Biometrika (2026), **113**, 3, asag036 https://doi.org/10.1093/biomet/asag036
Advance Access publication 11 June 2026

Doubly robust and efficient calibration of prediction sets for right-censored time-to-event outcomes

By REBECCA FARINA

Department of Statistics and Data Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.
rfarina@andrew.cmu.edu

ERIC TCHETGEN TCHETGEN

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, 265 South 37th Street, Philadelphia, Pennsylvania 19104, U.S.A.
ett@wharton.upenn.edu

and ARUN KUMAR KUCHIBHOTLA

Department of Statistics and Data Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.
arunku@stat.cmu.edu

Summary

Our objective is to construct well-calibrated prediction sets for a time-to-event outcome subject to right censoring with guaranteed coverage. Inspired by modern conformal inference, our approach avoids the need for a well-specified parametric or semiparametric survival model. Unlike existing conformal methods for survival data, which assume Type-I censoring with fully observed censoring times, we consider the more common right-censoring setting in which only the censoring time or the event time is observed, whichever comes first. Under a standard conditional independence censoring condition, we propose and analyse several lower prediction bounds for the survival time of a future observation, including inverse-probability-of-censoring weighting and its augmented version based on the semiparametric efficient influence function for the relevant marginal quantile of the outcome, accounting for dependent censoring. We formally establish asymptotic coverage guarantees for the proposed methods and demonstrate, both theoretically and through empirical experiments, that the augmented approach substantially improves efficiency over all other proposed methods. Specifically, its coverage error bound is doubly robust, and therefore of second order, thus ensuring that it is asymptotically negligible relative to the coverage error of the other methods.

Some key words: Censored data; Conformal prediction; Double robustness; Inverse-probability-of-censoring weighting; Prediction set; Semiparametric efficiency theory; Time-to-event outcome.

© 2026 Biometrika Trust
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

1. INTRODUCTION

1.1. *Background*

A fundamental challenge of survival analysis is the potential for a person's event time to not be completely observed if she drops out of the study or if the study ends before she experiences the event of primary interest, a phenomenon commonly known as *right censoring* (see, e.g., Kleinbaum & Klein, 1996; Hosmer Jr et al., 2008; Collett, 2023). While classical tools such as the Kaplan–Meier estimator of the marginal survival curve (Kaplan & Meier, 1958) and the Cox proportional-hazards regression model (Cox, 1972) are well established, there remains a lack of flexible frameworks that provide rigorous uncertainty quantification for predictive inference of right-censored time-to-event outcomes.

Conformal predictive inference, introduced by Vovk et al. (2005), provides a general blueprint for constructing well-calibrated prediction intervals, with marginal coverage guarantees under relatively mild conditions. The main advantage of conformal prediction is that it provides valid finite-sample prediction intervals for a new observation exchangeable with the training sample, without imposing strong distributional assumptions on the underlying data-generating mechanism. Conformal prediction methods have developed rapidly for a wide variety of settings that depart from the canonical independent and identically distributed setting, including covariate shift (Tibshirani et al., 2019; Yang et al., 2024), missing covariates (Zaffran et al., 2024), time series analysis (Stankeviciute et al., 2021) and high-dimensional regression (Chen et al., 2016), to name a few. However, conformal predictive inference for censored time-to-event outcomes remains underdeveloped.

In this paper, we propose several new methods to construct well-calibrated and efficient prediction bounds for survival times under standard right censoring. Notably, to calibrate our prediction intervals, we propose two separate approaches based on inverse probability of censoring (Robins & Rotnitzky, 1992; Robins, 1993; Robins & Finkelstein, 2000). The first approach (a) constructs a consistent inverse-probability-of-censoring weighting (IPCW) estimator of the marginal probability that the failure times exceed an estimated covariate-dependent lower bound, while the second approach (b) is based on an augmented-inverse-probability-of-censoring weighting (AIPCW) estimator for the same marginal probability. Although IPCW and AIPCW methods are well developed for estimation and inference in standard survival analysis settings, (a) and (b) appear to provide their first use for obtaining well-calibrated prediction intervals. In addition, we consider two so-called outcome regression-based methods that obviate the need for estimating the censoring mechanism: the outcome regression method, which entails estimating the conditional quantile function for the outcome at the desired nominal level, and the calibrated outcome regression method, which adjusts the quantile level to match the desired nominal marginal coverage.

1.2. *Problem set-up*

Let $X \in \mathbb{R}^d$ denote a unit's vector of covariates, $T \in \mathbb{R}_+$ the unit's survival time and $C \in \mathbb{R}_+$ the unit's potential censoring time. We denote by $P_{(X,T)}$ the joint distribution of (X, T), and by P_X, P_T, $P_{T|X}$, etc., the corresponding marginal and conditional distributions. While X is fully observed, the survival time T is right censored in the sense that only T or C is observed, whichever comes first, that is, $Y = \min\{T, C\}$ is observed, together with the censoring indicator $\Delta = \mathbf{1}\{T \leqslant C\}$. Thus, the fully observed data are given by the independent and identically distributed sample $\mathcal{D} = (X_i, Y_i, \Delta_i)_{i=1}^n$. We denote the underlying (unobserved) full data by $\mathcal{Z} = (X_i, T_i)_{i=1}^n$.

We now introduce additional notation used throughout. Given a generic function f, whenever necessary, we denote by $f^\star$ the true unknown function that generated the sample and by $\hat{f}$ a corresponding estimator obtained from the observed data. Let $S_{T|X}(\cdot \mid X)$ denote the survival function of the time-to-event outcome T, given X, i.e., $S_{T|X}(t \mid X) = \mathbb{P}(T > t \mid X)$, and $S_{C|X}(\cdot \mid X)$ the survival function of C, given X. For any $\gamma \in (0,1)$, the conditional quantile function of $T \mid X$ is defined as $q_{T|X}(\gamma \mid X) = \inf\{t \in \mathbb{R}_+ \colon S_{T|X}(t \mid X) \leqslant 1-\gamma\}$.

We aim to obtain predictive bounds $\hat{L}(X)$ and $\hat{U}(X)$ for time-to-event outcome T leveraging the covariate vector X with a marginal coverage guarantee, i.e., $\mathbb{P}\{\hat{L}(X) \leqslant T \leqslant \hat{U}(X)\} \geqslant 1-\alpha$. For comparison with most of the existing literature, we focus on lower predictive bounds (LPBs), even though all our methods can provide two-sided predictive bounds; see (1) below and the subsequent discussion.

DEFINITION 1. *Let $\hat{L}(\cdot)$ be a function estimated from the observed data $\mathcal{D}$, and let $\alpha \in [0,1]$ be a nonstochastic level. We say that $\hat{L}$ is a marginally calibrated LPB for T at level α if*

$$\mathbb{P}\{T \geqslant \hat{L}(X)\} \geqslant 1-\alpha,$$

where the probability is taken with respect to a new data point $(X,T) \sim P_{(X,T)}$ and the observed data $\mathcal{D}$. We say that $\hat{L}(\cdot)$ is an asymptotically marginally calibrated LPB for T at level α if

$$\mathbb{P}\{T \geqslant \hat{L}(X)\} \geqslant 1-\alpha-o(1) \quad \text{as } n \to \infty.$$

Given the distribution of (X,T), there are several possibilities for an oracle LPB. For instance, one can take $L(\cdot)$ to be the αth quantile of the marginal distribution T, which is a constant independent of X. Observing that a coverage guarantee conditional on the covariate vector X would be much stronger and more broadly applicable, we consider the oracle LPB as the αth conditional quantile of T, given X, i.e., $q^\star_{T|X}(\alpha \mid \cdot)$.

Given that we often have one observed dataset, one might require coverage conditional on the observed data. In this case, a probably approximately correct (PAC)-type LPB can provide a more relevant goal for the coverage guarantee.

DEFINITION 2. *Let $\hat{L}(\cdot)$ be a function estimated from the observed data $\mathcal{D}$, and let $\alpha, \epsilon \in [0,1]$ be fixed levels. We say that $\hat{L}$ is an (α,ϵ)-PAC LPB for T if*

$$\mathbb{P}[\mathbb{P}\{T \geqslant \hat{L}(X) \mid \mathcal{D}\} \geqslant 1-\alpha] \geqslant 1-\epsilon,$$

where the inner probability is calculated with respect to the new data point $(X,T) \sim P_{(X,T)}$, independent of $\mathcal{D}$. We say that $\hat{L}(\cdot)$ is an asymptotic (α,ϵ)-PAC LPB for T if

$$\mathbb{P}[\mathbb{P}\{T \geqslant \hat{L}(X) \mid \mathcal{D}\} \geqslant 1-\alpha-o_p(1)] \geqslant 1-\epsilon \quad \text{as } n \to \infty,$$

where the $o_p(1)$ random variable can depend on ϵ.

As Candès et al. (2023) noted, by definition, Y cannot be larger than T; thus, any valid LPB for Y is trivially an LPB for T. Nevertheless, a valid lower bound for Y may yield an overly conservative lower bound for T; for this reason, in this work, we focus on constructing bounds directly for the underlying survival time T of interest, rather than the observed censored time Y.

1.3. *Related work*

Recent conformal prediction methods for survival data include Cox regression-based (Teng et al., 2021) and random survival forest-based (Boström et al., 2017, 2019) methods, as well as those of Candès et al. (2023) and Gui et al. (2023), who proposed constructing covariate-dependent lower predictive bounds. However, these methods assume that censoring times are observed for all units in the training and validation samples, a setting also known as *Type-I censoring*. Formally, the observed data consist of tuples (X, C, Y). Under this setting, Candès et al. (2023) established that any distribution-free, well-calibrated LPB for T must necessarily be an LPB for Y. This motivated their proposed approach, which entails discarding any observation with an early censoring time, below a user-specified constant c_0, combined with weighted conformal inference (Tibshirani et al., 2019) to correct for the distribution shift introduced by the initial filtering step. Based on this approach, Gui et al. (2023) incorporated a covariate-dependent cut-off, making the procedure substantially less conservative. Under Type-I censoring, their method yields an LPB that is asymptotically PAC-calibrated in the sense of Definition 2, with the $o_p(1)$ term being the maximum of the estimation error of the conditional survival function of the censoring time, given X and an $\{\log(1/\epsilon)/n\}^{1/2}$-order term; see Theorem 3 of Gui et al. (2023).

By requiring that censoring times are fully observed, the methods of Candès et al. (2023) and Gui et al. (2023) are primarily useful in applications where potential censoring times are determined at the start of a unit's follow-up, such as experimental settings in which administrative censoring times are predetermined. In contrast, other common forms of censoring, such as loss to follow-up and related protocol departures, can seldom, if ever, be prevented by design. Another apparent limitation of the methods of Candès et al. (2023) and Gui et al. (2023) is their over-reliance on a cut-off tuning parameter that induces artificial censoring of units with censoring time in the left tail of their distribution, possibly censoring units with observed failure times. This form of artificial censoring introduces a bias-variance trade-off that may be challenging to balance in practice. Our proposed methods target an asymptotic PAC guarantee analogous to theirs. However, as detailed in the next section, we achieve this goal while (i) accommodating a more common form of right censoring; (ii) obviating the need for artificial censoring or a corresponding tuning parameter; and (iii) obtaining a doubly robust coverage error guarantee, in the sense that our coverage error slack is, not only of second order, but is actually of the mixed-bias product form of the estimation biases of the censoring and failure-time laws given, covariates, plus an irreducible root-n error term.

Recent works have considered alternative strategies for obtaining well-calibrated predictions under conventional censoring. In the Type-I censoring setting, Wan & Yi (2025) proposed distribution-free lower predictive bounds by combining hybrid quantile learning with a data-filtered, adaptive conformal calibration scheme. Moving beyond Type-I censoring, Sesia & Svetnik (2025b) extended the conformalized survival analysis framework of Candès et al. (2023) and Gui et al. (2023) to general right censoring by first imputing unobserved censoring times under conditionally independent censoring and then applying weighted conformal inference to the imputed data. A recent work by Qin et al. (2025) proposes a bootstrap-resampling approach for constructing conformal predictive intervals for general right-censored data, achieving asymptotic marginal validity under conditionally independent censoring. Building on our work, Si & Qiu (2025) studied lower prediction bounds with training-set-conditional APAC-type guarantees, meaning that the approximation is placed on the confidence level, rather than on the miscoverage probability as in

our asymptotic PAC formulation, while conditioning on the training sample. They adopted the efficient influence-function correction introduced by our approach. Moreover, Yi et al. (2025) constructed conformal prediction intervals using inverse-censoring weighting and quantile regression as a flexible base learner. A different line of work focuses on survival curves rather than on the time-to-event outcome: Sesia & Svetnik (2025a) constructed conformal bands for individual survival curves for screening-type tasks by combining inverse-censoring weighting with multiple-testing false discovery rate control. We also note the work of Park et al. (2025), which performs post hoc calibration of predicted survival curves by reducing a Kolmogorov–Smirnov discrepancy between predicted and empirical survival probabilities. Finally, Meixide et al. (2025) introduced a prediction inference framework for interval-censored data, an important but different focus to ours. While these papers substantially advance conformal uncertainty quantification for censored outcomes, they do not simultaneously meet our specific desiderata (i)–(iii). In particular, for the setting and goal we consider, we provide a doubly robust procedure that attains semiparametric efficiency (under standard regularity conditions) and yields second-order coverage error control.

In a separate research strand, IPCW and AIPCW have played a pivotal role in the development of robust and efficient methodologies for censored data, primarily in the context of inference for a finite-dimensional functional of interest. The foundational semiparametric theory for robust and efficient estimation in the presence of censoring, based on IPCW and AIPCW, was pioneered by Robins and colleagues (Robins & Rotnitzky, 1992; Robins, 1993). See van der Laan & Robins (2003) and Tsiatis (2006) for an introductory textbook overview of modern semiparametric efficiency theory for censored and other missing data problems. Our work extends the recent paper of Yang et al. (2024), which, to the best of our knowledge, provides the first application of semiparametric efficiency theory to a conformal prediction setting subject to covariate shift. While their proposed approach can equivalently be viewed as an outcome missing at random, their proposed methods do not directly apply to the censored survival outcome setting that presents a number of new challenges we address in this work.

1.4. *Our contributions*

Motivated by the limitations of existing methods, we develop a new approach for well-calibrated predictive inference for standard right-censored survival outcomes, whereby either only the censoring or the primary event time is observed, whichever comes first. Our goal is to derive predictive bounds while avoiding unnecessary modelling assumptions. The first of our two proposed methods incorporates IPCW, which data adaptively accounts for the unknown censoring mechanism, provided that measured covariates are sufficiently rich to explain any dependence between censoring and the event time of primary interest. This completely obviates the need for artificial censoring and the required tuning of a cut-off parameter, while recovering robust (asymptotically) well-calibrated predictive bounds for survival times.

A well-known limitation of IPCW is that it is potentially inefficient due to discarding information contained in censored observations. Thus, in addition, we introduce an augmented IPCW approach, which recovers information from censored observations with the potential for significant efficiency improvements over IPCW. Using AIPCW also leads to a desirable double robustness property, which ensures that the error in the coverage bound of the AIPCW prediction sets, due to estimation of nuisance parameters (mainly of the conditional survival curve for the primary outcome and for the censoring mechanism given covariates,

respectively), is of mixed bias product form (Rotnitzky et al., 2021). This ensures that our method can provide valid bounds even if one of the required survival curves is misspecified, offering enhanced flexibility and reliability in practical applications.

Importantly, the proposed methodology for deriving predictive bounds does not depend on the specific choice of a nonconformity score. The resulting bounds preserve their validity, and the AIPCW-based approach remains doubly robust, regardless of the choice of nonconformity score. This flexibility highlights the generality and robustness of our proposed framework. Furthermore, if the nonconformity score is properly chosen (see (2)) so as to align with the interpretation of the conditional quantile as the oracle LPB, then our methods also recover approximate conditional coverage; see Theorem 4 of Gui et al. (2023) for a related result.

As a complementary contribution, we introduce two simple yet effective approaches for constructing prediction bounds based on outcome regression: outcome regression (OR) and calibrated outcome regression (COR). The first estimates the conditional quantile function of the primary event time at the desired nominal level, while the second calibrates the estimated conditional quantile function of the survival time to achieve the desired marginal coverage guarantee. Although the validity of this method is contingent on a consistent estimator of the survival function model, it provides a computationally efficient alternative to the IPCW and AIPCW approaches. To the best of our knowledge, this simple calibration-based technique has not been previously explored in the context of conformal predictive inference for censored time-to-event data.

2. IPCW AND AIPCW METHODS

2.1. *IPCW method*

In order to ensure identifiability, throughout, we restrict the degree of dependence between the survival and censoring times, by invoking a standard conditional independence censoring assumption (Kalbfleisch & Prentice, 2002).

Assumption 1 (*Conditional independent censoring*). Event times and censoring times are independent conditional on the observed covariates, i.e., $T \perp\!\!\!\perp C \mid X$.

This assumption is analogous to missingness at random in the missing data literature and to unconfoundness in causal inference. It is appropriate in practice only to the extent that baseline covariates reasonably capture the dependence between the primary outcome and the censoring mechanism.

We propose a novel approach to construct valid PAC-type LPBs for the survival time of future observations, satisfying Definition 2. The key idea behind our method is to calibrate the coverage of prediction sets for the underlying time-to-event outcome of interest by incorporating IPCW to complete cases to account for selection bias due to censoring. Specifically, we use weights given by the reciprocal of the estimated conditional survival function of the censoring mechanism given covariates, evaluated at the observed failure time, thereby allowing us to properly estimate, and hence calibrate the coverage probability. In fact, for a given function $R(x,t)$ possibly based on an independent sample, and for any β, under Assumption 1,

$$\mathbb{P}\{R(X,T) \geqslant \beta\} = \mathbb{E}\left[\frac{\boldsymbol{\Delta}\mathbf{1}\{R(X,T) \geqslant \beta\}}{S^{\star}_{C|X}(T \mid X)}\right]. \tag{1}$$

Robins & Rotnitzky (1992) proved this result for a general full-data estimating function. For completeness, the proof in our setting is given in the Supplementary Material. If $R(X,T) = \hat{L}(X) - T$ for some function $\hat{L}$ estimated from the observed data, then a proper choice of β yields an LPB on T of the form $\hat{L}(X) - \beta$. Similarly, if $R(X,T) = T - \hat{U}(X)$ for some function $\hat{U}$ estimated from the observed data, then we recover an upper predictive bound $\hat{U}(X) + \beta$. Moreover, two-sided predictive bounds can be constructed using, for example, $R(X,T) = \max\{\hat{L}(X) - T, T - \hat{U}(X)\}$. In general, $R(X,T)$ can take any user-specified form; however, to simplify the exposition, consistent with prior works, we focus on LPBs.

Implementing the proposed IPCW estimator of the coverage probability first requires obtaining an estimator of the conditional survival curve for the censoring time $\hat{S}_{C|X}(\cdot \mid \cdot)$, and identifying a suitable, potentially covariate-dependent, candidate for the LPB $\hat{L}$. A natural candidate is given by an estimate of the conditional quantile function $\hat{q}_{T|X}(\cdot \mid \cdot)$, which can be obtained from the fitted survival curve for $T \mid X$, $\hat{S}_{T|X}(\cdot \mid \cdot)$. Empirically, this translates into solving for the 'pseudo-quantile' level $\beta \in [0,1]$, defined as the largest number to satisfy

$$\mathbb{E}\left[\frac{\Delta \mathbf{1}\{T \geqslant \hat{q}_{T|X}(\beta \mid X)\}}{\hat{S}_{C|X}(T \mid X)}\right] \geqslant 1 - \alpha.$$

We refer to β as a *pseudo-quantile* level because the approach does not, in fact, require $\hat{q}_{T|X}(\cdot \mid \cdot)$ to be consistent for the true conditional quantile function in order for the resulting prediction interval to be well calibrated. Rather, any working model for the conditional quantile function, even if incorrect, is guaranteed to produce a prediction set with correct marginal coverage. However, assuming that the survival curve of the censoring distribution used to define the weights is correctly specified, the optimal, or oracle, choice of $\hat{q}_{T|X}(\beta \mid X)$ entails the true conditional quantile function because, by definition of the conditional quantile, $\beta \equiv \alpha$. This ensures, not only correct marginal coverage, but also correct conditional coverage, representing a clear improvement in coverage control. This robustness property is a key contribution of our proposed approach.

For the implementation of this method, we consider the computationally efficient variant of conformal prediction, namely split conformal prediction, which relies on dividing the data into two independent subsets. Specifically, we split the data $\mathcal{D}$ into a training set $\mathcal{D}_1$ and a calibration set $\mathcal{D}_2$, with respective sets of indices $\mathcal{I}_1$ and $\mathcal{I}_2$. Although our methods and results are formally stated for arbitrary sizes of $\mathcal{D}_1$ and $\mathcal{D}_2$, we assume henceforth that the two subsets are of comparable sizes when discussing the rates of convergence of error terms. For example, by assuming that $|\mathcal{D}_1|/|\mathcal{D}_2| \in (0,\infty)$ as $n \to \infty$, it becomes feasible to express various error convergence rates in terms of n.

On the training set, we fit the conditional survival curves for both the survival time and the censoring time. We then estimate the coverage probability on the calibration set using the IPCW estimator

$$\hat{P}_{\text{HT-IPCW}}(\beta) = \frac{1}{|\mathcal{D}_2|} \sum_{i \in \mathcal{I}_2} \frac{\Delta_i \mathbf{1}\{T_i \geqslant \hat{q}_{T|X}(\beta \mid X_i)\}}{\hat{S}_{C|X}(T_i \mid X_i)}$$

for all $\beta \in [0,1]$. Finally, we select the optimal β as

$$\hat{\beta}_{\text{HT-IPCW}} = \sup\{\beta \in [0,1] \colon \hat{P}_{\text{HT-IPCW}}(\beta) \geqslant 1 - \alpha\},$$

and obtain the LPB as $\hat{L}(\cdot) = \hat{q}_{T|X}(\hat{\beta}_{\text{HT-IPCW}} \mid \cdot)$. The classical IPCW estimator can be viewed as a form of the Horvitz–Thompson estimator (HT-IPCW), a well-known inverse-probability-weighted estimator commonly used in the survey-sampling literature, which computes the sample average by normalizing with the sample size (Horvitz & Thompson, 1952). However, if estimated survival probabilities defining the weights $\hat{S}_{C|X}(T_i \mid X_i)$ are relatively small, the estimator may become unstable, leading to increased variability. To address this issue, we consider the Hájek version of the IPCW estimator (Basu, 1970; Hájek, 1971), a well-known ratio-adjusted version of the Horvitz–Thompson estimator. The Háek IPCW estimator normalizes by the sum of the inverse-probability weights, such that the coverage probability is estimated as

$$\hat{P}_{\text{IPCW}}(\beta) = \frac{\sum_{i \in \mathcal{I}_2} \Delta_i \mathbf{1}\{T_i \geqslant \hat{q}_{T|X}(\beta \mid X_i)\} / \hat{S}_{C|X}(T_i \mid X_i)}{\sum_{i \in \mathcal{I}_2} \Delta_i / \hat{S}_{C|X}(T_i \mid X_i)},$$

and the optimal β is given by $\hat{\beta}_{\text{IPCW}} = \sup\{\beta \in [0,1] \colon \hat{P}_{\text{IPCW}}(\beta) \geqslant 1-\alpha\}$. The Hájek estimator is approximately unbiased and often yields lower variance than the Horvitz–Thompson estimator (Datta & Polson, 2025). An equivalent estimating equation representation of the Hájek estimator is

$$\mathbb{E}\left[\frac{\Delta[\mathbf{1}\{T \geqslant \hat{q}_{T|X}(\beta \mid X)\} - (1-\alpha)]}{\hat{S}_{C|X}(T \mid X)}\right] \geqslant 0.$$

Thus, we may estimate the optimal β as $\hat{\beta}_{\text{IPCW}} = \sup\{\beta \in [0,1] \colon \hat{W}(\beta) \geqslant 0\}$, where

$$\hat{W}(\beta) = \frac{1}{|\mathcal{D}_2|} \sum_{i \in \mathcal{I}_2} \frac{\Delta_i[\mathbf{1}\{T_i \geqslant \hat{q}_{T|X}(\beta \mid X_i)\} - (1-\alpha)]}{\hat{S}_{C|X}(T_i \mid X_i)}.$$

Algorithm A.1 of the Supplementary Material provides a detailed description of this procedure.

The method described above can be extended to a general nonconformity score computed on an independent sample. Let $R \colon \mathbb{R}^{d+1} \to \mathbb{R}$ be an arbitrary function defining a given nonconformity score. One may think of R as a fixed nonstochastic function. This framework allows us to select R to be of a simple form, such as T, or of a more general form determined by the data, depending on the problem. If the LPB $\hat{L}$ is chosen to be the estimated conditional quantile of T, given X, then R is defined through

$$\{R(X,T) \geqslant \beta\} = \{T \geqslant \hat{q}_{T|X}(\beta \mid X)\}, \tag{2}$$

implying that, under correct specification of $\hat{q}_{T|X}$, $R(X,T)$ matches the estimated conditional cumulative distribution function $\hat{F}_{T|X}(T \mid X)$. In general, for any nonconformity score R, there is a corresponding LPB $\hat{L}$, which one may recover from an expression for R. Hereafter, we use R to indicate that our methodology allows for an arbitrary choice of nonconformity score. The detailed description of this procedure for a general nonconformity score R is given in Algorithm A.2 of the Supplementary Material.

In Theorem 1, we establish the asymptotic marginal coverage guarantee for the LPB constructed through the procedure described above. We present the result in the case of a general

nonconformity score R. In addition, when the nonconformity score is chosen as in (2), we further obtain an asymptotic conditional coverage guarantee. The following positivity assumption is required on the conditional survival curve of $C \mid X$ and its estimator.

Assumption 2 (*Positivity*). There exists $0 < s_0 < \infty$ such that $\hat{S}_{C|X}(T \mid X) \wedge S^{\star}_{C|X}(T \mid X) > 1/s_0$ almost surely under $P_{(X,T)}$.

This is a standard positivity condition on which much of the semiparametric efficiency theory for censored data is based. It ensures that, for all values of T and X under consideration, there is a positive probability of observing an individual with $C \geqslant T$. For a detailed discussion, we refer the reader to § 3.3 of van der Laan & Robins (2003). Let L^2 denote the space of all square-integrable functions with respect to the probability measure μ of a random variable Z, where $\mu(B) = \mathbb{P}(Z \in B)$ for any Borel set B. We denote by $\|f\|_{L^2}$ the L^2 norm of a function $f(\cdot)$, defined as

$$\|f\|_{L^2} = \left(\int f^2(z) \mu(\mathrm{d}z) \right)^{1/2} = (\mathbb{E}[f^2(Z)])^{1/2}.$$

Let $\|f\|_{\infty} = \sup_z |f(z)|$ denote the sup norm of a function $f(\cdot)$.

THEOREM 1. *Let $\epsilon \in (0,1)$ be fixed. There exists a universal constant K such that, under Assumptions 1 and 2, with probability at least $1-\epsilon$ over $\mathcal{D}$,*

$$\mathbb{P}\{R(X,T) \geqslant \hat{\beta}_{\mathrm{IPCW}} \mid \mathcal{D}\} \geqslant 1 - \alpha - s_0^2 \|\hat{S}_{C|X} - S^{\star}_{C|X}\|_{L^2} - s_0 \frac{(-\log \epsilon/2)^{1/2} + K}{|\mathcal{D}_2|^{1/2}}, \tag{3}$$

where the probability $\mathbb{P}$ is taken with respect to a new data point $(X,T) \sim P_{(X,T)}$. Furthermore, if the nonconformity score is chosen to satisfy (2), and if the distribution of $T \mid X$ is continuous, then, with probability at least $1-\epsilon$ over $\mathcal{D}$,

$$\mathbb{P}\{T \geqslant \hat{q}_{T|X}(\hat{\beta}_{\mathrm{IPCW}} \mid X) \mid X, \mathcal{D}\} \geqslant 1 - \alpha - s_0^2 \|\hat{S}_{C|X} - S^{\star}_{C|X}\|_{L^2} - 2\|\hat{S}_{T|X} - S^{\star}_{T|X}\|_{\infty} - s_0 \frac{(-\log \epsilon/2)^{1/2} + K}{|\mathcal{D}_2|^{1/2}},$$

where the probability $\mathbb{P}$ is taken with respect to a new data point $(X,T) \sim P_{(X,T)}$.

The proof of Theorem 1 is deferred to the Supplementary Material. As the size of the calibration set $|\mathcal{D}_2|$ goes to ∞, the last term on the right-hand side of (3) vanishes. The second error term converges to zero if $\hat{S}_{C|X}$ is consistent for $S^{\star}_{C|X}$. In other words, the prediction region for a new X, given by $\{t\colon R(X,t) \geqslant \hat{\beta}_{\mathrm{IPCW}}\}$, is asymptotically (α,ϵ)-PAC, implying that the corresponding $\hat{L}$ is an asymptotic (α,ϵ)-PAC LPB for T. Moreover, the prediction set enjoys asymptotic conditional coverage when the nonconformity score is defined as in (2). In other words, asymptotic conditional validity is achieved provided

that both conditional survival functions for the censoring mechanism and the time-to-event outcome are consistently estimated.

Remark 1. In the special case of completely independent censoring, where the censoring time C is independent of (T, X), the estimation of the censoring mechanism simplifies significantly. In this scenario, the censoring survival function $S^{\star}_{C}(t)$ depends only on t and can be consistently estimated at a root-n rate using standard nonparametric methods, such as the Kaplan–Meier estimator (Kaplan & Meier, 1958; Fleming & Harrington, 2013). This assumption, while more restrictive, is commonly satisfied in randomized trials, at least within treatment arms, where the censoring mechanism may reasonably be assumed to be independent of other variables by careful study design (Pocock, 2013; Collett, 2023). Consequently, the IPCW framework becomes even more robust and efficient in such settings, further strengthening its practical applicability.

In short, the procedure outlined in this section constructs an asymptotically (α, ϵ)-PAC LPB $\hat{L}(\cdot)$ for the right-censored survival time T. Moreover, two-sided bounds can also be obtained if the nonconformity score is chosen appropriately: $R(X, T) = |2\hat{S}_{T|X}(T \mid X) - 1|$ (Chernozhukov et al., 2021).

2.2. *AIPCW method*

The IPCW-based approach described in the previous section relies on weighting uncensored observations by the inverse of the probability of remaining uncensored up to the time at which they experience the primary outcome, thereby correcting for any potential bias due to censoring. While this method can be effective under fairly reasonable conditions, it can be inefficient because it does not make efficient use of information contained in censored observations and may not produce well-calibrated prediction sets if either the estimator of the censoring mechanism converges at a slow rate or fails to be consistent. To address these limitations, we extend our framework by introducing an augmentation to IPCW that recovers information from censored observations by leveraging a model for the outcome of interest.

The AIPCW method has attractive theoretical appeal over the IPCW approach, by producing a doubly robust estimator, which means that the estimator remains consistent if either the survival model or the censoring model is correctly specified (but not necessarily both). This is an immediate consequence of its mixed bias error, as shown in the result below. This robustness can be particularly valuable in survival analysis, even when machine learning or nonparametric methods are used to estimate unknown regression functions, as the AIPCW error rate is then guaranteed to be of a smaller order of magnitude than that of IPCW.

The efficiency and robustness properties of AIPCW are rooted in modern semiparametric efficiency theory, in which influence functions are central to the construction of efficient estimators. Briefly, according to a result due to Robins & Rotnitzky (1992), the efficient influence function of a regular parameter defined on a nonparametric model subject to censoring at random can be obtained as the projection of the IPCW version of the full-data efficient influence function onto the orthogonal complement of the tangent space for the censoring mechanism, assumed to satisfy Assumption 2 and otherwise unrestricted. The tangent space for a semiparametric model is technically defined as the closed linear span of scores of all regular parametric submodels contained in the model. Robins & Rotnitzky (1992) proved

that, under Assumptions 1 and 2, the tangent space for the censoring mechanism is given by

$$\mathcal{T} = \text{closure}\left\{ \int h(u, X)\, \mathrm{d}M^{\star}_{C|X}(u \mid X) \colon h \in L^2 \right\},$$

where $\mathrm{d}M^{\star}_{C|X}(u \mid X) = \mathrm{d}N_C(u) - \mathbf{1}\{Y \geqslant u\}\, \mathrm{d}\Lambda^{\star}_{C|X}(u \mid X)$ is the martingale difference process, with $\mathrm{d}N_C(u) = \mathbf{1}\{Y \in \mathrm{d}u, \Delta = 0\}$ the increment of the censoring counting process and $\Lambda^{\star}_{C|X}(u \mid X) = -\log\{S^{\star}_{C|X}(u \mid X)\}$ the true censoring cumulative hazard function. Let $W(\cdot)$ be the IPCW estimating function:

$$W(\beta) = \frac{\Delta\{\mathbf{1}\{R(X, T) \geqslant \beta\} - (1 - \alpha)\}}{S^{\star}_{C|X}(T \mid X)}.$$

Then, as shown by Robins & Rotnitzky (1992), the orthogonal projection of W onto the censoring tangent space is given by

$$\Pi(W(\beta) \mid \mathcal{T}) = -\int \frac{\mathbb{E}[\mathbf{1}\{R(X, T) \geqslant \beta\} \mid X, T \geqslant u] - (1 - \alpha)}{S^{\star}_{C|X}(u \mid X)}\, \mathrm{d}M^{\star}_{C|X}(u).$$

A self-contained proof of this result is provided in the Supplementary Material. The resulting AIPCW moment equation is thus given by the residual of the projection of the IPCW estimating function onto the censoring tangent space. As a result, we propose to identify the optimal quantile value $\beta \in [0, 1]$ as the largest number that satisfies

$$\mathbb{E}\left[\frac{\Delta[\mathbf{1}\{R(X, T) \geqslant \beta\} - (1 - \alpha)]}{\hat{S}_{C|X}(T \mid X)} + \int \frac{\hat{\eta}(\beta, u \mid X) - (1 - \alpha)}{\hat{S}_{C|X}(u \mid X)}\, \mathrm{d}\hat{M}_{C|X}(u \mid X) \right] \geqslant 0, \tag{4}$$

where $\hat{\eta}(\beta, u \mid X)$ is an estimator of $\eta^{\star}(\beta, u \mid X) = \mathbb{E}[\mathbf{1}\{R(X, T) \geqslant \beta\} \mid X, T \geqslant u]$. There are two aspects of the AIPCW moment equation that require further elaboration for practical implementation: (i) estimation of the additional nuisance component $\hat{\eta}$; and (ii) computation of the integral with respect to the martingale difference process. As with the IPCW method, the data $\mathcal{D}$ are divided into a training set $\mathcal{D}_1$ and a calibration set $\mathcal{D}_2$, with respective index sets $\mathcal{I}_1$ and $\mathcal{I}_2$. We use $\mathcal{D}_1$ to estimate all nuisance functions.

For (i), note that when, as we assume in the following exposition, the estimated conditional quantile of $T \mid X$ (obtained from an independent sample) is used to define the nonconformity score (as in (2)),

$$\begin{aligned}
\eta^{\star}(\beta, u \mid X) &= \mathbb{E}[\mathbf{1}\{T \geqslant \hat{q}_{T|X}(\beta \mid X)\} \mid X, T \geqslant u, \mathcal{D}_1] \\
&= \mathbb{P}\{T \geqslant \hat{q}_{T|X}(\beta \mid X) \mid X, T \geqslant u, \mathcal{D}_1\} \\
&= \frac{\mathbb{P}\{T \geqslant \hat{q}_{T|X}(\beta \mid X), T \geqslant u \mid X, \mathcal{D}_1\}}{\mathbb{P}(T \geqslant u \mid X, \mathcal{D}_1)} \\
&= \frac{S^{\star}_{T|X}[\max\{\hat{q}_{T|X}(\beta \mid X), u\} \mid X]}{S^{\star}_{T|X}(u \mid X)}.
\end{aligned} \tag{5}$$

Thus, in this case, an estimator of $\eta^\star(\beta, u \mid X)$ is conveniently expressed as

$$\hat{\eta}(\beta, u \mid X) = \frac{\hat{S}_{T|X}[\max\{\hat{q}_{T|X}(\beta \mid X), u\} \mid X]}{\hat{S}_{T|X}(u \mid X)}.$$

For (ii), note that when evaluated in the observed sample, the integral in (4) often reduces to a finite sum. In fact, for several estimators, such as random survival forests, $\hat{S}_{C|X}$ is piecewise constant with jumps at observed censoring times, implying that $\hat{\Lambda}_{C|X}$ is also piecewise constant and the martingale difference process is a discrete measure. In such cases, denoting the observed ordered censoring times as $u_{(1)}, \ldots, u_{(Q)}$ for some finite Q, and setting $u_{(0)} = 0$, we have

$$\begin{aligned}
&\int \frac{\hat{\eta}(\beta, u \mid X) - (1-\alpha)}{\hat{S}_{C|X}(u \mid X)} \, \mathrm{d}\hat{M}_{C|X}(u \mid X) \\
&= \sum_{k=1}^{Q} \frac{\hat{\eta}(\beta, u_{(k)} \mid X) - (1-\alpha)}{\hat{S}_{C|X}(u_{(k)} \mid X)} \{\hat{M}_{C|X}(u_{(k)} \mid X) - \hat{M}_{C|X}(u_{(k-1)} \mid X)\},
\end{aligned}$$

where $\hat{M}_{C|X}(u \mid X) = \mathbf{1}\{Y \leqslant u, \Delta = 0\} - \hat{\Lambda}_{C|X}(Y \wedge u \mid X)$. More generally, for estimators of the censoring mechanism that rely on smoothing the baseline hazard, such as kernel smoothing, the integral may not naturally reduce to a finite sum over censoring times; however, the latter may still be approximated by an alternative (weighted) finite sum via, say, Gaussian quadrature or analogous numerical integration techniques.

On the training set $\mathcal{D}_1$, the conditional survival curves for both the survival time and the censoring time are estimated. Subsequently, on the calibration set $\mathcal{D}_2$, we compute the IPCW term as

$$\hat{W}(\beta) = \frac{1}{|\mathcal{D}_2|} \sum_{i \in \mathcal{I}_2} \frac{\Delta_i[\mathbf{1}\{T_i \geqslant \hat{q}_{T|X}(\beta \mid X_i)\} - (1-\alpha)]}{\hat{S}_{C|X}(T_i \mid X_i)}$$

and the augmentation term as

$$\begin{aligned}
\hat{\Pi}(\beta) = \frac{1}{|\mathcal{D}_2|} \sum_{i \in |\mathcal{D}_2|} \sum_{k=1}^{Q} &\frac{\hat{\eta}(\beta, u_{(k)} \mid X_i) - (1-\alpha)}{\hat{S}_{C|X}(u_{(k)} \mid X_i)} \{\hat{M}_{C|X}(u_{(k)} \mid X_i) \\
&- \hat{M}_{C|X}(u_{(k-1)} \mid X_i)\}
\end{aligned}$$

for all $\beta \in [0, 1]$. The optimal β is then estimated as $\hat{\beta}_{\mathrm{AIPCW}} = \sup\{\beta \in [0, 1] \colon \hat{W}(\beta) + \hat{\Pi}(\beta) \geqslant 0\}$. The detailed steps of this procedure are provided in Algorithm A.3 of the Supplementary Material. As with the IPCW-based method, this procedure can be generalized to use any arbitrary nonconformity score $R(X, T)$, with the corresponding steps given in Algorithm A.4 of the Supplementary Material.

The remainder of this section establishes the asymptotic validity and double robustness of our proposed AIPCW-based approach. The efficient influence function can be readily

derived analogously to the derivation above, as the projection of an IPCW moment function onto the orthogonal complement of the censoring tangent space:

$$\begin{aligned}\mathrm{IF}_\beta(S_{C|X}, \eta) = \frac{\Delta}{S_{C|X}(T \mid X)}\{\mathbf{1}\{R(X,T) \geqslant \beta\} - (1-\alpha)\} \\ + \int \{\eta(\beta, u \mid X) - (1-\alpha)\}\frac{\mathrm{d}M_{C|X}(u \mid X)}{S_{C|X}(u \mid X)}.\end{aligned}$$

For clarity, the explicit dependence on X, T and R is suppressed. Additionally, note that the dependence on $\mathrm{d}M_{C|X}$ is captured through the dependence on $S_{C|X}$. In Lemma 1 below, we provide an alternative expression for the efficient influence function of β, which is instrumental in proving the key double robustness result outlined in Theorem 2 below.

LEMMA 1. *Under Assumption 1,*

$$\begin{aligned}\mathrm{IF}_\beta(S_{C|X}, \eta) = \mathbf{1}\{R(X,T) \geqslant \beta\} - (1-\alpha) \\ - \int [\mathbf{1}\{R(X,T) \geqslant \beta\} - \eta(\beta, u \mid X)]\frac{\mathrm{d}M_{C|X}(u \mid X)}{S_{C|X}(u \mid X)}.\end{aligned}$$

This expression holds for any choice of functions η and $S_{C|X}$, whether estimated or evaluated at their true value. The proof of Lemma 1 can be found in the Supplementary Material.

The main double robustness property of the derived efficient influence function of β is established below. Before presenting the result, we introduce some additional notation. For any function $f^\star \colon \mathbb{R}^{d+1} \to \mathbb{R}$ with estimate $\hat{f}$, define

$$\mathrm{Err}_{1,f} := \|\hat{f}(Y \mid X) - f^\star(Y \mid X)\|_{L^2}, \qquad \mathrm{Err}_{2,f}(u) := \|\hat{f}(u \mid X) - f^\star(u \mid X)\|_{L^2},$$
$$\mathrm{Err}_{3,f}(u) := \left\| \frac{\mathrm{d}\hat{f}(u \mid X)}{\mathrm{d}u} - \frac{\mathrm{d}f^\star(u \mid X)}{\mathrm{d}u} \right\|_{L^2}.$$

Let $\mathrm{Err}_{j,\eta^\star_\beta}(\cdot)$, $1 \leqslant j \leqslant 3$ denote the errors in estimating $\eta^\star(\beta, \cdot \mid \cdot)$, and, similarly, let $\mathrm{Err}_{j,\Lambda^\star_{C|X}}(\cdot), 1 \leqslant j \leqslant 3$ denote the errors in estimating $\Lambda^\star_{C|X}(\cdot \mid \cdot)$.

THEOREM 2. *Under Assumptions 1 and 2, for any estimated (from $\mathcal{D}_1$) functions $\hat{S}_{C|X}$ and $\hat{\eta}$, and any $\beta \in [0,1]$, the following inequality holds:*

$$\begin{aligned}&|\mathbb{E}[\mathrm{IF}_\beta(\hat{S}_{C|X}, \hat{\eta}) \mid \mathcal{D}_1] - \mathbb{E}[\mathrm{IF}_\beta(S^\star_{C|X}, \eta^\star)]| \\ &< s_0 \min\left\{ \int \mathrm{Err}_{2,\eta^\star_\beta}(u)\mathrm{Err}_{3,\Lambda^\star_{C|X}}(u)\,\mathrm{d}u, \right. \\ &\qquad \left. \mathrm{Err}_{1,\eta^\star_\beta}\mathrm{Err}_{1,\Lambda^\star_{C|X}} + \int \mathrm{Err}_{3,\eta^\star_\beta}(u)\mathrm{Err}_{2,\Lambda^\star_{C|X}}(u)\,\mathrm{d}u \right\}.\end{aligned}$$

This result implies that the influence function $\mathrm{IF}_\beta(\hat{S}_{C|X}, \hat{\eta})$ is asymptotically unbiased as long as either $S^\star_{C|X}$ or $\eta^\star$ is estimated consistently. The proof of Theorem 2 is detailed in the Supplementary Material.

Remark 2. As mentioned in Remark 1, under completely independent censoring, the censoring survival curve $S^\star_C(u)$, and therefore the cumulative hazard function $\Lambda^\star_C(u)$, can be consistently estimated at a root-n rate. Hence, $\|\hat{\Lambda}_C(u) - \Lambda_C(u)\|_{L^2} = O_p(n^{-1/2})$, which implies that the error bound of the AIPCW estimator is negligible and is guaranteed to be better than that of the IPCW estimator.

Remark 3. When the estimated conditional quantile of $T \mid X$ is used to define the nonconformity score (as in (2)), $\eta^\star$ takes the specific form described in (5), which involves the ratio of $S^\star_{T|X}$ evaluated at two different time-points. Therefore, under a positivity assumption for the time-to-event outcome, analogous to that of the censoring mechanism in Assumption 2, the error in estimating $\eta^\star$ can be bounded by the corresponding error in estimating $S^\star_{T|X}$, scaled by the positivity constant.

Remark 4. The coverage error bound for the AIPCW estimator depends, not only on the estimation errors of the nuisance functions themselves, but also on the estimation accuracy of their derivatives. In general, the convergence rate of a derivative of a function is slower than that of the function itself, depending on the smoothness of the function. Observe that $\eta^\star$ (for most scores) can be written in terms of $S^\star_{T|X}$, as shown in (5). Also, $\Lambda^\star_{C|X}$ is a smooth function of $S^\star_{C|X}$. Hence, the typical rates of convergence of $\mathrm{Err}_{j,\eta^\star}$ and $\mathrm{Err}_{j,\Lambda^\star}$ are the same for each $j \in \{1, 2, 3\}$. As with any nonparametric function estimation problem, the actual rates of convergence depend on the smoothness assumptions on these functions. Under (semi)parametric models, faster rates are possible. For instance, in the Cox proportional-hazards model, with the cumulative baseline hazard estimated via the Breslow estimator, the cumulative hazard converges at the standard parametric rate $O_p(n^{-1/2})$ for a fixed time point, under standard regularity conditions (Tsiatis, 1981). Its derivative, the hazard function, does not have a well-defined convergence rate since the Breslow estimator is a step function that only increases at observed event times. To address the nonsmoothness of the Breslow estimator, smooth estimators are typically used. They can have different convergence rates, depending on the specific smoothing technique and the smoothness of the underlying true function. Under twice differentiability, the rate is $O_p(n^{-2/(d+4)})$ (Lopuhaä & Musta, 2017).

The asymptotic coverage guarantee for the LPB constructed using the AIPCW-based approach, for a general nonconformity score R, is given in the following theorem. To the best of our knowledge, this result represents a key novelty of the paper, providing the first marginal coverage guarantee for predictive inference on a right-censored time-to-event outcome with a mixed-bias coverage error rate.

THEOREM 3. *Let $\epsilon \in (0, 1)$ be fixed. There exists a universal constant K such that, under Assumptions 1 and 2, with probability at least $1 - \epsilon$ over $\mathcal{D}$,*

$$\begin{aligned}&\mathbb{P}\{R(X, T) \geqslant \hat{\beta}_{\mathrm{AIPCW}} \mid \mathcal{D}\} \\ &> 1 - \alpha - (s_0 + 2\max\{1, s_0 - 1\})\frac{(-\log \epsilon/2)^{1/2} + K}{|\mathcal{D}_2|^{1/2}}\end{aligned}$$

$$- s_0 \sup_{\beta \in [0,1]} \min \Big\{ \int \mathrm{Err}_{2,\eta^\star_\beta}(u) \mathrm{Err}_{3,\Lambda^\star_{C|X}}(u)\, \mathrm{d}u,$$

$$\mathrm{Err}_{1,\eta^\star_\beta} \mathrm{Err}_{1,\Lambda^\star_{C|X}} + \int \mathrm{Err}_{3,\eta^\star_\beta}(u) \mathrm{Err}_{2,\Lambda^\star_{C|X}}(u)\, \mathrm{d}u \Big\},$$

where the probability $\mathbb{P}$ *is taken with respect to a new data point* $(X, T) \sim P_{(X,T)}$.

The proof of Theorem 3 is presented in the Supplementary Material.

Remark 5. While the coverage guarantees for IPCW (Theorem 1) and for AIPCW (Theorem 3) appear to involve different types of estimation errors for the censoring mechanism, with the former expressed in terms of the survival curve $S^\star_{C|X}$ and the latter in terms of the cumulative hazard function $\Lambda^\star_{C|X}$, these errors are closely related due to their smooth relationship, i.e., $\Lambda^\star_{C|X}(u \mid X) = -\log\{S^\star_{C|X}(u \mid X)\}$. This implies that we can bound the error in estimating $S^\star_{C|X}$ with the error in estimating $\Lambda^\star_{C|X}$, and vice versa, ensuring that the rates of convergence for these estimation errors are of the same order.

2.3. *Calibrated outcome regression method*

As discussed in § 1.2, the oracle LPB for the survival time is given by the conditional quantile of $T \mid X$ at level α, $q^\star_{T|X}(\alpha \mid \cdot)$. A naive approach to estimate an LPB is OR, which consists of directly estimating the conditional quantile $\hat{q}_{T|X}(\alpha \mid X)$. However, such an approach does not necessarily guarantee the desired marginal coverage probability. We introduce the COR method, which adjusts the quantile level to ensure proper calibration. This approach ensures that the estimated LPB satisfies the desired probabilistic coverage constraint by leveraging the estimated survival function of $T \mid X$. The implementation is straightforward and similar to that of IPCW and AIPCW methods. First, on the training set $\mathcal{D}_1$, we estimate the survival function $\hat{S}_{T|X}(\cdot \mid X)$ and use it to compute the empirical conditional quantiles $\hat{q}_{T|X}(\beta \mid X)$ over a grid of values for β. Second, on the calibration set $\mathcal{D}_2$, we determine the optimal quantile level by solving

$$\hat{\beta}_{\mathrm{COR}} = \sup \Big\{ \beta \in [0,1] \colon \frac{1}{|\mathcal{D}_2|} \sum_{i \in \mathcal{I}_2} \hat{S}_{T|X}\{\hat{q}_{T|X}(\beta \mid X_i) \mid X_i\} \geqslant 1 - \alpha \Big\}.$$

Finally, we output the LPB $\hat{L}(\cdot) = \hat{q}_{T|X}(\hat{\beta}_{\mathrm{COR}} \mid \cdot)$.

The OR and COR methods provide a computationally efficient way to estimate an LPB while ensuring calibrated marginal and conditional coverage. However, their validity relies on the assumption that $\hat{S}_{T|X}(\cdot \mid X)$ is a consistent estimator of the true survival function. This approach is similar in spirit to maximum likelihood estimation for survival analysis, as it relies directly on modelling the conditional survival function. If the model is misspecified, the resulting predictions may be biased. The OR and COR methods account for dependence between censoring and the target outcome by conditioning on X without directly modelling the censoring distribution. In contrast, IPCW and AIPCW methods explicitly account for censoring through inverse-probability weighting. Although COR has a first-order bias, AIPCW benefits from a second-order mixed bias, making it theoretically more robust. In summary, COR offers a straightforward way to obtain well-calibrated prediction bounds using outcome regression, whereas AIPCW provides additional robustness by leveraging

information from both the censoring and survival models. Importantly, to the best of our knowledge, both OR and COR appear to be novel methods.

3. SIMULATION STUDIES

3.1. *Simulation design and benchmark comparison methods*

We conduct a comprehensive series of simulation studies to evaluate the performance of our proposed methods. The code for reproducing the results of our simulations can be found at https://github.com/rebyfa98/DR_conformal_censored. For each simulation scenario, we generate 100 independent and identically distributed datasets. Each dataset is divided into three sets: the training set $\mathcal{D}_1$, the calibration set $\mathcal{D}_2$ and the test set $\mathcal{D}_3$, with $|\mathcal{D}_1| = |\mathcal{D}_2| = |\mathcal{D}_3| = 1000$. We generate the full data, in which the event time T is observed for all individuals in the test set $\mathcal{D}_3$. Although the full data are not used to implement our methods, they allow us to accurately compute the empirical coverage rate in the test set $\mathcal{D}_3$ as $|\mathcal{D}_3|^{-1} \sum_{i \in \mathcal{I}_3} \mathbf{1}\{T_i \geqslant \hat{L}(X_i)\}$. Moreover, we compute the average estimated LPB in the test set $\mathcal{D}_3$ as $|\mathcal{D}_3|^{-1} \sum_{i \in \mathcal{I}_3} \hat{L}(X_i)$. In all experiments, we set the target coverage level to $1 - \alpha = 0.9$. The grid of points β_j is defined as equally spaced points $\{0.001j,\ j = 0, \ldots, 1000\}$.

Both proposed methods are implemented three times, using three different estimators for the conditional survival curves: the Cox proportional-hazards regression model, random survival forests (RSF) and super learner (SL). The Cox model, widely used in survival analysis, serves as a benchmark because of its popularity, with the cumulative baseline hazard function estimated using the Breslow estimator (Breslow, 1972). However, our primary focus is on nonparametric approaches, which do not impose restrictive modelling assumptions such as proportional hazards. RSF is a suitable choice for this purpose, owing to its flexibility in modelling survival data. For additional flexibility, we also employ SL, an ensemble method that optimally combines multiple survival algorithms by minimizing the cross-validated risk, thereby reducing the risk of overfitting in the final model Golmakani & Polley (2020). To implement these estimators, we use available `R` packages: `survival` for the Cox model, `randomForestSRC` (Ishwaran & Kogalur, 2021) for RSF and `survSuperLearner` for SL (R Development Core Team, 2026). This SL algorithm optimally combines candidate models to estimate both the conditional survival and censoring functions simultaneously. The set of candidate learners includes Kaplan–Meier, Cox proportional hazards, exponential regression, Weibull regression, log-logistic regression and random survival forests. For the COR method and its noncalibrated version (OR), we apply the SL model directly to ensure greater flexibility.

We compare our methods against two groups of benchmarks. First, we consider 'naive' quantile-based baselines that ignore censoring: quantile regression forests fit to the observed time Y (QR-Y) and their conformalized version (CQR-Y), as well as quantile regression forests fitted to the event time T using only uncensored observations (QR-T), and their conformalized version (CQR-T). These approaches serve as reference points because they do not technically account for the censoring mechanism, despite relying on the same assumptions as our IPCW and AIPCW methods. Second, we compare against recent conformal survival methods, namely the approaches proposed by Meixide et al. (2025), Qin et al. (2025), Sesia & Svetnik (2025b), Si & Qiu (2025) and Yi et al. (2025). A brief description of each competing method, together with the implementation details used in our experiments, is provided in the Supplementary Material.

3.2. *Synthetic data*

We examine six synthetic experiments, outlined in Table T.1 of the Supplementary Material, designed to capture a range of censoring regimes and model complexities. In Settings 1–2, both the event time T and censoring time C follow covariate-dependent exponential models, implying that the Cox proportional-hazards assumption holds. In Setting 1, we have completely independent censoring, i.e., C is independent of (T, X), and the observed censoring rate is about 30%. In Setting 2, both T and C depend on nonlinear transformations and interactions of the covariates, yielding informative censoring, and the observed censoring rate is about 50%. In Settings 3–6 we move beyond proportional hazards by using lognormal models, which allow us to assess the performance of nonparametric models such as RSF. In these settings, censoring is covariate dependent, with observed censoring rates of approximately 62% in Setting 3, 70% in Setting 4, 66% in Setting 5 and 50% in Setting 6. Furthermore, Setting 5 includes nonlinear and interaction effects in both the event and the censoring mechanisms. Finally, Setting 6 introduces heteroscedasticity in the time-to-event model.

3.3. *Results*

Figure 1 shows boxplots illustrating the empirical coverage rates obtained from the 100 datasets in each setting, providing a visual comparison of the performance of all methods. For a closer examination of each simulation study, we refer the reader to the enlarged plots in Figs. F.1 and F.2 of the Supplementary Material. Our theoretical results are asymptotic, whereas these experiments are conducted on finite samples; thus, some deviation from the target coverage is anticipated. In addition, Fig. 2 displays boxplots of the average estimated LPB values across the 100 datasets in each setting. In all plots, we highlight our recommended method AIPCW-SL.

Settings 1–2 satisfy proportional hazards, so the Cox-based nuisance models are appropriate. Indeed, in Setting 1 the IPCW and AIPCW methods attain coverage close to the target level, with similar LPBs. The outcome-regression approaches (OR and COR) also perform well because the data-generating model is simple and can be captured by a flexible learner such as SL. Setting 2 is more challenging since censoring is informative and both $T \mid X$ and $C \mid X$ depend on nonlinear transformations or interactions. Here, the IPCW and AIPCW methods remain well calibrated, whereas OR and COR lead to undercoverage and overcoverage, respectively, reflecting the sensitivity of outcome-regression approaches to errors in estimating the survival curve in this nonlinear informative-censoring regime.

In the misspecified Settings 3–6, the Cox-based IPCW and AIPCW procedures break down, producing substantial undercoverage and overcoverage, respectively. When paired with flexible nuisance learners (RSF or SL), both IPCW and AIPCW recover near-nominal coverage. Moreover, AIPCW typically achieves coverage closer to the target and larger average LPBs than IPCW, which aligns with our theoretical results: the augmented construction is doubly robust, and its coverage error is governed by a second-order mixed-bias remainder. Outcome-regression methods perform well in Setting 3, but not in Settings 4–6, again owing to their dependence on accurate estimation of the conditional survival curve, which becomes more challenging in high-dimensional, nonlinear or heteroscedastic regimes.

The naive methods (QR-Y, CQR-Y, QR-T, CQR-T) produce overly conservative bounds across all settings, as evidenced by high coverage rates and low estimated LPBs. Among recent competing methods, Qin et al. (2025) and Sesia & Svetnik (2025b) attained near-nominal coverage in Settings 1–3. However, their performance deteriorates in the

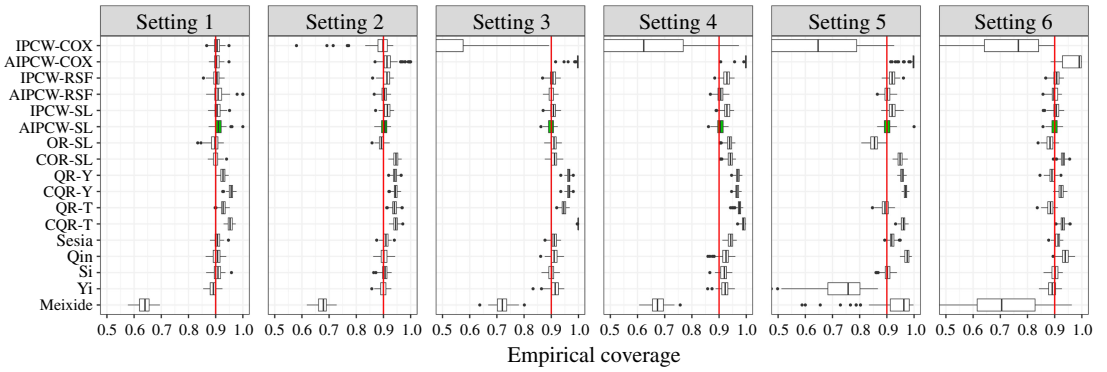


Fig. 1. Distribution of empirical marginal coverage rates across 100 simulated test datasets for each method under Settings 1–6 (Table T.1 in the Supplementary Material). The vertical line denotes the 90% target coverage.

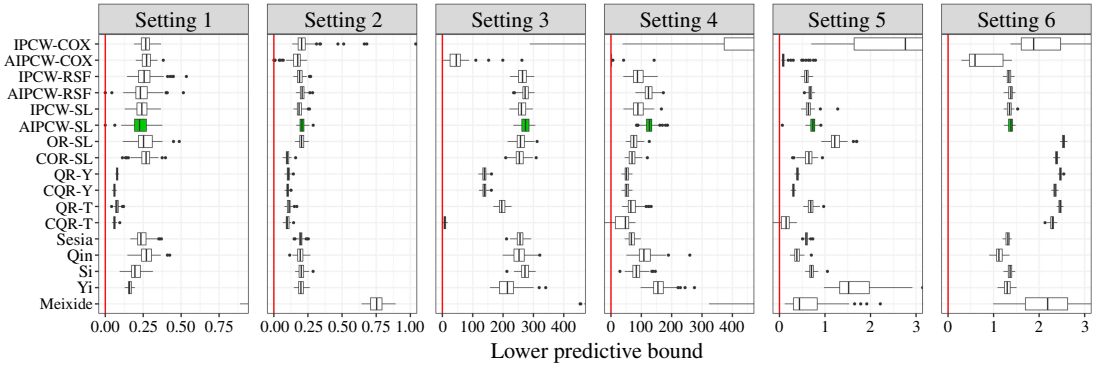


Fig. 2. Distribution of the average estimated LPB in the test set across 100 simulated datasets for each method under Settings 1–6 (Table T.1 in the Supplementary Material). The vertical line at 0 marks the lower limit of the support of the time-to-event outcome.

more complicated Settings 4–6, where the structure and heavier censoring make it more difficult (i) to reliably perform imputation-based approaches (Sesia & Svetnik, 2025b), and (ii) to stably bootstrap-calibrate prediction under limited tail information (Qin et al., 2025), leading to more conservative bounds. The concurrent EIF-based approach of Si & Qiu (2025) provides a valid alternative and behaves similarly to ours in most settings, but its more conservative APAC-style calibration can yield overcoverage in more challenging regimes, as in Setting 4, or slightly less informative LPBs than our AIPCW-SL method, as in Settings 5–6. Likewise, the approach proposed by Yi et al. (2025) performs well when the IPCW-quantile-regression combination adequately captures the target quantile, but it becomes less reliable under more complex censoring mechanisms, as in Setting 5, where it exhibits noticeable undercoverage. Finally, the method proposed by Meixide et al. (2025), which is designed for interval-censoring data, is not directly competitive here.

Overall, our results indicate that the proposed AIPCW approach offers a favourable and comparatively stable trade-off between validity and bound sharpness across the range of scenarios considered. As expected, naive methods that ignore censoring tend to be conservative, producing high coverage and smaller LPBs. Competing methods can match the performance of our methods in particular scenarios; however, in our experiments none simultaneously improves upon AIPCW-SL in terms of calibration while also yielding more informative

bounds across all settings. These findings support recommending AIPCW-SL as a practical default, as it maintains near-nominal coverage and informative LPBs.

Additional simulation results, including studies across different target coverage levels and training sample sizes, conditional coverage experiments and further results on the performance of competing methods, are provided in the Supplementary Material.

4. Real data application

4.1. *Data description*

We use data from a cohort of 1240 patients with rheumatoid arthritis from the Wichita Arthritis Center, an outpatient rheumatology facility (Choi et al., 2002). The original study aimed to evaluate the impact of methotrexate, the most commonly used disease-modifying antirheumatic drug (DMARD), on patient mortality. Our goal is to apply our methods to estimate predictive bounds for the survival time to death.

During the data collection, each patient underwent longitudinal follow-up, with multiple assessments recorded over time. For our analysis, we focus on baseline covariates to account for patient characteristics at the start of the study. Among the 1240 patients, 191 died during follow-up, resulting in an observed event rate of 15.4%. The average follow-up time was 72.3 months. The covariates considered in our analysis include age, sex, disease duration, education level, source of enrolment, prednisone use, DMARD use, treatment contraindications, rheumatoid factor positivity, smoking status, and several clinical measures such as functional disability score, global assessment of disease score and tender joint count. This dataset provides a robust framework for evaluating survival prediction methods in a real-world clinical setting where censoring is common.

4.2. *Performance assessment*

To evaluate the performance of our methods on this dataset, we estimate the empirical coverage of the predictive lower bounds using two different approaches to account for censoring: one based on inverse-probability-of-censoring weighting (IPCW and AIPCW) and the other based on outcome regression. Unlike in the simulation study, where all event times were available for validation (for both uncensored and censored observations), here we only have access to event times for uncensored outcomes, thus requiring different techniques for estimating coverage.

The IPCW/AIPCW-based empirical coverage is computed by reweighting the observed failures to correct for censoring. Specifically, given an estimated optimal level β and a test dataset $\mathcal{D}_3$ with index set $\mathcal{I}_3$, the IPCW-based empirical coverage is computed as

$$\frac{\sum_{i\in\mathcal{I}_3} \Delta_i \mathbf{1}\{R(X_i,T_i) \geqslant \hat{\beta}\}/\hat{S}_{C|X}(T_i \mid X_i)}{\sum_{i\in\mathcal{I}_3} \Delta_i/\hat{S}_{C|X}(T_i \mid X_i)}.$$

In other words, the empirical coverage is estimated as the weighted proportion of observed failure times that exceed the lower predictive bound, using inverse probability weights derived from the estimated censoring survival function. Similarly, the AIPCW-based coverage computation extends this by incorporating an augmentation term.

The OR-based empirical coverage is computed as $|\mathcal{D}_3|^{-1}\sum_{i\in\mathcal{I}_3}\hat{\eta}(\hat{\beta},0,X_i)$. In fact, $\hat{\eta}(\beta,u \mid X)$ is an estimator of $\eta^\star(\beta,u \mid X) = \mathbb{E}[\mathbf{1}\{R(X,T) \geqslant \beta\} \mid X, T \geqslant u]$. When

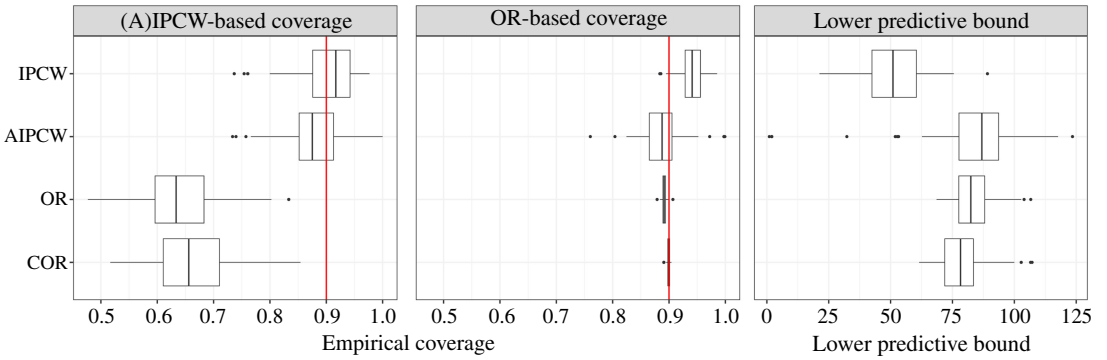


Fig. 3. Evaluation on the real-world dataset across 100 random data splits. The left and centre panels display the distribution of empirical coverage using (A)IPCW-based coverage evaluation (left) and OR-based coverage evaluation (centre), with vertical lines marking the 90% target coverage. The right panel displays the average LPB.

the nonconformity score is defined through the conditional quantile function of $T \mid X$, the OR-based empirical coverage reduces to $|\mathcal{D}_3|^{-1} \sum_{i \in \mathcal{I}_3} \hat{S}_{T|X}\{\hat{q}_{T|X}(\hat{\beta} \mid X_i) \mid X_i\}$.

Both metrics provide valid but complementary assessments of predictive coverage. By using both evaluation strategies, we obtain a more comprehensive understanding of the calibration and reliability of our methods in this real-world setting.

4.3. *Results*

We implement our methods (OR, COR, IPCW, AIPCW) across 100 different random splits of the training, calibration and test sets, using the SL algorithm. The estimated coefficients of the candidate models used within the SL algorithm, are provided in the Supplementary Material. Figure 3 presents boxplots of the empirical coverage in the test set for each of these 100 splits.

In the evaluation based on the (A)IPCW metric (left panel), the AIPCW estimator is assessed using AIPCW-based coverage, while the other three methods are evaluated through IPCW-based coverage. The results indicate that IPCW and AIPCW outperform the regression-based methods, though they exhibit slight deviations from the nominal level, with IPCW tending to overcover and AIPCW slightly undercovering, on average. These deviations are expected, given the relatively small sample size, as our theoretical results hold asymptotically. Moreover, considering the low event rate in this dataset, the performance of our methods is strong.

The OR-based evaluation approach (centre panel) produces almost perfect coverage for the OR and COR methods. However, this result is a consequence of the fact that the coverage is computed using the same metric employed to estimate the LPBs. Specifically, for the OR method, by construction, $\hat{S}_{T|X}\{\hat{q}_{T|X}(\hat{\beta} \mid X_i) \mid X_i\} = 1 - \alpha$, since the conditional quantile is directly derived from the estimated survival curve. For COR, we observe a slight deviation from the target level due to the additional calibration step. Meanwhile, the IPCW and AIPCW methods continue to exhibit reasonable coverage even when evaluated using this alternative approach, further reinforcing their reliability.

Additionally, we compute the average estimated LPB values in the test set across the 100 cross-validation splits. The right panel of Fig. 3 displays the distribution of these mean LPB estimates for each of the four methods, showing comparable results across methods. IPCW tends to produce smaller LPBs than the others, reflecting a more conservative approach.

Acknowledgement

The authors thank the associate editor and the reviewers for helpful comments. A.K.K. and E.T.T. were supported by the NSF (DMS-2210662). E.T.T. is grateful for funding from the NIH (AG065276).

Supplementary material

The Supplementary Material includes detailed proofs of all the main results presented in the paper, along with supporting lemmas. It also provides algorithms for the proposed procedures for a general nonconformity score, a summary table of the simulation settings, a description of the competing methods and their implementations, and additional simulation results.

References

Basu, D. (1970). An essay on the logical foundations of survey sampling, part 1. In *Foundations of Statistical Inference*, Ed. V. Godambe and D. A. Sprott, pp. 203–42. Toronto: Holt, Rinehart and Winston.

Boström, H., Asker, L., Gurung, R., Karlsson, I., Lindgren, T. & Papapetrou, P. (2017). Conformal prediction using random survival forests. In *2017 16th IEEE Conf. Mach. Learn. Applic.*, pp. 812–17. Piscataway, NJ: IEEE Press.

Boström, H., Johansson, U. & Vesterberg, A. (2019). Predicting with confidence from survival data. In *Proc. 8th Symp. Conformal Prob. Prediction Applic*. (Proc. Mach. Learn. Res. 105), pp. 123–41. JMLR.

Breslow, N. E. (1972). Contribution to discussion of paper by Dr Cox. *J. R. Statist. Soc. B* **34**, 216–17.

Candès, E., Lei, L. & Ren, Z. (2023). Conformalized survival analysis. *J. R. Statist. Soc. B* **85**, 24–45.

Chen, W., Wang, Z., Ha, W. & Barber, R. F. (2016). Trimmed conformal prediction for high-dimensional models. *arXiv:* 1611.09933v1.

Chernozhukov, V., Wüthrich, K. & Zhu, Y. (2021). Distributional conformal prediction. *Proc. Nat. Acad. Sci*. **118**, e2107794118.

Choi, H. K., Hernán, M. A., Seeger, J. D., Robins, J. M. & Wolfe, F. (2002). Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *The Lancet* **359**, 1173–7.

Collett, D. (2023). *Modelling Survival Data in Medical Research*. Boca Raton, FL: Chapman and Hall/CRC Press.

Cox, D. R. (1972). Regression models and life-tables. *J. R. Statist. Soc. B* **34**, 187–220.

Datta, J. & Polson, N. (2025). Inverse probability weighting: from survey sampling to evidence estimation. *arXiv*: 2204.14121v3.

Fleming, T. R. & Harrington, D. P. (2013). *Counting Processes and Survival Analysis*. New York: John Wiley and Sons.

Golmakani M. K. & Polley E. C. (2020). Super learner for survival data prediction. *Int. J. Biostatist*. **16**, 20190065.

Gui, Y., Hore, R., Ren, Z. & Barber, R. F. (2023). Conformalized survival analysis with adaptive cut-offs. *Biometrika* **111**, 459–77.

Hájek, J. (1971). Comment on 'An essay on the logical foundations of survey sampling, part 1'. In *Foundations of Statistical Inference*, Ed. V. Godambe and D. A. Sprott, p. 236. Toronto: Holt, Rinehart and Winston.

Horvitz, D. G. & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Assoc*. **47**, 663–85.

Hosmer, D. W. Jr, Lemeshow, S. & May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. New York: John Wiley and Sons.

Ishwaran, H. & Kogalur, U. (2021). randomForestSRC - Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edn. New York: John Wiley and Sons.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc*. **53**, 457–81.

Kleinbaum, D. G. & Klein, M. (1996). *Survival Analysis a Self-Learning Text*. New York: Springer.

Lopuhaä, H. P. & Musta, E. (2017). Smooth estimation of a monotone hazard and a monotone density under random censoring. *Statist. Neer*. **71**, 58–82.

Meixide, C. G., Kosorok, M. R. & Matabuena, M. (2025). Uncertainty quantification for intervals. *arXiv:* 2408.16381v2.

PARK, J., KIM, D., KIM, C., PARK, H., KANG, S. & KIM, G. (2025). KSP: Kolmogorov-Smirnov metric-based post-hoc calibration for survival analysis. In *39th Ann. Conf. Neural Info. Proces. Syst.*

POCOCK, S. J. (2013). *Clinical Trials: A Practical Approach*. New York: John Wiley and Sons.

QIN, J., PIAO, J., NING, J. & SHEN, Y. (2025). Conformal predictive intervals in survival analysis: a resampling approach. *Biometrics* **81**, ujaf063.

R DEVELOPMENT CORE TEAM (2026). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

ROBINS, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proc. Biopharm. Sect., Am. Statist. Assoc*., pp. 24–33. Alexandria, VA: American Statistical Association.

ROBINS, J. M. & FINKELSTEIN, D. M. (2000). Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56**, 779–88.

ROBINS, J. M. & ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*, Ed. N. P. Jewell, K. Dietz and V. T. Farewell, pp. 297–331. Boston: Birkhäuser.

ROTNITZKY, A., SMUCLER, E. & ROBINS, J. M. (2021). Characterization of parameters with a mixed bias property. *Biometrika* **108**, 231–8.

SESIA, M. & SVETNIK, V. (2025a). Conformal survival bands for risk screening under right-censoring. In *Proc. 14th Symp. Conformal Prob. Prediction Applic*. (Proc. Mach. Learn. Res. **266**), Ed. K. A. Nguyen, Z. Luo, H. Papadopoulos, T. Löfström, L. Carlsson and H. Boström, pp. 464–514. JMLR.

SESIA, M. & SVETNIK, V. (2025b). Doubly robust conformalized survival analysis with right-censored data. In *Proc. 42nd Int. Conf. Mach. Learn.* (Proc. Mach. Learn. Res. **267**), pp. 53989–4029. JMLR.

SI, W. & QIU, H. (2025). Training-set conditionally valid prediction sets with right-censored data. *arXiv:* 2505.02213v1.

STANKEVICIUTE, K., M ALAA, A. & VAN DER SCHAAR, M. (2021). Conformal time-series forecasting. In *Proc. 35th Int. Conf. Neural Info. Proces. Syst*., pp. 6216–28. Red Hook, NY: Curran Associates.

TENG, J., TAN, Z. & YUAN, Y. (2021). T-SCI: a two-stage conformal inference algorithm with guaranteed coverage for Cox-MLP. In *Proc. 38th Int. Conf. Mach. Learn.* (Proc. Mach. Learn. Res. **139**), pp. 10203–13. JMLR.

TIBSHIRANI, R. J., FOYGEL BARBER, R., CANDÈS, E. & RAMDAS, A. (2019). Conformal prediction under covariate shift. In *Proc. 33rd Int. Conf. Neural Info. Proces. Syst*., pp. 2530–40. Red Hook, NY: Curran Associates.

TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108.

TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

VAN DER LAAN, M. & ROBINS, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer.

VOVK V. GAMMERMAN A. SHAFER G. 2005. *Algorithmic Learning in a Random World*. Boston, MANew York. Springer. Springer.

WAN, Y. & YI, G. (2025). Distribution-free lower predictive bounds for right-censored time-to-event data via hybrid quantile learning and DFT-adaptive conformal calibration. https://openreview.net/forum?id=aMXVp1QK2Q

YANG, Y., KUCHIBHOTLA, A. K. & TCHETGEN TCHETGEN, E. (2024). Doubly robust calibration of prediction sets under covariate shift. *J. R. Statist. Soc. B* **86**, 943–65.

YI, M., XIAO, Z., JUDY WANG, H. & TANG, Y. (2025). Survival conformal prediction under random censoring. *Stat* **14**, e70052.

ZAFFRAN, M., JOSSE, J., ROMANO, Y. & DIEULEVEUT, A. (2024). Predictive uncertainty quantification with missing covariates. *arXiv:* 2405.15641v1.

[*Received on* 12 *March* 2025. *Accepted on* 1 *June* 2026]