to important bias [9]. Other causal inference methods could be preferred [6]. For instance, we chose the IPCW method

rather than a Rank Preserving Structural Failure Time Model because the assumption of "common treatment effect", which is crucial in this approach, is questionable. Indeed, in the SHIVA01 protocol, patients could switch only if they had had a disease progression. Thus, the treatment effect could not be the same before or after the switch, preventing those patients from the "common treatment effect". As mentioned above, the use of the IPCW method requires unverifiable assumptions, which represent a barrier to its use. The expertise of clinicians is then essential to collect the confounding factors of switching in the most exhaustive way. At last, not at least, although the IPCW method provides a formal way to handle treatment switches, it cannot replace the intent-to-treat analysis that defines the primary analysis of RCTs.

5. Conclusion

This paper aimed to present a R package and some practical guidance in the estimation of treatment effect on survival in the RCT setting to take into account the treatment switches, commonly observed in oncology trials. Transferring causal methodology from observational studies to the clinical trial setting seems appealing. It allows providing an estimate of the treatment effect on survival in the hypothetical case where all patients had remained on their randomized arm. Thus, it accounts for treatment switches, usually motivated by the failure of the front-line treatment, that is, disease progression.

The IPCW approach could be of interest in the clinical setting outside the field of non-compliance. It allows to estimate an "hypothetical" causal

effect on the outcome of any treatment of interest when treatment decisions are possibly based on confounders. However, it could not handle every settings. For instance, when the front line treatment fails, patient could be re-randomized to another treatment. With such a practice, planning a sequential trial is the key to determine the effect of the sequence of treatment [25]. However, open labelled trial with many regimens available, as in the SHIVA01 trial, are not feasible as sequential trial.

Disclosure

The authors declare that they have no conflict of interest.

Conflicts of interest

None declared.

Acknowledgements

We thank the patients who volunteered to participate in this study for their dedication and the study-site staff who cared for them. This work is supported by grant ANR-10-EQPX-03 from the Agence Nationale de la Recherche (Investissements d'avenir) and Site de Recherche Intégré contre le Cancer (SiRIC). High-throughput sequencing was done by the NGS platform of the Institut Curie, supported by grants ANR-10-EQPX-03 and ANR-10-INBS-09-08 from the Agence Nationale de la Recherche (Investissements d'avenir) and the Canceropôle Ile-de-France.

Appendix A. Package description

The following functions will facilitate the data formatting and the computation of the weights.

The function `timesTokeep` can be used with a dataframe in the wide format to retain times when changes occur for the time-dependent covariates. It is used as:

```
timesTokeep(data, id, tstart, tstop, mes.cov, time.cov)
```

and takes the following parameters:

- `data` is the dataframe containing the following variables,
- `id` is the patient's id,
- `tstart` is the date of the beginning of the follow-up (in Date format),
- `tstop` is the date of the end of the follow-up (in Date format),
- `mes.cov` is a list of vectors, each of them must contain the names (in character format) of the repeated measurements related to one time-dependent covariate,
- `time.cov` is a list of vectors, each of them must contain the times (in `Date` format) when the above-mentioned measurements were done.

The output is a list of two lists, one in Date format the other in numeric format. Each of them contains, for each patient, the event time and the times of changes in time-varying covariates.

The function `wideToLongTDC` builds the long format from a dataframe in the wide format given some times defined by the user. It is used as:

```
wideToLongTDC(data, id, tstart, tstop, event, bas.cov, mes.cov, time.cov, times)
```

and takes the same parameters as the previous function added to:

- `event` is the indicator of failure (a death is denoted by 1 at the end of the follow-up),
- `bas.cov` is a vector containing the names (in character format) of the baseline covariates,
- `times` is a list of vectors, each of them must contain, for each patient, the event time and the times of changes in time-varying covariates.

Thus, the `times` parameter can be obtained from the `timesTokeep` function applied on the same dataframe in the wide format.
The output is the long format version of the initial dataframe `data`. The repeated values included in each vector of the list `mes.cov` are aggregated in a variable named as the name of the corresponding list member. This function is not included in the following one because it allows to do some intermediate analyses, such as a time-dependent Cox model adjusted on both fixed and time-dependent covariates.

The key point of the IPCW method is to censor the patients who stopped the randomized treatment and build a censoring indicator, namely `cens`. We provide the function `cens.ipw` that can be used as:

```
cens.ipw(data, id, tstart, tstop, event, arm, realtrt, censTime, trt.start, trt.stop)
```

and that takes the following parameters:

- `data` is the dataframe containing the following variables,

- `id` is the patient's id,
- `tstart` is the date of the beginning of the follow-up (in numeric format),
- `tstop` is the date of the end of the follow-up (in numeric format),
- `arm` is the randomized treatment (2-levels factor),
- `realtrt` is a boolean indicating if treatment really allocated should be introduced (FALSE by default),
- `censTime` is the chosen time to censor the patients (in numeric format),
- `trt.start` is the time of initiation of the randomized treatment (NULL by default),
- `trt.stop` is the time of termination of the randomized treatment (NULL by default).

Note that this function provides the option to include in the data the treatment really taken with the corresponding dates. Then, the treatment really taken is a 3-levels factor, i.e., the two from the randomized arms and a third indicating the no-treatment case (`None`).
The output is a dataframe in the long format, with the data being censored according to the input date, `censTime`.

The following step of data formatting consists in replicating the rows so that each patients' follow-up is split according to all event times. The function `replicRows`, based on the `survSplit` function [14], is used as:

```
replicRows(data, tstart, tstop, event, cens, times1, times2, arm)
```

and has the following parameters:

- `data` is the dataframe containing the following variables,
- `tstart` is the date of the beginning of the follow-up (in numeric format, with the first being equal at 0),
- `tstop` is the date of the end of the follow-up (in numeric format),
- `event` is the indicator of failure (a death is denoted by 1 at the end of the follow-up),
- `cens` is the indicator of censoring (denoted by 1 at the end of the follow-up),
- `times1` is a vector of times (in numeric format) indicating the times according to which the rows have to be split for patients in the first arm,
- `times2` is a vector of times (in numeric format) indicating the times according to which the rows have to be split for patients in the second arm,
- `arm` is the randomized treatment (2-levels factor).

The function `ipcw` allows to compute the stabilized weights on the data formatted according to the previous steps. It is used as:

```
ipcw(data, id, tstart, tstop, cens, arm, bas.cov, conf, trunc = NULL,
type = 'kaplan-meier')
```

and takes the following parameters:

- `data` is the dataframe containing the following variables,
- `id` is the patient's id,
- `tstart` is the date of the beginning of the follow-up (in numeric format, with the first being equal at 0),
- `tstop` is the date of the end of the follow-up (in numeric format),
- `cens` is the indicator of censoring (denoted by 1 at the end of the follow-up),
- `bas.cov` is a vector of baseline covariates,
- `conf` is a vector of time-dependent confounders,
- `trunc` is an optional fraction for the weights. For instance, when `trunc = 0.01`, the left tail is truncated to the 1st percentile and the right tail is truncated to the 99th percentile,
- `type` is a character string specifying the type of survival curve. The default is `type='kaplan-meier'`.

The output appends 2 columns to the input: the stabilized weights, and optionally the stabilized truncated weights. The choice of the truncation threshold can be made rather roughly by comparing the distributions of the different truncated weights. This can be done with the `ipwplot` function of the `ipw` package [20]. Then, the mean of the weights should be as close to 1 as possible and the range should not be too large. This approach can be considered too crude. Thus, some authors proposed to compare both the

bias and precision obtained with different levels of truncation, assuming the marginal structural model estimate being unbiased [26].

References

[1] R. Porcher, V. Levy, S. Chevret, Sample size correction for treatment crossovers in randomized clinical trials with a survival endpoint, Controlled Clinical Trials 23 (2002) 650–661. doi:10.1016/S0197-2456(02)00239-8.

[2] A. Davies, A. Briggs, J. Schneider, A. Levy, O. Ebeid, S. Wagner, S. Kotapati, S. Ramsey, The ends justify the mean: outcome measures for estimating the value of new cancer therapies, Health Outcomes Research in Medicine (2012) e25–e36. doi:10.1016/j.ehrm.2012.01.001.

[3] J. P. Morden, P. C. Lambert, N. Latimer, K. R. Abrams, A. J. Wailoo, Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study, BMC Medical Research Methodology 11 (2011) 4. doi:10.1186/1471-2288-11-4.

[4] N. R. Latimer, K. R. Abrams, P. C. Lambert, M. J. Crowther, A. J. Wailoo, J. P. Morden, R. L. Akehurst, M. J. Campbell, Adjusting for treatment switching in randomised controlled trials – a simulation study and a simplified two-stage method, Statistical Methods in Medical Research 26 (2017) 724–751. doi:10.1177/0962280214557578.

[5] N. R. Latimer, K. R. Abrams, P. C. Lambert, M. J. Crowther, A. J. Wailoo, J. P. Morden, R. L. Akehurst, M. J. Campbell, Adjusting survival time estimates to account for treatment switching in randomized

controlled trials–an economic evaluation context: methods, limitations, and recommendations, Medical Decision Making 34 (2014) 387–402. doi:10.1177/0272989X13520192.

[6] N. R. Latimer, I. R. White, K. R. Abrams, U. Siebert, Causal inference for long-term survival in randomised trials with treatment switching: should re-censoring be applied when estimating counterfactual survival times?, Statistical Methods in Medical Research (2018). doi:10.1177/0962280218780856.

[7] A. S. Walker, I. R. White, A. G. Babiker, Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment, Statistics in Medicine 23 (2004) 571–590. doi:10.1002/sim.1618.

[8] J. M. Robins, Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers, in: Proceedings of the Biopharmaceutical Section, American Statistical Association, volume 24, San Francisco CA, 1993, p. 3.

[9] C. Watkins, X. Huang, N. Latimer, Y. Tang, E. J. Wright, Adjusting overall survival for treatment switches: commonly used methods and practical application, Pharmaceutical Statistics 12 (2013) 348–357. doi:10.1002/pst.1602.

[10] N. R. Latimer, K. R. Abrams, NICE DSU Technical Support Document 16: adjusting survival time estimates in the presence of treat-

ment switching (2014). URL: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD16_Treatment_Switching.pdf.

[11] S. Willems, A. Schat, M. S. van Noorden, M. Fiocco, Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator, Statistical Methods in Medical Research 27 (2018) 323–335. doi:10.1177/0962280216628900.

[12] S. Dodd, P. Williamson, I. R. White, Adjustment for treatment changes in epilepsy trials: a comparison of causal methods for time-to-event outcomes, Statistical Methods in Medical Research 28 (2019) 717–733. doi:10.1177/0962280217735560.

[13] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2018. URL: https://www.R-project.org/.

[14] T. M. Therneau, A package for survival analysis in S, 2015. URL: https://CRAN.R-project.org/package=survival, version 2.38.

[15] J. Fox, M. S. Carvalho, The RcmdrPlugin.survival package: extending the R Commander interface to survival analysis, Journal of Statistical Software 49 (2012) 1–32. doi:10.18637/jss.v049.i07.

[16] L. Thomas, E. M. Reyes, Tutorial: survival estimation for Cox regression models with time-varying coefficients using SAS and R, Journal of Statistical Software 61 (2014) 1–23. doi:10.18637/jss.v061.c01.

[17] J. M. Robins, M. A. Hernán, B. Brumback, Marginal structural models

and causal inference in epidemiology, Epidemiology 11 (2000) 550–560. doi:10.1097/00001648-200009000-00011.

[18] J. M. Robins, D. M. Finkelstein, Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests, Biometrics 56 (2000) 779–788. doi:10.1111/j.0006-341X.2000.00779.x.

[19] J. M. Robins, A. Rotnitzky, Recovery of information and adjustment for dependent censoring using surrogate markers, in: Jewell N.P., Dietz K., Farewell V.T. (eds) AIDS Epidemiology, Birkhäuser, Boston, MA, 1992. doi:10.1007/978-1-4757-1229-2_14.

[20] W. M. van der Wal, R. B. Geskus, ipw: an R package for inverse probability weighting, Journal of Statistical Software 43 (2011) 1–23. doi:10.18637/jss.v043.i13.

[21] C. Le Tourneau, J. P. Delord, A. Goncalves, C. Gavoille, C. Dubot, N. Isambert, M. Campone, O. Tredan, M. A. Massiani, C. Mauborgne, S. Armanet, N. Servant, I. Bieche, V. Bernard, D. Gentien, P. Jezequel, V. Attignon, S. Boyault, A. Vincent-Salomon, V. Servois, M. P. Sablin, M. Kamal, X. Paoletti, Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial, The Lancet Oncology 16 (2015) 1324–1334. doi:10.1016/S1470-2045(15)00188-6.

[22] G. Maldonado, S. Greenland, Simulation study of confounder-selection

strategies, American Journal of Epidemiology 138 (1993) 923–936. doi:10.1093/oxfordjournals.aje.a116813.

[23] K. Skaltsa, C. Ivanescu, S. Naidoo, D. Phung, S. Holmstrom, N. Latimer, Adjusting overall survival estimates after treatment switching: a case study in metastatic castration-resistant prostate cancer, Targeted Oncology 12 (2017) 111–121. doi:10.1007/s11523-016-0472-3.

[24] R. E. Farmer, D. Kounali, A. S. Walker, J. Savović, A. Richards, M. T. May, D. Ford, Application of causal inference methods in the analyses of randomised controlled trials: a systematic review, Trials 19 (2018) 23. doi:10.1186/s13063-017-2381-x.

[25] K. M. Kidwell, SMART designs in cancer research: past, present, and future, Clinical Trials 11 (2014) 445–456. doi:10.1177/1740774514525691.

[26] S. R. Cole, M. A. Hernán, Constructing inverse probability weights for marginal structural models, American Journal of Epidemiology 168 (2008) 656–664. doi:10.1093/aje/kwn164.

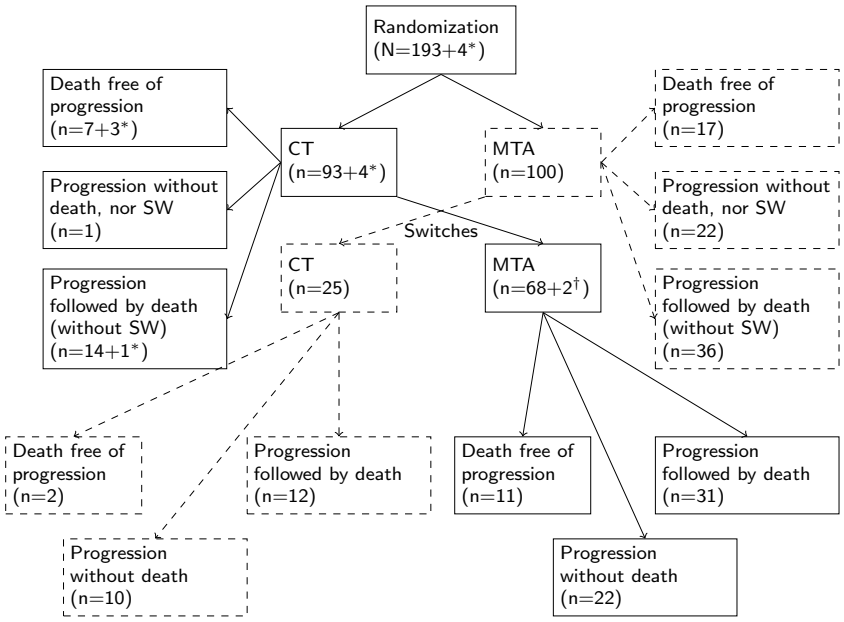


Figure 1: Flow chart for the 197 randomized patients in the SHIVA01 trial. Patients denoted by * did not initiate their randomized treatment (molecularly targeted therapy (MTA) or conventional therapy (CT)), and were not included in the analysis using the IPCW method. SW refers to patient's switch to the other arm, which may occur in case of disease progression. Patients denoted by † were considered as switchers but did not progress nor initiate the other arm treatment, and were not treatment censored.

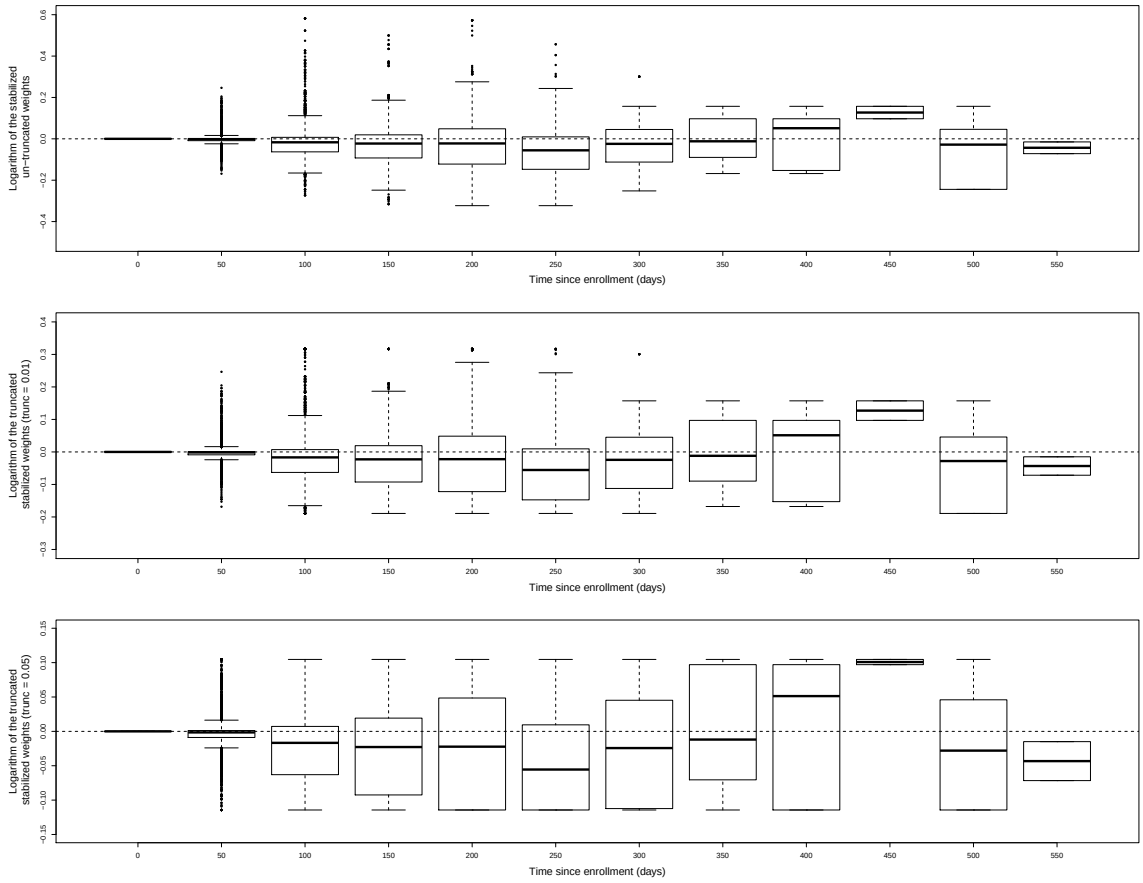


Figure 2: Distribution of the log–stabilized weights either un–truncated or at 0.01 and 0.05, levels of truncation, using different scales for the y-axis.