Received: 19 May 2016 | Revised: 1 June 2017 | Accepted: 18 September 2017

DOI: 10.1002/pst.1834

MAIN PAPER

WILEY

# A simple, doubly robust, efficient estimator for survival functions using pseudo observations

Jixian Wang

Celgene International Sarl, Boudry, Switzerland

**Correspondence**
Jixian Wang, Celgene International Sarl, Route de Perreux 1, Boudry, Basel, Switzerland.
Email: jixwang@celgene.com

Survival functions are often estimated by nonparametric estimators such as the Kaplan-Meier estimator. For valid estimation, proper adjustment for confounding factors is needed when treatment assignment may depend on confounding factors. Inverse probability weighting is a commonly used approach, especially when there is a large number of potential confounders to adjust for. Direct adjustment may also be used if the relationship between the time-to-event and all confounders can be modeled. However, either approach requires a correctly specified model for the relationship between confounders and treatment allocation or between confounders and the time-to-event. We propose a pseudo-observation–based doubly robust estimator, which is valid when either the treatment allocation model or the time-to-event model is correctly specified and is generally more efficient than the inverse probability weighting approach. The approach can be easily implemented using standard software. A simulation study was conducted to evaluate this approach under a number of scenarios, and the results are presented and discussed. The results confirm robustness and efficiency of the proposed approach. A real data example is also provided for illustration.

**KEYWORDS**

causal inference, doubly robust, Kaplan-Meier estimator, inverse probability weighting, pseudo observation

## 1 | INTRODUCTION

Nonparametric estimators such as the Kaplan-Meier (KM) estimator are commonly used to estimate survival functions for a specific population under given treatments.[1] However, when treatment assignment is not randomized, confounding often occurs since the assignment may depend on individual subject's baseline characteristics and prognosis. Consequently the estimator may be biased as the subpopulation receiving a particular treatment may be quite different from the whole population. There are mainly 2 types of approaches to eliminating or reducing confounding biases. When all confounding factors are known and their relationship with treatment assignment can be modeled correctly, the inverse probability weighting (IPW) approach based on the propensity of receiving a given treatment, known as the propensity score (PS),[2] can be used to weight individual event times in the KM estimator to eliminate confounding biases. The PS is a function of potential confounding factors and can be estimated by fitting a model to treatment assignment data. For using IPW in nonparametric estimation of survival function, see Robins and Finkelstein, Hubbard et al, Cole and Hernn, and Xie and Liu.[3-6] Another approach is direct confounding adjustment, based on a time-to-event model with confounding factors as covariates, which will be referred to hereafter as the outcome regression (OR) model. The survival function

  

is estimated by averaging the predictions of individual subjects under a given treatment over the specific population. This approach is also known as g-computation.[7] Each approach is valid only if the model upon which it is based is correctly specified, but none can be verified by the data alone.

Doubly robust (DR) estimators[7-10] were developed to estimate mean treatment effect and parameters in different types of regression models, including those for censored time-to-event data. They are based on a combination of the IPW and the direct adjustment approaches but are robust to model misspecification, as they are valid when at least one of the PS and OR models is correct. At the same time, they are also more efficient than the IPW approach when both models are correct, and even when the OR model is slightly misspecified. For a brief introduction, see Kang and Schafer[10] and Funk et al,[11] while the original development can be found in Robins[7] and Robins et al,[8] in addition to a number of papers by Robins and his colleagues. Doubly robust estimators for mean treatment effect estimation are relatively simple and easy to apply. However, those estimators developed for the estimation of survival functions are practically more difficult to implement.[4] To develop a simple DR estimator for survival functions, we explore the possibility of transforming survival function estimation to that of the mean treatment effect such that DR estimators can be constructed and implemented easily.

To this end, the pseudo-observation approach[12] is a powerful tool. The pseudo-observation approach constructs "pseudo observations" of the survival function, or its functions, for individual subjects such that they can be used for further analyses without censoring issues. Furthermore, there exist a number of applications such as the estimation of survival probability at given time points, restricted mean survival times, and cumulative incidence functions in the setting of competing risks, as well as an alternative approach of fitting a Cox regression model.[13-18] Additionally, their asymptotic properties have also been thoroughly investigated.[19,20] Our approach follows pseudo-observation approaches for survival function estimation and Cox regression[21] in combination with a DR approach for the mean treatment effect estimation.

Our approach consists of the following steps. First, a model for treatment assignment with confounding factors as covariates (the PS model) is fitted to treatment assignment and confounder data. The pseudo observations of survival function for individual subjects, ie, the probability of survival, at given time points are generated from time-to-event data. The IPW estimator for the survival function can be constructed as a weighted mean of the individual pseudo observations. In addition, individual survival probabilities at these time points are predicted by fitting an OR model with confounders as covariates. Finally, the DR estimator is constructed based on the IPW estimator and the OR model–based prediction. Although the approach consists of a few steps, none of them is complex in terms of theory or implementation.

The paper is organized as follows. The next section provides a road map, including the key elements of pseudo observations, and IPW and DR estimators in their simple forms for population mean effect estimation, leading to the development of our approach. Our proposed DR estimator is described in Section 3. Section 4 describes a simulation study and reports the results. For illustration, Section 5 describes an application of the proposed approach to the estimation of survival functions in a clinical study where treatments were not randomized.

## 2 | POPULATION AVERAGE SURVIVAL, CONFOUNDING, AND ADJUSTMENTS

In this section, we present a road map leading to the development of our approach, connecting some key elements including a formal specification of population average survival function, an IPW approach that can be made DR in Section 3, and the motivation of using pseudo observations along with an introduction to them.

### 2.1 | Population average survival function and confounding

Although individual patient's survival depends on their characteristics, we often are interested in the average survival of a patient population under a given treatment. Let $(T_i, C_i, D_i)$ be the survival time, censoring time, and a binary treatment indicator, respectively, and $\mathbf{Z}_i$ be a set of covariates for subject $i, i = 1, \ldots, n$, characterized by the distribution $\mathbf{Z}_i \sim F(\mathbf{Z})$. The population average survival functions under treatments $D_i = 0$ (control) and $D_i = 1$ (active) can be written as

$$S_d(t) = \int S_d(t|\mathbf{Z}_i) dF(\mathbf{Z}_i), \tag{1}$$

where $S_d(t|\mathbf{Z}_i)$ is the survival function for patients with $\mathbf{Z}_i$ under treatment $d$. When $D_i$ is independent of $\mathbf{Z}_i$, (eg, is randomized), the KM estimator[1] based on $(T_i, C_i, D_i)$ for those with $D_i = d$ is valid for $S_d(t)$. However, when the $D_i$ and $\mathbf{Z}_i$ are correlated, confounding bias occurs in the KM estimator, as it estimates

$$\int S_d(t|\mathbf{Z}_i)dF(\mathbf{Z}_i|D_i = d), \tag{2}$$

where $F(\mathbf{Z}_i|D_i = d)$ is the conditional distribution of $\mathbf{Z}_i$ for those with $D_i = d$, which may be significantly different from $F(\mathbf{Z}_i)$. For example, if $\mathbf{Z}_i$ represents age and the probability of receiving the active treatment $D_i = 1$ depends on age, then $F(\mathbf{Z}_i|D_i = 1)$ represents the age distribution among those receiving the active treatment, which is different from the overall age distribution $F(\mathbf{Z}_i)$ or that receiving the control $F(\mathbf{Z}_i|D_i = 0)$. The difference in $F(\mathbf{Z}_i|D_i = d)$ leads to bias in the KM estimator, as it averages over a different age distribution from $F(\mathbf{Z}_i)$.

## 2.2 | Inverse probability weighting

The IPW approach is commonly used for confounding bias adjustment. To introduce the IPW approach, and also the DR approach subsequently, we consider its use in a simple situation. Suppose that $y_i$ is the response measure of patient $i$, and we are interested in its population mean $\mu_1 \equiv E(y_i(1))$, where $y_i(1)$ is the response of subject $i$ when treatment $D_i = 1$ is given, although in observed data $D_i$ may not equal to 1. Therefore, when $D_i$ is not randomized, $\mu_1$ may be very different from $E(y_i|D_i = 1)$: the mean among those who received $D_i = 1$. The IPW estimator

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i y_i}{p_{1i}}, \tag{3}$$

where $p_{1i} = P(D_i = 1|\mathbf{Z}_i)$ is the PS, can be used to estimate $\mu_1$. Use the fact that for binary $D_i$, $y_i$ can be written as $y_i = D_i y_i(1) + (1 - D_i)y_i(0)$, it is easy to verify that

$$E(\hat{\mu}_1) = E(E(y_i D_i p_{1i}^{-1}|\mathbf{Z}_i)) = E(y_i(1)) = \mu_1; \tag{4}$$

hence, $\hat{\mu}_1$ is unbiased. In practice, $p_{1i}$ can be estimated by a logistic regression with

$$p_{1i}(\boldsymbol{\gamma}) = \frac{\exp(\boldsymbol{\gamma}^T\mathbf{Z}_i)}{1 + \exp(\boldsymbol{\gamma}^T\mathbf{Z}_i)}, \tag{5}$$

where $\boldsymbol{\gamma}$ are parameters to be estimated. When model (5) is correct, $\hat{\mu}_1$ is consistent, based on (4). One important advantage of estimator (3) is, being a weighted sum, it can be made DR easily, as we do in the next section. However, adapting (3) for estimating $S_d(t)$ is difficult as we do not observe individual patients survival curve $S_d(t|\mathbf{Z}_i)$.

## 2.3 | Pseudo-observation–based IPW

Although individual patients survival curves cannot be observed, they can be estimated by the pseudo-observation approach. This approach was introduced by Andersen et al[13] and applied to estimating the survival curve at a single time point by Klein et al.[18] The pseudo observation for the survival function of subject $i$ is estimated as

$$\hat{S}_d^i(t) = n\hat{S}_d(t) - (n-1)\hat{S}_d^{-i}(t), \tag{6}$$

where $\hat{S}_d(t)$ and $\hat{S}_d^{-i}(t)$ are the KM estimators using all samples and that leaving out subject $i$, respectively. The asymptotic properties of the pseudo observations have been examined by Graw et al[19] for competing risk models, which also apply to $\hat{S}_d^i(t)$ here. They showed that when $n \to \infty$,

$$E(\hat{S}_d^i(t)|\mathbf{Z}_i) = S_d(t|\mathbf{Z}_i) + o(1), \tag{7}$$

where $o(1)$ is a term that tends to 0. A higher-order term correction has been given in Jacobsen and Martinussen. [20]

According to property (7), the IPW approaches can be easily applied to estimate $S_d(t)$ based on pseudo observations $\hat{S}_d^i(t)$, as, for a fixed $t$, one can treat $\hat{S}_d^i(t)$ as $y_i$ in estimator (3). Therefore, we can construct the IPW estimator for, eg, $S_1(t)$ by replacing $y_i$ with $\hat{S}_1^i(t)$ in (3) and obtain an estimator:

$$\hat{S}_1^{IPW}(t) = n^{-1}\sum_{i=1}^{n}\frac{D_i\hat{S}_1^i(t)}{p_{1i}(\hat{\boldsymbol{\gamma}})}, \tag{8}$$

where $\hat{\boldsymbol{\gamma}}$ is obtained by fitting the PS model (5). Using the property in (7), we get $E(\hat{S}_1^{IPW}(t)) \to S_d(t)$. As shown in the next section, the estimator (8) can be used to construct DR estimators that are easier to implement than, eg, those in Hubbard et al.[4] Note that despite the nice property in (7), $\hat{S}_d^i(t)$ may not necessarily be in the range of (0,1). This issue does not affect the consistency of the DR estimators in Section 3 based on $\hat{S}_d^i(t)$, as it only depends on property (7). However, when the sample size is small, the DR estimates might be out of the (0,1) range for some $t$ values and may be truncated.

As a final remark, the IPW can also be applied to the original KM estimator[1] to adjust confounding biases. However, unlike estimator (8), this weighted original KM estimator is not a weighted sum and hence cannot be made DR easily.

## 2.4 | Model-based direct adjustments

Another approach for eliminating the bias is to directly adjust the confounding in an OR model for the relationship between the event time, confounding factors and treatments, and to predict $S_d(t|\mathbf{Z}_i)$ for $\mathbf{Z}_i$s in the data or generated by $F(\mathbf{Z}_i)$. Then (1) can be estimated by averaging over these predictions. For this purpose, the most commonly used model is the proportional hazard (Cox) model[21] with hazard function

$$h_d(t|\mathbf{Z}_i) = h_{0d}(t)\exp(\boldsymbol{\beta}_d^T\mathbf{Z}_i), \tag{9}$$

where $h_{0d}(t)$ is the baseline hazard function and $\boldsymbol{\beta}_d$ is a set of parameters under treatment $d$. In practice, simpler proportional hazard models for treatment effects and shared parameters for $\mathbf{Z}_i$ are often used. According to model (9), one can write

$$S_d(t|\mathbf{Z}_i) = S_{d0}(t)^{\exp(\boldsymbol{\beta}_d^T\mathbf{Z}_i)} \tag{10}$$

in which $S_{d0}(t)$ and $\boldsymbol{\beta}_d$ can be estimated by fitting the Cox model. Then (1) can be approximated by

$$\hat{S}_d(t) = n^{-1}\sum_{i=1}^{n}\hat{S}_{d0}(t)^{\exp(\hat{\boldsymbol{\beta}}_d^T\mathbf{Z}_i)} \tag{11}$$

as a population average of estimated $S_d(t|\mathbf{Z}_i)$s. The approach is also known as the g-computation.[7] The approach is valid only if model (9) is correctly specified, which is often difficult to verify.

Pseudo observations can also be used for model-based prediction to construct the DR estimator in the next section. For example, the Cox model can be converted to a generalized linear model for pseudo observations.[12] For this model and a given time $t$, the complementary log-log link function $g(x) = \log(-\log(x))$ leads to

$$E(g^{-1}(\hat{S}_d^i(t)|\mathbf{Z}_i)) \approx \beta_{dt} + \boldsymbol{\beta}_d^T\mathbf{Z}_i, \tag{12}$$

where $\beta_{dt} = \int_0^t h_{d0}(\tau)d\tau$. This model can be used to estimate $\boldsymbol{\beta}_d$ in the Cox model as well as the marginal survival function for a specific population. Note that, although $\beta_{dt}$ are time specific, $\boldsymbol{\beta}_d$ are shared between different times $t$s.

The model (12) can be fitted to $\hat{S}_d^i(\mathbf{t})$: a set of $\hat{S}_d^i(t)$s at a set of times $\mathbf{t} = (t_1, \ldots, t_K)$ by solving the following generalized estimating equation (GEE):

$$U(\boldsymbol{\beta}_d^*) = \sum_{i=1}^{n}\frac{\partial\mu_i(\mathbf{t},\boldsymbol{\beta}_d^*,\mathbf{Z}_i)}{\partial\boldsymbol{\beta}_d^*}\mathbf{V}_i(\hat{S}_d^i(\mathbf{t}) - \mu_i(\mathbf{t},\boldsymbol{\beta}_d^*,\mathbf{Z}_i)) = 0, \tag{13}$$

where $\mu_i(\mathbf{t},\boldsymbol{\beta}_d^*,\mathbf{Z}_i) = g^{-1}(\boldsymbol{\beta}_{dt} + \boldsymbol{\beta}_d^T\mathbf{Z}_i)$, $\mathbf{V}_i$ is a working covariance matrix, $\boldsymbol{\beta}_{dt} = (\beta_{d1}, \ldots, \beta_{dK})$ and $\boldsymbol{\beta}_d^* = (\boldsymbol{\beta}_{dt}, \boldsymbol{\beta}_d)$. Parameters $\boldsymbol{\beta}_d^*$ can be estimated by GEE software including a dummy categorical variable for $(t_1, \ldots, t_K)$ as well as $\mathbf{Z}_i$. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}_d^*$ can be estimated by a sandwich estimator.[12,13] We omit the details here as the variance estimator for the DR estimator we proposed in the next section does not depend on it.

# 3 | THE PROPOSED DR ESTIMATOR

Both IPW and direct adjustment approaches are sensitive to the misspecification of the models they depend on. However, DR estimators can be constructed based on the IPW estimator (8) and model-based prediction using either (11) or (12) such that they are consistent as long as one of the models is correct. Doubly robust estimators for population means are simple to construct and easy to implement. Therefore, before presenting the proposed DR estimator for $S_d(t)$, we give a brief introduction to the DR estimation for the mean treatment effect $E(y_i(d))$. See Kang and Schafer[10] and Funk[11] for more details. A DR estimator for $E(y_i(d))$ can be constructed as a weighted sum of the IPW estimator (3) and OR model

$E(y_i|D_i, \mathbf{Z}_i) = m_{D_i}(\mathbf{Z}_i, \boldsymbol{\beta})$ for $y_i, D_i$, and $\mathbf{Z}_i$. After fitting this model and the PS model (5) to obtain parameter estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, a DR estimator for $E(y_i(1))$ can be constructed as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i D_i - (D_i - p_{1i}(\hat{\boldsymbol{\gamma}}))m_1(\mathbf{Z}_i, \hat{\boldsymbol{\beta}})}{p_{1i}(\hat{\boldsymbol{\gamma}})}, \tag{14}$$

which can also be written as

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i(1) + \frac{(D_i - p_{1i}(\hat{\boldsymbol{\gamma}}))(y_i - m_1(\mathbf{Z}_i, \hat{\boldsymbol{\beta}}))}{p_{1i}(\hat{\boldsymbol{\gamma}})}\right]. \tag{15}$$

Its DR property comes from the fact that the second term has approximately a zero mean if either the model for $p_{1i}(\boldsymbol{\gamma})$ or the model for $m_1(\mathbf{Z}_i, \boldsymbol{\beta})$ is correctly specified. It has also been demonstrated that when both models are correct, the estimator is semiparametric efficient.[10,11]

In the same way as we construct the estimator (8), for a given time $t$, a DR estimator for $S_d(t)$ can be constructed by replacing $y_i$ by $\hat{S}_d^i(t)$ in (14). This leads to our proposed DR estimator for, for example, $S_1(t)$:

$$\hat{S}_{1DR}(t) = n^{-1}\sum_{i=1}^{n}\frac{D_i\hat{S}_1^i(t) - (D_i - p_{1i}(\hat{\boldsymbol{\gamma}}))m_1(t, \mathbf{Z}_i, \hat{\boldsymbol{\beta}}^*)}{p_{1i}(\hat{\boldsymbol{\gamma}})}, \tag{16}$$

where $m_1(t, \mathbf{Z}_i, \hat{\boldsymbol{\beta}}^*) = \exp(-\exp(\hat{\beta}_{1t} + \hat{\boldsymbol{\beta}}_1^T\mathbf{Z}_i))$ is the estimate of $S_1^i(t)$ based on the OR model (12). Alternatively the Cox model based prediction (11) can also be used for $m_1(t, \mathbf{Z}_i, \hat{\boldsymbol{\beta}})$. The first term is $\hat{S}_1^{IPW}(t)$ given in the previous section. The variance of $\hat{S}_{1DR}(t)$ is rather complex,[9] but can be estimated as

$$n^{-2}\sum_{i=1}^{n}\left(\frac{D_i\hat{S}_1^i(t) - (D_i - p_{1i}(\hat{\boldsymbol{\gamma}}))m_1(t, \mathbf{Z}_i, \hat{\boldsymbol{\beta}}^*)}{p_{1i}(\hat{\boldsymbol{\gamma}})} - \hat{S}_{1DR}(t)\right)^2. \tag{17}$$

An alternative approach is bootstrapping,[22] which has been used for other DR approaches.

The formula (16) suggests that if $m_1(t, \mathbf{Z}_i, \hat{\boldsymbol{\beta}}^*)$ has no or weak correlation with $D_i - p_{1i}$, the DR estimator is approximately the same as $\hat{S}_1^{IPW}(t)$, since $E((D_i - p_{i1})/p_{1i}) = 0$. This happens when $m_1(t, \mathbf{Z}_i, \boldsymbol{\beta})$ only contains covariates different from those in $p_{1i}$.

## 4 | A SIMULATION STUDY

We conducted a simulation study to evaluate the performance of the proposed DR estimator. Assuming that there were 3 normally distributed covariates $\mathbf{Z}_i = (Z_{1i}, Z_{2i}, Z_{3i})$ with unit variance and covariance 0.2 between $Z_{1i}$ and $Z_{2i}$, while $Z_{3i}$ was independent of the others. For the PS model, we assumed that

$$\text{logit}(p_{i1})(\boldsymbol{\gamma}) = (Z_{1i} + Z_{2i} + Z_{3i})/3. \tag{18}$$

$T_i$ was generated from an exponential distribution under 2 scenarios with a hazard in the OR model

$$h_{D_i}(\mathbf{Z}_i) = \exp(-3 + D_i + 0.5Z_{1i} - 0.5Z_{2i}) \tag{19}$$

or

$$h_{D_i}(\mathbf{Z}_i) = \exp(-3 + D_i + 0.5Z_{1i} + 0.5Z_{2i}), \tag{20}$$

among them (19) has weak confounding (as the confounding effects of $Z_{1i}$ and $Z_{2i}$ are partially canceled out) and (20) has strong confounding. An independent censoring with constant risk exp(−3.5), which led to about 28% censored data, was implemented. The setting is similar to that in Zhang and Schaubel.[23] We were interested in estimating $S_1(t)$ at $t = 5, 15, 30$, which gave $S_1(t)$ values 0.49, 0.23, and 0.10, respectively. The IPW estimator (8) and the DR estimator (16) were used to estimate $S_1(t)$ and compared with the unadjusted KM estimator $\sum_{i=1}^{n}\hat{S}_1^i(t)/\sum_{i=1}^{n}D_i$, which is a special case of (8) without IPW.

We considered 4 scenarios: (1) Both the hazard in the OR model (Equations 19 and 20) and the PS model (18), which will be referred to as the $T_i$ and $D_i$ models for simplicity, were correct, (2) the $T_i$ model was correct but the $D_i$ model was not, (3) the $T_i$ model was not correct but the $D_i$ model was correct, and (4) both models were incorrect. In the incorrect

$T_i$ model $h_{D_i}(\mathbf{Z}_i)$ included $Z_{1i}, Z_{3i}$ instead of $Z_{1i}, Z_{2i}$, and the incorrect $D_i$ model did not include $Z_{2i}$. For each scenario, 1000 simulations were run with sample sizes $n = 200$ and $n = 500$. For pseudo-observation estimation, the R–package "pseudo" was used, and for solving (13), R–function "geese" was used. The relative bias as the mean of $(\hat{S}(t) - S(t))/S(t)$, the relative SD: $SD(\hat{S}(t))/S(t)$ and the relative MSE were calculated. The results under the weak confounding scenario are presented in Table 1. When both models were correct, the IPW and DR estimators were similar, almost unbiased with significantly lower MSEs than that of the unadjusted KM estimator. But the DR estimator also had a slightly lower SD than that of the IPW estimator. Their SDs when $n = 200$ were higher than those when $n = 500$, as expected. When the $T_i$ model was correct but the $D_i$ model was not; the IPW estimator was biased, also as expected, for both $n = 500$ and $n = 200$. The biases in the DR estimator were very low, except a slightly higher one −0.019 at $t = 30$ and $n = 200$, compared with −0.0128 when both models were correct. To further investigate, a simulation with 5000 runs was repeated for this scenario and the resulting bias was −0.0129, which suggests that the slightly higher value seemed due to simulation errors. The IPW and DR estimators were quite similar regardless of sample sizes, with almost no bias and slightly lower SD when the $T_i$ model was not correct but the $D_i$ model was correct. When both models were incorrect, the IPW and DR estimators were also quite similar. They were biased, but their biases were lower than those of the unadjusted KM estimator, and the SDs were also slightly lower.

Table 2 gives the relative bias, SD, and MSE for the scenario of strong confounding under which the unadjusted estimator had up to about 30% bias. Similar patterns as seen in Table 1 were found. When both models were correct, IPW and DR estimates were almost unbiased and had significantly lower MSEs, and sometimes significantly lower SDs, eg, when

**TABLE 1** Relative bias, SD, and mean squared error (MSE) of unadjusted, inverse probability weighting (IPW), and doubly robust (DR) estimates for $S_1(t)$ for $t = 5, 15, 30$ and $n = 200$ and $n = 500$ sample sizes, with weak confounding. Under column Models, $D_i$ and $T_i$denote the models for $p_{1i}$ and $h_{D_i}(\mathbf{Z}_i)$, respectively, and C and N are shorthands for correct and incorrect models

| Models | | | Time points | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $t = 5$ | | | $t = 15$ | | | $t = 30$ | | |
| $D_i$ | $T_i$ | Statistic | Unadj | IPW | DR | Unadj | IPW | DR | Unadj | IPW | DR |
| $n = 200$ | | | | | | | | | | | |
| C | C | Bias | 0.0522 | −5e−04 | −3e−04 | 0.0721 | −0.0110 | −0.0107 | 0.1025 | −0.0135 | −0.0128 |
| | | SD | 0.1002 | 0.1050 | 0.1023 | 0.1504 | 0.1492 | 0.1435 | 0.2650 | 0.2511 | 0.2439 |
| | | MSE | 0.0128 | 0.0110 | 0.0105 | 0.0278 | 0.0224 | 0.0207 | 0.0808 | 0.0633 | 0.0597 |
| N | C | Bias | 0.0347 | −0.0432 | 0.0012 | 0.0445 | −0.0660 | −0.0056 | 0.0475 | −0.0993 | −0.0192 |
| | | SD | 0.0957 | 0.0934 | 0.0905 | 0.1458 | 0.1324 | 0.1344 | 0.2313 | 0.2066 | 0.2110 |
| | | MSE | 0.0104 | 0.0106 | 0.0082 | 0.0232 | 0.0219 | 0.0181 | 0.0558 | 0.0526 | 0.0449 |
| C | N | Bias | 0.0505 | −0.0012 | −7e−04 | 0.0695 | 0.0070 | 0.0075 | 0.0885 | 0.0147 | 0.0154 |
| | | SD | 0.1055 | 0.1018 | 0.1013 | 0.1597 | 0.1548 | 0.1544 | 0.2844 | 0.2718 | 0.2705 |
| | | MSE | 0.0137 | 0.0104 | 0.0103 | 0.0304 | 0.024 | 0.0239 | 0.0887 | 0.0741 | 0.0734 |
| N | N | Bias | 0.0755 | −0.0389 | −0.0402 | 0.1056 | −0.0526 | −0.0527 | 0.1289 | −0.0672 | −0.0661 |
| | | SD | 0.0995 | 0.0963 | 0.0957 | 0.1499 | 0.1334 | 0.1325 | 0.2483 | 0.2098 | 0.2084 |
| | | MSE | 0.0156 | 0.0108 | 0.0108 | 0.0336 | 0.0206 | 0.0203 | 0.0783 | 0.0486 | 0.0478 |
| $n = 500$ | | | | | | | | | | | |
| C | C | Bias | 0.0603 | −9e−04 | −0.0012 | 0.0925 | −1e−04 | −7e−04 | 0.1274 | −0.0027 | −0.0035 |
| | | SD | 0.0639 | 0.0631 | 0.062 | 0.1046 | 0.0992 | 0.0974 | 0.1804 | 0.165 | 0.162 |
| | | MSE | 0.0077 | 0.004 | 0.0038 | 0.0195 | 0.0098 | 0.0095 | 0.0488 | 0.0272 | 0.0263 |
| N | C | Bias | 0.0650 | −0.0392 | −8e−04 | 0.1060 | −0.0467 | 0.0019 | 0.1390 | −0.0699 | −0.0087 |
| | | SD | 0.0647 | 0.0615 | 0.0608 | 0.1023 | 0.0912 | 0.0911 | 0.1770 | 0.1494 | 0.1510 |
| | | MSE | 0.0084 | 0.0053 | 0.0037 | 0.0217 | 0.0105 | 0.0083 | 0.0506 | 0.0272 | 0.0229 |
| C | N | Bias | 0.0512 | −0.0028 | −0.0028 | 0.0807 | 0.0010 | 0.0010 | 0.1056 | −0.0027 | −0.0027 |
| | | SD | 0.0669 | 0.0646 | 0.0643 | 0.1015 | 0.0956 | 0.0953 | 0.1710 | 0.1592 | 0.1587 |
| | | MSE | 0.0071 | 0.0042 | 0.0041 | 0.0168 | 0.0091 | 0.0091 | 0.0404 | 0.0254 | 0.0252 |
| N | N | Bias | 0.0504 | −0.0404 | −0.0404 | 0.0636 | −0.0585 | −0.0589 | 0.0739 | −0.0802 | −0.0808 |
| | | SD | 0.0596 | 0.0569 | 0.0570 | 0.0923 | 0.0841 | 0.0840 | 0.1569 | 0.1380 | 0.1377 |
| | | MSE | 0.0061 | 0.0049 | 0.0049 | 0.0126 | 0.0105 | 0.0105 | 0.0301 | 0.0255 | 0.0255 |

**TABLE 2** Relative bias, SD, and mean squared error (MSE) of unadjusted, inverse probability weighting (IPW), and doubly robust (DR) estimates for $S_1(t)$for $t = 5, 10, 20$ and sample sizes $n = 200$ and $n = 500$, with strong confounding. Under column Models, $D_i$ and $T_i$ denote the models for $p_{1i}$and $h_{D_i}(\mathbf{Z}_i)$, respectively, and C and N are shorthands for correct and incorrect models

| Models | | | Time points | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $t = 5$ | | | $t = 15$ | | | $t = 30$ | | |
| $D_i$ | $T_i$ | Statistic | Unadj | IPW | DR | Unadj | IPW | DR | Unadj | IPW | DR |
| $n = 200$ | | | | | | | | | | | |
| C | C | Bias | 0.0987 | −0.0044 | −0.0046 | 0.1594 | −0.0039 | −0.0042 | 0.2432 | 0.0000 | −6e−04 |
| | | SD | 0.1019 | 0.1016 | 0.1014 | 0.1716 | 0.1578 | 0.1577 | 0.3441 | 0.2857 | 0.2842 |
| | | MSE | 0.0201 | 0.0104 | 0.0103 | 0.0548 | 0.0249 | 0.0249 | 0.1775 | 0.0816 | 0.0807 |
| N | C | Bias | 0.1004 | 0.0529 | −0.0046 | 0.1643 | 0.0838 | 0.0065 | 0.2390 | 0.1150 | 0.0128 |
| | | SD | 0.0978 | 0.0998 | 0.0976 | 0.1617 | 0.1573 | 0.1493 | 0.3184 | 0.3005 | 0.2794 |
| | | MSE | 0.0196 | 0.0128 | 0.0095 | 0.0531 | 0.0318 | 0.0223 | 0.1585 | 0.1035 | 0.0782 |
| C | N | Bias | 0.1048 | −0.0035 | −0.0028 | 0.1734 | 0.0049 | 0.0057 | 0.2620 | 0.0144 | 0.0153 |
| | | SD | 0.1015 | 0.0984 | 0.0992 | 0.1670 | 0.1498 | 0.1503 | 0.3232 | 0.2667 | 0.2670 |
| | | MSE | 0.0213 | 0.0097 | 0.0099 | 0.0580 | 0.0225 | 0.0226 | 0.1731 | 0.0714 | 0.0715 |
| N | N | Bias | 0.1253 | 0.0546 | 0.0530 | 0.2022 | 0.0790 | 0.0781 | 0.2971 | 0.1048 | 0.1048 |
| | | SD | 0.1038 | 0.1074 | 0.1063 | 0.1748 | 0.1682 | 0.1671 | 0.3397 | 0.3082 | 0.3069 |
| | | MSE | 0.0265 | 0.0145 | 0.0141 | 0.0715 | 0.0345 | 0.0340 | 0.2037 | 0.1060 | 0.1052 |
| $n = 500$ | | | | | | | | | | | |
| C | C | Bias | 0.1222 | −1e−04 | −3e−04 | 0.2026 | 0.0048 | 0.0047 | 0.2937 | 9e−04 | 0.0010 |
| | | SD | 0.0684 | 0.0678 | 0.0676 | 0.1114 | 0.0992 | 0.099 | 0.2131 | 0.1747 | 0.1741 |
| | | MSE | 0.0196 | 0.0046 | 0.0046 | 0.0534 | 0.0099 | 0.0098 | 0.1317 | 0.0305 | 0.0303 |
| N | C | Bias | 0.1133 | 0.0457 | 7e−04 | 0.1717 | 0.0592 | 0.0029 | 0.2561 | 0.0809 | 0.0103 |
| | | SD | 0.0638 | 0.0642 | 0.062 | 0.1099 | 0.1032 | 0.0984 | 0.2062 | 0.1824 | 0.1724 |
| | | MSE | 0.0169 | 0.0062 | 0.0038 | 0.0416 | 0.0141 | 0.0097 | 0.1081 | 0.0398 | 0.0298 |
| C | N | Bias | 0.1161 | −0.0014 | −0.0015 | 0.1843 | 5e−04 | 7e−04 | 0.2579 | −0.0087 | −0.0085 |
| | | SD | 0.0686 | 0.0661 | 0.0661 | 0.1118 | 0.0991 | 0.0993 | 0.2168 | 0.1774 | 0.1775 |
| | | MSE | 0.0182 | 0.0044 | 0.0044 | 0.0465 | 0.0098 | 0.0099 | 0.1135 | 0.0315 | 0.0316 |
| N | N | Bias | 0.1071 | 0.0500 | 0.0499 | 0.1706 | 0.0774 | 0.0772 | 0.2500 | 0.1089 | 0.1086 |
| | | SD | 0.0623 | 0.0642 | 0.0642 | 0.0999 | 0.0971 | 0.0973 | 0.1914 | 0.1764 | 0.1764 |
| | | MSE | 0.0154 | 0.0066 | 0.0066 | 0.0391 | 0.0154 | 0.0154 | 0.0991 | 0.0430 | 0.0429 |

$t = 30$, than that of the unadjusted one. Doubly robustness of the DR approach was also confirmed, as noticeable biases only occurred when both models were incorrect.

We also investigated the empirical SE estimator based on (17), by comparing their means with the SD of the DR estimates of the 1000 runs. The results are compared in Table 3. In general, the empirical estimator overestimated the variation, with upward biases ranging from 2% to 13%, the highest bias occurred when $t = 5$, $n = 500$, and both models were correct. There was no clear trend in the bias with sample size changes. Nevertheless the bias under all situations led to an overestimated variability, hence directly using (17), although being slightly conservative, will not invalidate statistical inference. This situation was further investigated by replacing the pseudo observations with random samples from a normal distribution with confounded means. The bias was not significantly reduced (results not presented). This suggests that the problem may be due to the empirical SE estimator for the IPW estimator, which is the main source of variation, rather than due to the use of pseudo observations. When a more accurate estimator is preferred, a bootstrap approach, which has been used widely for the IPW estimator can be used.

## 5 | AN APPLICATION TO A BREAST CANCER STUDY

The proposed DR estimator was applied to the cohort GSE6532 of a breast cancer study, reported in Loi et al,[24] to compare Tamoxifen with a control for treating patients with breast carcinomas. Zhu et al[25] used this dataset to illustrate their approach to finding the optimal treatment using a machine learning approach and implemented in R-package RLT. The treatments were not randomized, hence a direct comparison between the 2 treatment groups may not be valid. The dataset

**TABLE 3** Relative biases of empirical SE of doubly robust estimates for $S_1(t)$for $t = 5, 15, 30$. Under column Models, $D_i$ and $T_i$ denote the models for $p_{1i}$ and $h_{D_i}(\mathbf{Z}_i)$, respectively, and C and N are shorthands for correct and incorrect models

| | | Models | | Time Points | | |
|---|---|---|---|---|---|---|
| **Confounding weak** | **Sample size** | $D_i$ | $T_i$ | $t = 5$ | $t = 15$ | $t = 30$ |
| | 200 | C | C | 0.0523 | 0.0463 | 0.0616 |
| | | N | C | 0.0430 | 0.0871 | 0.0719 |
| | | C | N | 0.0867 | 0.0872 | 0.0764 |
| | | N | N | 0.0405 | 0.0377 | 0.0202 |
| | 500 | C | C | 0.0251 | 0.0727 | 0.0758 |
| | | N | C | 0.0774 | 0.0711 | 0.1154 |
| | | C | N | 0.0772 | 0.0997 | 0.0761 |
| | | N | N | 0.0187 | 0.0647 | 0.0624 |
| Strong | | | | $t = 5$ | $t = 15$ | $t = 30$ |
| | 200 | C | C | 0.0371 | 0.0983 | 0.0849 |
| | | N | C | 0.0630 | 0.0387 | 0.0673 |
| | | C | N | 0.0780 | 0.0648 | 0.0525 |
| | | N | N | 0.0752 | 0.0951 | 0.0902 |
| | 500 | C | C | 0.1290 | 0.0698 | 0.0691 |
| | | N | C | 0.0789 | 0.0582 | 0.0305 |
| | | C | N | 0.0700 | 0.0787 | 0.0674 |
| | | N | N | 0.0895 | 0.0690 | 0.0566 |

in the RLT package includes baseline covariates age, tumor grade and size, node involvement, and estrogen receptor positive ($ER^+$). These covariates were used for the adjustment of survival curve estimates, using either the IPW and the proposed DR approaches. Some patients had missing values in some covariates, hence were excluded in the analysis. The complete dataset had 228 and 104 patients in the Tamoxifen and the control groups, respectively. Node involvement was not included in the model, as no patient in the control group had node involvement. The fitted PS model is

$$\text{logit}(p_i) = -10.5 + 0.12(0.02) \times \text{Age} + 0.22(0.24) \times \text{Grade} + 0.44(0.18) \times \text{Size} + 3.19(0.66) \times \text{ER}^+, \tag{21}$$

where numbers in brackets are the SEs. Three factors significant at the 5% level were age ($P < .001$), ER positive ($P < .001$), and tumor size (0.014). The log transformation was tried on age and tumor size but did not result in a better model fitting. The survival time data were fitted to Cox models. It was found that the log transformation was appropriate to age and tumor size. The fitted model had a log-hazard ratio (HR)

$$\begin{aligned}\text{logHR} = &-0.10(0.60) \times \text{log-Age} + 0.12(0.15) \times \text{Grade} + 1.05(0.27) \times (\text{log-Size}+0.1) \\ &- 0.40(0.32) \times \text{ER}^+ - 0.22(0.31) \times \text{Tamo},\end{aligned} \tag{22}$$

where a constant 0.1 was added to the log-tumor size, as some patients had a zero tumor size. The fitted model is insensitive to moderate changes in this constant. Only the effect of log size was significant at the 5% level.

To construct the DR estimator, we used 2 prediction approaches. One directly used the Cox regression on the raw survival times, the other fitted pseudo observations to the GEE model (12). They resulted in very similar estimated curves; hence, we will only show those based on the pseudo observations. To reduce computation, $\hat{S}_d^i(t)$ was calculated yearly. The DR survival curves by treatment are shown in Figure 1, together with the corresponding IPW adjusted and unadjusted curves. Their CIs were not shown in the Figure due to space limitation. For the active treatment, the adjusted survival curves were similar to the unadjusted one. The adjusted survival curves for the control group were much lower than the unadjusted one, suggesting a much higher treatment effect after the adjustments. For both groups, the IPW and DR estimates were very similar, as the second term in the DR estimator (16) did not have much impact. As discussed in Section 3, the impact of the second term may be small when there is a lack of common significant factors in both the PS and OR models. This might be a reason for the similarity between the IPW and DR estimates, as only tumor size was significant at the 5% level in both models (not significant at the 1% level in the PS model).

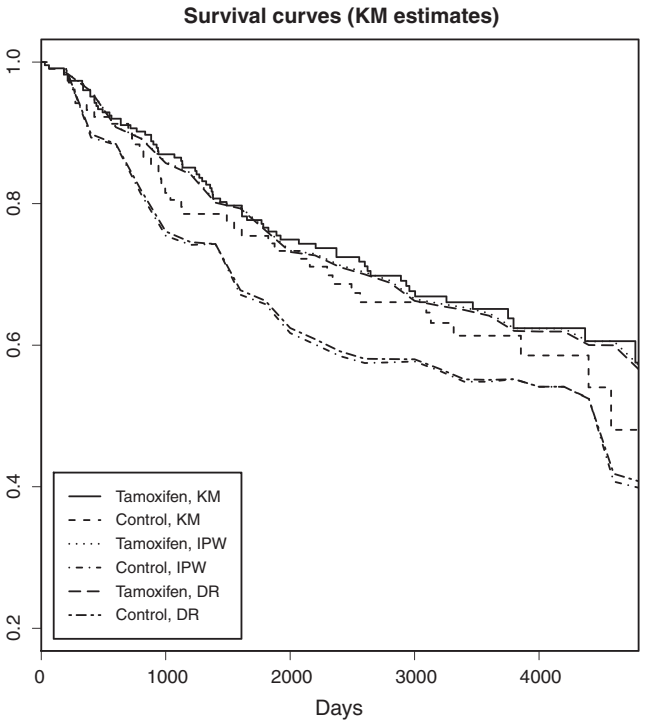


**FIGURE 1** Adjusted survival curves for the Tamoxifen and control groups by the inverse probability weighting (IPW) and doubly robust (DR) approach and compared with unadjusted Kaplan-Meier (KM) curves

As a comparison, the marginal log-HR was estimated by a Cox model without any covariate and the estimate was $-0.023$(SE = 0.20, $P$ value = .24), similar to the covariate adjusted one. Applying IPW to the same Cox model, the estimated log-HR was $-0.40$ (SE = 0.13, $P$ value = .003) also suggested a much higher treatment effect after the IPW adjustment. We did not use DR estimators for estimating log-HR, as they are not easy to implement.

## 6 | DISCUSSION

We have proposed a pseudo-observation–based confounding adjustment approach to the estimation of survival functions. The approach is DR and often more efficient than the common IPW approach, yet it is simple to implement. The approach is also applicable to estimating measures derived from the survival function, such as restricted mean survival times. We have also evaluated the easy-to-use empirical SE estimator via simulation. The simulation results suggest that using this estimator is valid for inference, although slightly conservative. Extending this approach to multiple treatments is straightforward. The formulas applying to treatment $D_i = d$ can be used for multiple $d$ values. One only needs to obtain estimates for $P(D_i = d|\mathbf{Z}_i)$ by, eg, multinomial regression.[26]

Jiang et al[27] used an approach of a similar nature. They grouped subjects into multiple groups according to their confounding factors, then used weighted least squares to fit survival frequency data at multiple time points. Their approach is simpler than the proposed approach but is not DR and cannot be used for adjusting multiple continuous confounders unless it can be used for partitioning subjects into a small number of groups.

The pseudo-observation–based approach does have some drawbacks. When using model (12) as the OR model, by default, the pseudo package generates $\hat{S}_d^i(t)$ at all event times, which may require a large number of parameters in $\boldsymbol{\beta}_t$. Consequently, solving Equation 13 using function geese may take a long time. Alternatives to including a large number of parameters in the model are to fit smooth time functions such as the spline functions or to specify a small number of time points to generate $\hat{S}_d^i(t)$. Each method leads to some arbitrariness in the analysis. In this case, using the Cox model to fit the raw survival data is a more efficient. Also the DR estimate $\hat{S}_{dDR}(t)$ may not necessarily be a monotone function of $t$. In situations where this property is important, a monotone function may be fitted to $\hat{S}_{dDR}(t)$s to obtain a monotonic estimate for $S_d(t)$.

Not all software for generalized linear models can be used to solve GEE (13). Some software use algorithms which require the calculation of $\log(-\log(\hat{S}_d^i(t)))$, which is invalid when $\hat{S}_d^i(t)$ is not within (0,1). This fact should be considered when selecting software. For example, the R function glm() in the standard R package cannot be used in our approach due to this issue. For large dataset analyses or simulations, more computationally efficient software than geese are desirable.

## ACKNOWLEDGEMENT

The author would like to thank the editor and 2 referees for their helpful comments and suggestions and Dr C Di Casoli's suggestions and proof reading, which led to significant improvement to the manuscript.

## ORCID

*Jixian Wang* http://orcid.org/0000-0001-9963-6022

## REFERENCES

1. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457-81.
2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41-55.
3. Robins JM, Finkelstein D. Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56:779-788.
4. Hubbard AE, van der Laan MJ, Robins JM. Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies, U.C. Berkeley Division of Biostatistics Working Paper Series 68; 2001.
5. Cole SR, Hernn MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75:45-9.
6. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24:3089-110.
7. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect. *Math Modell*. 1986;7:1393-1512.
8. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc*. 1994;89:846-866.
9. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med*. 2004;23:2937-2960.
10. Kang JD, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). *Stat Sci*. 2008;22:523-580.
11. Funk MJ, Westreich D, Wiesen C, Strmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *Am J Epidemiol*. 2011;173:761-767.
12. Andersen PK, Hansen MG, Perme MP. Pseudoobservations in survival analysis. *Stat Methods Med Res*. 2010;19:71-99.
13. Andersen PK, Klein J, Rosthoj S. Generalised Linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*. 2003;90:15-27.
14. Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudoobservations. *Lifetime Data Anal*. 2004;10:335-350.
15. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005;61:223-229.
16. Andrei AC, Murray S. Regression models for the mean of quality-of-life adjusted restricted survival time using pseudo-observations. *Biometrics*. 2007;63:398-404.
17. Ambrogi F, Andersen PK. Predicting smooth survival curves through pseudo-values. Research report 14/1, Department of Biostatistics, University of Copenhagen; 2014.
18. Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. *Stat Med*. 2007;26:4505-4519.
19. Graw F, Gerds T, Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal*. 2009;15:241-255.
20. Jacobsen M, 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. https://doi.org/10.1111/sjos.12212.
21. Cox DR. Regression models and life tables. *J R Stat Soc Series B*. 1972;34:187-220.
22. Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press; 1997.
23. Zhang M, Schaubel DE. Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. *Biometrics*. 2012;68:999-1009.
24. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor positive breast carcinomas through genomic grade. *J Clin Oncol*. 2007;25:1239-1246.
25. Zhu R, Zhao YQ, Chen G, Ma S, Zhao H. *Greedy Outcome Weighted Tree Learning of Optimal Personalized Treatment Rules*; 2017;73:391-400. https://doi.org/10.1111/biom.12593.

26. McCulloch CE, Searle SR, Neuhaus JM. *Generalized, Linear, and Mixed Models*. 2nd ed. Hoboken: John Wiley & Sons; 2008.
27. Jiang H, Symanowski J, Qu Y, Ni X, Wang Y. Covariate-adjusted non-parametric survival curve estimation. *Stat Med*. 2011;30:1243-1253.

**How to cite this article:** Wang J. A simple, doubly robust, efficient estimator for survival functions using pseudo observations. *Pharmaceutical Statistics*. 2018;17:38–48. https://doi.org/10.1002/pst.1834