Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(10): 7707–7738.
DOI: 10.3934/era.2026332
Received: 20 June 2026
Revised: 02 September 2026
Accepted: 03 September 2026
Published: 08 September 2026

Research article

Auxiliary-survival-probabilities-informed survival analysis under Cox model

Jie Ding[1,*] **and Bo Han**[2]

[1] School of Science, Dalian Maritime University, Liaoning 116026, China
[2] School of Mathematics and Statistics, Yunnan University, Yunnan 650091, China

* **Correspondence:** Email: stat.jie@dlmu.edu.cn.

Abstract: Transferring summary-level auxiliary information from source domains to enhance statistical analysis in target domains is increasingly important, yet settings involving right-censored survival data remain underexplored. To address this gap, we developed a general methodology for incorporating subgroup Kaplan–Meier survival probabilities at multiple time points, a widely reported form of auxiliary information. Our framework handles both estimation of an abstractly defined target parameter, such as a regression coefficient or a causal estimand, and hypothesis testing for it. We first constructed a target-only estimator that possesses a unified structure applicable to a broad class of target parameters. This estimator was then linked, via a leveraging factor, to a transitional element computed from the target dataset in a manner analogous to the available auxiliary information. We proposed and compared two versions of this element. The model-free version is universally applicable and robust, whereas the model-based version delivers improved efficiency under a partly independent censoring assumption that has often been overlooked in existing literature. Our method requires no individual-level source data and maintains low computational complexity. Theoretical derivations and numerical experiments together demonstrated gains in estimation efficiency and testing power. In addition, Python implementations of the proposed approaches were provided.

Keywords: Cox model; partly independent censoring; summary-level auxiliary information; transfer learning; transitional element

1. Introduction

Survival data, which involve time-to-event outcomes, are widely encountered across diverse fields such as precision medicine and financial risk management. Statistical methods developed for this type of data have proven highly valuable in practice, effectively supporting core analytical tasks. For instance, they can evaluate whether the effects of potential risk factors on the event time of interest

are statistically significant [1, 2]. However, due to practical constraints such as difficulties in recruiting patients with rare diseases, limited funding for clinical trials, and the high cost of long-term follow-up, survival data often suffer from small sample sizes. This small-sample dilemma can severely undermine the performance of traditional survival analysis methods, highlighting an urgent need for novel strategies that improve statistical performance when survival data are scarce.

With the diversification of information storage and access channels, external studies addressing identical or similar problems have been increasing. These studies are often presented to the public in various forms, such as public databases maintained by authoritative institutions and published research literature. They frequently contain information of potential value for analyzing the target problem and are collectively referred to as auxiliary information. As one of the cutting-edge data analysis techniques, transfer learning has shown great potential in integrating data from the target domain with auxiliary information in source domains, offering novel ideas for achieving more precise statistical modeling [3]. There is a growing need to enrich the toolbox of methods available in the context of transfer learning.

Transfer learning based on individual-level data from the source domain has attracted considerable attention from statisticians in recent years due to the directness of such auxiliary information. For instance, in the context of non-survival data in the target domain, Cai et al. [4] and Li et al. [5] investigated transfer learning methods based on linear models. Subsequently, Tian and Feng [6] and Li et al. [7] extended this line of research to generalized linear models. Inspired by this series of pioneering works, Li et al. [8] and Pei et al. [9] respectively addressed the problem of transfer learning under right-censored survival data based on the renowned Cox proportional hazards model [10], often referred to simply as the Cox model, and the accelerated failure time model, while Xie et al. [11] further extended it to interval-censored data. Additionally, Deng et al. [12] proposed a method capable of integrating interval-censored individual-level data from the source domain with right-censored target domain data. All the aforementioned methods rely on the premise that individual-level data in the source domain can be directly accessed. However, in practice, it is often difficult to extract comprehensive raw data directly, due to constraints such as commercial value and personal privacy, among others, restricting the application of the aforementioned methods in practice.

Transferring auxiliary information in forms of summary-level statistics extracted from the source domain can circumvent direct reliance on the individual-level source data itself, but the construction of corresponding transfer learning methods is more challenging [13]. Synthesizing this type of auxiliary information constitutes a recent critical research focus from statisticians [14, 15]. For example, as a type of auxiliary information unique to and widely present in survival analysis, the transfer problem of subgroup survival probabilities at given time points under different survival models has been a key topic of discussion in numerous studies. Starting from the empirical likelihood technique [16–18], Huang et al. [19] and Chen et al. [20] proposed new estimators under the Cox model. As an extension, He et al. [21], Han et al. [22], and Cheng et al. [23] respectively discussed the additive hazards model, non-mixture cure model and the transformation model. Based on the generalized method of moments [24], Shang and Wang [25], Sheng et al. [26], and Ding et al. [27] respectively investigated methods for transferring this auxiliary information under the additive multiplicative hazards model, accelerated failure time model, and additive hazards model. Notably, Ding et al. [28] and Ding et al. [29] adopted the control variate method to implement a transfer strategy for the same type of Euclidean auxiliary information under the mixture cure model. Other Euclidean auxiliary information

with specific practical meanings has also garnered widespread attention, including but not limited to coefficients from source models [30–32], incidence rates [33], and restricted mean survival times [34], among others.

The main contributions of this paper are summarized as follows. First, we develop a transfer learning framework for right-censored survival data that incorporates aggregate auxiliary information in the form of subgroup Kaplan–Meier survival probabilities at (possibly) multiple time points. Unlike earlier works such as Huang et al. [19], which was limited to estimating regression coefficients using subgroup survival probabilities at a single time point, our framework accommodates a broad class of target parameters, including regression coefficients, personalized survival probabilities, and causal estimands as special cases, and can handle auxiliary information from multiple subgroups and time points simultaneously. Second, we introduce a novel transfer mechanism based on a transitional element computed from the target dataset and a leveraging factor that optimally combines the target-only estimator with the auxiliary information. This mechanism requires no access to individual-level source records and thus preserves privacy. Third, we propose and compare two types of transitional elements. The model-free version is universally applicable and robust against violations of the censoring assumption, whereas the model-based version achieves higher estimation efficiency under a partly independent censoring assumption that has been largely overlooked in the existing literature. We demonstrate that neglecting this assumption can lead to biased estimation, a cautionary point that distinguishes our study from prior work. Fourth, we establish the asymptotic normality of the proposed estimators and derive closed-form expressions for the efficiency gain. Extensive numerical experiments confirm that our method yields substantial improvements in both estimation efficiency and testing power over target-only approaches. Collectively, these contributions provide a principled and flexible solution for integrating summary-level external information into survival analysis.

The remainder of this article is structured as follows. Section 2 introduces the model setup and our focused tasks. Our focused auxiliary information is formally introduced in Section 3. We present the methodology without auxiliary information in Section 4, while Section 5 describes the methodology that takes the auxiliary information into consideration. Two common types of transitional elements are shown in Section 6. Section 7 reports simulation studies under various settings. Finally, Section 8 offers concluding remarks. Numerical studies are implemented in the Python environment, and we have created a Python package, aisurv, which is publicly available.

2. Model and problem setup

In this section, we introduce the model setup and clarify our focused statistical tasks. Our accessible observations from the target domain are highlighted as well.

2.1. The Cox model and target dataset

Let T be a non-negative continuous random variable representing the time-to-event outcome, and let $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ be a p-dimensional vector of covariates. To model the relationship between T and $\boldsymbol{X}$, the Cox model has long served as a cornerstone across diverse applied fields [35]. Specifically, we assume that the conditional survival function of T given $\boldsymbol{X} = \boldsymbol{x}$ follows the Cox model as follows:

$$S(t|\boldsymbol{x}) = \mathbb{P}(T > t|\boldsymbol{X} = \boldsymbol{x}) = \exp\left\{-\Lambda_0^\star(t)\exp(\boldsymbol{x}^\top\boldsymbol{\beta}^\star)\right\}, \tag{2.1}$$

where $\boldsymbol{\beta}^\star = (\beta_1^\star, \ldots, \beta_p^\star)^\top$ is an unknown p-dimensional vector of true regression coefficients and $\Lambda_0^\star(t)$ is a completely unspecified true cumulative baseline hazard function. To distinguish it from the true $\boldsymbol{\beta}^\star$, we also introduce the generic symbol $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ (without the asterisk) to denote a variable argument in our subsequent discussion, for instance, when constructing likelihood functions. Given the central role of the Cox model in survival analysis, many key problems have been explored starting from it, thereby offering crucial insights that facilitate the development of more complex models.

In the routine of survival analysis, a target dataset within the target domain, denoted by $\mathcal{O}_n$, is often accessible. To define $\mathcal{O}_n$ more formally, we first note that survival times generated from T are not always observed due to censoring [36,37]. We focus on the most commonly encountered random right censoring mechanism, under which, instead of directly observing T, we observe $Y = \min\{T, C\}$ and $\Delta = \mathbb{I}_{\{T \le C\}}$ for each subject, where C is a random variable denoting the right censoring time and $\mathbb{I}_{\{\cdot\}}$ denotes an indicator function. Consequently, $\mathcal{O}_n$ is defined as

$$\lfloor \text{Target Dataset} \rfloor \quad \mathcal{O}_n = \{(Y_i, \Delta_i, \boldsymbol{X}_i) : i \in [n]\},$$

which consists of n independent and identically distributed (I.I.D.) observations of $(Y, \Delta, \boldsymbol{X})$, where for any positive integer d, we have defined the set $[d] = \{1, \ldots, d\}$.

Remark 1. The censoring time C partially obscures information about T. To enable meaningful analysis, constraints on the dependence structure between T and C are necessary. A common strategy adopted in the existing literature is to impose the following conditional independence assumption:

$$\lfloor \text{Conditionally Independent Censoring} \rfloor \quad T \perp C \mid \boldsymbol{X}. \tag{2.2}$$

This assumption requires that the event time and the censoring time are independent within each covariate stratum. It effectively rules out the presence of any unmeasured confounder that influences both the survival time and the censoring time.

2.2. Target parameter of interest

Consider a scenario in which there is a target parameter (that is, a parameter of interest) regarding the Cox model in Eq (2.1), abstractly denoted by $\alpha^\star$. Throughout this paper, we focus on developing valid estimation and inferential procedures for $\alpha^\star$. Specifically, we aim to accomplish two tasks:

Task 1. Estimate $\alpha^\star$ and provide confidence intervals for the corresponding estimator.

Task 2. Conduct statistical inference on the hypothesis testing problem $\mathcal{H}_0 : \alpha^\star = a$ versus $\mathcal{H}_1 : \alpha^\star \neq a$, where a is a pre-specified real number.

Although we focus on a one-dimensional target parameter $\alpha^\star$ for ease of presentation, the subsequently proposed methodology can be readily extended to a multi-dimensional vector of parameters as well in a similar manner. Below, we provide three representative examples of $\alpha^\star$.

Example 1. *When the objective is to assess the effect of a particular covariate on the survival time T, where the covariate may represent, for instance, whether a patient received a certain treatment or whether a specific economic policy was implemented, we can define $\alpha^\star$ as $\alpha^\star = \beta_j^\star$ for some $j \in [p]$. Notably, the exponential of $\alpha^\star$ defined here is the widely used hazard ratio.*

Example 2. *When the focus is on evaluating a personalized survival probability, that is, the likelihood that a given subject will survive up to a future time $t^\star$, we can take $\alpha^\star$ to be $\alpha^\star = S(t^\star|\boldsymbol{x}^\star)$, where $\boldsymbol{x}^\star$ denotes the covariate profile of the individual in question.*

Example 3. *One may be interested in characterizing the causal effect of a particular variable contained in $\boldsymbol{X}$. Specifically, write $\boldsymbol{X} = (W, \boldsymbol{Z}^\top)^\top$, where W is a binary treatment variable with $W = 1$ for treatment and $W = 0$ for control. Also let $T(w)$ be the potential survival outcome under treatment $W = w$. Then, we can take $\alpha^\star$ as the contrasts of survival probabilities:*

$$\alpha^\star = S_1(t^\star) - S_0(t^\star), \tag{2.3}$$

where $S_w(t) = \mathbb{P}(T(w) > t)$ for $t \geq 0$ is the marginal survival function of the potential outcomes for $w = 0$ and 1, and $t^\star$ is a pre-specified time point.

In practice, different target parameters may arise in different contexts, so that the specific form of $\alpha^\star$ varies depending on the statistical question under investigation. We aim to provide a unified methodology that can handle Tasks 1 and 2 concerning commonly encountered choices of $\alpha^\star$, including Examples 1 to 3, as well as others as special cases.

3. Auxiliary information and its uncertainty

To address Tasks 1 and 2, conventional approaches rely solely on the target dataset $\mathcal{O}_n$. In contrast to this setting, we assume that, in addition to the individual-level target dataset $\mathcal{O}_n$, we have access to auxiliary information obtained from source domains that are independent of the target domain. In this section, we introduce the auxiliary information of interest and examine its properties.

3.1. Kaplan–Meier estimators across subgroups

In the context of survival analysis, the applied literature frequently reports estimated survival probabilities at representative time points for subgroups of interest. For example, one may encounter the 5-year survival probability for lung cancer patients aged over 55 years, which helps readers gain a general understanding of the dataset under consideration. These quantities are typically derived from Kaplan–Meier estimators [38] and are commonly reported in published studies.

Suppose for simplicity that all accessible survival probabilities are extracted from a single source domain with a source dataset $\mathcal{O}^\circ_{n^\circ}$, where n° denotes the sample size. The extension to multiple source domains can be handled similarly. Then, define $\mathcal{O}^\circ_{n^\circ}$ as follows:

$$\mathcal{O}^\circ_{n^\circ} = \{(Y^\circ_i, \Delta^\circ_i, \boldsymbol{X}^\circ_i) : i \in [n^\circ]\}, \tag{3.1}$$

which consists of n° I.I.D. observations from $(Y, \Delta, \boldsymbol{X})$. Moreover, let $\mathbb{X}$ be the range space of $\boldsymbol{X}$. Consequently, for each given $k \in [K]$, define

$$\hat{\phi}^\circ_k = \prod_{i=1}^{n^\circ} \left[1 - \frac{\Delta^\circ_i \mathbb{I}_{\{\boldsymbol{X}^\circ_i \in \Omega_k\}} \mathbb{I}_{\{Y^\circ_i \leq t^*_k\}}}{\sum_{j=1}^{n^\circ} \mathbb{I}_{\{\boldsymbol{X}^\circ_j \in \Omega_k\}} \mathbb{I}_{\{Y^\circ_j \geq Y^\circ_i\}}} \right], \tag{3.2}$$

where t^*_k is a pre-specified time point and $\Omega_k \subseteq \mathbb{X}$ is a set that characterizes a subgroup. Essentially, for each $k \in [K]$, $\hat{\phi}^\circ_k$ is a Kaplan–Meier estimator computed using only $\{(Y^\circ_i, \Delta^\circ_i) : \boldsymbol{X}^\circ_i \in \Omega_k, i \in [n^\circ]\}$,

a subset of the source dataset $\mathcal{O}^{\circ}_{n^{\circ}}$ in Eq (3.1). For ease of presentation, let us formulate the auxiliary information as follows:

$$\lfloor \text{Auxiliary Information} \rfloor \quad \hat{\boldsymbol{\phi}}^{\circ} = (\hat{\phi}^{\circ}_1, \ldots, \hat{\phi}^{\circ}_K)^{\top} = \mathcal{M}(\mathcal{O}^{\circ}_{n^{\circ}}), \tag{3.3}$$

where $\mathcal{M}$ is a completely known mapping defined by Eq (3.2) above that links the source dataset $\mathcal{O}^{\circ}_{n^{\circ}}$ to the accessible auxiliary information $\hat{\boldsymbol{\phi}}^{\circ}$. Throughout this paper, we assume that the integer K, the number of accessible auxiliary values, is a fixed constant.

Remark 2. For each given $k \in [K]$, the auxiliary information $\hat{\phi}^{\circ}_k$ is intended to estimate the survival probability at t^*_k for the subgroup characterized by Ω_k. In other words, the target quantity is

$$\phi^{\#}_k = \mathbb{P}(T > t^*_k | \boldsymbol{X} \in \Omega_k), \tag{3.4}$$

for $k \in [K]$. Under standard regularity conditions, $\hat{\boldsymbol{\phi}}^{\circ}$ is indeed a consistent estimator for $\boldsymbol{\phi}^{\#} = (\phi^{\#}_1, \ldots, \phi^{\#}_K)^{\top}$, but a key assumption on the censoring mechanism is required beyond the conditionally independent censoring assumption (2.2). Specifically, one needs to establish that within each subgroup Ω_k, the censoring time C is independent of the survival time T, as will be discussed in detail in Section 3.2. This stronger condition does not follow from our earlier assumption unless the subgroup Ω_k ($k \in [K]$) is defined by a level of $\boldsymbol{X}$ that renders the conditional independence automatic. This subtle but important distinction has been overlooked in the existing literature on synthesizing auxiliary survival probabilities. Many works proceed by directly assuming that $\hat{\boldsymbol{\phi}}^{\circ}$ consistently estimates $\boldsymbol{\phi}^{\#}$ without verifying the extra censoring condition, potentially leading to biased inference when the condition fails. We will discuss the asymptotic behavior of $\hat{\boldsymbol{\phi}}^{\circ}$ under different censoring scenarios in Section 3.2.

Remark 3. Although it appears that the subgroups are defined by evaluating the entire covariate vector $\boldsymbol{X}$, only some components of them are sometimes sufficient. More generally, for $k \in [K]$, we have:

$$\boldsymbol{X} \in \Omega_k \quad \Longleftrightarrow \quad \boldsymbol{g}_k(\boldsymbol{X}) \in \Omega^{\circ}_k,$$

where $\boldsymbol{g}_k$ is vector-valued function taking values in $\mathbb{R}^{p_k}$ for some integer p_k and $\Omega^{\circ}_k \subseteq \mathbb{R}^{p_k}$. As a special case, for some sets $\mathcal{A} \subset [p]$, let $\boldsymbol{X}_{\mathcal{A}}$ denote a sub-vector of $\boldsymbol{X}$ containing the components whose indices belong to $\mathcal{A}$ and set $\Omega^{\circ}_k \subseteq \mathbb{X}_{\mathcal{A}}$ with $\mathbb{X}_{\mathcal{A}}$ being the range space of $\boldsymbol{X}_{\mathcal{A}}$, where we use the notation $\boldsymbol{a}_{\mathcal{A}}$ to denote a subvector of the p-dimensional vector $\boldsymbol{a}$. Consequently, the source dataset $\mathcal{O}^{\circ}_{n^{\circ}}$ defined in Eq (3.1) can be replaced by $\{(Y_i, \Delta_i, \boldsymbol{X}_{i,\mathcal{A}}) : i \in [n^{\circ}]\}$, which indicates that the source domain is allowed to contain only the covariates with indices in $\mathcal{A}$.

Our key objective is to investigate how to incorporate $\hat{\boldsymbol{\phi}}^{\circ}$ into the analysis of the target dataset $\mathcal{O}_n$, with the goal of estimating the target parameter $\alpha^{\star}$ and conducting inference on it, while achieving improved estimation efficiency or testing power relative to target-only methods. We emphasize that our focus is on the setting where only summary-level statistics derived from the source dataset $\mathcal{O}^{\circ}_{n^{\circ}}$, as specified in Eq (3.3), are available as auxiliary information. This setup distinguishes our work from much of the existing literature, as noted in Section 1, because it eliminates the need to access the raw individual-level records in $\mathcal{O}^{\circ}_{n^{\circ}}$.

Remark 4. The auxiliary information considered in this paper generalizes the setting of Huang et al. [19], who mainly focused on subgroup survival probabilities at a fixed time point for the purpose

of estimating regression coefficients under the Cox model. In contrast, our framework accommodates multiple time points and multiple subgroups simultaneously, and the target parameter $\alpha^\star$ is allowed to be a wide variety of quantities, including regression coefficients, personalized survival probabilities, and causal estimands as special cases. This flexibility is achieved by the unified target-only estimator introduced in Section 4 and the transitional element construction in Section 6, which together enable information transfer without restricting the form of the target parameter. The ability to handle more complex structures and broader target parameters highlights the advantage of our method.

3.2. Characterization of randomness

When the sample size n° of the source dataset is sufficiently large, the randomness in estimating $\hat{\boldsymbol{\phi}}^\circ$ becomes negligible and we may directly treat $\hat{\boldsymbol{\phi}}^\circ$ as a non-random quantity. However, when n° is not substantially larger than the target sample size n, the randomness is non-negligible and must be accounted for. To better understand this issue, we proceed to examine the asymptotic properties of the accessible auxiliary information. Notably, each component of $\hat{\boldsymbol{\phi}}^\circ$ is a Kaplan–Meier estimator, so one might attempt to directly apply standard asymptotic results from the literature. Those results typically require the following assumption:

$$\lfloor\text{Partly Independent Censoring}\rfloor \quad T \perp C \mid \boldsymbol{X} \in \Omega_k \quad \text{for} \quad k \in [K], \tag{3.5}$$

which is stronger than our original conditionally independent censoring assumption displayed in Eq (2.2). Indeed, the latter does not imply the former unless the subgroup Ω_k is defined solely by levels of $\boldsymbol{X}$ that preserve conditional independence. Unfortunately, this additional restriction is often overlooked in existing work that synthesizes auxiliary survival probabilities. We proceed to establish the asymptotic properties of the auxiliary information $\hat{\boldsymbol{\phi}}^\circ$ without this assumption.

To avoid relying on this restrictive assumption, we characterize the dependence between T and C given $\boldsymbol{X} \in \Omega_k$ for each $k \in [K]$ using copula techniques. By Sklar's theorem [39], for each $k \in [K]$, there should exist a unique copula C_k such that

$$\mathbb{P}(T > t, C > c|\boldsymbol{X} \in \Omega_k) = C_k(S_{T,k}(t), S_{C,k}(c))$$

holds for any $t, c \geq 0$, where C is a bivariate function with uniform margins on $[0,1]^2$, $S_{T,k}(t) = \mathbb{P}(T > t|\boldsymbol{X} \in \Omega_k)$, and $S_{C,k}(c) = \mathbb{P}(C > c|\boldsymbol{X} \in \Omega_k)$. Then, for each $k \in [K]$, let us further define

$$\lambda_k^{\#}(t) = \frac{-\frac{\partial}{\partial u}\mathbb{P}(T > u, C > t|\boldsymbol{X} \in \Omega_k)\big|_{u=t}}{\mathbb{P}(T > u, C > t|\boldsymbol{X} \in \Omega_k)} = \lambda_k(t)\frac{S_{T,k}(t)h_{C|T}(S_{C,k}(t)|S_{T,k}(t))}{C_k(S_{T,k}(t), S_{C,k}(t))},$$

where $\lambda_k(t) = \lim_{h\to 0}\mathbb{P}(t \leq T \leq t + h|T \geq t, \boldsymbol{X} \in \Omega_k)$ is the ordinary hazard function of T given $\boldsymbol{X} \in \Omega_k$ and $h_{C|T}(v|u) = (\partial/\partial u)C(u, v)$. Moreover, define

$$\boldsymbol{\phi}^\circ = (\phi_1^\circ, \ldots, \phi_K^\circ)^\top,$$

where $\phi_k^\circ = \exp\{-\Lambda_k^{\#}(t_k^*)\}$ and $\Lambda_k^{\#}(t) = \int_0^t \lambda_k^{\#}(u)du$ for $k \in [K]$. Then, we have the following theorem.

Theorem 1. *Suppose that for $k \in [K]$, it holds that $\mathbb{P}(Y \geq t_k^*, \boldsymbol{X} \in \Omega_k) > 0$. Then, $\hat{\boldsymbol{\phi}}^\circ$ converges in probability to $\boldsymbol{\phi}^\circ$ and admits the following asymptotic linear representation:*

$$\hat{\boldsymbol{\phi}}^\circ - \boldsymbol{\phi}^\circ = \frac{1}{n^\circ}\sum_{i=1}^{n^\circ}\boldsymbol{\zeta}^\circ(Y_i^\circ, \Delta_i^\circ, \boldsymbol{X}_i^\circ) + o_{\mathbb{P}}(n^{\circ -1/2}),$$

where $\boldsymbol{\zeta}^{\circ}(Y, \Delta, \boldsymbol{X})$ is a K-dimensional random vector that has an expectation of zero with its explicit expression presented in Section B of the Appendix.

The proof of this theorem (as well as all subsequent theorems) is given in Section B of the Appendix. Theorem 1 demonstrates that even without the partly independent censoring assumption in Eq (3.5), the auxiliary information $\hat{\boldsymbol{\phi}}^{\circ}$ still has a well-defined probability limit $\boldsymbol{\phi}^{\circ}$. Importantly, when the partly independent censoring assumption does hold, we have $\boldsymbol{\phi}^{\circ} = \boldsymbol{\phi}^{\#}$, as has already been mentioned in Remark 2. When the assumption fails, however, $\boldsymbol{\phi}^{\circ} \neq \boldsymbol{\phi}^{\#}$, and this discrepancy must be carefully accounted for in developing transfer learning methodologies. Existing literature on synthesizing auxiliary survival probabilities often proceeds by directly starting from $\boldsymbol{\phi}^{\#}$ without acknowledging this gap, which can lead to incorrect conclusions.

Remark 5. The proposed methodology shown later relies on the assumption that the source domain and the target domain are homogeneous, meaning that the auxiliary survival probabilities $\hat{\boldsymbol{\phi}}^{\circ}$ are computed from a population that follows the same data generating process as the target dataset O_n. Under this assumption, the probability limit $\boldsymbol{\phi}^{\circ}$ of $\hat{\boldsymbol{\phi}}^{\circ}$ coincides with the limit of the transitional element $\hat{\boldsymbol{\phi}}$ constructed from the target data (defined later in Section 5), ensuring the validity of the transfer procedure. If the source and target populations differ substantially, for example, due to differences in covariate distributions or baseline hazard functions, then $\boldsymbol{\phi}^{\circ}$ may deviate from the target counterpart, and the resulting estimator shown later could exhibit bias or even deteriorate relative to the target-only estimator. This phenomenon is known as negative transfer in the transfer learning literature. Addressing domain heterogeneity is beyond the scope of the current paper and we refer the reader to Section 8 for a brief discussion and relevant references.

4. Target-only methodology

In this section, we address Tasks 1 and 2 without the use of auxiliary information. Existing literature typically specifies the target parameter $\alpha^{\star}$ in a concrete manner, for example, as a scalar or a vector of coefficients, and then develops tailored methods accordingly. In contrast, we adopt an abstract definition of $\alpha^{\star}$ without committing to a specific interpretation, aiming to derive a unified methodology applicable to a wide variety of instances.

4.1. Generally formulated estimator

The partial likelihood under the Cox model is widely used, and many estimation tasks begin with it. The scaled log-partial-likelihood function [40] is

$$\hat{l}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \left[\boldsymbol{X}_i^{\top} \boldsymbol{\beta} - \log \left(\hat{\mu}^{(0)}(Y_i, \boldsymbol{\beta}) \right) \right],$$

where $\hat{\mu}^{(0)}(t, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} R_i(t) \exp(\boldsymbol{X}_i^{\top} \boldsymbol{\beta})$ with $R_i(t) = \mathbb{I}_{\{Y_i \geq t\}}$ for $i \in [n]$. Typically, an estimator of $\alpha^{\star}$ can be expressed in the general form as follows:

$$\hat{\alpha} = \hat{h}(\hat{\boldsymbol{\beta}}), \tag{4.1}$$

where $\hat{h}(\boldsymbol{\beta})$ is a (possibly) random function and $\hat{\boldsymbol{\beta}}$ is the renowned maximum partial likelihood estimator of the coefficient vector $\boldsymbol{\beta}^\star$, obtained from the following optimization:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^p} \hat{l}(\boldsymbol{\beta}).$$

Note that for all specific examples of $\alpha^\star$ presented in Section 2.2, the associated estimators can indeed be formulated as in Eq (4.1), as shown in Remark 6.

Remark 6. A natural estimator of the target parameter $\alpha^\star$ defined in Example 1 is $\hat{\alpha} = \hat{\beta}_j$, the jth component of $\hat{\boldsymbol{\beta}}$ with $j \in [p]$. Consequently, we can set $\hat{h}(\boldsymbol{\beta}) = \boldsymbol{e}_j^\top \boldsymbol{\beta}$, where $\boldsymbol{e}_j$ is the vector whose jth element is 1 and all other elements are 0. For $\alpha^\star$ introduced in Example 2, we can first define

$$\hat{\Lambda}_0(t,\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i \mathbb{I}_{\{Y_i \le t\}}}{\hat{\mu}^{(0)}(Y_i,\boldsymbol{\beta})},$$

and then set $\hat{\alpha} = \exp\{-\hat{\Lambda}_0(t^\star,\hat{\boldsymbol{\beta}})\exp(\boldsymbol{x}^{\star\top}\hat{\boldsymbol{\beta}})\}$. Note that $\hat{\Lambda}_0(t) = \hat{\Lambda}_0(t,\hat{\boldsymbol{\beta}})$ is the Breslow-type estimator of $\Lambda_0^\star(t)$. Hence, $\hat{h}(\boldsymbol{\beta}) = \exp\{-\hat{\Lambda}_0(t^\star,\boldsymbol{\beta})\exp(\boldsymbol{x}^{\star\top}\boldsymbol{\beta})\}$. Similar discussions for $\alpha^\star$ defined in Example 3 are deferred to the later section.

The asymptotic properties of $\hat{\boldsymbol{\beta}}$ have been extensively studied in [41–43]. Instead of studying $\hat{\alpha}$ separately for different concrete specifications, we proceed to discuss its asymptotic properties in a general manner. To this end, define:

$$\mu^{(0)}(t,\boldsymbol{\beta}) = \mathbb{E}\{R(t)\exp(\boldsymbol{X}^\top\boldsymbol{\beta})\} \quad \text{and} \quad \boldsymbol{\mu}^{(l)}(t,\boldsymbol{\beta}) = \mathbb{E}\left\{R(t)\exp(\boldsymbol{X}^\top\boldsymbol{\beta})\boldsymbol{X}^{\otimes l}\right\},$$

for $l = 1$ and 2, where $R(t) = \mathbb{I}_{\{Y\ge t\}}$ and for any vector $\boldsymbol{a}$, we write $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$ and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^\top$. Furthermore, let $\boldsymbol{\eta}(t,\boldsymbol{\beta}) = \boldsymbol{\mu}^{(1)}(t,\boldsymbol{\beta})/\mu^{(0)}(t,\boldsymbol{\beta})$ and define the information matrix $\boldsymbol{H}^\star$ as follows:

$$\boldsymbol{H}^\star = \int_0^\tau \left\{\frac{\boldsymbol{\mu}^{(2)}(t,\boldsymbol{\beta}^\star)}{\mu^{(0)}(t,\boldsymbol{\beta}^\star)} - \boldsymbol{\eta}(t,\boldsymbol{\beta}^\star)^{\otimes 2}\right\}\mu^{(0)}(t,\boldsymbol{\beta}^\star)d\Lambda_0^\star(t).$$

To facilitate subsequent deductions, we impose a series of regularity conditions, namely, Conditions S1 to S3 listed in Section A of the Appendix. These conditions are standard in the literature for ensuring the asymptotic properties of the estimator $\hat{\boldsymbol{\beta}}$, as illustrated by Andersen and Gill [41]. Under Conditions S1 to S2 and the conditionally independent censoring assumption, if we define:

$$\boldsymbol{\xi}_{\boldsymbol{U}}(Y,\Delta,\boldsymbol{X}) = \int_0^\tau \left\{\boldsymbol{X} - \boldsymbol{\eta}(s,\boldsymbol{\beta}^\star)\right\} dM(s),$$

then it can be deduced that the empirical partial likelihood score function $\hat{\boldsymbol{U}}(\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})\hat{l}(\boldsymbol{\beta})$ evaluated at $\boldsymbol{\beta}^\star$ can be expressed as [42, 43]

$$\hat{\boldsymbol{U}}(\boldsymbol{\beta}^\star) = n^{-1}\sum_{i=1}^{n} \boldsymbol{\xi}_{\boldsymbol{U}}(Y_i,\Delta_i,\boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}), \tag{4.2}$$

where $N(t) = \Delta\mathbb{I}_{\{Y\le t\}}$, $M(t) = N(t) - \int_0^t R(u)\exp(\boldsymbol{X}^\top\boldsymbol{\beta}^\star)d\Lambda_0^\star(u)$, and $\tau = \inf\{t : \mathbb{P}(Y > t) = 0\}$ is the upper bound of the observation time Y. If $\boldsymbol{H}^\star$ is further assumed to be invertible, or equivalently,

Condition S3 holds, we also have the equality $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star = \boldsymbol{H}^{\star -1}\hat{\boldsymbol{U}}(\boldsymbol{\beta}^\star) + o_{\mathbb{P}}(n^{-1/2})$. Consequently, by the mean value theorem, we have the following decomposition:

$$\hat{\alpha} - \alpha^\star = \hat{h}(\boldsymbol{\beta}^\star) - \alpha^\star + \hat{\boldsymbol{h}}_\bullet(\tilde{\boldsymbol{\beta}})^\top \boldsymbol{H}^{\star -1}\hat{\boldsymbol{U}}(\boldsymbol{\beta}^\star) + o_{\mathbb{P}}(n^{-1/2}), \tag{4.3}$$

where $\hat{\boldsymbol{h}}_\bullet(\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})\hat{h}(\boldsymbol{\beta})$ and $\tilde{\boldsymbol{\beta}}$ is a random vector lying between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^\star$. Motivated by Eqs (4.2) and (4.3), we further impose the following regularity conditions.

Condition 1. *There exists a random variable $\xi_h(Y, \Delta, \boldsymbol{X})$ that has an expectation of zero such that $\hat{h}(\boldsymbol{\beta}^\star)$ is asymptotically equivalent to an I.I.D. summation as follows:*

$$\hat{h}(\boldsymbol{\beta}^\star) - \alpha^\star = \frac{1}{n}\sum_{i=1}^{n} \xi_h(Y_i, \Delta_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}).$$

Condition 2. *There exists a p-dimensional non-random vector $\boldsymbol{h}_\bullet^\star$ such that for any p-dimensional random vector $\tilde{\boldsymbol{\beta}}$ with $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\| = o_{\mathbb{P}}(1)$, it holds that $\|\hat{\boldsymbol{h}}_\bullet(\tilde{\boldsymbol{\beta}}) - \boldsymbol{h}_\bullet^\star\| = o_{\mathbb{P}}(1)$.*

Both Conditions 1 and 2 pertain to $\hat{h}(\boldsymbol{\beta})$. It can be demonstrated that the associated functions arising in our Examples 1–3 do satisfy these two conditions. Recalling Eqs (4.2) and (4.3), we can immediately proceed to the next theorem.

Theorem 2. *Suppose that the conditionally independent censoring assumption, Conditions 1 and 2, and Conditions S1 to S3 in Section A of the Appendix hold. Then, $\hat{\alpha}$ is asymptotically equivalent to an I.I.D. summation as follows:*

$$\hat{\alpha} - \alpha^\star = \frac{1}{n}\sum_{i=1}^{n} \xi(Y_i, \Delta_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),$$

where $\xi(Y, \Delta, \boldsymbol{X})$ is a random variable that has an expectation of zero and defined as

$$\xi(Y, \Delta, \boldsymbol{X}) = \xi_h(Y, \Delta, \boldsymbol{X}) + \boldsymbol{\omega}^{\star\top}\boldsymbol{\xi}_{\boldsymbol{U}}(Y, \Delta, \boldsymbol{X}),$$

with $\boldsymbol{\omega}^\star = \boldsymbol{H}^{\star -1}\boldsymbol{h}_\bullet^\star$. Moreover, the asymptotic normality $\sqrt{n}(\hat{\alpha} - \alpha^\star)/\sqrt{v} \xrightarrow{d} \mathcal{N}(0, 1)$ is satisfied if the scalar $v = \mathbb{E}\{\xi(Y, \Delta, \boldsymbol{X})^2\}$ is positive and finite.

Theorem 2 establishes that the estimator $\hat{\alpha}$ is regular and asymptotically linear [44], a standard concept in estimation theory. Based on the asymptotic normality, a $100(1-\eta)\%$ confidence interval for the target parameter $\alpha^\star$ can be constructed as $[\hat{\alpha} - z_{1-\eta/2}\sqrt{\hat{v}/n}, \hat{\alpha} + z_{1-\eta/2}\sqrt{\hat{v}/n}]$, where $\eta \in (0, 1)$ is a pre-specified significance level, $z_{1-\eta/2}$ is the $(1-\eta/2)$th quantile of $\mathcal{N}(0, 1)$, and $\hat{v}$ is a consistent estimator of v, which completes Task 1. For the hypothesis testing in Task 2, a Wald test statistic can be defined by $\hat{\mathcal{T}} = n(\hat{\alpha} - a)^2/\hat{v}$, which converges in distribution to a chi-square distribution with 1 degree of freedom, denoted by χ_1^2, under the null hypothesis. Consequently, the null hypothesis is rejected if and only if $\hat{\mathcal{T}} > \chi_1^2(1-\eta)$, where $\chi_1^2(1-\eta)$ denotes the $(1-\eta)$th quantile of χ_1^2.

4.2. Detailed illustration via the causal estimand

Our generally formulated estimator and the associated theorem presented in Section 4.1 offer a unified treatment that encompasses different specific examples of target parameters. Estimation and inference for common examples, such as Examples 1–3, can indeed be discussed within this framework. We proceed to illustrate this via Example 3, where the target parameter is defined as a causal estimand. Notably, one of the primary goals of many observational studies in medicine, epidemiology, and economics is to estimate the causal effect of a treatment or exposure on a time-to-event outcome. For decades, the Cox model has been the default analytical tool, with the hazard ratio, as introduced in Example 1, serving as the primary summary measure. However, a robust methodological literature has established that the hazard ratio is a noncollapsible measure of association that generally lacks a direct causal interpretation, even under conditions of no unmeasured confounding and correctly specified models [45, 46]. This critical insight underscores the necessity of adopting formal causal inference frameworks, grounded in the potential outcomes paradigm, to define, identify, and estimate causal effects from observational survival data [47].

To identity marginal survival functions $S_1(t)$ and $S_0(t)$ needed in defining $\alpha^\star$ in Eq (2.3), several standard assumptions frequently employed in the causal inference literature are required. These are Conditions S4 to S6 in Section A of the Appendix. Under these assumptions, the marginal survival function $S_w(t)$ for both $w = 0$ and $w = 1$ is identified by the G-formula [48]:

$$S_w(t) = \mathbb{E}[S(t|\boldsymbol{X}_w)], \tag{4.4}$$

where $\boldsymbol{X}_w = (w, \boldsymbol{Z}^\top)^\top$ and the expectation is taken over the marginal distribution of $\boldsymbol{Z}$. Eq (4.4) shows that the marginal survival functions can be obtained by averaging the conditional survival function over the covariate distribution.

Let us decompose $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\boldsymbol{\beta}}_2^\top)^\top$ with $\hat{\beta}_1$ being 1-dimensional. Similarly, for $i \in [n]$, write $\boldsymbol{X}_i = (W_i, \boldsymbol{Z}_i^\top)^\top$ with W_i being 1-dimensional. For the ith subject ($i \in [n]$) in the target dataset, we can predict their marginal survival functions under treatment $w = 0$ and $w = 1$ by plugging in the estimates from the Cox model as follows:

$$\hat{S}(t|\boldsymbol{X}_{w,i}) = \exp\{-\hat{\Lambda}_0(t)\exp(w\hat{\beta}_1 + \boldsymbol{Z}_i^\top\hat{\boldsymbol{\beta}}_2)\},$$

where $\boldsymbol{X}_{w,i} = (w, \boldsymbol{Z}_i^\top)^\top$ for $i \in [n]$ and $w \in \{0, 1\}$. Then, the marginal survival function $S_w(t)$ for $w = 0$ or 1 can be estimated by averaging these individual predictions over the empirical distribution of $\boldsymbol{Z}$, that is, the sample average, as follows:

$$\hat{S}_w(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{S}(t|\boldsymbol{X}_{w,i}).$$

This is the model-based standardization (or G-computation) estimator. Note that for each subject, we use their own covariates $\boldsymbol{X}_i$ ($i \in [n]$), but set the treatment to the desired value w. Finally, the causal estimand $\alpha^\star$ can be estimated by

$$\hat{\alpha} = \hat{S}_1(t^\star) - \hat{S}_0(t^\star). \tag{4.5}$$

The estimator in Eq (4.5) can be formulated via the generally structured estimator in Eq (4.1). Specifically, we can define the function $\hat{h}$ needed there as follows:

$$\hat{h}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left[q(\boldsymbol{X}_{1,i}, \boldsymbol{\beta}, \hat{\Lambda}_0(t^\star, \boldsymbol{\beta})) - q(\boldsymbol{X}_{0,i}, \boldsymbol{\beta}, \hat{\Lambda}_0(t^\star, \boldsymbol{\beta}))\right],$$

where $q(\boldsymbol{X}, \boldsymbol{\beta}, u) = \exp\{-u\exp(W\beta_1 + \boldsymbol{Z}^\top\boldsymbol{\beta}_2)\}$. To establish the asymptotic properties of $\hat{\alpha}$ defined in Eq (4.5), we can directly follow our discussions in Section 4.1 and verify Conditions 1 and 2.

Remark 7. It is worth noting that we can also measure the effect of the treatment W via other causal estimands, such as a deterministic transformation of the survival time T:

$$\alpha^\star = \mathbb{E}[g(T(1)) - g(T(0))],$$

where $g(\cdot)$ is a known transformation. Our focused causal estimand defined in Eq (2.3) is a special case with $g(x) = \mathbb{I}_{\{x>t^\star\}}$. If $g(x) = \min\{x, t^\star\}$, the causal estimand becomes the difference in restricted mean survival time up to $t^\star$:

$$\alpha^\star_{\text{RMST}} = \int_0^{t^\star} S_1(u)\,du - \int_0^{t^\star} S_0(u)\,du.$$

Throughout this paper, we focus exclusively on the estimation and inference on $\alpha^\star$ defined in Eq (2.3) for illustration. Nevertheless, our subsequently developed approach can be extended smoothly to other commonly encountered causal estimands as well.

Remark 8. The estimation procedure does not require modeling the propensity score $\mathbb{P}(W = 1|\boldsymbol{Z})$. This is because the identification formula in Eq (4.4) depends only on the conditional survival function and the marginal distribution of $\boldsymbol{Z}$. Under the assumption of no unmeasured confounding and a correctly specified Cox model, the G-formula provides a consistent estimator of the causal effect without needing to model the treatment mechanism. However, if the Cox model is misspecified, the estimator may be biased. In such cases, doubly robust estimators that combine the outcome model with a propensity score model may be preferred. This line of research deserves future investigation.

5. Auxiliary information transfer

We next formulate our methodology that can efficiently combine O_n with $\hat{\boldsymbol{\phi}}^\circ$ to derive estimators and test statistics that can outperform their target-only counterparts.

5.1. Intermediate estimator

Following our discussions in Section 4, our key challenge for tackling Tasks 1 and 2 in the presence of the auxiliary information $\hat{\boldsymbol{\phi}}^\circ$ lies in first constructing an estimator that takes $\hat{\boldsymbol{\phi}}^\circ$ into consideration. To this end, we propose to construct an estimator for $\alpha^\star$ by the following three steps:

Step 1. Calculate the classical target-only estimator $\hat{\alpha}$ for $\alpha^\star$ as in Eq (4.1).

Step 2. Based solely on the target dataset O_n, construct a random vector $\hat{\boldsymbol{\phi}}$, referred to as a transitional element, that has the same probability limit with the focused auxiliary information $\hat{\boldsymbol{\phi}}^\circ$. The specific constructions of $\hat{\boldsymbol{\phi}}$ is deferred to Section 6.

Step 3. Subtract $\hat{\boldsymbol{\phi}}$ defined in Step 2 with $\hat{\boldsymbol{\phi}}^\circ$ and linearly link the resulting difference to $\hat{\alpha}$ defined in Step 1, leading to an estimator structured as follows:

$$\hat{\alpha}^\dagger(\boldsymbol{b}) = \hat{\alpha} - \boldsymbol{b}^\top(\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^\circ), \tag{5.1}$$

where $\boldsymbol{b}$ is a K-dimensional vector and we call it the leveraging factor.

The novel estimator $\hat{\alpha}^{\dagger}(\boldsymbol{b})$ obtained by conducting these three steps is used for estimating the target parameter $\alpha^{\star}$ and we call it the intermediate estimator from here on.

Since the target-only estimator $\hat{\alpha}$ is consistent with the target parameter $\alpha^{\star}$ and the difference $\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^{\circ}$ converges in probability to zero, $\hat{\alpha}^{\dagger}(\boldsymbol{b})$ is always a consistent estimator for $\alpha^{\star}$, regardless of the specific choice of the leveraging factor $\boldsymbol{b}$. Hence, the transitional element bridges the target and source domains. Moreover, when the leveraging factor $\boldsymbol{b}$ equals zero, $\hat{\alpha}^{\dagger}(\boldsymbol{b})$ reduces to the target-only estimator $\hat{\alpha} = \hat{\alpha}^{\dagger}(\boldsymbol{0})$. Therefore, it is expected that by appropriately specifying the leveraging factor $\boldsymbol{b}$, the intermediate estimator $\hat{\alpha}^{\dagger}(\boldsymbol{b})$ can be no less efficient than the target-only estimator $\hat{\alpha}$.

We proceed by investigating the asymptotic normality of the immediate estimator $\hat{\alpha}^{\dagger}(\boldsymbol{b})$ for any given $\boldsymbol{b} \in \mathbb{R}^K$, as these results will be instrumental in determining the final concrete form of the leveraging factor. We impose the following regularity condition on the transitional element $\hat{\boldsymbol{\phi}}$ defined in Step 2.

Condition 3. *The transitional element $\hat{\boldsymbol{\phi}}$ can be asymptotically expressed as follows:*

$$\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{\circ} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\zeta}(Y_i, \Delta_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}), \tag{5.2}$$

where $\boldsymbol{\zeta}(Y, \Delta, \boldsymbol{X})$ is a K-dimensional random vector that has an expectation of zero.

Our specifically defined transitional elements shown in Section 6 do satisfy Condition 3. Denote $\boldsymbol{\gamma} = \mathbb{E}\{\boldsymbol{\zeta}(Y, \Delta, \boldsymbol{X})\xi(Y, \Delta, \boldsymbol{X})\}$. Furthermore, let $\boldsymbol{V} = \mathbb{E}\{\boldsymbol{\zeta}(Y, \Delta, \boldsymbol{X})^{\otimes 2}\}$ and $\boldsymbol{V}^{\circ} = \mathbb{E}\{\boldsymbol{\zeta}^{\circ}(Y, \Delta, \boldsymbol{X})^{\otimes 2}\}$. Now we have the next theorem.

Theorem 3. *Suppose that Condition 3 and all conditions imposed in Theorems 1 and 2 hold. Moreover, assume that there exists a constant $\rho \in [0, \infty)$ such that $n/n^{\circ} \to \rho$ as $n \to \infty$. Then, for any $\boldsymbol{b} \in \mathbb{R}^K$, if we define:*

$$v^{\dagger}(\boldsymbol{b}) = v - 2\boldsymbol{\gamma}^{\top}\boldsymbol{b} + \boldsymbol{b}^{\top}(\boldsymbol{V} + \rho\boldsymbol{V}^{\circ})\,\boldsymbol{b},$$

it follows that the asymptotic normality $\sqrt{n}(\hat{\alpha}^{\dagger}(\boldsymbol{b}) - \alpha^{\star}) \xrightarrow{d} \mathcal{N}(0, v^{\dagger}(\boldsymbol{b}))$ is satisfied.

It is important to note that the case where $\rho = \infty$ is outside the conditions of Theorem 3, which assumes a finite limit ρ. This case is discussed here only as a separate limiting situation. When $\rho = \infty$, the external sample size is asymptotically negligible relative to the target sample size. In that degenerate situation, the auxiliary information transfer (AIT) estimator reduces to the target-only estimator and is not of primary interest in the transfer learning context. From Theorem 3, we can conclude that although the intermediate estimator $\hat{\alpha}^{\dagger}(\boldsymbol{b})$ is consistent, its estimation efficiency, or equivalently, its asymptotic variance $v^{\dagger}(\boldsymbol{b})$, varies with different choices of $\boldsymbol{b}$. Consequently, the key challenge is to identify a suitable leveraging factor $\boldsymbol{b}$ such that the resulting estimator, which shares the same structure as the new estimator defined in Eq (5.1), attains the optimal estimation efficiency.

5.2. Feasible estimator with optimal estimation efficiency

Recalling the issue highlighted at the end of the previous subsection, we proceed by determining the specific definition of the leveraging factor. Notably, the estimator attaining the minimum asymptotic variance is $\hat{\alpha}^{\dagger}(\boldsymbol{b}^{\dagger})$, where we define

$$\boldsymbol{b}^{\dagger} = (\boldsymbol{V} + \rho\boldsymbol{V}^{\circ})^{-1}\boldsymbol{\gamma} = \underset{\boldsymbol{b}\in\mathbb{R}^K}{\arg\min}\; v^{\dagger}(\boldsymbol{b}), \tag{5.3}$$

assume that $\boldsymbol{V} + \rho\boldsymbol{V}^\circ$ is invertible thanks to Theorem 3, and have that $\sqrt{n}(\hat{\alpha}^\dagger(\boldsymbol{b}^\dagger) - \alpha^\star)$ converges in distribution to a normal random variable with a mean of zero and variance $v^\dagger(\boldsymbol{b}^\dagger) = v - \boldsymbol{\gamma}^\top(\boldsymbol{V} + \rho\boldsymbol{V}^\circ)^{-1}\boldsymbol{\gamma}$. Although $\hat{\alpha}^\dagger(\boldsymbol{b}^\dagger)$ achieves the optimal estimation efficiency, it is infeasible in practice since $\boldsymbol{b}^\dagger$ is unknown.

Motivated by the expression shown in Eq (5.3), a natural approach to defining a feasible leveraging factor is to take $\boldsymbol{b}$ as an appropriate estimator of $\boldsymbol{b}^\dagger$. According to the specific definitions of $\boldsymbol{\gamma}$ and $\boldsymbol{V}$, we can naturally define the following estimators:

$$\hat{\boldsymbol{\gamma}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\xi}(Y_i, \Delta_i, \boldsymbol{X}_i)\hat{\boldsymbol{\zeta}}(Y_i, \Delta_i, \boldsymbol{X}_i) \quad \text{and} \quad \hat{\boldsymbol{V}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{\zeta}}(Y_i, \Delta_i, \boldsymbol{X}_i)^{\otimes 2}, \tag{5.4}$$

where $\hat{\xi}(Y, \Delta, \boldsymbol{X})$ and $\hat{\boldsymbol{\zeta}}(Y, \Delta, \boldsymbol{X})$ are estimators of $\xi(Y, \Delta, \boldsymbol{X})$ and $\boldsymbol{\zeta}(Y, \Delta, \boldsymbol{X})$, respectively, and can typically be obtained by evaluating the later population-level quantities using the target dataset $\mathcal{O}_n$, provided that their specific forms are known. More computational details can be found in Section C of the Appendix. Assume without loss of generality that $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{V}}$ are consistent. Moreover, define $\hat{\rho} = n/n^\circ$ and let $\hat{\boldsymbol{V}}^\circ$ be a consistent estimator for $\boldsymbol{V}^\circ$, which can be either extracted from the source domain or estimated via the target dataset $\mathcal{O}_n$. Consequently, we substitute the leveraging factor $\boldsymbol{b}$ needed in Eq (5.1) by $\hat{\boldsymbol{b}} = (\hat{\boldsymbol{V}} + \hat{\rho}\hat{\boldsymbol{V}}^\circ)^{-1}\hat{\boldsymbol{\gamma}}$ to obtain the final estimator

$$\hat{\alpha}_{\mathrm{AIT}} = \hat{\alpha}^\dagger(\hat{\boldsymbol{b}}) = \hat{\alpha} - \hat{\boldsymbol{b}}^\top(\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^\circ). \tag{5.5}$$

Notably, as emphasized in Section 3.1, the calculation of the estimator $\hat{\alpha}_{\mathrm{AIT}}$ only requires $\mathcal{O}_n$ and the auxiliary information $\hat{\boldsymbol{\phi}}^\circ$, with no need to access the individual-level records contained in the source dataset $\mathcal{O}^\circ_{n^\circ}$.

We next establish the asymptotic properties of the estimator $\hat{\alpha}_{\mathrm{AIT}}$ and demonstrate its efficiency gain compared to the classical target-only estimator $\hat{\alpha}$. Notably, it holds that:

$$\begin{aligned}\sqrt{n}(\hat{\alpha}_{\mathrm{AIT}} - \alpha^\star) &= \sqrt{n}(\hat{\alpha}^\dagger(\boldsymbol{b}^\dagger) - \alpha^\star) - \sqrt{n}(\hat{\boldsymbol{b}} - \boldsymbol{b}^\dagger)^\top(\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^\circ) \\ &= \sqrt{n}(\hat{\alpha}^\dagger(\boldsymbol{b}^\dagger) - \alpha^\star) + o_{\mathbb{P}}(1),\end{aligned}$$

since $|\sqrt{n}(\hat{\boldsymbol{b}} - \boldsymbol{b}^\dagger)^\top(\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^\circ)| \le \|\hat{\boldsymbol{b}} - \boldsymbol{b}^\dagger\|\|\sqrt{n}(\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^\circ)\| = o_{\mathbb{P}}(1)$. Consequently, we can conclude that the estimator $\hat{\alpha}_{\mathrm{AIT}}$ and the optimal but infeasible estimator $\hat{\alpha}^\dagger(\boldsymbol{b}^\dagger)$ have the same estimation efficiency. We have the following theorem regarding $\hat{\alpha}_{\mathrm{AIT}}$.

Theorem 4. *Suppose that all conditions imposed in Theorem 3 hold and $\boldsymbol{V} + \rho\boldsymbol{V}^\circ$ is invertible. Then, the asymptotic normality $\sqrt{n}(\hat{\alpha}_{\mathrm{AIT}} - \alpha^\star) \xrightarrow{d} \mathcal{N}(0, v_{\mathrm{AIT}})$ is satisfied, where $v_{\mathrm{AIT}} = v - \boldsymbol{\gamma}^\top(\boldsymbol{V} + \rho\boldsymbol{V}^\circ)^{-1}\boldsymbol{\gamma}$.*

Theorem 4 indicates that the proposed estimator $\hat{\alpha}_{\mathrm{AIT}}$ not only is asymptotically normal, but also can achieve the optimal estimation efficiency asymptotically. It is worth emphasizing that $\hat{\alpha}_{\mathrm{AIT}}$ can always perform at least as efficiently as the classical target-only estimator $\hat{\alpha}$ in terms of the estimation efficiency since $v_{\mathrm{AIT}} \le v$. It can also be seen that the gain of estimation efficiency, that is, $v - v_{\mathrm{AIT}} = \boldsymbol{\gamma}^\top(\boldsymbol{V} + \rho\boldsymbol{V}^\circ)^{-1}\boldsymbol{\gamma}$, depends on the value of ρ. When $\rho = 0$, the efficiency gain is maximized, whereas as $\rho \to \infty$, the efficiency gain approaches zero. This observation is intuitive as a larger n° (for fixed n) implies higher-quality auxiliary information since the randomness of $\hat{\phi}^\circ$ is reduced. Conversely, for fixed n°, the target dataset itself becomes more informative as n increases, making the auxiliary information relatively less necessary.

Remark 9. Following our discussions at the end of Section 4, the answers for Tasks 1 and 2 can be provided immediately based on the novel estimator $\hat{\alpha}_{\mathrm{AIT}}$. Specifically, starting from the asymptotic normality established in Theorem 4, a $100(1-\eta)\%$ confidence interval for the target parameter $\alpha^\star$ can be constructed as follows:

$$\left[\hat{\alpha}_{\mathrm{AIT}} - z_{1-\eta/2}\sqrt{\hat{v}_{\mathrm{AIT}}/n}, \hat{\alpha}_{\mathrm{AIT}} + z_{1-\eta/2}\sqrt{\hat{v}_{\mathrm{AIT}}/n}\right],$$

where $\hat{v}_{\mathrm{AIT}}$ is an estimator of v_{AIT}, which completes Task 1. From the perspective of hypothesis testing for Task 2, a Wald test statistic can be defined by $\hat{\mathcal{T}}_{\mathrm{AIT}} = n(\hat{\alpha}_{\mathrm{AIT}} - a)^2/\hat{v}_{\mathrm{AIT}}$. Consequently, the null hypothesis is rejected if and only if $\hat{\mathcal{T}}_{\mathrm{AIT}} > \chi_1^2(1-\eta)$ for some pre-specified significance level η.

To sum up, the whole implementation of the proposed transfer learning framework proceeds through the following six steps:

Step 1. Collect the target dataset $\mathcal{O}_n$, the accessible auxiliary information $\hat{\boldsymbol{\phi}}^\circ$ (subgroup survival probabilities at pre-specified time points), and the external sample size n°.

Step 2. Specify the focused target parameter $\alpha^\star$, such as in the examples in Section 2.2.

Step 3. Obtain the target-only estimator $\hat{\alpha}$ for $\alpha^\star$ selected in Step 2 using Eq (4.1).

Step 4. Choose either the model-free version (Section 6.1) or the model-based version (Section 6.2) to compute the transitional element $\hat{\boldsymbol{\phi}}$ from the target dataset.

Step 5. Compute the leveraging factor $\hat{\boldsymbol{b}}$ and the final AIT estimator $\hat{\boldsymbol{\alpha}}_{\mathrm{AIT}}$ according to Eq (5.5).

Step 6. Follow Remark 9 to obtain the confidence interval (Task 1) and the test decision (Task 2).

6. Concrete transitional element

The feasibility of our auxiliary information transfer estimator depends on the specific definition of the transitional element $\hat{\boldsymbol{\phi}}$ first introduced in Section 5.1. In this subsection, we discuss two specific constructions of $\hat{\boldsymbol{\phi}}$. The first, presented in Section 6.1, is model-free and always applicable in practice. The second, discussed in Section 6.2, is model-based and yields to an estimator with higher estimation efficiency, but it requires an additional restrictive assumption.

6.1. Identical mapping inherited from auxiliary information

Recalling the definition of the auxiliary information $\hat{\boldsymbol{\phi}}^\circ$ in Eq (3.3), we propose to define a random vector in an identical manner, but with the source dataset $\mathcal{O}_{n^\circ}^\circ$ input to the mapping $\mathcal{M}$ replaced by its counterpart derived from the target dataset $\mathcal{O}_n$. Specifically, let us define:

$$\lfloor\text{Transitional Element}\rfloor \quad \hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_K)^\top = \mathcal{M}(\mathcal{O}_n), \tag{6.1}$$

referred to as the transitional element, which is model-free, as no model assumptions are needed. Equivalently, for each $k \in [K]$, we can write

$$\hat{\phi}_k = \prod_{i=1}^{n}\left[1 - \frac{\Delta_i \mathbb{I}_{\{\boldsymbol{X}_i \in \Omega_k\}}\mathbb{I}_{\{Y_i \le t_k^*\}}}{\sum_{j=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_j \in \Omega_k\}}\mathbb{I}_{\{Y_j \ge Y_i\}}}\right],$$

where t_k^* is the same pre-specified time point used in defining the auxiliary information $\hat{\phi}_k^\circ$. Because the transitional element $\hat{\boldsymbol{\phi}}$ in Eq (6.1) and the auxiliary information $\hat{\boldsymbol{\phi}}^\circ$ in Eq (3.3) are constructed via the same mapping $\mathcal{M}$, they are expected to share a close relationship.

Regarding asymptotic properties, Theorem 1 also applies to the transitional element $\hat{\boldsymbol{\phi}}$. Therefore, $\hat{\boldsymbol{\phi}}$ converges in probability to the same limit $\boldsymbol{\phi}^\circ$ and Condition 3 is satisfied. Consequently, we can conclude straightforwardly that their difference $\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^\circ$ converges in probability to zero. It is worth emphasizing that this property holds without requiring the partly independent censoring assumption in Eq (3.5).

6.2. Partly independent censoring-assisted refinement

In Section 6.1, we provided a specific construction of the transitional element. This construction is model-free and always feasible, as it does not require additional assumptions regarding the dependence between T and C. Naturally, other definitions are possible, and comparing different transitional elements is meaningful. In what follows, we propose an alternative construction that has both advantages and disadvantages relative to the previously defined one.

As indicated in Remark 2, $\hat{\phi}_k^\circ$ is not necessarily a consistent estimator of $\phi_k^\#$ for $k \in [K]$. However, under the partly independent censoring assumption stated in Eq (3.5), consistency can be ensured. The following theorem formally establishes that, under this assumption, the probability limit of the auxiliary information coincides with the target quantity in Eq (3.4).

Theorem 5. *Suppose that the partly independent censoring assumption in Eq* (3.5) *and all conditions imposed in Theorem 1 hold. Then,* $\boldsymbol{\phi}^\circ = \boldsymbol{\phi}^\#$ *and* $\hat{\boldsymbol{\phi}}^\circ$ *converges in probability to* $\boldsymbol{\phi}^\#$.

Starting from the expression in Eq (3.4), we can adopt the formula for conditional probability and deduce that, for $k \in [K]$, it holds that

$$\phi_k^\circ = \frac{\psi_k}{\pi_k},$$

where

$$\psi_k = \mathbb{P}(T > t_k^*, \boldsymbol{X} \in \Omega_k) \quad \text{and} \quad \pi_k = \mathbb{P}(\boldsymbol{X} \in \Omega_k).$$

The denominator can be naturally estimated by $\hat{\pi}_k = n^{-1} \sum_{i=1}^n \mathbb{I}_{\{\boldsymbol{X}_i \in \Omega_k\}}$. For the numerator, note that the following equality holds by the law of double expectation:

$$\psi_k = \mathbb{E}\left\{\mathbb{I}_{\{\boldsymbol{X}\in\Omega_k\}} S(t_k^*|\boldsymbol{X})\right\} = \mathbb{E}\left\{\psi(\boldsymbol{X}, \boldsymbol{\beta}^\star, \Lambda_0^\star(t_k^*), \Omega_k)\right\}, \tag{6.2}$$

where $\psi(\boldsymbol{X}, \boldsymbol{\beta}, u, \Omega) = \mathbb{I}_{\{\boldsymbol{X}\in\Omega\}} \times \exp\left(-u \exp(\boldsymbol{X}^\top \boldsymbol{\beta})\right)$. Motivated by this expression, we can empirically evaluate the left-hand side based on the target dataset $\mathcal{O}_n$. Specifically, for any $k \in [K]$, we can define an estimator of ψ_k as:

$$\hat{\psi}_k = \hat{h}_k(\hat{\boldsymbol{\beta}}) \quad \text{with} \quad \hat{h}_k(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \psi(\boldsymbol{X}_i, \boldsymbol{\beta}, \hat{\Lambda}_0(t_k^*, \boldsymbol{\beta}), \Omega_k).$$

We then propose to define

$$\hat{\phi}_k = \frac{\hat{\psi}_k}{\hat{\pi}_k}, \tag{6.3}$$

for $k \in [K]$, yielding a new transitional element defined as $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_K)^\top$. This construction is model-based because it utilizes the fact that the underlying model is the Cox model.

We next establish the asymptotic properties of the model-based transitional element defined in Eq (6.3). To this end, define

$$\xi_\Lambda(t, Y, \Delta, \boldsymbol{X}) = \int_0^t \frac{dM(s)}{\mu^{(0)}(s, \boldsymbol{\beta}^\star)}.$$

Additionally, for $k \in [K]$, denote $g^\star_{\bullet,k} = \mathbb{E}\{\bar{g}_{\bullet,k}(\boldsymbol{X})\}$, where:

$$\begin{aligned} \bar{g}_{\bullet,k}(\boldsymbol{X}) &= \left[\frac{\partial}{\partial u}\psi(\boldsymbol{X}, \boldsymbol{\beta}^\star, u, \Omega_k)\right]\Bigg|_{u=\Lambda_0^\star(t_k^*)} \\ &= -\mathbb{I}_{\{\boldsymbol{X}\in\Omega_k\}} \exp(\boldsymbol{X}^\top\boldsymbol{\beta}^\star) \exp\left(-\Lambda_0^\star(t_k^*)\exp(\boldsymbol{X}^\top\boldsymbol{\beta}^\star)\right). \end{aligned}$$

Also write $\boldsymbol{h}^\star_{\bullet,k} = g^\star_{\bullet,k}\boldsymbol{\Lambda}^\star_{0,\bullet}(t_k^*) + \Lambda_0^\star(t_k^*)\mathbb{E}[\bar{g}_{\bullet,k}(\boldsymbol{X})\boldsymbol{X}]$ for $k \in [K]$, where

$$\boldsymbol{\Lambda}^\star_{0,\bullet}(t) = -\int_0^t \frac{\boldsymbol{\mu}^{(1)}(s, \boldsymbol{\beta}^\star)}{\mu^{(0)}(s, \boldsymbol{\beta}^\star)} d\Lambda_0^\star(s).$$

Further denote $\psi_k^c(\boldsymbol{X}) = \psi(\boldsymbol{X}, \boldsymbol{\beta}^\star, \Lambda_0^\star(t_k^*), \Omega_k) - \pi_k\phi_k^\#$ and

$$\xi_{\psi,k}(Y, \Delta, \boldsymbol{X}) = \psi_k^c(\boldsymbol{X}) + \boldsymbol{h}^{\star\top}_{\bullet,k}\boldsymbol{H}^{-1}\boldsymbol{\xi}_{\boldsymbol{U}}(Y, \Delta, \boldsymbol{X}) + g^\star_{\bullet,k}\xi_\Lambda(t_k^*, Y, \Delta, \boldsymbol{X}),$$

for $t \in \mathbb{T}_k$ and $k \in [K]$. Consequently, we have the next theorem.

Theorem 6. *Suppose that the conditionally independent censoring assumption, Conditions S1 to S3 in Section A of the Appendix, and all conditions imposed in Theorem 5 hold. Then, the representation in Eq* (5.2) *is satisfied for the currently defined transitional element if for $k \in [K]$, we define the kth component of $\boldsymbol{\zeta}(Y, \Delta, \boldsymbol{X})$ to be $\pi_k^{-1}\xi_{\psi,k}(Y, \Delta, \boldsymbol{X})$.*

Theorem 6 shows that the transitional element defined in Eq (6.3) admits an asymptotic I.I.D. representation, thereby satisfying Condition 3, which is required for our proposed approach. Recall that the transitional element defined earlier in Eq (6.1) is fully nonparametric, as it imposes no model assumptions. In contrast, the newly defined transitional element is model-based and is therefore expected to yield a new estimator with higher estimation efficiency. Simulation studies in Section 7 compare these two definitions and confirm this advantage. It is important to emphasize that the model-based transitional element relies on the partly independent censoring assumption in Eq (3.5). When this assumption is violated, the estimator $\hat{\alpha}_{\text{AIT}}$ constructed from the model-based transitional element may be biased, and the corresponding test may have an inflated Type I error rate. Consequently, any apparent improvement in testing power under such settings should be interpreted with caution, as it may reflect size distortion rather than genuine superiority. This caveat will be further illustrated in the numerical experiments of Section 7.4.

Remark 10. When the time points are equal to each other and we focus on Example 1, it can be deduced that this estimator is equivalent to the double empirical likelihood-based estimator proposed by Huang et al. [19] in terms of the asymptotic estimation efficiency. Nevertheless, our estimator enjoys significantly higher computational efficiency, so that it should be more preferable in practice.

7. Numerical experiments

We conduct numerical experiments to evaluate the finite-sample performance of our methodology. All subsequent numerical experiments are carried out in the Python environment using the aisurv package mentioned in Section 1.

7.1. Data generating process

The survival times for the target datasets are simulated from the Cox model in Eq (2.1). For the covariate design, we first draw a 2-dimensional latent vector $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2)^\top$ from a multivariate normal distribution with a mean of $\mathbf{0}$ and a first-order auto-regressive covariance matrix having a marginal variance of 1 and the auto-correlation coefficient 0.2. Then, we define $\boldsymbol{X} = (X_1, X_2)^\top$, where $X_1 = \mathbb{I}_{\{\tilde{X}_1 \geq 0\}}$ and X_2 is constructed by truncating $\tilde{X}_2$ at ± 3. Notably, the binary indicator X_1 behaves essentially as a Bernoulli variable with a success probability of 0.5, mimicking a clinical treatment assignment in practice. We set the true coefficient vector to $\boldsymbol{\beta}^\star = (1, 1)^\top$ and take the cumulative baseline hazard $\Lambda_0^\star(t) = t$ for $t \geq 0$, that is, an exponential baseline with a rate of 1. The censoring time for each target observation is drawn from a uniform distribution on the interval $[0, r]$, where the right end point r is defined as

$$r = c|\boldsymbol{X}^\top \boldsymbol{\beta}^\star|^d,$$

with constants $c > 0$ and $d \geq 0$. When $d = 0$, we have $r = c$, so that C is generated independently of $\boldsymbol{X}$ (hence $T \perp C$ and $T \perp C|\boldsymbol{X}$ both hold), and the partly independent censoring assumption is automatically satisfied. Nevertheless, when $d \neq 0$, the censoring upper bound depends on $\boldsymbol{X}$, indicating that only the weaker $T \perp C|\boldsymbol{X}$ is enforced, while the subgroup-level condition fails in general.

In each replication, the auxiliary survival probabilities $\hat{\boldsymbol{\phi}}^\circ$ are produced from the same underlying population as the target, so that the auxiliary information is faithful to the truth rather than misaligned. We fix two subgroups

$$\Omega_1 = \{\boldsymbol{X} : X_1 = 0\} \quad \text{and} \quad \Omega_2 = \{\boldsymbol{X} : X_2 = 1\},$$

and then consider the following summary-level quantities:

Part 1. Generate one Kaplan–Meier survival probability at $t_1^* = 0.4$ using Ω_1.

Part 2. Generate another two Kaplan–Meier survival probabilities at $t_2^* = 0.2$ and $t_3^* = 0.6$ using Ω_2.

The total number of accessible auxiliary information is $K = 3$. In our simulation studies, these values are generated from an independent sample with a sample size n° chosen between 1000 and 5000, which simulates two scenarios with different scales of the source datasets.

7.2. Performances of estimation efficiency

We first evaluate the improvement in estimation efficiency achieved by our proposed methodology for the target parameter of interest. Subsequently, we focus on solving Tasks 1 and 2 for the target parameter defined in Example 3 and detailed in Section 4.2. This choice serves as illustration and other target parameters would yield similar conclusions. Specifically, we set the treatment variable to $W = X_1$ and fix the time point $t^\star$ in Eq (2.3) at 0.5. The sample size n takes the values 300 and 600.

We first consider the setting where $d = 0$, where T and C are independent. The constant c is specified at different values to yield the percentages of censored samples to be approximately 25% ($c = 4$) and 50% ($c = 1$). All our subsequent simulation results are summarized after conducting 500 repetitions.

We compare our method with the classical target-only (TO) estimator, which uses no auxiliary information. We denote the results from our proposed methodology in Section 5 as "AIT". Results are summarized using the average bias (Bias), sample standard deviation (SD), average model-based standard error (SE), and empirical coverage probability (CP) of 95% confidence intervals. Table 1 presents the estimation results for the target parameter. The column labeled "Model" indicates which definition of the transitional element is used, with "Model-Free" representing the model-free version from Section 6.1 and "Model-Based" representing the model-based version from Section 6.2. From Table 1, we observe that regardless of n, n°, and the censoring rate, the methods incorporating auxiliary information achieve substantial efficiency gains, reflected in smaller SD and SE values compared to the TO estimator. The efficiency gain is more pronounced when using the model-based transitional element and when n° is larger. Furthermore, the coverage probabilities remain close to the 95% nominal level.

Table 1. Estimation for the target parameter under $d = 0$.

				Censoring Rate = 25%				Censoring Rate = 50%			
n	Method	n°	Model	BIAS	SD	SE	CP	BIAS	SD	SE	CP
300	TO		—	0.000	0.036	0.035	0.944	0.002	0.046	0.044	0.932
	AIT	1000	Model-Free	0.000	0.031	0.030	0.940	0.002	0.036	0.035	0.930
			Model-Based	0.000	0.030	0.028	0.928	0.002	0.034	0.033	0.946
		5000	Model-Free	0.002	0.030	0.028	0.934	0.003	0.034	0.032	0.928
			Model-Based	0.001	0.028	0.025	0.922	0.003	0.030	0.029	0.932
600	TO		—	0.003	0.024	0.025	0.954	0.003	0.032	0.031	0.936
	AIT	1000	Model-Free	0.003	0.021	0.022	0.958	0.004	0.027	0.026	0.938
			Model-Based	0.003	0.020	0.021	0.964	0.004	0.026	0.025	0.942
		5000	Model-Free	0.002	0.020	0.020	0.952	0.004	0.024	0.024	0.950
			Model-Based	0.002	0.018	0.019	0.952	0.003	0.022	0.022	0.950

7.3. Performances of testing power

We next investigate hypothesis testing based on the proposed estimator to assess the impact of auxiliary information on testing power. We focus on conducting statistical inference for the hypothesis test in Task 2 for the target parameter. The performance of the Wald test statistic is evaluated under various settings. Specifically, we set $a = \alpha^\star + 0.04 \times \nu$ with $\nu = 1$, 2, or 3. Notably, we consider only scenarios where $\nu \neq 0$, because evaluating the test size under $\nu = 0$ (where the null hypothesis is true) is equivalent to examining one minus the coverage probability of the parameter estimator. Table 2 summarizes the testing power for the target parameter. From the table, we observe that the power increases with ν and that the gains are generally larger when n° is larger, which is a desirable outcome.

Table 2. Powers of hypothesis testing for the target parameter under $d = 0$.

				Censoring Rate = 25%			Censoring Rate = 50%		
n	Method	n°	Model	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 1$	$\nu = 2$	$\nu = 3$
300	TO		—	0.198	0.618	0.920	0.164	0.408	0.770
	AIT	1000	Model-Free	0.298	0.754	0.980	0.188	0.616	0.920
			Model-Based	0.312	0.790	0.980	0.212	0.648	0.942
		5000	Model-Free	0.302	0.754	0.982	0.218	0.666	0.928
			Model-Based	0.350	0.844	0.990	0.266	0.740	0.968
600	TO		—	0.284	0.890	0.998	0.216	0.712	0.958
	AIT	1000	Model-Free	0.376	0.960	1.000	0.274	0.822	0.986
			Model-Based	0.394	0.974	1.000	0.278	0.838	0.994
		5000	Model-Free	0.434	0.984	1.000	0.342	0.890	0.996
			Model-Based	0.532	0.988	1.000	0.412	0.936	0.998

7.4. Violation to the partly independent censoring assumption

We next consider a nonzero value of d so that T and C are conditionally independent given $\boldsymbol{X}$ but the partly independent censoring assumption no longer holds. The target sample size n and constant c are set to be 300 and 4, respectively. Notably, although the same value of $c = 4$ is used as in the setting where $d = 0$, the censoring rate changes with d because the censoring distribution now depends on the covariates. For $d = 1$ and $d = 2$, the approximate censoring rates are 34% and 41%, respectively. Table 3 presents the estimation results for the target parameter, while Table 4 summarizes the testing power under the same setup as in Section 7.3. From Tables 3 and 4, we draw similar conclusions for the model-free definition of the transitional element, indicating that it continues to yield trustworthy results even when the partly independent censoring assumption is violated. However, Table 3 shows that the model-based definition of the transitional element exhibits substantially larger estimation bias compared to the other two methods, and the coverage probabilities are no longer close to the nominal level. These issues become more severe as d increases, meaning that the conditional dependence between T and C within the subgroups grows stronger. This finding further supports our remarks at the end of Section 6.2. Caution must be exercised when adopting a model-based transitional element unless the partly independent censoring assumption can be confidently justified. In Table 4, the model-based AIT method exhibits substantially higher empirical power than the model-free version when $d \neq 0$. However, as shown in Table 3, the model-based estimator suffers from notable bias and its confidence interval coverage deviates from the nominal level under these settings. The elevated power in Table 4 is therefore likely driven by size inflation rather than a genuine improvement in detection ability. This observation aligns with the cautionary remark at the end of Section 6.2: the model-based transitional element should only be trusted when the partly independent censoring assumption can be confidently justified.

Table 3. Estimation for the target parameter with varied d under target sample size $n = 300$.

			$d = 1$ (Censoring Rate $\approx 34\%$)				$d = 2$ (Censoring Rate $\approx 41\%$)			
Method	n°	Model	BIAS	SD	SE	CP	BIAS	SD	SE	CP
TO		—	0.000	0.040	0.039	0.932	0.001	0.044	0.043	0.940
AIT	1000	Model-Free	0.002	0.034	0.032	0.934	0.003	0.037	0.036	0.940
		Model-Based	-0.009	0.031	0.030	0.926	-0.022	0.033	0.032	0.882
	5000	Model-Free	0.002	0.031	0.031	0.946	0.004	0.036	0.034	0.944
		Model-Based	-0.012	0.027	0.027	0.914	-0.027	0.031	0.029	0.812

Table 4. Powers of hypothesis testing for the target parameter with varied d under target sample size $n = 300$.

			$d = 1$ (Censoring Rate $\approx 34\%$)			$d = 2$ (Censoring Rate $\approx 41\%$)		
Method	n°	Model	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 1$	$\nu = 2$	$\nu = 3$
TO		—	0.166	0.538	0.866	0.166	0.418	0.792
AIT	1000	Model-Free	0.232	0.666	0.942	0.200	0.550	0.904
		Model-Based	0.366	0.830	0.984	0.456	0.874	0.988
	5000	Model-Free	0.250	0.710	0.962	0.210	0.584	0.932
		Model-Based	0.484	0.912	0.996	0.630	0.946	1.000

7.5. Simulation results beyond Example 3 of the target parameter

In this subsection, we provide additional numerical evidence to demonstrate that the proposed framework applies to a broad class of target parameters beyond the causal survival contrast in Example 3. We consider the regression coefficient in Example 1 with $j = 1$. The data generating mechanism is identical to that described in Section 7.1, and the simulation configurations are the same as those in the main simulation. The target parameter is actually the log hazard ratio of the covariate of interest. We compare the target-only estimator, the AIT estimator with the model-free transitional element, and the AIT estimator with the model-based transitional element.

Table 5 reports the results under independent censoring with $d = 0$. The results show the same pattern as those for the causal survival contrast, that is, Example 3 of the target parameter reported previously. Both the model-based AIT and the model-free AIT reduce the values of SD relative to TO, and the model-based AIT yields the smallest SD values. The values of SE for all methods closely track the empirical standard deviations, and the values of CP are close to the nominal level. These findings confirm that the proposed method improves estimation efficiency for the regression coefficient as well.

We then examine the sensitivity of the estimators to cases with $d \neq 0$, in which the partly independent censoring assumption does not hold. Table 6 reports the results for $d = 1$ and $d = 2$ (as have been used previously) with the target sample size n and constant c fixed at 300 and 4, respectively. The target-only estimator (TO) remains approximately unbiased. The model-free AIT estimator is also approximately unbiased and achieves a noticeable reduction in the values of SD. In contrast, the model-based AIT estimator exhibits an increasing bias as d moves away from zero, and its values

of CP deteriorate. This is expected because the model-based transitional element relies on the partly independent censoring assumption, which is violated when d is nonzero. The model-free transitional element does not rely on this assumption and therefore remains robust. These results agree with the conclusions drawn from the causal survival contrast in Section 7.4.

Table 5. Estimation for Example 1 of the target parameter under $d = 0$.

				Censoring Rate = 25%				Censoring Rate = 50%			
n	Method	n°	Model	BIAS	SD	SE	CP	BIAS	SD	SE	CP
300	TO		—	0.004	0.144	0.140	0.940	0.004	0.174	0.173	0.946
	AIT	1000	Model-Free	-0.001	0.120	0.117	0.946	-0.002	0.137	0.137	0.950
			Model-Based	-0.003	0.111	0.111	0.952	0.002	0.127	0.130	0.954
		5000	Model-Free	0.001	0.117	0.112	0.936	0.001	0.127	0.127	0.960
			Model-Based	-0.001	0.105	0.101	0.932	0.007	0.111	0.115	0.962
600	TO		—	0.001	0.102	0.099	0.954	0.010	0.117	0.122	0.956
	AIT	1000	Model-Free	-0.003	0.090	0.086	0.946	0.005	0.101	0.102	0.948
			Model-Based	-0.004	0.086	0.084	0.952	0.005	0.099	0.099	0.944
		5000	Model-Free	-0.003	0.086	0.080	0.918	-0.001	0.092	0.093	0.948
			Model-Based	-0.005	0.079	0.074	0.914	-0.002	0.086	0.085	0.948

Table 6. Estimation for Example 1 of the target parameter with varied d under target sample size $n = 300$.

			$d = 1$ (Censoring Rate $\approx 34\%$)				$d = 2$ (Censoring Rate $\approx 41\%$)			
Method	n°	Model	BIAS	SD	SE	CP	BIAS	SD	SE	CP
TO		—	-0.002	0.152	0.154	0.950	-0.001	0.161	0.168	0.960
AIT	1000	Model-Free	-0.007	0.133	0.127	0.938	-0.009	0.144	0.140	0.948
		Model-Based	0.024	0.120	0.119	0.934	0.073	0.126	0.127	0.910
	5000	Model-Free	-0.007	0.126	0.120	0.938	-0.012	0.135	0.133	0.948
		Model-Based	0.034	0.109	0.106	0.938	0.093	0.110	0.112	0.870

8. Discussion

We have developed a transfer learning framework for survival analysis that leverages aggregate auxiliary information in the form of subgroup survival probabilities derived from Kaplan–Meier estimators. The proposed methodology addresses two fundamental tasks, estimation of a prespecified target parameter and construction of hypothesis tests for it. A key innovation is the introduction of a transitional element, a random vector computed from the target dataset analogously to the auxiliary information. This element is combined with a unified target-only estimator through a leveraging factor, enabling efficient information transfer without requiring individual-level source data. We have investigated two types of transitional elements. The model-free version is universally applicable and

robust, whereas the model-based version offers improved estimation efficiency at the cost of imposing a partly independent censoring assumption. A critical insight is that this assumption, often overlooked in existing literature, can lead to biased inference when violated. We have provided rigorous theoretical justification for the asymptotic properties of the proposed estimators and validated their finite-sample performance through numerical experiments. These contributions establish a principled and flexible framework for incorporating summary-level external information into survival analysis, improving statistical inference in settings with limited target data.

Our methodology has many advantages but also several important limitations that warrant further investigation. First, we focus on the scenario where the external information consists of subgroup survival probabilities. Other forms of aggregate information may exist, and our proposed framework can be readily generalized to accommodate such situations, provided that a valid transitional element can be constructed. In-depth studies on the specific derivations of the asymptotic properties of the resulting estimator are warranted. Second, as noted in Remark 5, the proposed method assumes homogeneity between the source and target domains. When this assumption is violated, the performance of the estimator may deteriorate, a phenomenon often referred to as negative transfer in [14]. Several recent works on transfer learning have attempted to overcome this barrier, for example, by characterizing heterogeneity through auxiliary source models with drifted parameters [5, 8, 49]. In the context of aggregate transfer learning, various shrinkage estimation approaches have been explored in different settings [20, 27–29, 50]. Potentially undesirable consequences caused by negative transfer should always be addressed appropriately to avoid spurious improvements in statistical analysis. These nontrivial limitations merit further research.

Appendix

A. Supplementary regularity conditions

This section provides supplementary regularity conditions needed in our main text. All of them are extracted from existing literature and commonly used for specific tasks. First, we impose the following series of regularity conditions.

Condition S1. *The point* $\tau > 0$ *is finite and* $\mathbb{P}(R(t) = 1 \ \forall \ t \in [0, \tau]) > 0$ *and* $\Lambda_0^\star(\tau) < \infty$.

Condition S2. *Covariates are uniformly bounded, that is,* $\|\boldsymbol{X}\|_\infty \leq C_{\boldsymbol{X}}$ *for* $C_{\boldsymbol{X}} < \infty$, *where* $\|\boldsymbol{a}\|_\infty = \max_{j\in[p]} |a_j|$ *for any p-dimensional vector* $\boldsymbol{a} = (a_1, \ldots, a_p)^\top$.

Condition S3. *The population-level information matrix* $\boldsymbol{H}^\star$ *is positive definite.*

Conditions S1 to S3 are routinely used in guaranteeing the asymptotic properties of the estimator $\hat{\boldsymbol{\beta}}$, as illustrated by Andersen and Gill [41]. To identify these marginal quantities from the observed data, we make the following standard assumptions in causal inference.

Condition S4 (Consistency). *The observed outcome is the potential outcome under the actual treatment received, that is,* $T = T(W)$.

Condition S5 (Conditional exchangeability). *The treatment assignment is independent of the potential outcomes given the covariates, that is,* $W \perp \{T(0), T(1)\}|\boldsymbol{Z}$.

Condition S6 (Positivity). *The probability of receiving either treatment, given the covariates, is positive, that is,* $0 < \mathbb{P}(W = 1|\boldsymbol{Z}) < 1$ *almost surely.*

B. Proofs of theorems

All the proofs of theorems listed in the main text are provided in this part.

Proof of Theorem 1

Proof. For each given $k \in [K]$, the auxiliary survival probability is essentially a Kaplan–Meier estimator, which has been shown in the literature to be asymptotic equivalent to the Nelson–Aalen estimator [51] defined as follows:

$$\hat{\phi}^{\circ}_{\mathrm{NA},k} = \exp\left\{-\sum_{i=1}^{n^{\circ}} \frac{\Delta_i^{\circ}\mathbb{I}_{\{\boldsymbol{X}_i^{\circ}\in\Omega_k\}}\mathbb{I}_{\{Y_i^{\circ}\le t_k^*\}}}{\sum_{j=1}^{n^{\circ}}\mathbb{I}_{\{\boldsymbol{X}_j^{\circ}\in\Omega_k\}}\mathbb{I}_{\{Y_j^{\circ}\ge Y_i^{\circ}\}}}\right\}. \tag{B.1}$$

It can be deduced that we can further write $\hat{\phi}^{\circ}_{\mathrm{NA},k}$ as follows:

$$-\log\left(\hat{\phi}^{\circ}_{\mathrm{NA},k}\right) = \sum_{i=1}^{n^{\circ}}\int_0^{t^*} \frac{1}{\sum_{j=1}^{n^{\circ}} R^{\circ}_{k,j}(t)} dN^{\circ}_{k,i}(t), \tag{B.2}$$

where $N^{\circ}_{k,i}(t) = \Delta_i^{\circ}\mathbb{I}_{\{Y_i^{\circ}\le t,\boldsymbol{X}_i^{\circ}\in\Omega_k\}}$ and $R^{\circ}_{k,i}(t) = \mathbb{I}_{\{Y_i^{\circ}\ge t,\boldsymbol{X}_j^{\circ}\in\Omega_k\}}$ for $i \in [n^{\circ}]$ and $k \in [K]$. Following the discussions in [52], if we further define

$$M^{\circ}_{k,i}(t) = N^{\circ}_{k,i}(t) - \int_0^t R^{\circ}_{k,i}(u)d\Lambda^{\#}_k(u),$$

we also have

$$\log\left(\phi^{\circ}_k\right) - \log\left(\hat{\phi}^{\circ}_{\mathrm{NA},k}\right) = \sum_{i=1}^{n^{\circ}}\int_0^{t^*} \frac{1}{\sum_{j=1}^{n^{\circ}} R^{\circ}_{k,j}(t)} dM^{\circ}_{k,i}(t).$$

The proof that the right-hand-side of this equation goes to zero in probability is analogous to that of Theorem 3.4.2 in [51]. Consequently, for $k \in [K]$, $\hat{\phi}^{\circ}_{\mathrm{NA},k}$ converges in probability to ϕ°_k, and so does $\hat{\phi}^{\circ}_k$. As in the proof of Theorem 6.2.1 of [51], we have

$$\log\left(\phi^{\circ}_k\right) - \log\left(\hat{\phi}^{\circ}_{\mathrm{NA},k}\right) = \frac{1}{n^{\circ}}\sum_{i=1}^{n^{\circ}}\int_0^{t_k^*} \frac{1}{\bar{R}_k(t)} dM^{\circ}_{k,i}(t) + o_{\mathbb{P}}(n^{\circ-1/2}),$$

where $\bar{R}_k(t) = \mathbb{P}(Y \ge t, \boldsymbol{X} \in \Omega_k)$. Motivated by this expression, we can then define $\boldsymbol{\zeta}^{\circ}(Y,\Delta,\boldsymbol{X}) = (\zeta^{\circ}_1(Y,\Delta,\boldsymbol{X}),\ldots,\zeta^{\circ}_K(Y,\Delta,\boldsymbol{X}))^{\top}$, where for $k \in [K]$, $\zeta^{\circ}_k(Y,\Delta,\boldsymbol{X}) = -\phi^{\circ}_k\int_0^{t_k^*}\bar{R}_k(t)^{-1}dM^{\circ}_k(t)$ and $M^{\circ}_k(t)$ is defined similarly to $M^{\circ}_{k,i}(t)$ with $(Y^{\circ}_i,\Delta^{\circ}_i,\boldsymbol{X}^{\circ}_i)$ replaced by $(Y,\Delta,\boldsymbol{X})$. Hence, we complete the proof. □

Proof of Theorem 2

Proof. As has been shown in our main text, it can be shown that we have the decomposition by applying the mean value theorem:

$$\begin{aligned}\hat{\alpha} - \alpha^{\star} &= \hat{h}(\boldsymbol{\beta}^{\star}) - \alpha^{\star} + \hat{\boldsymbol{h}}_{\bullet}(\tilde{\boldsymbol{\beta}})^{\top}\boldsymbol{H}^{\star-1}\hat{\boldsymbol{U}}(\boldsymbol{\beta}^{\star}) \\ &= \hat{h}(\boldsymbol{\beta}^{\star}) - \alpha^{\star} + \boldsymbol{h}_{\bullet}^{\top}\boldsymbol{H}^{\star-1}\hat{\boldsymbol{U}}(\boldsymbol{\beta}^{\star}) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

in which the second equality is due to Condition 2 and the fact that $\sqrt{n}\hat{\boldsymbol{U}}(\boldsymbol{\beta}^{\star}) = O_{\mathbb{P}}(1)$ [41]. By further adopting Condition 1 and the expression shown in Eq (4.2), we have:

$$\hat{\alpha} - \alpha^{\star} = \frac{1}{n}\sum_{i=1}^{n}\Big[\xi_h(Y_i, \Delta_i, \boldsymbol{X}_i) + \boldsymbol{h}_{\bullet}^{\star\top}\boldsymbol{H}^{\star-1}\boldsymbol{\xi}_U(Y_i, \Delta_i, \boldsymbol{X}_i)\Big],$$

which implies that the assertion of Eq (2) in this theorem holds. The second equality, that is, the asymptotic normality $\sqrt{n}(\hat{\alpha} - \alpha^{\star})/\sqrt{v} \xrightarrow{d} \mathcal{N}(0, 1)$ follows immediately from the central limit theorem [44]. Therefore, we complete the proof. □

Proof of Theorem 3

Proof. According to the definition of $\hat{\alpha}^{\dagger}(\boldsymbol{b})$, we can deduce that the following equalities are satisfied:

$$\begin{aligned}
\sqrt{n}(\hat{\alpha}^{\dagger}(\boldsymbol{b}) - \alpha^{\star}) &= \sqrt{n}(\hat{\alpha} - \alpha^{\star}) - \boldsymbol{b}^{\top}\Big[\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{\circ}\Big] + \boldsymbol{b}^{\top}\Big[\sqrt{n}(\hat{\boldsymbol{\phi}}^{\circ} - \boldsymbol{\phi}^{\circ})\Big] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\xi(Y_i, \Delta_i, \boldsymbol{X}_i) - \boldsymbol{b}^{\top}\boldsymbol{\zeta}(Y_i, \Delta_i, \boldsymbol{X}_i)\Big] + \boldsymbol{b}^{\top}\Big[\sqrt{n}(\hat{\boldsymbol{\phi}}^{\circ} - \boldsymbol{\phi}^{\circ})\Big] + o_{\mathbb{P}}(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi^{\dagger}(\boldsymbol{b}, Y_i, \Delta_i, \boldsymbol{X}_i) + \boldsymbol{b}^{\top}\Big[\sqrt{n}(\hat{\boldsymbol{\phi}}^{\circ} - \boldsymbol{\phi}^{\circ})\Big] + o_{\mathbb{P}}(1),
\end{aligned}$$

where we set $\xi^{\dagger}(\boldsymbol{b}, Y, \Delta, \boldsymbol{X}) = \xi(Y, \Delta, \boldsymbol{X}) - \boldsymbol{b}^{\top}\boldsymbol{\zeta}(Y, \Delta, \boldsymbol{X})$. The asymptotic normality of the first part follows immediately from the central limit theorem [44] and the resulting variance-covariance matrix is $\mathbb{E}\Big[\xi^{\dagger}(\boldsymbol{b}, Y, \Delta, \boldsymbol{X})\Big] = v - 2\boldsymbol{\gamma}^{\top}\boldsymbol{b} + \boldsymbol{b}^{\top}\boldsymbol{V}\boldsymbol{b}$. The asymptotic normality of the second part follows similarly by recalling Theorem 1 and the variance-covariance matrix is $\rho\boldsymbol{V}^{\circ}$. Since the first part is independent of the second part, we can then conclude that the required asymptotic normality list in this theorem holds. This completes the proof. □

Proof of Theorem 4

Proof. By our discussions shown in Section 5.2, we have $\sqrt{n}(\hat{\alpha}^{\dagger}_{\mathrm{AIT}} - \alpha^{\star}) = \sqrt{n}(\hat{\alpha}^{\dagger}(\boldsymbol{b}^{\dagger}) - \alpha^{\star}) + o_{\mathbb{P}}(1)$. Invoking the result of Theorem 3, the asserted joint normality follows immediately by fixing $\boldsymbol{b} = \boldsymbol{b}^{\dagger}$, which completes the proof. □

Proof of Theorem 5

Proof. When the partly independent censoring assumption holds, we know that the copula function C_k can be written as $C_k(u, v) = uv$ for $k \in [K]$. Consequently, according to the specific definition of $\boldsymbol{\phi}^{\circ}$, we can conclude directly that $\boldsymbol{\phi}^{\circ} = \boldsymbol{\phi}^{\#}$. The convergence in probability stated in the theorem follows immediately from Theorem 1, which completes the proof. □

Proof of Theorem 6

Proof. Let the index of subscript $k = 1, \ldots, K$ be fixed. By the law of large numbers, we have that $\hat{\pi}_k$ converges in probability to π_k as n tends to infinity. Consequently, according to the specific definition of the transitional element $\hat{\phi}_k$, our key idea of proving its I.I.D. expression is to first deduce the I.I.D expression for $\hat{\psi}_k$. Let us first decompose $\hat{\psi}_k$ into several parts as follows:

$$\hat{\psi}_k = \hat{h}_k(\hat{\boldsymbol{\beta}}) = \hat{h}_k(\hat{\boldsymbol{\beta}}) - \hat{h}_k(\boldsymbol{\beta}^{\star}) + \hat{h}_k(\boldsymbol{\beta}^{\star}).$$

We next deal with the two parts $\hat{h}_k(\hat{\boldsymbol{\beta}}) - \hat{h}_k(\boldsymbol{\beta}^\star)$ and $\hat{h}_k(\boldsymbol{\beta}^\star)$ separately. We will demonstrate that these two parts can respectively be written into I.I.D. expressions.

We first focus on $\hat{h}_k(\hat{\boldsymbol{\beta}}) - \hat{h}_k(\boldsymbol{\beta}^\star)$. Let us define $\hat{\boldsymbol{h}}_{\bullet,k}(\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})\hat{h}_k(\boldsymbol{\beta})$ and it can be shown that it has the following explicit expression:

$$-\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{I}_{\{\boldsymbol{X}_i\in\Omega_k\}}\exp(\boldsymbol{X}_i^\top\boldsymbol{\beta})\exp(-\hat{\Lambda}_0(t_k^*,\boldsymbol{\beta})\exp(\boldsymbol{X}_i^\top\boldsymbol{\beta}))\left[\hat{\boldsymbol{\Lambda}}_{0,\bullet}(t_k^*,\boldsymbol{\beta}) + \hat{\Lambda}_0(t_k^*,\boldsymbol{\beta})\boldsymbol{X}_i\right]\right\},$$

in which we denote $\hat{\boldsymbol{\Lambda}}_{0,\bullet}(t,\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})\hat{\Lambda}_0(t,\boldsymbol{\beta}) = -n^{-1}\sum_{i=1}^{n}\Delta_i\mathbb{I}_{\{Y_i\le t\}}\hat{\boldsymbol{\eta}}(Y_i,\boldsymbol{\beta})/\hat{\mu}^{(0)}(Y_i,\boldsymbol{\beta})$. Then, by the mean value theorem, it holds that:

$$\hat{h}_k(\hat{\boldsymbol{\beta}}) - \hat{h}_k(\boldsymbol{\beta}^\star) = \hat{\boldsymbol{h}}_{\bullet,k}(\tilde{\boldsymbol{\beta}})^\top(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star),$$

for some random vector $\tilde{\boldsymbol{\beta}}$ that lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^\star$. Based on the statistical convergences of $\hat{\Lambda}_0(t,\boldsymbol{\beta})$ and $\hat{\boldsymbol{\Lambda}}_{0,\bullet}(t,\boldsymbol{\beta})$, it can be deduced that $\hat{\boldsymbol{h}}_{\bullet,k}(\tilde{\boldsymbol{\beta}}_t)$ should converge in probability to $\boldsymbol{h}^\star_{\bullet,k}$ since the estimator $\tilde{\boldsymbol{\beta}}$ satisfies $\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star\| = o_{\mathbb{P}}(1)$ due to its definition and the continuity of $\hat{h}_k(\boldsymbol{\beta})$ in $\boldsymbol{\beta}$. Consequently, we have

$$\begin{aligned}\hat{h}_k(\hat{\boldsymbol{\beta}}) - \hat{h}_k(\boldsymbol{\beta}^\star) &= \boldsymbol{h}^{\star\top}_{\bullet,k}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + o_{\mathbb{P}}(n^{-1/2})\\ &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{h}^{\star\top}_{\bullet,k}\boldsymbol{H}^{\star-1}\boldsymbol{\xi}_{\boldsymbol{U}}(Y_i,\Delta_i,\boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

in which we have used the fact that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star = n^{-1}\sum_{i=1}^{n}\boldsymbol{H}^{\star-1}\boldsymbol{\xi}_{\boldsymbol{U}}(Y_i,\Delta_i,\boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2})$ under Conditions S1 to S3 and the conditionally independent censoring assumption, a property that has been discussed in the literature [41–43].

We then come to the term $\hat{h}_k(\boldsymbol{\beta}^\star)$. Starting from this essential quantity, we have the following decomposition:

$$\hat{h}_k(\boldsymbol{\beta}^\star) - \pi_k\phi_k^\circ = \hat{I} + \frac{1}{n}\sum_{i=1}^{n}\left[\psi(\boldsymbol{X}_i,\boldsymbol{\beta}^\star,\Lambda_0^\star(t_k^*),\Omega_k) - \pi_k\phi_k^\circ\right],$$

where we write

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n}\psi(\boldsymbol{X}_i,\boldsymbol{\beta}^\star,\hat{\Lambda}_0(t_k^*,\boldsymbol{\beta}^\star),\Omega_k) - \frac{1}{n}\sum_{i=1}^{n}\psi(\boldsymbol{X}_i,\boldsymbol{\beta}^\star,\Lambda_0^\star(t_k^*),\Omega_k).$$

Note that $\hat{h}_k(\boldsymbol{\beta}^\star) - \pi_k\phi_k^\circ - \hat{I}$ is already in an I.I.D. expression. Denote

$$\begin{aligned}\hat{g}_{\bullet,k}(u) &= \frac{1}{n}\frac{\partial}{\partial u}\left[\sum_{i=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_i\in\Omega_k\}}\exp(-u\exp(\boldsymbol{X}_i^\top\boldsymbol{\beta}^\star))\right]\\ &= -\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_i\in\Omega_k\}}\exp(\boldsymbol{X}_i^\top\boldsymbol{\beta}^\star)\exp\left(-u\exp(\boldsymbol{X}_i^\top\boldsymbol{\beta}^\star)\right).\end{aligned}$$

Again using the mean value theorem, it holds that:

$$\hat{I} = \hat{g}_{\bullet,k}(\tilde{\Lambda}_{0,k})(\hat{\Lambda}_0(t_k^*,\boldsymbol{\beta}^\star) - \Lambda_0^\star(t_k^*)),$$

for some random variable $\tilde{\Lambda}_{0,k}$ that lies between $\hat{\Lambda}_0(t_k^*,\boldsymbol{\beta}^\star)$ and $\Lambda_0^\star(t_k^*)$. Based on the statistical convergence of $\hat{\Lambda}_0(t_k^*,\boldsymbol{\beta})$, $\hat{g}_{\bullet,k}(\tilde{\Lambda}_{0,k})$ should converge in probability to $g^\star_{\bullet,k}$ since the estimator $\tilde{\Lambda}_{0,k}$ satisfies $|\tilde{\Lambda}_{0,k}(t_k^*) - \Lambda_0^\star(t_k^*)| = o_{\mathbb{P}}(1)$ due to its definition and the continuity of $\hat{g}_{\bullet,k}(u)$ in u. Consequently, we have the following equation:

$$\begin{aligned}\hat{I} &= g^{\star}_{\bullet,k}(\hat{\Lambda}_0(t^*_k,\boldsymbol{\beta}^{\star}) - \Lambda^{\star}_0(t^*_k)) + o_{\mathbb{P}}(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^{n} g^{\star}_{\bullet,k}\xi_{\Lambda}(t^*_k, Y_i, \Delta_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

in which we have used the fact that $\hat{\Lambda}_0(t,\boldsymbol{\beta}^{\star}) - \Lambda^{\star}_0(t) = n^{-1}\sum_{i=1}^{n}\xi_{\Lambda}(t, Y_i, \Delta_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2})$ under Conditions S1 to S3 and the conditionally independent censoring assumption, a property that has been discussed in the literature [43]. Therefore, we complete the proof. □

C. Computational details

We provide further details on the concrete formulas needed to construct the AIT estimator $\hat{\alpha}_{\mathrm{AIT}}$ and its empirical variance $\hat{v}_{\mathrm{AIT}}$. As discussed in Section 5.2, the key steps lie in the construction of $\hat{\xi}$ and $\hat{\boldsymbol{\zeta}}$. All other quantities, such as the leveraging factor $\hat{\boldsymbol{b}}$ required for $\hat{\alpha}_{\mathrm{AIT}}$ and the aforementioned $\hat{v}_{\mathrm{AIT}}$, can be obtained accordingly, as will be shown later.

Let $\hat{\boldsymbol{\mu}}^{(1)}(t,\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} R_i(t)\exp(\boldsymbol{X}_i^{\top}\boldsymbol{\beta})\boldsymbol{X}_i$ and $\hat{\boldsymbol{\eta}}(t,\boldsymbol{\beta}) = \hat{\boldsymbol{\mu}}^{(1)}(t,\boldsymbol{\beta})/\hat{\mu}^{(0)}(t,\boldsymbol{\beta})$. For $i \in [n]$, define

$$\hat{M}_i(t) = N_i(t) - \int_0^t R_i(u)\exp(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\beta}})d\hat{\Lambda}_0(u),$$

where $N_i(t) = \Delta_i\mathbb{I}_{\{Y_i \le t\}}$. Then, for $i \in [n]$, define

$$\hat{\boldsymbol{\xi}}_{\boldsymbol{U},i} = \hat{\boldsymbol{\xi}}_{\boldsymbol{U}}(Y_i, \Delta_i, \boldsymbol{X}_i) = \int_0^{\infty}\left\{\boldsymbol{X}_i - \hat{\boldsymbol{\eta}}(s,\hat{\boldsymbol{\beta}})\right\}d\hat{M}_i(s).$$

Consequently, we can naturally define $\hat{\xi}$ as follows:

$$\hat{\xi}(Y,\Delta,\boldsymbol{X}) = \hat{\xi}_h(Y,\Delta,\boldsymbol{X}) + \hat{\boldsymbol{h}}_{\bullet}(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{H}}^{-1}\hat{\boldsymbol{\xi}}_{\boldsymbol{U}}(Y,\Delta,\boldsymbol{X}), \tag{C.1}$$

where $\hat{\boldsymbol{H}} = -(\partial^2/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\top})\hat{l}(\boldsymbol{\beta})$ estimates $\boldsymbol{H}^{\star}$ and $\hat{\xi}_h(Y,\Delta,\boldsymbol{X})$ is an estimator of $\xi_h(Y,\Delta,\boldsymbol{X})$.

The specific form of $\hat{\xi}_h(Y,\Delta,\boldsymbol{X})$ depends on the target parameters. We next discuss its computation.

Details for Example 1. For Example 1, we simply have $\hat{\xi}_h(Y,\Delta,\boldsymbol{X}) = \xi_h(Y,\Delta,\boldsymbol{X}) = 0$.

Details for Example 2. For Example 2, note that

$$\begin{aligned}\hat{h}(\boldsymbol{\beta}^{\star}) - \alpha^{\star} &= \exp\{-\hat{\Lambda}_0(t^{\star},\boldsymbol{\beta}^{\star})\exp(\boldsymbol{x}^{\star\top}\boldsymbol{\beta}^{\star})\} - \exp\{-\Lambda^{\star}_0(t^{\star})\exp(\boldsymbol{x}^{\star\top}\boldsymbol{\beta}^{\star})\} \\ &= -\alpha^{\star}\exp(\boldsymbol{x}^{\star\top}\boldsymbol{\beta}^{\star})(\hat{\Lambda}_0(t^{\star},\boldsymbol{\beta}^{\star}) - \Lambda^{\star}_0(t^{\star})) + o_{\mathbb{P}}(n^{-1/2}) \\ &= -\sum_{i=1}^{n}\alpha^{\star}\exp(\boldsymbol{x}^{\star\top}\boldsymbol{\beta}^{\star})\xi_{\Lambda}(t^{\star}, Y_i, \Delta_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

which yields the influence function ξ_h. Further denote

$$\hat{\xi}_{\Lambda,i}(t) = \hat{\xi}_{\Lambda}(t, Y_i, \Delta_i, \boldsymbol{X}_i) = \int_0^t \frac{d\hat{M}_i(s)}{\hat{\mu}_0(s,\hat{\boldsymbol{\beta}})},$$

for $i \in [n]$. Consequently, an estimator of ξ_h under Example 2 can be defined by

$$\hat{\xi}_h(Y,\Delta,\boldsymbol{X}) = -\hat{\alpha}\exp(\boldsymbol{x}^{\star\top}\hat{\boldsymbol{\beta}})\hat{\xi}_{\Lambda}(t^{\star}, Y, \Delta, \boldsymbol{X}).$$

Details for Example 3. We next consider Example 3. For $w = 1$ or 0, it holds that

$$\frac{1}{n}\sum_{i=1}^{n} q(\boldsymbol{X}_{w,i}, \boldsymbol{\beta}^\star, \hat{\Lambda}_0(t^\star, \boldsymbol{\beta}^\star)) - S_w(t^\star) = \hat{I}_w + \frac{1}{n}\sum_{i=1}^{n}\left[S(t^\star|\boldsymbol{X}_{w,i}) - S_w(t^\star)\right], \tag{C.2}$$

where we write

$$\hat{I}_w = \frac{1}{n}\sum_{i=1}^{n} q(\boldsymbol{X}_{w,i}, \boldsymbol{\beta}^\star, \hat{\Lambda}_0(t^\star, \boldsymbol{\beta}^\star)) - \frac{1}{n}\sum_{i=1}^{n} q(\boldsymbol{X}_{w,i}, \boldsymbol{\beta}^\star, \Lambda_0(t^\star)).$$

It can be further deduced that

$$\begin{aligned}\hat{I}_w &= -\frac{1}{n}\sum_{i=1}^{n} q(\boldsymbol{X}_{w,i}, \boldsymbol{\beta}^\star, \Lambda_0(t^\star)) \exp(\boldsymbol{X}_{w,i}^\top \boldsymbol{\beta}^\star)(\hat{\Lambda}_0(t^\star, \boldsymbol{\beta}^\star) - \Lambda_0^\star(t^\star)) + o_{\mathbb{P}}(n^{-1/2}) \\ &= -\sum_{i=1}^{n} S(t^\star|\boldsymbol{X}_{w,i}) \exp(\boldsymbol{X}_{w,i}^\top \boldsymbol{\beta}^\star)\xi_\Lambda(t^\star, Y_i, \Delta_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}).\end{aligned}$$

This formula, together with Equation (C.2), jointly determines the specific expression for ξ_h and we can then estimate it via $\hat{\xi}_h(Y, \Delta, \boldsymbol{X}) = \hat{\xi}_{1,h}(Y, \Delta, \boldsymbol{X}) - \hat{\xi}_{0,h}(Y, \Delta, \boldsymbol{X})$, where for $w = 1$ or 0, we let

$$\hat{\xi}_{w,h}(Y, \Delta, \boldsymbol{X}) = \hat{S}(t^\star|\boldsymbol{X}_w) - \hat{S}_w(t^\star) - \hat{S}(t^\star|\boldsymbol{X}_{w,i}) \exp(\boldsymbol{X}_w^\top \hat{\boldsymbol{\beta}})\hat{\xi}_\Lambda(t^\star, Y, \Delta, \boldsymbol{X}).$$

Based on the preceding discussion and Eq (C.1), the empirical version of ξ can be obtained in practice for the three examples considered. The empirical version $\hat{\boldsymbol{\zeta}}$ does not depend on the specific target parameter, as it is derived from the influence function of the transitional element, regardless of whether the model-free or model-based version is used. From the derivation of $\hat{\xi}$ under different scenarios, we find that the key step is obtaining the explicit expression for ξ. Once the expression is known, the subsequent derivation follows the same pattern: replace all theoretical quantities with their empirical counterparts. For example, in constructing $\hat{\xi}$, we replaced $\mu^{(0)}$ and $\boldsymbol{\mu}^{(1)}$ with $\hat{\mu}^{(0)}$ and $\hat{\boldsymbol{\mu}}^{(1)}$, and $\Lambda_0^\star$ with $\hat{\Lambda}_0$, among others. Since the explicit expressions for $\boldsymbol{\zeta}$ under both the model-free and model-based versions have already been derived, the empirical version $\boldsymbol{\zeta}$ can be obtained in the same manner and we omit the repetitive details here.

Having defined $\hat{\xi}$ and $\hat{\boldsymbol{\zeta}}$, we then obtain $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{V}}$ as in Eq (5.4). Consequently, the empirical variance $\hat{v}_{\text{AIT}}$, which estimates v_{AIT} as given in Theorem 4, is given by

$$\hat{v}_{\text{AIT}} = \hat{v} - \hat{\boldsymbol{\gamma}}^\top(\hat{\boldsymbol{V}} + \hat{\rho}\hat{\boldsymbol{V}}^\circ)^{-1}\hat{\boldsymbol{\gamma}}.$$

Ideally, $\hat{\boldsymbol{V}}^\circ$ would be obtained directly from the source domain. Nevertheless, we do not assume access to individual-level data in the source domain, and the literature typically does not report $\hat{\boldsymbol{V}}^\circ$. To address this issue, we can estimate it using the target dataset $\mathcal{O}_n$. Notably, the model-free transitional element shares the same asymptotic properties as the auxiliary information, so we can simply set $\hat{\boldsymbol{V}}^\circ = \hat{\boldsymbol{V}}$. When the model-based transitional element is used instead, the influence function of the transitional element is different from that of the external Kaplan–Meier estimator. In this case, $\hat{\boldsymbol{V}}$ as in Eq (5.4) is based on the influence function of the model-based transitional element and therefore does not provide an estimator of $\boldsymbol{V}^\circ$. Nevertheless, a consistent estimator of $\boldsymbol{V}^\circ$ can still be obtained from the target data by evaluating Eq (5.4) with the estimated influence function of the model-free transitional element. Although such a quantity is not the covariance matrix associated with the model-based transitional element, it serves as the required estimator of $\boldsymbol{V}^\circ$.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Jie Ding's work is supported by the Fundamental Research Funds for the Central Universities (Grant 3132026192). Bo Han gratefully acknowledges funding from the Yunnan Fundamental Research Projects (202501AU070055), the Young Scientists Fund of the National Natural Science Foundation of China (Category C) (12501389), the Ministry of Education Humanities and Social Sciences Research Project (25YJC910001), and the Xing Dian Ying Cai Support Program Project.

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. D. M. Witten, R. Tibshirani, Survival analysis with high-dimensional covariates, *Stat. Methods Med. Res.*, **19** (2010), 29–51. https://doi.org/10.1177/0962280209105024
2. J. Li, S. Ma, *Survival Analysis in Medicine and Genetics*, Chapman & Hall CRC Press, New York, 2013.
3. K. Weiss, T. M. Khoshgoftaar, D. Wang, A survey of transfer learning, *J. Big Data*, **3** (2016), 9. https://doi.org/10.1186/s40537-016-0043-6
4. T. Cai, M. Liu, Y. Xia, Individual data protected integrative regression analysis of high-dimensional heterogeneous data, *J. Am. Stat. Assoc.*, **117** (2022), 2105–2119. https://doi.org/10.1080/01621459.2021.1904958
5. S. Li, T. T. Cai, H. Li, Transfer learning for high-dimensional linear regression: prediction, estimation and minimax optimality, *J. R. Stat. Soc. B*, **84** (2022), 149–173. https://doi.org/10.1111/rssb.12479
6. Y. Tian, Y. Feng, Transfer learning under high-dimensional generalized linear models, *J. Am. Stat. Assoc.*, **118** (2023), 2684–2697. https://doi.org/10.1080/01621459.2022.2071278
7. S. Li, L. Zhang, T. T. Cai, H. Li, Estimation and inference for high-dimensional generalized linear models with knowledge transfer, *J. Am. Stat. Assoc.*, **119** (2024), 1274–1285. https://doi.org/10.1080/01621459.2023.2184373
8. Z. Li, Y. Shen, J. Ning, Accommodating time-varying heterogeneity in risk estimation under the Cox model: a transfer learning approach, *J. Am. Stat. Assoc.*, **118** (2023), 2276–2287. https://doi.org/10.1080/01621459.2023.2210336
9. Y. B. Pei, Z. Y. Yu, J. S. Shen, Transfer learning for accelerated failure time model with microarray data, *BMC Bioinf.*, **26** (2025), 84. https://doi.org/10.1186/s12859-025-06056-w
10. D. R. Cox, Regression models and life tables, *J. R. Stat. Soc. B*, **34** (1972), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

11. M. Xie, T. Hu, J. Zhou, Transfer learning under the Cox model with interval-censored data, *Stat. Anal. Data Min.*, **17** (2024), e11680. https://doi.org/10.1002/sam.11680
12. D. Deng, L. Zhang, H. Feng, V. M. Chinchilli, C. Chen, M. Wang, Improving estimation efficiency for survival data analysis by integrating a coarsened time-to-event outcome from an external study, *Biometrics*, **81** (2025), ujae168. https://doi.org/10.1093/biomtc/ujae168
13. J. Qin, Y. Liu, P. Li, A selective review of statistical methods using calibration information from similar studies, *Stat. Theory Relat. Fields*, **6** (2022), 175–190. https://doi.org/10.1080/24754269.2022.2037201
14. S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, **22** (2010), 1345–1359. https://doi.org/10.1109/TKDE.2009.191
15. X. Hu, X. Zhang, Optimal parameter-transfer learning by semiparametric model averaging, *J. Mach. Learn. Res.*, **24** (2023), 1–53. Available from: https://jmlr.org/papers/v24/23-0030.html.
16. J. Qin, J. Lawless, Empirical likelihood and general estimating equations, *Ann. Stat.*, **22** (1994), 300–325. https://doi.org/10.1214/aos/1176325370
17. N. Chatterjee, Y. H. Chen, P. Maas, R. J. Carroll, Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources, *J. Am. Stat. Assoc.*, **111** (2016), 107–117. https://doi.org/10.1080/01621459.2015.1123157
18. H. Zhang, L. Deng, M. Schiffman, J. Qin, K. Yu, Generalized integration model for improved statistical inference by leveraging external summary data, *Biometrika*, **107** (2020), 689–703. https://doi.org/10.1093/biomet/asaa014
19. C. Y. Huang, J. Qin, H. T. Tsai, Efficient estimation of the Cox model with auxiliary subgroup survival information, *J. Am. Stat. Assoc.*, **111** (2016), 787–799. https://doi.org/10.1080/01621459.2015.1044090
20. Z. Chen, J. Ning, Y. Shen, J. Qin, Combining primary cohort data with external aggregate information without assuming comparability, *Biometrics*, **77** (2021), 1024–1036. https://doi.org/10.1111/biom.13356
21. J. He, H. Li, S. Zhang, X. Duan, Additive hazards model with auxiliary subgroup survival information, *Lifetime Data Anal.*, **25** (2019), 128–149. https://doi.org/10.1007/s10985-018-9426-7
22. B. Han, I. van Keilegom, X. Wang, Semiparametric estimation of the non-mixture cure model with auxiliary survival information, *Biometrics*, **78** (2022), 448–459. https://doi.org/10.1111/biom.13450
23. Y. J. Cheng, Y. C. Liu, C. Y. Tsai, C. Y. Huang, Semiparametric estimation of the transformation model by leveraging external aggregate data in the presence of population heterogeneity, *Biometrics*, **79** (2023), 1996–2009. https://doi.org/10.1111/biom.13778
24. L. P. Hansen, Large sample properties of generalized method of moments estimators, *Econometrica*, **50** (1982), 1029–1054. https://doi.org/10.2307/1912775
25. W. Shang, X. Wang, The generalized moment estimation of the additive-multiplicative hazard model with auxiliary survival information, *Comput. Stat. Data Anal.*, **112** (2017), 154–169. https://doi.org/10.1016/j.csda.2017.03.013

26. Y. Sheng, Y. Sun, D. Deng, C. Y. Huang, Censored linear regression in the presence or absence of auxiliary survival information, *Biometrics*, **76** (2020), 734–745. https://www.jstor.org/stable/48584672
27. J. Ding, J. Li, Y. Han, I. W. McKeague, X. Wang, Fitting additive risk models using auxiliary information, *Stat. Med.*, **42** (2023), 894–916. https://doi.org/10.1002/sim.9649
28. J. Ding, J. Li, M. Zhang, X. Wang, CureAuxSP: An R package for estimating mixture cure models with auxiliary survival probabilities, *Comput. Methods Programs Biomed.*, **251** (2024), 108212. https://doi.org/10.1016/j.cmpb.2024.108212
29. J. Ding, J. Li, X. Wang, Efficient auxiliary information synthesis for cure rate model, *J. R. Stat. Soc. C*, **73** (2024), 497–521. https://doi.org/10.1093/jrsssc/qlad106
30. C. Y. Huang, J. Qin, A unified approach for synthesizing population-level covariate effect information in semiparametric estimation with survival data, *Stat. Med.*, **39** (2020), 1573–1590. https://doi.org/10.1002/sim.8499
31. P. Xie, J. Ding, X. Wang, Leveraging external aggregated information for the marginal accelerated failure time model, *Stat. Med.*, **43** (2024), 5203–5216. https://doi.org/10.1002/sim.10224
32. Z. Chen, Y. Shen, J. Qin, J. Ning, Likelihood adaptively incorporated external aggregate information with uncertainty for survival data, *Biometrics*, **80** (2024), ujae120. https://doi.org/10.1093/biomtc/ujae120
33. P. F. Su, J. Zhong, Y. C. Liu, T. H. Lin, H. T. Ou, Efficient estimation of a Cox model when integrating the subgroup incidence rate information, *J. Appl. Stat.*, **50** (2023), 2151–2170. https://doi.org/10.1080/02664763.2022.2068512
34. J. Y. Hung, J. Zhong, H. T. Ou, P. F. Su, Efficient estimation of the Cox model when incorporating the subgroup restricted mean survival time, *J. Biopharm. Stat.*, **36** (2025), 527–548. https://doi.org/10.1080/10543406.2024.2444242
35. J. D. Kalbfleisch, D. E. Schaubel, Fifty years of the Cox model, *Annu. Rev. Stat. Appl.*, **10** (2023), 1–23. https://doi.org/10.1146/annurev-statistics-033021-014043
36. K. M. Leung, R. M. Elashoff, A. A. Afifi, Censoring issues in survival analysis, *Annu. Rev. Public Health*, **18** (1997), 83–104. https://doi.org/10.1146/annurev.publhealth.18.1.83
37. J. P. Klein, M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd edition, Springer-Verlag, New York, 2003. https://doi.org/10.1007/b97377
38. E. L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, *J. Am. Stat. Assoc.*, **53** (1958), 457–481. https://doi.org/10.1080/01621459.1958.10501452
39. A. Sklar, Fonctions de répartition à *n* dimensions et leurs marges, *Publ. Inst. Stat. Univ. Paris*, **8** (1959), 229–231.
40. D. R. Cox, Partial likelihood, *Biometrika*, **62** (1975), 269–276. https://doi.org/10.1093/biomet/62.2.269
41. P. K. Andersen, R. D. Gill, Cox's regression model for counting processes: a large sample study, *Ann. Stat.*, **10** (1982), 1100–1120. https://doi.org/10.1214/aos/1176345976
42. N. Reid, H. Crépeau, Influence functions for proportional hazards regression, *Biometrika*, **72** (1985), 1–9. https://doi.org/10.1093/biomet/72.1.1

43. M. R. Kosorok, *Introduction to Empirical Processes and Semiparametric Inference*, Springer, New York, 2008. https://doi.org/10.1007/978-0-387-74978-5
44. A. W. Van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 2012. https://doi.org/10.1017/CBO9780511802256
45. T. Martinussen, S. Vansteelandt, P. K. Andersen, Subtleties in the interpretation of hazard ratios, *Lifetime Data Anal.*, **26** (2020), 833–855. https://doi.org/10.1007/s10985-020-09501-5
46. M. J. Stensrud, M. A. Hernán, Why test for proportional hazards? *JAMA*, **323** (2020), 1401–1402. https://doi.org/10.1001/jama.2020.1267
47. M. A. Hernán, J. M. Robins, *Causal Inference: What If*, Chapman & Hall/CRC, Boca Raton, 2020.
48. J. M. Robins, A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect, *Math. Modell.*, **7** (1986), 1393–1512. https://doi.org/10.1016/0270-0255(86)90088-6
49. S. Li, T. Cai, R. Duan, Targeting underrepresented populations in precision medicine: a federated transfer learning approach, *Ann. Appl. Stat.*, **17** (2023), 2970–2992. https://doi.org/10.1214/23-AOAS1747
50. Y. Zhai, P. Han, Data integration with oracle use of external information from heterogeneous populations, *J. Comput. Graphical Stat.*, **31** (2022), 1001–1012. https://doi.org/10.1080/10618600.2022.2050248
51. T. R. Fleming, D. P. Harrington, *Counting Processes and Survival Analysis*, Wiley, New York, 1991.
52. L. P. Rivest, M. T. Wells, A martingale approach to the copula-graphic estimator for the survival function under dependent censoring, *J. Multivar. Anal.*, **79** (2001), 138–155. https://doi.org/10.1006/jmva.2000.1959

AIMS Press