

MAIN PAPER

Estimating survival parameters under conditionally independent left truncation

Arjun Sondhi Flatiron Health, Inc, New York,
New York, USA**Correspondence**Arjun Sondhi, Flatiron Health, 233 Spring
St., New York, NY 10013, USA.Email: arjun@flatiron.com**Funding information**This study was sponsored by Flatiron
Health Inc., which is an independent
subsidiary of the Roche group.

[Correction added on 29 March 2022, after first online publication: The 'conditional probability of failure at time X_j ' in Section 2.3 and 'Z₁ confounding variable' in Section 3.1 have been corrected.]

Abstract

Databases derived from electronic health records (EHRs) are commonly subject to left truncation, a type of selection bias that occurs when patients need to survive long enough to satisfy certain entry criteria. Standard methods to adjust for left truncation bias rely on an assumption of marginal independence between entry and survival times, which may not always be satisfied in practice. In this work, we examine how a weaker assumption of conditional independence can result in unbiased estimation of common statistical parameters. In particular, we show the estimability of conditional parameters in a truncated dataset, and of marginal parameters that leverage reference data containing non-truncated data on confounders. The latter is complementary to observational causal inference methodology applied to real-world external comparators, which is a common use case for real-world databases. We implement our proposed methods in simulation studies, demonstrating unbiased estimation and valid statistical inference. We also illustrate estimation of a survival distribution under conditionally independent left truncation in a real-world clinico-genomic database.

KEYWORDS

left truncation, real world data, survival analysis

1 | INTRODUCTION

In time-to-event analyses, the outcome variable of interest is defined as the time from an initiating event to a terminal event, which is often subject to censoring. Time-to-event data may additionally be truncated, meaning that subjects whose time-to-event falls outside a certain interval cannot be observed.¹ Truncation is commonly found in healthcare data collected outside of clinical trials, such as databases derived from electronic health records (EHRs), where patients need to satisfy certain entry criteria in order to be observed.² Then, when analyzing survival, any patients who died before satisfying the entry criteria (e.g., undergoing a biomarker testing procedure) are not observed; this is known as left truncation, and is a selection bias since patients observed in the database had to live long enough to qualify for entry.^{3,4}

Given left truncated data, standard methods for estimating the marginal survival distribution are instead estimating survival in the target population conditional on surviving up to entry time. Applying risk set adjustment, where patients

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Author. *Pharmaceutical Statistics* published by John Wiley & Sons Ltd.

are only counted at risk for death once they have satisfied the entry criteria, is the most common approach to estimate the marginal distribution. This can be implemented with both Kaplan–Meier survival estimators and Cox proportional hazards regression models.⁵ However, risk set adjustment relies on the assumption of independence between the time to event T and the time to entry E (both measured from a common initiating event). This assumption is known as independent left truncation, and is testable from observed truncated data.^{6,7}

In practice, the left truncation mechanism is not always independent. Consider an EHR-derived database of patients with cancer who undergo comprehensive genomic profiling (CGP). This is known as real-world data (RWD), collected observationally outside of a clinical trial. If patients are tested later in their treatment course due to worsening of disease and exhaustion of standard therapies, this would result in dependency between the survival and entry time (i.e., time to CGP). Given this dependent left truncation, risk set adjustment is no longer guaranteed to unbiasedly recover the survival distribution of interest. Although some proposed methods claim to estimate marginal survival in the presence of dependent left truncation, they make strong parametric assumptions that are not testable from observed data. They also generally do not extend to multivariate regression modeling.⁸

In this work, we considered the setting of conditionally independent left truncation, where survival time T and entry time E are independent conditional on a set of confounding variables Z . This is weaker than the marginally independent left truncation assumption described above, and can also be easily tested. Intuitively, this is similar to missing at random (MAR) data, where the probability of an observation being missing is independent of its true value, conditional on a set of observed variables.

Likewise, data where an observation's true value directly affects its missingness probability (missing not at random or nonignorable missingness) is analogous to dependently truncated data.⁹ We showed that under conditional independence of T and E given Z , unbiased estimates of certain survival parameters can be obtained. Our methods were motivated by common analyses performed with RWD from EHR-derived sources.

We first considered estimation of survival regression parameters, by observing that the conditional survival distribution $T|Z$ in the target population was estimable. Then, by adjusting for confounders Z , a Cox model's regression parameters can be estimated and interpreted as the conditional log hazard ratios in the correct target population. In order to extend this to estimating parameters of the marginal survival distribution T , we required additional reference data on the marginal confounder distribution Z , for example, non-truncated observations from the target population. Note that the existence of a non-truncated dataset would not remove the need to conduct estimation in the truncated dataset if it was of interest to analyze non-truncated data (such as from a clinical trial) and truncated data together, for example, to compare their health outcomes. We considered the setting of comparing an external cohort from RWD to a single-arm treated cohort from a clinical trial. The latter group may be treated with novel therapy, and the former with standard-of-care treatments. By weighing the external cohort to match the confounder distribution in the trial cohort, we showed that the causal hazard ratio comparing the treatment to the control can be estimated. This weighting procedure corresponds to that used in estimation of the average treatment effect on the treated (ATT). Using a similar method, we can also estimate the marginal survival distribution in a dependently truncated cohort by weighting toward a reference dataset with confounder variables observed. This requires the additional assumption that the reference distribution corresponds to the target population of interest, which is true by definition for the ATT.

The rest of this paper is organized as follows. We first detail our methodology for estimating survival parameters under dependent left truncation. We then demonstrate in simulation studies that our methods can consistently estimate the parameters of interest. Finally, we illustrate estimation of a marginal survival distribution in an analysis of data from a real world EHR-derived database, and conclude with a short discussion.

2 | METHODS FOR CONDITIONALLY INDEPENDENT LEFT TRUNCATION

2.1 | Background

As defined previously, let T denote the survival time, E denote the entry time, and C denote the censoring time from a common initiating event. Then, let $Y = \min(T, C)$ denote the observed time-to-event, and $\delta = I(T \leq C)$ indicate whether a death event was observed. We assumed throughout that T and C are independent. The observed data then consists of $(Y_i, \delta_i, E_i) | Y_i > E_i, i = 1, \dots, n$. With these variables observed, it is only possible to determine whether T and E are marginally independent or dependent. Under independent left truncation, applying risk set adjustment to a Kaplan–Meier analysis can recover the true marginal distribution of T .

Now suppose a set of confounding variables Z are also observed at baseline, such that T is independent of E conditional on Z . Throughout this paper, we defined ‘baseline’ to be the time of the initiating event or start of follow-up from which the times T and E are measured. For example, if Z is a single categorical variable (e.g., sex), then among patients with a particular $Z = z$, survival and entry times are independent. As such, applying risk set adjustment in a single strata z would consistently estimate the distribution $T|Z = z$, without the truncation selection bias $Y > E$. More generally, under conditionally independent left truncation, the distribution $T|Z$ is estimable from data. We exploited this property in order to estimate various survival parameters in the true target population of interest.

As with marginally independent left truncation, it is possible to test for conditionally independent left truncation. This can be done by fitting a risk set adjusted Cox proportional hazards model, with entry time E and confounders Z as covariates. If the coefficient for entry time is significantly different from the null value, this provides evidence that the survival and entry times are dependent, conditional on confounders.

2.2 | Estimating conditional hazard ratios

Suppose we are interested in estimating the covariate-adjusted effect of a binary treatment, denoted by the variable trt , on survival. Causal graphs corresponding to different possible left truncation scenarios are shown in Figure 1.

Given conditionally independent left truncation, we can fit the Cox proportional hazards model:

$$\lambda_T(t, trt|Z) = \lambda_{bh}(t) \exp(\beta trt + \beta'_z Z),$$

where $\lambda_{bh}(t)$ is an unspecified baseline hazard function. Then, the parameter β corresponds to:

$$\exp(\beta) = \frac{\lambda_T(t|trt = 1, Z)}{\lambda_T(t|trt = 0, Z)}$$

which is a ratio of two functionals of the conditional distribution $T|Z$. Because this distribution can be estimated from our observed data, we can also estimate the log hazard ratio β unbiasedly. For simplicity in Figure 1, trt and Z are shown as independent. However, estimability would still hold given an edge from Z to trt , because E and T would remain conditionally independent in that scenario. This also extends to regression parameters of any other covariates X that may be included in the model, as long as conditioning on them would not induce an association between entry and survival times. Practically, we expect most variables of interest that are adjusted for would be collected at baseline (at the time of the initiating event, or start of follow-up), and therefore are unlikely to result in this association.

2.3 | Estimating marginal hazard ratios and survival distributions

We have demonstrated how conditionally independent left truncation can yield unbiased estimation of certain conditional hazard ratio parameters using the conditional distribution $T|Z$. In practice, however, marginal parameters that describe population-averaged effects are often of primary interest. In order to estimate these, we require the marginal survival distribution T . We can express the density of T as:

$$\pi_T(t) = \int_z \pi_{T|Z}(t|z)\pi_Z(z)dz$$

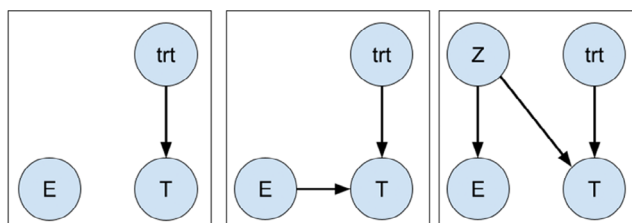


FIGURE 1 Graphical illustrations of independent (left), dependent (center), and conditionally independent (right) left truncation

which provides insight into how to recover the marginal distribution. The first component $T|Z$ is estimable from observed data; however, in our truncated dataset, we only observe $Z|Y > E$, rather than the marginal distribution Z . Now suppose we are able to obtain weights $w = \frac{\pi(z)}{\pi(z|y > e)}$, corresponding to the density ratio between Z and $Z|Y > E$. We can then weigh our observed data distribution to match the true marginal distribution:

$$\begin{aligned}\pi(y, e, z | y > e)w &= \pi(y, e | z, y > e)\pi(z|y > e)w \\ &= \pi(y, e | z, y > e)\pi(z) \\ &= \pi(y | z, y > e)\pi(e | z, y > e)\pi(z) \\ &= \pi(y|z)\pi(e|z)\pi(z) \\ &= \pi(y, e | z)\pi(z) \\ &= \pi(y, e, z)\end{aligned}$$

In the above decomposition, the third and fourth equalities follow as a consequence of conditionally independent left truncation. Intuitively, we want to weigh the distribution of Z in our sample to make it similar to the correct target population distribution.

In order to estimate these weights, we require a reference sample of Z drawn from a population that is non-truncated, and therefore representative. Then, the density ratio can be estimated through the following procedure:

1. Vertically concatenate the non-truncated (Z) and truncated ($Z|Y > E$) confounder observations.
2. Label each observation $J = 1$ if it is from the non-truncated dataset or $J = 0$ if it is from the truncated dataset.
3. Train a probabilistic classifier that estimates $P(J = 1|Z)$.
4. For a patient in the truncated dataset having confounders $Z = z$, the estimated weight is then

$$\hat{w}_i = \frac{\hat{P}(J = 1|Z = z)\hat{P}(J = 0)}{\hat{P}(J = 0|Z = z)\hat{P}(J = 1)}$$

which targets the density ratio:

$$\frac{\pi(z)}{\pi(z|y > e)}$$

This result follows from a simple application of Bayes rule.^{10,11} In practice, software such as the WeightIt package in R¹² can compute the term

$$\frac{\hat{P}(J = 1|Z = z)}{\hat{P}(J = 0|Z = z)}$$

by estimating balancing weights targeting the average treatment effect on the treated (ATT). Here, patients in the non-truncated sample have weights set to 1.

Although access to a reference sample may seem like a strong condition for estimation of marginal parameters, this naturally occurs in analyses where truncated data is analyzed together with a non-truncated data source. An example of this is comparing survival of patients in a single-arm clinical trial to that of real-world patients as an external comparator arm.^{13–15} This may be done to demonstrate treatment effectiveness, or to contextualize early-phase results for rare diseases or biomarkers when enrolling a trial control arm can be difficult. Here, the trial arm is not subject to left truncation, while the real-world arm potentially is. Given conditionally independent left truncation, we can weigh the RWD toward the confounder distribution in the trial arm as described above. Then, fitting a weighted and risk set adjusted Cox proportional hazards model with the arm indicator as the sole covariate will estimate a marginal hazard ratio comparing the trial treatment and real-world cohorts. The corresponding log partial likelihood is:

$$\ell(\beta) = \sum_{i:\delta_i=1} \widehat{w}_i \left[\beta \text{trt}_i - \log \left(\sum_{j \in R_i} \exp(\beta \text{trt}_j) \right) \right]$$

where \widehat{w}_i is the estimated weight for patient i , and R_i denotes the set of patients in the risk set at the observed event time T_i . In fact, this is the same procedure used to estimate the ATT hazard ratio (which requires the causal assumption that Z contains sufficient information to satisfy treatment assignment ignorability). In other words, we have shown that standard causal inference methodology can provide unbiased estimation when applied to left truncated data, given conditional independence of survival and entry times.

For other analyses, it may be of interest to estimate the marginal survival distribution in a RWD set subject to left truncation, without making any statistical comparison. Under conditionally independent left truncation, this can be done in a similar manner by estimating weights given a reference dataset, as above. This requires an additional assumption that the reference confounder distribution is representative of the target population of interest, which is definitionally true when estimating the ATT.

The weights can then be used by fitting a weighted and risk set adjusted Kaplan–Meier estimator for the survival distribution. Here, the conditional probability of failure at time X_j is estimated as:

$$\widehat{F}(x_j) = \frac{\sum_{i=1}^n I(E_i \leq x_j, Y_i = x_j) \delta_i \widehat{w}_i}{\sum_{i=1}^n I(E_i \leq x_j \leq Y_i) \widehat{w}_i}$$

where \widehat{w}_i is the estimated weight for subject i . A consistency proof of this estimator is sketched in the supplement.

Statistical inference for the ATT estimated by the weighted Cox proportional hazards model should be done using robust standard errors (which are fit by default when the `coxph` function in R is given weights), since a weighted analysis induces dependence among observations. For the same reason, inference for the marginal survival distribution estimated by the weighted Kaplan–Meier estimator can be done by bootstrapping from the weighted empirical distribution.

3 | SIMULATION STUDIES

3.1 | Design

We simulated a non-truncated treatment arm (such as from a single-arm clinical trial) and a real-world control arm that was subject to conditionally independent left truncation. First, we generated two confounding variables as:

$$Z_1 \sim \text{Bernoulli}(1 - p_E)$$

$$Z_2 \sim \text{Normal}(0, 0.5),$$

where p_E was the probability of entering the cohort at or before the start of follow-up (i.e., without delayed entry) in the real-world arm. We set this parameter to be 0.2. Then, cohort entry times for the real world patients were generated as:

$$E \sim \begin{cases} 0 & Z_1 = 0 \\ \text{Exponential}(\lambda_{\text{entry}}) & Z_1 = 1 \end{cases}$$

$$\lambda_{\text{entry}} = \lambda_{\text{ebh}} \exp(\beta_{\text{entry}} Z_2)$$

indicating that patients with $Z_1 = 0$ did not have delayed entry (i.e., an entry time of 0). The constant parameter λ_{ebh} was set via root-finding in order to achieve a certain truncation probability, given all the other parameters. We varied the truncation probability $P(Y > E | \text{trt} = 0)$ from 0.1 to 0.7 by increments of 0.1. The parameter β_{entry} was varied among

$\log(0.5, 0.8, 0.1)$; for $\beta_{\text{entry}} \neq 0$, patients with higher values of Z_2 were more likely to enter the cohort later. The cohort entry times for the non-truncated patients were all 0.

We used the following model for survival times among two randomly assigned treatment groups:

$$\text{trt} \sim \text{Bernoulli}(0.5)$$

$$T \sim \text{Exponential}(\lambda_T)$$

$$\lambda_T = \lambda_{bh} \exp(\beta \text{trt} + \beta_Z Z_1 + \beta_Z Z_2),$$

where the parameters were set as follows:

- We set λ_{bh} to be $1/12$, which corresponded to an average survival time of 12 months for patients on the $\text{trt} = 0$ arm with confounders $Z_1 = Z_2 = 0$.
- The parameter β was set to $\log(0.8)$, which implied that patients on the $\text{trt} = 1$ arm have 80% of the hazard of death as do the patients on the $\text{trt} = 0$ arm with the same confounder values.
- β_Z was the parameter describing the association between survival and both confounders. We varied this parameter among $\log(1, 1.5, 2)$; for $\beta_Z \neq 0$, patients with higher values of Z_1 and Z_2 were more likely to die sooner.

This design considered a positive association between the confounders Z and the hazard of death combined with a negative association between Z and the hazard of entry. This resulted in a spurious marginal association of decreased survival with later time to entry, which was commonly observed in our motivating application of real-world clinico-genomic data; patients who are genomic tested later in their disease course tend to have worse survival.

Finally, the censoring model used was

$$C \sim \text{Exponential}(\lambda_C)$$

$$\lambda_C = \lambda_T$$

corresponding to a censoring probability of 0.5 in the complete dataset (before truncation is applied). Given this data-generating process, the observed time-to-event was $Y = \min(T, C)$. Truncation was applied by filtering out observations where $Y < E$ in the real-world $\text{trt} = 0$ arm.

We fixed the expected real-world arm sample size at 250 by generating a larger dataset with sample size

$$\frac{250}{P(Y > E | \text{trt} = 0)}$$

and applying the truncation. The non-truncated treatment arm had a fixed sample size of 250. For each parameter configuration, we conducted 1000 simulation iterations.

3.2 | Conditional parameters

Using the simulation framework described above, we demonstrated how the conditional hazard ratio comparing treatment arms can be estimated unbiasedly under conditionally independent left truncation. In each simulation iteration, we estimated this parameter on the complete dataset without truncation applied by fitting a Cox proportional hazards model adjusting for the treatment arm and the confounders; this estimate was considered the ground truth. We then fitted the same model on the dataset where the real-world arm was left truncated, with and without risk set adjustment. To evaluate the performance of these estimators, we reported the relative bias of the hazard ratio estimate with respect to the true hazard ratio (Figure 2), and coverage of the 95% confidence intervals (Figure 3). We saw that the risk set adjusted Cox estimator was able to provide unbiased estimates and valid coverage, despite the marginal dependence between entry time and survival that was induced through

the confounding variables. As expected, the naive model fit without adjusting for delayed entry results in severe bias. In general, this was unaffected by the relationship between survival and the other variables, as shown by similar results for all parameter configurations.

3.3 | Marginal parameters

We then demonstrated our proposed methodology for estimating the marginal hazard ratio to compare the non-truncated and truncated arms. In each simulation iteration, we estimated this parameter on the complete dataset without truncation applied. This estimate was treated as the ground truth, and compared to those produced under the following methods applied to the truncated dataset:

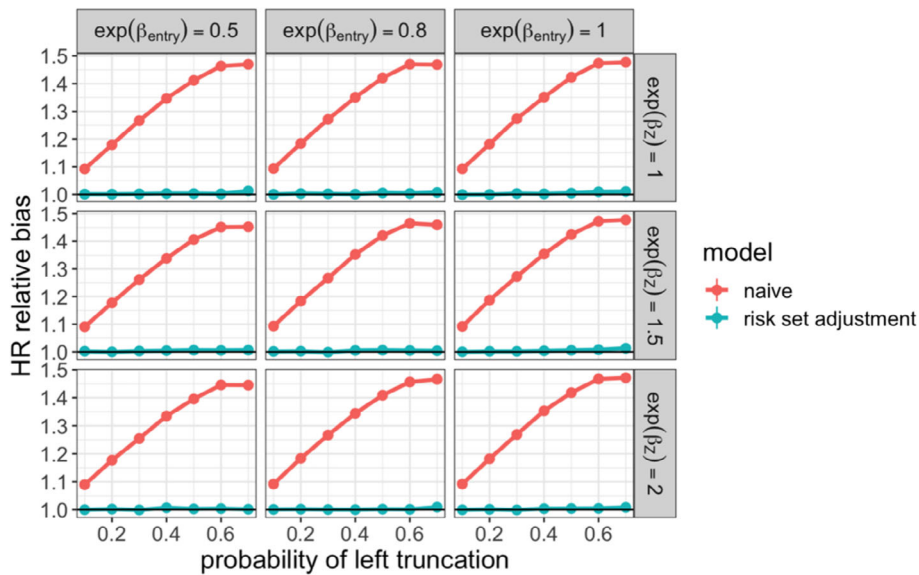


FIGURE 2 Relative bias for estimated conditional hazard ratio comparing non-truncated arm (ground truth fit) to real-world treatment arm across simulation settings

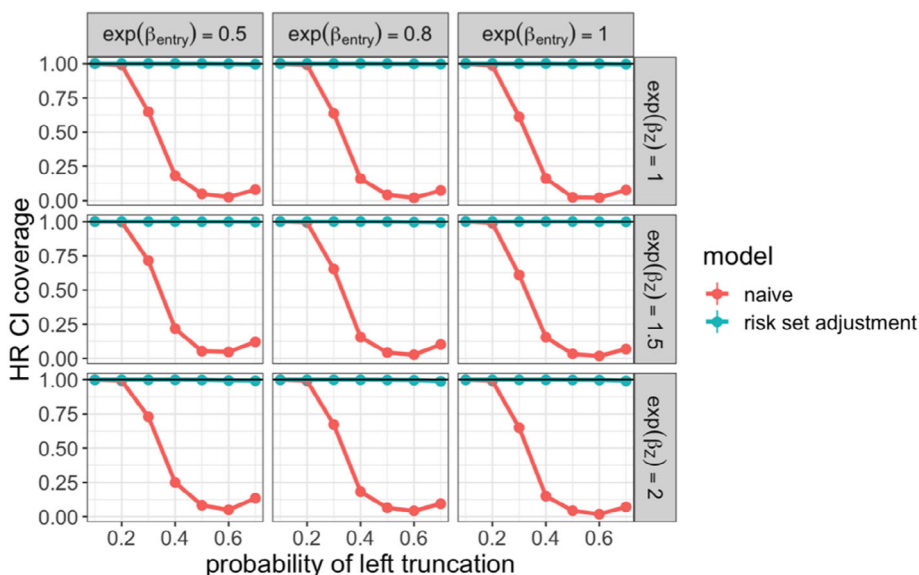


FIGURE 3 95% confidence interval coverage for estimated conditional hazard ratio comparing non-truncated arm (ground truth fit) to real-world treatment arm across simulation settings

- **naive:** Cox model estimator that did not account for delayed entry
- **risk set adjustment:** Cox model estimator that did account for delayed entry
- **weighted Cox:** Risk set adjusted Cox model estimator with real-world arm weighted toward confounder distribution in non-truncated arm

To evaluate the performance of these estimators, we examined the relative bias of the hazard ratio estimate with respect to the true hazard ratio. We also examined the coverage of the associated 95% confidence intervals. The results are shown in Figures 4 and 5. We saw that the weighted Cox model estimator had low bias and maintained valid coverage at all levels of truncation. When $\beta_Z = 0$, both the risk set adjusted and weighted Cox models were unbiased, since Z_1 and Z_2 did not affect survival, and were therefore not confounders. As the strength of confounding increased, the risk set adjusted Cox estimator showed increasing bias and lost valid coverage under high-truncation probability.

We also examined estimation of the marginal survival distribution in the real-world arm. In each simulation iteration, we estimated the median survival time on the complete real-world cohort. This estimate was treated as the ground truth, and compared to those produced under the following methods applied to the truncated cohort:

- **naive:** Kaplan–Meier estimator that did not account for delayed entry
- **risk set adjustment:** Kaplan–Meier estimator that did account for delayed entry
- **weighted Cox:** Risk set adjusted Kaplan–Meier estimator with real-world arm weighted toward confounder distribution in non-truncated arm

To evaluate the performance of these estimators, we examined the relative bias of the median survival time estimate, shown in Figure 6. As with the hazard ratio, we saw that the weighted estimator had low bias at all levels of truncation, while the unweighted risk set adjusted estimator was biased in the presence of confounding.

4 | APPLICATION: OVERALL SURVIVAL IN NGS-TESTED PATIENTS WITH METASTATIC PROSTATE CANCER

We applied the weighting method described to estimate the survival distribution of patients with metastatic prostate cancer who received a Foundation Medicine next generation sequencing (NGS) test, taken from the nationwide (US-based) de-identified Flatiron Health-Foundation Medicine Clinico-Genomic Database (CGDB). The de-identified data originated from approximately 280 US cancer clinics (representing around 800 sites of care). Retrospective longitudinal

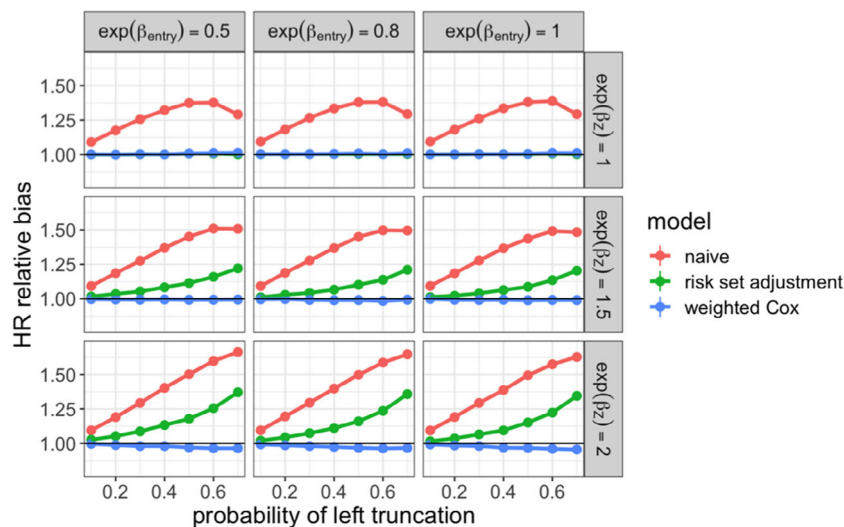


FIGURE 4 Relative bias for estimated marginal hazard ratio comparing non-truncated arm to real-world treatment arm across simulation settings

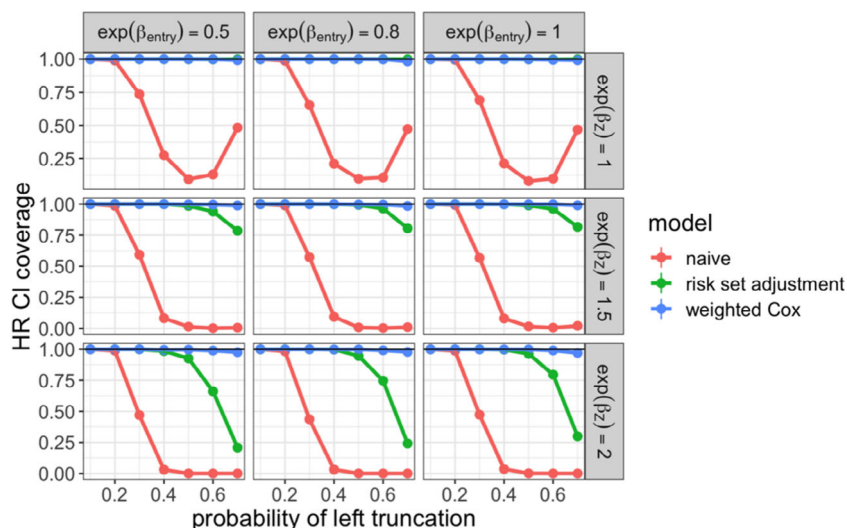


FIGURE 5 95% confidence interval coverage for estimated marginal hazard ratio comparing non-truncated arm to real-world treatment arm across simulation settings

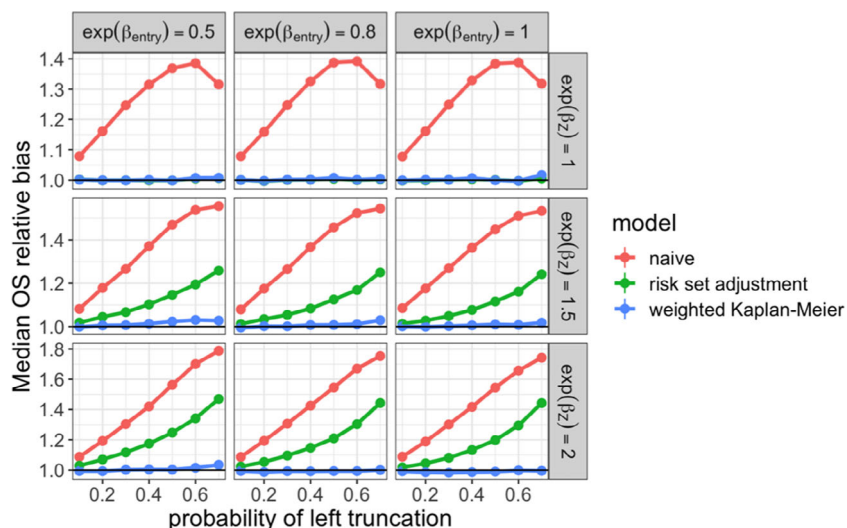


FIGURE 6 Relative bias for estimated median survival time in real-world treatment arm across simulation settings

clinical data were derived from EHR data, comprising patient-level structured and unstructured data, curated via technology-enabled abstraction, and were linked to genomic data derived from FMI CGP tests in the FH-FMI CGDB by de-identified, deterministic matching.¹⁶ Genomic alterations were identified via CGP of over 300 cancer-related genes on Foundation Medicine’s next-generation sequencing (NGS) test.¹⁷

We were specifically interested in estimating overall survival time from the start of first-line (1L) therapy in the metastatic setting. The NGS test report date was considered the cohort entry date, so a patient who was tested after their 1L start date would have delayed entry. Patients who died before being tested were left truncated from the cohort. The observed event time for each patient was either their date of death or date of last follow-up activity (at which time the patient was censored). Patients with a gap of greater than 90 days between their metastatic diagnosis date and their first subsequent structured EHR activity were excluded from the analysis. This resulted in a final cohort size of 1256.

For our initial analysis, we fitted a risk set adjusted Kaplan–Meier estimator to the observed entry, death, and censoring times. This yielded a median overall survival time of 17.8 months post-1L therapy start, with a 95% confidence interval of [16.5, 19.3]. We then tested for marginally dependent left truncation. This was done by fitting a risk set adjusted Cox proportional hazards model to estimate the association between survival time and time to NGS testing

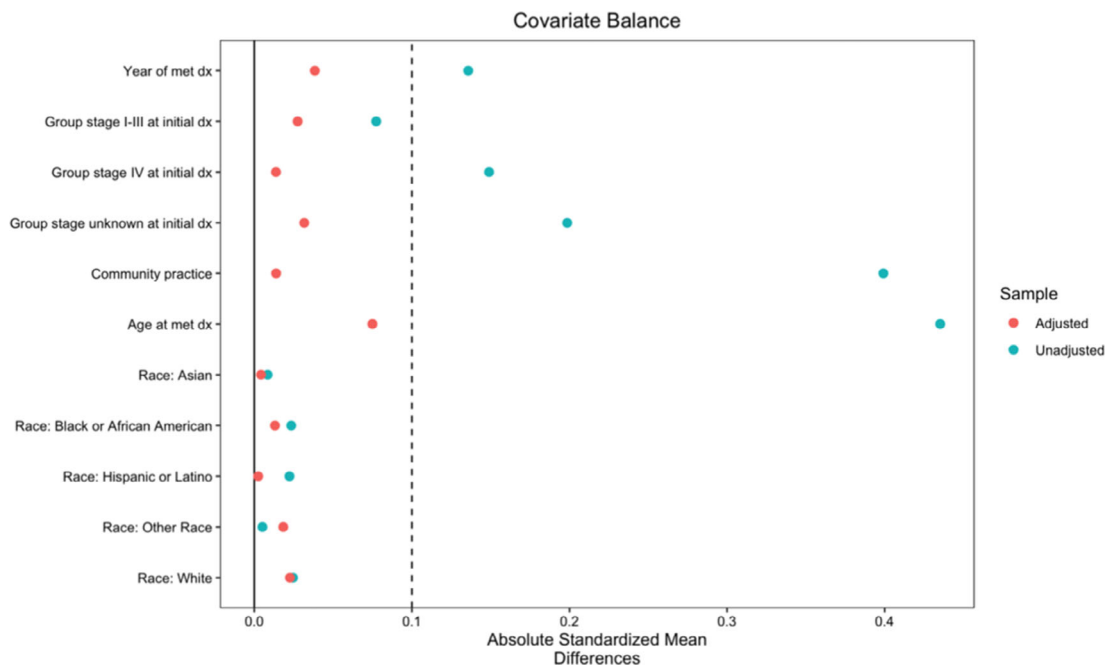


FIGURE 7 Covariate balance plot comparing FH and CGDB cohorts, displaying both weighted and unweighted absolute standardized mean differences. A threshold of 0.1 is used to indicate balance

from 1L therapy start. We estimated a hazard ratio of 1.01 with 95% confidence interval [1.0, 1.02] (p -value = 0.0075). This indicated a small degree of dependent left truncation; patients were estimated to have a 1% greater hazard of death for each additional later month from 1L start to NGS testing. Given this dependence, the median survival estimate given above may be inaccurate, though the bias would likely be low, since the magnitude of the association was small.

Next, we looked to see if this dependence can be explained by observed baseline covariates; these were selected a priori from among the variables measured in both databases. We fitted a risk set adjusted Cox model of overall survival on entry time that controlled for year of metastatic diagnosis, clinic practice type (academic vs. community), patient's age at metastatic diagnosis, patient's race, tumor histology, and group stage at diagnosis. We estimated a conditional hazard ratio of 1.00 with 95% confidence interval [0.99, 1.01] (p -value = 0.17), providing evidence that left truncation was independent conditional on these covariates. Therefore, weighting to the appropriate distribution of these variables would recover the correct survival distribution.

We obtained a reference distribution of the required variables from the nationwide de-identified, EHR-derived Flatiron Health (FH) metastatic prostate cancer database. Unlike in the CGDB, an NGS test is not required for a patient to qualify for inclusion in this database. Therefore, the distribution of observed variables would not be truncated by test time. We applied the same inclusion–exclusion criteria to this database, and extracted the adjustment variables from the resulting cohort. Then, we vertically concatenated the confounder datasets, labeling the FH database (non-truncated) observations as 1 and the CGDB (truncated) observations as 0. The density ratio weights were computed using the WeightIt R package, specifying inverse propensity weights targeting the ATT estimand, with the ‘treatment’ being the truncated versus non-truncated cohort label. The weights for the CGDB patients were also multiplied by the sample size adjustment factor $\frac{\hat{P}(J=0)}{\hat{P}(J=1)}$.

The weights for the FH patients were all 1. The weighted distribution of confounders in the CGDB dataset was similar to the distribution in the FH dataset, as visualized in Figure 7.

Finally, we fitted a weighted and risk set adjusted Kaplan–Meier estimator to the CGDB dataset, using the computed density ratio weights. This resulted in an estimated survival time of 16.5 months post-1L therapy start. Using the weighted nonparametric bootstrap with 1000 resamples, we computed a 95% confidence interval of [15.5, 17.8]. This suggested that the dependent left truncation led to a slightly upwardly biased estimate of survival in these patients, assuming that the non-truncated cohort would have the baseline confounder distribution as observed in the FH cohort. Under this assumption, we interpreted this estimate as the median survival time for the population of FH metastatic prostate cancer patients who would receive Foundation Medicine NGS testing.

5 | DISCUSSION

Real world EHR-derived datasets may be subject to left truncation, which can present complications for making statistical inferences on populations of scientific interest. Applying risk set adjustment in survival analyses is a straightforward solution, but requires independence between survival and cohort entry times. In the literature, testing for this condition is generally restricted to marginal dependence. If this is found to hold, then practitioners are faced with a lack of commonly accepted methods to analyze dependently truncated data.

In this work, we extended the conditions for valid inference beyond marginally independent left truncation by considering conditional independence. This occurs when confounding variables induce a spurious relationship between entry and survival times, and can be tested similarly as with marginal independence. Under conditionally independent left truncation, we showed that certain parameters that are commonly of interest in real-world evidence can be estimated unbiasedly. In particular, the methods we described are complementary to analyses that compare real-world cohorts to non-truncated trial cohorts, which we treated as reference data for learning the true confounder distribution. We demonstrated the effectiveness of our methods in simulation studies, and illustrated potential use by analyzing survival in a real-world left truncated cohort. More broadly, these methods are applicable whenever the true distribution of confounders in the target population is known or estimable; as long as the truncated data can be appropriately weighted, the decomposition in Section 2.3 allows estimation of marginal parameters.

A limitation of this methodology is that it relies on the existence of baseline confounders yielding conditionally independent left truncation. In practice, these may be difficult to obtain. Using the CGDB as an example, it may be the case that NGS testing is usually performed for patients who have a worsening prognosis or cancer progression. These factors would not be measured at the start of a therapy, but would affect both survival and test timing. This type of treatment journey is particularly plausible for disease settings where NGS testing is not standard-of-care. An interesting future direction for this work would be to appropriately incorporate post-baseline or time-varying confounders. It would also be of interest to quantify how much bias remains when conditioning with a set of variables that reduces the dependence between survival and entry without fully eliminating it.

In addition to the correct set of confounders, valid estimation and inference for marginal parameters also requires the weights to be well-estimated. To this end, standard diagnostics from the observational causal inference literature can be applied to assess the balancing quality of the weights. Moreover, when applying classifier-based density ratio estimation as we suggest in Section 2.3, cross-validation can also help to select an appropriate model class to train, since minimizing classifier loss directly corresponds to optimizing balance.¹¹

Despite these limitations, our work relaxed the necessary assumptions for valid inference under left truncation, allowing for a broader range of analyses to be conducted. We showed that conditionally independent left truncation can be easily tested for, and result in unbiased estimates from common survival analyses. In practice, we recommend that researchers adopt this methodology when analyzing RWD subject to left truncation.

ACKNOWLEDGMENTS

The author would like to acknowledge and thank Olivier Humblet, Emily Castellanos, Katherine Tan, and Sandy Griffith for review and critical input of this manuscript, and Julia Saiz-Shimosato (Flatiron Health, Inc.) and Hannah Gilham (Flatiron Health, Inc.) for publication management and writing support.

CONFLICT OF INTEREST

AS reports employment in Flatiron Health, which is an independent subsidiary of the Roche group and stock ownership in Roche.

AUTHOR CONTRIBUTIONS

Design and concept: Arjun Sondhi. *Data collection:* Flatiron Health. *Analysis and interpretation:* Arjun Sondhi. *Manuscript writing, review, and approval:* Arjun Sondhi.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study have been originated by Flatiron Health, Inc. and Foundation Medicine, Inc. These de-identified data may be made available upon request, and are subject to a license agreement with Flatiron

Health and Foundation Medicine; interested researchers should contact <cgdb-fmi@flatiron.com> and <dataaccess@flatiron.com> to determine licensing terms.

ORCID

Arjun Sondhi  <https://orcid.org/0000-0002-7997-0477>

REFERENCES

1. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. Vol 1230. Springer; 2003.
2. Agarwala V, Khozin S, Singal G, et al. Real-world evidence in support of precision medicine: clinico-genomic cancer data as a case study. *Health Aff*. 2018;37(5):765-772.
3. Chubak J, Boudreau DM, Wirtz HS, McKnight B, Weiss NS. Threats to validity of nonrandomized studies of postdiagnosis exposures on cancer recurrence and survival. *J Natl Cancer Inst*. 2013;105(19):1456-1462.
4. Cain KC, Harlow SD, Little RJ, et al. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *Am J Epidemiol*. 2011;173(9):1078-1084.
5. Tsai WY, Jewell NP, Wang MC. A note on the product-limit estimator under right censoring and left truncation. *Biometrika*. 1987;74(4):883-886.
6. Mackenzie T. Survival curve estimation with dependent left truncated data using Cox's model. *Int J Biostat*. 2012;8(1):1-18.
7. Martin EC, Betensky RA. Testing quasi-independence of failure and truncation times via conditional Kendall's tau. *J Am Stat Assoc*. 2005;100(470):484-492.
8. Chaieb LL, Rivest LP, Abdous B. Estimating survival under a dependent truncation. *Biometrika*. 2006;93(3):655-669.
9. Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? *Int J Epidemiol*. 2014; 43(4):1336-1339.
10. Sugiyama M, Suzuki T, Kanamori T. Density-ratio matching under the Bregman divergence: a unified framework of density ratio estimation. *Ann Inst Stat Math*. 2012;64(5):1009-1044.
11. Sondhi A, Arbour D, Dimmery D. Balanced off-policy evaluation in general action spaces. *PMLR*. 2020;108:2413-2423.
12. Greifer N. *Package 'WeightIt'*. CRAN; 2019.
13. Carrigan G, Whipple S, Taylor MD, et al. An evaluation of the impact of missing deaths on overall survival analyses of advanced non-small cell lung cancer patients conducted in an electronic health records database. *Pharmacoepidemiol Drug Saf*. 2019;28(5):572-581.
14. Davies J, Martinec M, Delmar P, et al. Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. *J Compar Effect Res*. 2018;7(09):855-865.
15. Carrigan G, Whipple S, Capra WB, et al. Using electronic health records to derive control arms for early phase single-arm lung cancer trials: proof-of-concept in randomized controlled trials. *Clin Pharmacol Ther*. 2020;107(2):369-377.
16. Singal G, Miller PG, Agarwala V, et al. Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database. *JAMA*. 2019;321(14):1391-1399.
17. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*. 2013;31(11):1023-1031.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Sondhi A. Estimating survival parameters under conditionally independent left truncation. *Pharmaceutical Statistics*. 2022;21(5):895-906. doi:10.1002/pst.2202