Published in final edited form as:
J Am Stat Assoc. 2023 ; 118(544): 2276–2287. doi:10.1080/01621459.2023.2210336.

Accommodating time-varying heterogeneity in risk estimation under the Cox model: a transfer learning approach

Ziyi Li,
Yu Shen,
Jing Ning[*]
Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

Transfer learning has attracted increasing attention in recent years for adaptively borrowing information across different data cohorts in various settings. Cancer registries have been widely used in clinical research because of their easy accessibility and large sample size. Our method is motivated by the question of how to utilize cancer registry data as a complement to improve the estimation precision of individual risks of death for inflammatory breast cancer (IBC) patients at The University of Texas MD Anderson Cancer Center. When transferring information for risk estimation based on the cancer registries (i.e., source cohort) to a single cancer center (i.e., target cohort), time-varying population heterogeneity needs to be appropriately acknowledged. However, there is no literature on how to adaptively transfer knowledge on risk estimation with time-to-event data from the source cohort to the target cohort while adjusting for time-varying differences in event risks between the two sources. Our goal is to address this statistical challenge by developing a transfer learning approach under the Cox proportional hazards model. To allow data-adaptive levels of information borrowing, we impose Lasso penalties on the discrepancies in regression coefficients and baseline hazard functions between the two cohorts, which are jointly solved in the proposed transfer learning algorithm. As shown in the extensive simulation studies, the proposed method yields more precise individualized risk estimation than using the target cohort alone. Meanwhile, our method demonstrates satisfactory robustness against cohort differences compared with the method that directly combines the target and source data in the Cox model. We develop a more accurate risk estimation model for the MD Anderson IBC cohort given various treatment and baseline covariates, while adaptively borrowing information from the National Cancer Database to improve risk assessment.

[*]To whom the correspondence should be addressed to: jning@mdanderson.org.

SUPPLEMENTARY MATERIAL

Supplementary: this file includes Figures S1–9 and Tables S1–3. (pdf)

R-package for Trans-Cox implementation: R-package TransCox is publicly available from https://github.com/ziyili20/TransCox with a detailed user manual and example. The code to reproduce the simulation figures and results in the manuscript is provided through https://github.com/ziyili20/TransCox_reproduceFigures.

R-package for Trans-Cox implementation: R-package TransCox is publicly available from https://github.com/ziyili20/TransCox with a detailed user manual and example code.

Keywords

Cox proportional hazards model; Inflammatory breast cancer; National Cancer Database; Risk assessment; Transfer learning

1 Introduction

Estimating the risk of a failure event is an important topic in clinical research for chronic diseases such as cardiovascular disease and cancer (Jiao et al., 2018; Kumar et al., 2020). If the population of interest comes from a single institution or a clinical trial, however, the limited sample size may prevent accurate individualized risk assessment, especially for rare diseases. Here, we consider a rare but aggressive type of cancer called inflammatory breast cancer (IBC) (Jaiyesimi et al., 1992). Although IBC constitutes only 1% to 6% of all breast cancer patients in the United States, the diagnosed patients have a worse prognosis with five-year survival of around 34% to 47% (Masuda et al., 2014). IBC patients seen at the Morgan Welch Inflammatory Breast Cancer Research Program and Clinic at The University of Texas MD Anderson (MDA) Cancer Center are the target population of this study. Although MD Anderson Cancer Center is a leading cancer care institute, the performance of risk assessment using the data of the MDA cohort alone is far from satisfactory due to the rarity of the disease. Large population-based registry data sources, including the Surveillance, Epidemiology and End Results (SEER) and the National Cancer Data Base (NCDB), have been increasingly used as complement data cohorts because of their easy accessibility and large sample size (Carvalho et al., 2005; Bilimoria et al., 2008).

Combining information from outside registry data (i.e., the source cohort) in the analysis of the target cohort (e.g., the MDA IBC cohort in the motivating example) has been a promising solution to the small sample size problem in risk estimation. NCDB data were obtained as our source cohort in the motivating example. Cancer registries have been used as auxiliary information to improve statistical inference in recent literature (Chatterjee et al., 2016; Antonelli et al., 2017; Chen et al., 2021). Specifically, Chatterjee et al. (2016) and Huang et al. (2016) developed likelihood-based frameworks to borrow information from external data sources in regression models. Most of these methods rely on an essential assumption, i.e., the two study populations are comparable (Chatterjee et al., 2016; Li et al., 2022). This assumption is often violated in practice. It is well noted that there is substantial referral bias in the MDA population compared to general patients from cancer registries, i.e., the MDA patients may have more complicated conditions or delayed diagnosis (Carlé et al., 2013). In our motivating example, the MDA-IBC cohort has a worse prognosis compared to the IBC patients of the NCDB cohort due to referral bias, suggesting substantial heterogeneity between the cohorts. An essential question here is how and to what extent we can learn from using cancer registry data in the risk estimation. Both Liu et al. (2014) and Huang et al. (2016) modified the Cox model by relaxing the cumulative hazard function in the target population and multiplying a constant factor. Still, their methods could not address the potential differences in the baseline hazards of the two populations. A recent work by Chen et al. (2021) developed an adaptive estimation procedure, which allows the source population to be incomparable with the target population. They used summary statistics to adjust the level of information borrowed from external sources. This is a promising approach, but it is also restricted in that only the summary level survival data, such as 10-year survival rate, can be borrowed from the external data.

Recently, a few transfer-learning-based statistical methods have been proposed to adaptively borrow information from the source population under different analysis frameworks. The advantage of these transfer-learning approaches is that, even when the external population is dramatically different from the target cohort, termed “negative transfer,” these methods can still provide reasonable estimations because the level of information borrowing was adaptively determined by the data similarities. The considered scenarios include transfer learning in Gaussian graphical models (Li et al., 2022), nonparametric classification (Cai and Wei, 2021), high-dimensional linear regression (Li et al., 2020), generalized linear models (GLM) (Tian and Feng, 2022), and federated learning with GLM (Li et al., 2021). However, none of these methods can be applied to the analysis of time-to-event outcomes, as considered in the motivating problem. Cox models are widely used for assessing risk with time-to-event data due to their easy interpretation and model flexibility (Cox, 1972). There are a number of additional statistical challenges in transferring the knowledge from the source to the target cohort in lifetime analysis under the Cox models. First, compared to the transfer learning in the regression setting in which models borrow information for the coefficients only, Cox models would be represented by different baseline hazard functions when the source and target cohorts have time-varying risk shifts. Additionally, the coefficients of baseline covariates and baseline hazards in the Cox models could have different but dependent levels of information sharing, and thus they should be controlled simultaneously. Second, the baseline cumulative hazards function is routinely estimated semi-parametrically in the framework of Cox models. The jump points of its estimation are decided by the event times in the observed data. As a result, the target and source populations can have distinct sets of breakpoints for the baseline hazard estimations, which makes information borrowing more challenging. Third and finally, individual-level data may not always be available for the source population due to privacy and logistical concerns (Platt and Kardia, 2015). The desired method should be able to allow both scenarios, i.e., using individual-level data or incorporating summary statistics obtained from the source population without the need to share individual-level data.

In this article, we address all the aforementioned challenges in transfer learning for time-to-event data under the Cox models. The proposed method overcomes the difficulty of sharing information with different sets of event times in the two cohorts. We allow different levels of information borrowing in the regression coefficients and baseline hazards through tuning parameters, and the resultant estimates are obtained in a unified framework simultaneously. As a result, the proposed method has great flexibility in that both covariate distributions and the associated coefficients, as well as baseline hazards are all allowed to be different between the two cohorts. As shown in our extensive simulation studies, the proposed method demonstrates satisfactory robustness, accuracy, and efficiency gained even when the source and target cohorts are heterogeneous in varying patterns and degrees. The applications to the MDA IBC cohort with NCDB data as a complement cohort also suggest improved precision of risk estimation compared with a regular Cox model with the MDA cohort alone. We present the notation, model, and algorithm in Section 2. The results from the simulation studies are present in Section 3 to evaluate the empirical performance. In Section 4, we present the data analysis results for the motivating example with IBC cohorts and interpret the results. We provide concluding remarks and discussions in Section 5.

2 Method

2.1 Notation and Model

Denote the outcome, the time from an initial event to an event of interest, by T. Let the covariates of interest be the p-dimensional vector $\boldsymbol{X}$. Given $\boldsymbol{X} = \boldsymbol{x}$, define the conditional density function and conditional survival function of T as $f(T \mid \boldsymbol{x})$ and $S(T \mid \boldsymbol{x})$. We denote the censoring time by C and the occurrence of the interested events (e.g., time to death) by $\delta = I(T \leq C)$. In the context of transfer learning, we have two sets of cohorts: a target cohort and a source cohort. The target cohort consists of N independent samples, $\{(Y_i, \delta_i, \boldsymbol{X}_i), i = 1, \cdots, N\}$, where $Y_i = \min\{T_i, C_i\}$, $\delta_i = I(T_i \leq C_i)$. Let the number of unique event time points in the target cohort be n_0.

In the target cohort, the Cox model assumes the covariate specific hazard function follows

$$h(t \big| \boldsymbol{X}_i) = h_0(t)\exp(\boldsymbol{X}_i^T\boldsymbol{\beta}), i = 1, \cdots, N .$$

Here, $\boldsymbol{\beta}$ is the p-dimensional vector of regression coefficients and $h_0(\cdot)$ is an unspecified baseline hazard function.

For the source cohort, we consider two scenarios when the individual-level data are or are not available. When the individual-level data are available, the source cohort is represented by N^S independent observations, $\left\{(Y_i^s, \delta_i^s, \boldsymbol{X}_i^s), i = 1, \cdots, N^S\right\}$ The number of unique event time points is n_0^s. The corresponding Cox model for the source population also has the proportional hazards assumption,

$$h^S(t \big| \boldsymbol{X}_i^s) = h_0^s(t)\exp\left\{(\boldsymbol{X}_i^s)^T\boldsymbol{\beta}^S\right\}, i = 1, \cdots, N^S,$$

where $h_0^s(\cdot)$ is an unspecified baseline hazard function. The cumulative baseline hazard function is denoted by $H_0^s(t) = \int_0^t h_0^s(u)du$. If the individual-level data cannot be shared due to privacy and logistical concerns, the proposed method still works given the estimators of coefficients $\boldsymbol{\beta}^S$ and baseline cumulative hazards $\widehat{H}_0^s(\cdot)$, which can be transferred by the source cohort site.

In addition to the proportional hazards assumption, we assume that the same set of covariates are available in both cohorts. However, the distributions of the covariates are allowed to be different in the two cohorts, which is termed as a "covariate shift" (Sugiyama et al., 2007; Jeong and Namkoong, 2020). Note that we do not need to assume the two cohorts are comparable, i.e., the regression coefficient $\boldsymbol{\beta}^S$ and baseline cumulative hazards $H_0^s(\cdot)$ in the source cohort can be different from $\boldsymbol{\beta}$ and $H_0(\cdot)$ in the target cohort, as well as the covariate distribution.

2.2 Transfer Learning Algorithm

With both the target and the source cohorts, a naïve approach to obtaining a more accurate risk assessment model is to combine data from the two cohorts and apply a Cox model. However, it is possible that the source cohort is different from the target cohort, which would lead to biased estimates (or risk estimation) of the target cohort. Even if the differences of the coefficients and baseline hazards in the two cohorts are small, the estimation bias by combining the two datasets can still be substantial when the sample size of the source cohort is much larger than that of the target cohort. To facilitate borrowing information from the source cohort, which may or may not be similar to the target cohort, our solution is to propose a transfer learning algorithm, called Trans-Cox, for improving the efficiency and accuracy of risk estimation for the target cohort.

When the individual-level data from the source cohort is available, we fit a Cox model using the source cohort only. Denote the ordered observed unique failure times in the target dataset by $\{\tilde{y}_1, \cdots, \tilde{y}_{n_0}\}$ and in the source dataset by $\{\tilde{y}_1^s, \cdots, \tilde{y}_{n_0^s}^s\}$. The log-likelihood for modeling the source cohort is

$$l(\boldsymbol{Y}^S, \boldsymbol{X}^S, \boldsymbol{\delta}^S; \boldsymbol{H}^S, \boldsymbol{\beta}^S) = \sum_{i=1}^{N^S} \delta_i^s[(\boldsymbol{X}_i^s)^T \boldsymbol{\beta}^S + \log\left\{h_i^s\right\}] - \sum_{i=1}^{N^S} H_0^s(Y_i^s) \exp\left\{(\boldsymbol{X}_i^s)^T \boldsymbol{\beta}^S\right\}, \tag{1}$$

where $h_i^s = \mathrm{d}H_0^s(Y_i^s)$ After inserting the Breslow estimator of the baseline hazard function (Breslow, 1972),

$$\widehat{h}_j^s(\boldsymbol{\beta}^S) = \left(\sum_{i \in R(\tilde{y}_j^s)} \exp\{(\boldsymbol{X}_i^s)^T \boldsymbol{\beta}^S\}\right)^{-1},$$

the baseline cumulative hazard function is estimated by $\widehat{H}_0^s(t) = \sum_{j=1}^{n_0^s} \hat{h}_j^s I(\tilde{y}_j^s \le t)$, where $R(t) = \{j : \tilde{y}_j^s \ge t\}$ is the risk set at t. We obtain the partial likelihood,

$$l(\boldsymbol{\beta}^S) = \sum_{i=1}^{N^S} \delta_i^s (X_i^s)^T \boldsymbol{\beta}^S - \sum_{i=1}^{N^S} \delta_i^s \log\left(\sum_{j \in \mathscr{R}(Y_i^s)} \exp\left\{(\boldsymbol{X}_j^s)^T \boldsymbol{\beta}^S\right\}\right). \tag{2}$$

The coefficients in the source cohort $\boldsymbol{\beta}^S$ can be estimated by maximizing (2), and the cumulative baseline hazards can be estimated from the Breslow estimator

$$\widehat{H}_0^s(t) = \sum_{i=1}^{N^S} \frac{I(Y_i^s \le t)\delta_i^s}{\sum_{j \in \mathscr{R}(Y_i^s)} \exp\left\{(\boldsymbol{X}_j^s)^T \boldsymbol{\beta}^s\right\}} . \tag{3}$$

In reality, sharing individual-level data can be challenging across multiple sites due to different data sharing policies for privacy, feasibility, and other concerns (Toh, 2020; Karr et al., 2007). Instead of sharing the individual-level data, a more practical approach is to share the summary-level statistics from external sites and draw conclusions using meta-analysis or distributed analysis (Lin and Zeng, 2010; Li et al., 2018). Our proposed method shares the same principle as the distributed analysis. When individual data are not available, our proposed method takes summary statistics in the form of estimated coefficients $\boldsymbol{\beta}^s$ and cumulative baseline hazards $\widehat{H}_0^s(\cdot)$ from the source cohort to achieve the same purpose. The whole analysis procedure (described below) demonstrates how the same outcomes are obtained with individual-level data or with summary statistics only.

Starting from the estimated coefficients $\boldsymbol{\beta}^s$ and the cumulative baseline hazard function $\widehat{H}_0^s(\cdot)$, we first obtain the “source” version of hazard estimations at the event times of the target cohort by creating a reference hazard estimation. We define the reference hazards $\Delta\widehat{H}_0^s(t)$ at $t = \tilde{y}_i$ by the difference of the baseline cumulative hazards function from the source dataset at the neighboring two consecutive time points of the target cohort, i.e.,

$$\Delta\widehat{H}_0^s(\tilde{y}_i) = \widehat{H}_0^s(\tilde{y}_i) - \widehat{H}_0^s(\tilde{y}_{i-1}),\ i = 1, \cdots, n_0 . \tag{4}$$

When inferring the risk estimation through the Cox model to the target cohort, we allow the two to be different by assuming that

$$\boldsymbol{\beta} = \boldsymbol{\beta}^S + \boldsymbol{\eta} \text{ and } dH_0(\tilde{y}_j) = \Delta\widehat{H}_0^s(\tilde{y}_j) + \xi_j,\ j = 1, \cdots, n_0 . \tag{5}$$

The two sets of parameters, $\boldsymbol{\eta}$ and $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_{n_0})$, are to quantify potential discrepancies in the covariate effects and the time-varying baseline hazard. When the two resources are comparable in terms of risk models, the two sets of parameters degenerate to zero, i.e., $\boldsymbol{\eta} = \boldsymbol{\xi} = 0$. Otherwise, there exists at least one non-zero component of the two vectors. Under this scenario, directly transferring the estimated Cox model from the source cohort to target cohort or combining the two cohorts for joint estimation would result in biased risk estimation. Our strategy is to estimate and identify a nonzero subset of these parameters adaptively for this scenario using information from the two resources.

Inspired by the penalized likelihood for variable selection in regression analysis, we add an L-1 penalty to the changing terms to control the sparsity and let the data drive the magnitude of source cohort information borrowing. With the formulation (5), the objective function to be minimized is:

$$
\begin{aligned}
O(\boldsymbol{\eta}, \boldsymbol{\xi}, \lambda_\eta, \lambda_\xi) = & - \sum_{i=1}^{N} \delta_i \left[\boldsymbol{x}_i^T \left(\boldsymbol{\beta}^S + \boldsymbol{\eta} \right) + \log \left\{ \Delta \widehat{H}_0^s(y_i) + \xi_i \right\} \right] \\
& + \sum_{i=1}^{N} \left[\sum_{j=1}^{n_0} \left\{ \Delta \widehat{H}_0^s(\tilde{y}_j) + \xi_j \right\} I\left(\tilde{y}_j \le y_i \right) \right] \exp\left\{ x_i^T \left(\boldsymbol{\beta}^S + \boldsymbol{\eta} \right) \right\} \\
& + \lambda_\eta \, \| \boldsymbol{\eta} \|_1 + \lambda_\xi \, \| \boldsymbol{\xi} \|_1 ,
\end{aligned}
\tag{6}
$$

where λ_η and λ_ξ are tuning parameters controlling the sparsity. The optimal values of λ_η and λ_ξ can be selected using the Bayesian Information Criterion (BIC) (Neath and Cavanaugh, 2012) with grid search. Our proposed Trans-Cox algorithm is formally presented in Algorithm 1.

Algorithm 1: Trans-Cox algorithm

```
Data: Individual-level data from target cohort $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\delta})$; Individual-level data $(\boldsymbol{X}^S, \boldsymbol{Y}^S, \boldsymbol{\delta}^S)$ or summary statistics $\{\boldsymbol{\beta}^S, \widehat{H}_0^s\}$ from the source cohort.
Result: $\boldsymbol{\beta}, \widehat{H}_0^s$
Step 1. Obtain $\boldsymbol{\beta}^S$ and $\widehat{H}_0^s$ from source cohort.
if $(\boldsymbol{X}^S, \boldsymbol{Y}^S, \boldsymbol{\delta}^S)$ available then
    Estimate coefficients $\boldsymbol{\beta}^S$ by maximizing the partial log-likelihood defined in (2);
    Estimate cumulative baseline hazards $\widehat{H}_0^s$ by (3).
else
    Take $\boldsymbol{\beta}^S$ and $\widehat{H}_0^s$ as inputs.
end
Step 2. Estimate reference hazards at the unique event time points of the target cohort by equation (4).
Step 3. Identify values for tuning parameters $\lambda_\eta$, $\lambda_\xi$, $T_{lr}$, and $T_{sp}$ using BIC.
Step 4. Estimate $(\boldsymbol{\eta}, \boldsymbol{\xi})$ by minimizing the objective function (6). The estimated coefficients and baseline hazards of the target cohort are
$\boldsymbol{\beta} = \boldsymbol{\beta}^S + \boldsymbol{\eta}$ and $d\widehat{H}_0(\boldsymbol{y}) = \Delta\widehat{H}_0^s(\boldsymbol{y}) + \boldsymbol{\xi}$.
```

The cumulative baseline hazards function is $\widehat{H}_0(t) = \sum_{i=1}^{n_0} d\hat{H}_0(\tilde{y}_i)\, I(\tilde{y}_i \leq t)$.

2.3 Remark on Optimization

Solving the objective function (6) is a challenging non-linear optimization problem. We implement the minimization using R 4.0.3 by invoking the TensorFlow (Dillon et al., 2017) solver from Python. This core optimization step is solved by the “tfp.math.minimize” function from TensorFlow probability (Dürr et al., 2020). After the required Python environment has been installed, users can use all the Trans-Cox functions we coded in R without additional coding in Python. To achieve optimal numerical performance, the TensorFlow functions need the inputs of two additional tuning parameters: learning rate T_{lr} and number of steps T_{sp}. For any given target data, we first select the combination of T_{lr} and T_{sp} with the smallest BIC values by fixing the other two tuning parameters λ_η and λ_ξ as 0.1. Then we fix T_{lr} and T_{sp} as the selected values, and we select the optimal set of λ_η and λ_ξ with BIC. Our method has been implemented in a user-friendly R package *Trans-Cox* with a detailed usage manual, and it is publicly available at https://github.com/ziyili20/TransCox.

3 Simulation

We conduct simulation studies to evaluate the finite sample performance of the proposed Trans-Cox algorithm. We also compare the performance of the Trans-Cox results with those of the standard Cox regression models using the target cohort only, the naïve combination of target and source cohorts, or the stratified Cox model.

3.1 Simulation set-ups

Without loss of generality, we assume that individual data of both cohorts are available. For both cohorts, we consider five covariates $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5)^T$. Of these, X_1, X_4 and X_5 are continuous covariates following a uniform distribution ranging from 0 to 1 and X_2 is a binary variable from a Bernoulli distribution with $p = 0.5$. Taking into account the potentially different distributions of covariates between the target and source cohorts, we assume that X_3 follows a standard uniform distribution in the target cohort, whereas it follows a *Beta*(1, 2) distribution in the source cohort. We generate the covariate-specific survival times from a Weibull distribution with a hazard function $h(t \mid \boldsymbol{X}) = \kappa t \cdot \exp(\boldsymbol{X}^T \boldsymbol{\beta})$ where $\boldsymbol{\beta} = (-0.5, 0.5, 0.2, 0.1, 0.1)^T$ and $k = 2$ for the target cohort. To mimic real-world scenarios in practice, we consider a total of four simulation settings for the source cohort. The source samples in the first setting are generated using the exact parameters as those of the target cohort. The source samples in the second setting are generated by changing $\boldsymbol{\beta}_2$ from 0.5 to 0.2, which enables us to evaluate the performance of Trans-Cox when the covariate X_2 has different effects on the survival time between two cohorts. In the third setting, the two cohorts share the same regression coefficients but have different time-varying hazard functions, which reflects a different baseline risk for the outcome of interest (e.g., overall worse or improved survival over time for the patients). Specifically, we set the baseline hazard function of the source cohort as $3t$ (i.e., $k = 3$), indicating there is a time-varying risk shift between two resources. The last setting allows both the regression coefficients and baseline hazard functions to be different, indicating less shared information between the two cohorts ($\beta_2 = 0.2$, $k = 3$).

It is worth noting that the cumulative baseline hazards in Settings 3 and 4 exhibit significant differences between the two cohorts, as shown in Figure S1, indicating that the amount of shared knowledge between the cohorts regarding baseline hazards is limited. For this simulation study, the sample size is fixed at $N = 250$ and $N_s = 6400$ to best mimic the observation amount in the real data application. In our evaluation, we compare the proposed Trans-Cox algorithm with two direct applications of Cox regressions with the target cohort only (i.e., Cox-Tonly) and with the combined cohort (i.e., Cox-Both), as well as stratified Cox with the combined cohort (i.e., Cox-Str). We further evaluate the bootstrap-based variance estimation using simulation study. The tuning parameters T_{lr}, T_{sp}, λ_η, and λ_ξ are selected using BIC (Burnham and Anderson, 2004).

3.2 Simulation results

Figure 1 summarizes the simulation results of the first three estimated regression coefficients (Panel A). To save space, the estimates for the fourth and fifth coefficients are not included as they have similar bias levels to the first one as the three covariates share the same uniform distribution in the two cohorts. We present the estimated biases using the four methods under four simulation settings. Note that the results from Cox-Tonly remain the same across the four settings as the differences across the settings only occur in the source cohort. For regression coefficients, the Trans-Cox results have smaller variations than Cox-Tonly while maintaining a similar level of biases. For example, the mean biases($\times 10^3$) for estimating $\boldsymbol{\beta}_1$ by Trans-Cox models are 14.4, 13.3, 3.3, and 54.2 in the four settings, respectively, while the mean bias($\times 10^3$) of the Cox model using the target cohort only is −27.1. The corresponding standard deviations ($SD \times 10^3$) of $\boldsymbol{\beta}_1$ by Trans-Cox (169.5, 170.3, 50.4, and 93.6, respectively) are about half the SD by Cox-Tonly for $\boldsymbol{\beta}_1$ (265.6). It is expected that Cox-Both tends to provide even smaller estimated variances and biases if the two cohorts have the same risk models. However, the biases by Cox-Both can be substantially larger when the two cohorts are heterogeneous in terms of regression coefficients or baseline hazards. In Settings 2 and 4, the estimation biases($\times 10^3$) of $\boldsymbol{\beta}_2$ by Cox-Both (−291.4 and −296.8) are four to six times the biases by Trans-Cox (−50.2 and −82.3). Stratified Cox presents similarly poor results as Cox-Both when the true coefficients are different in the two cohorts (Settings 2 and 4). The gray dotted lines in Figure 1(A) mark the place where the estimation biases are zero. It is clear that Cox-Both and Cox-Str have more biased boxes in Settings 2 and 4 (dark green and dark orange boxes), while the boxes representing Trans-Cox (red) and Cox-Tonly (blue) have similarly good performance (closer to the gray line) in all settings. It is also worth noting that different covariate coefficients in the two cohorts have a minimum impact on the estimations by Trans-Cox and Cox-Str, but affect the estimation by Cox-Both in Settings 3 and 4.

We report the biases of the cumulative baseline hazard estimators at two time points, $H_1 = H_0(t = 0.6)$ and $H_2 = H_0(t = 1.2)$, by the four methods in Figure 1(B). An interesting observation is that there is no major negative impact on the baseline hazards estimation when the regression coefficients differ between the two cohorts except for the stratified Cox model (Setting 2). Aside from this, we observe similar patterns for the estimator of cumulative baseline hazards as seen for regression coefficients. The biases from Trans-Cox and Cox-Tonly are comparable, while the SDs by Trans-Cox are about half of those by Cox-Tonly. The biases by Cox-Both are smaller and the SDs are about one fifth of those by Cox-Tonly. However, in Settings 3 and 4, when the risk models have different baseline hazards in the two cohorts, the biases of the cumulative baseline hazards at the two time points can be as high as ten times the biases by Trans-Cox or Cox-Tonly. In Figure 1(B), the results by Cox-Both (green boxes) show striking distances from the zero-bias line compared to the other two methods (red and blue boxes).

In Figure 2, we compare the personalized prediction accuracy using the absolute error between the individual predicted probability and the true survival probability (APPE1 and APPE2) across all subjects at time 0.6 and 1.2. Cox-Both exhibits the smallest absolute prediction error in the first setting, outperforming the other three methods. However, as expected, Cox-Both demonstrates substantially larger errors in Settings 3 and 4. Overall, Trans-Cox achieves the smallest errors in Settings 2, 3, and 4, and it is robust to cohort heterogeneity compared to the other methods. The corresponding numerical values for the mean biases and standard deviations are presented in Table S1 along with the restricted mean survival times (RMSTs) and mean squared errors (MSEs). RMST measures the average survival up to the specific time points ($t = 0.6$ and 1.2 in Table S1) for a fixed covariate combination ($\boldsymbol{X} = (0.5, 1.0, 1.0, 0.5, 0.5)$). MSE quantifies the mean squared estimation error of the baseline cumulative function at times 0.6 and 1.2. Trans-Cox generally has the smallest MSE for estimating the regression coefficients, as well as the smallest RMST and the smallest absolute personalized predictive error (APPE) compared to the other three methods when the two cohorts differ.

We perform additional simulation settings to further evaluate the impact of distribution shift, which is often observed in age-related data analysis. For this scenario, $\boldsymbol{X}_3$ follows normal distribution with mean 0 and variance 1 in the target cohort but mean 0.5 in the source cohort. Our findings, presented in Figure S2, demonstrate similar trends. Trans-Cox shows smaller estimation variance with comparable bias levels to the Cox regression with target data only. We provide the exact estimation biases with standard deviations for the regression coefficients $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$, cumulative baseline hazards at times 0.6 and 1.2, RMST, and MSE for the corresponding scenarios in Table S2.

Moreover, we visualize the estimated cumulative baseline hazard curves and the survival curves for our proposed method and the Cox regression using the combined cohort in Figure 3. The shaded areas represent the 95% empirical confidence intervals, which are obtained by taking the 0.025 and 0.975 percentiles from the repeated experiments. The survival curve is evaluated at the fixed covariate combination ($\boldsymbol{X} = (0.5, 1.0, 1.0, 0.5, 0.5)$). We do not include Cox-Tonly and Cox-Str on these figures. This is because the lines almost overlap with the Trans-Cox inference while having consistently wider confidence intervals. We show that the cumulative baseline hazards and the survival curves estimated by the Trans-Cox method are close to the true curves in all four settings. Cox-Both leads to substantial biases in Settings

3 and 4 where the true baseline cumulative hazards are different in the two cohorts. Similar observations can be found in the setting with a normal distribution shift in the two cohorts (Figure S3).

For the bootstrap-based variance estimation in Trans-Cox, Table S3 shows reasonable standard error estimations and coverage probability. As expected, the coverage probabilities generally are closer to the nominal value in the settings when the coefficients are the same in the two cohorts (Settings 1 and 3). Larger sample sizes do not offer a substantial improvement in the performance, indicating the need to tailor the standard bootstrapping method for more accurate inference in future research.

Lastly, we present the estimation sparsity of the parameters ($\boldsymbol{\eta}$, ξ) in Figure S5 and evaluate the computational cost of Trans-Cox in various settings in Figure S9. As expected, there are more zero or close-to-zero estimations in ξ than $\boldsymbol{\eta}$ due to the high dimension of the event time points in the target cohort. Our implementation offers superior computational performance, with the functions implemented in R and a core solver invoked from Python. A run with 200 patients in the target cohort on average takes around 0.6 seconds, and with 400 patients it takes about 1 second (Figure S9). The size of the source population has a minimum impact on the total computational time. The fast speed makes it feasible to construct a bootstrap-based variance estimation. For example, it takes about 21 minutes to complete 1000 bootstrap iterations in the IBC application (target cohort: 251 patients; source cohort: 6420 patients) of Section 4.

4 Data Application

The Morgan Welch Inflammatory Breast Cancer Clinic at MD Anderson Cancer Center is one of the largest breast cancer centers in the United States to treat IBC patients. As a rare but aggressive form of breast cancer, IBC accounts for less than 5% of breast cancer diagnoses with a five year survival rate of only around 40% (Van Uden et al., 2015). Compared to the general population of IBC patients, the IBC patients treated at MD Anderson usually have more complicated medical conditions or more severe symptoms. This is because many of them were referred to MD Anderson from local hospitals, noted as referral bias (Carlé et al., 2013). Recent studies have shown the survival advantage of the recommended therapy, trimodality treatment, for IBC patient populations (Rueth et al., 2014; Liauw et al., 2004).

To understand how trimodality, IBC stage, age, and other disease-associated factors impact patients' survival when they are cared for at MD Anderson, we analyze a cohort (MDA cohort) consisting of MD Anderson-treated IBC patients who were diagnosed with nonmetastatic IBC between 1992 and 2012. After removing six patients with missing tumor grade information, the analysis cohort includes 251 patients with a median follow up time of 5.19 years and a censoring rate of 57%. Because of the rarity of IBC, the MDA cohort can provide the estimated individualized survival risk with limited precision. This motivates us to transfer the information of risk assessment from other large population-level databases.

The National Cancer Data Base (NCDB) was collected collaboratively by the American College of Surgeons, the American Cancer Society, and the Commission on Cancer(Raval et al., 2009). This database serves as a comprehensive cancer care resource and has recorded patient demographic, tumor, treatment, and outcome variables from approximately 70% of all new cancer diagnoses in the US annually. The NCDB cohort consists of 9493 patients who underwent surgical treatment of nonmetastatic IBC from 1998 to 2010. We subsequently removed patients with missing grade levels, missing race information, or missing treatment information. The resulted non-metatasis IBC patients in NCDB cohort consists of 6420 patients with a median follow-up time of 7.05 years and a censoring rate 57%.

Figure 4 presents the patients' characteristics from the MDA and NCDB cohorts. We consider five prognosis factors in the Cox model: age, tumor grade (I/II or III), race (white, black, or other), clinical stage (IIIB or IIIC), and trimodality treatment (yes or no). A comparison of the distributions between the two cohorts (Figures 4A and B) demonstrates many differences in characteristics. MD Anderson patients tend to be younger with more advanced tumor grade and clinical stages, and fewer received trimodality. Additionally, a naïve comparison shows that the survival outcomes of the MDA patients are significantly worse than the NCDB cohort (Figure 4C).

Our goal is to borrow IBC patient data from NCDB (source cohort) to improve the individualized risk estimation of MD Anderson patients (target cohort). The observations from Figure 4 suggest a possible referral bias for data from large cancer cancers, such as MD Anderson, in which patients tend to be sicker than the general IBC patients represented by NCDB. Meanwhile, due to the substantial differences in mortality risk between the two cohorts, it is not appropriate to directly merge the two cohorts for the inference of the MDA cohort.

We apply both the proposed Trans-Cox algorithm and conventional Cox regression on the MDA and NCDB cohorts. Table 1 shows the estimated coefficients (log hazard ratios) from the five methods: Trans-Cox with MDA as the target cohort and NCDB as the source cohort, Cox model with MDA only, Cox model with NCDB only, Cox model with the combined data of MDA and NCDB, and Cox model stratified by the data sources. Age is standardized in both datasets (Figure S6). There are several interesting observations. First, the Cox model with the simple combined data and stratified approach has almost identical inference results as the Cox model with NCDB data only. This illustrates that when the sample size of the source cohort is much larger than the target cohort (6420 vs 251), the estimation results of the combined cohort are dominated by the source cohort. This observation is also confirmed by the estimation of cumulative baseline hazards presented in Figure S7 in which the estimated cumulative baseline hazards using the combined cohort are almost identical to those using NCDB only. Figure S8 shows the sparsity of the estimated η and ξ by TransCox. We observe that one estimated η (for "Race:Other vs White") and several ξ estimators are close to zero, indicating that the coefficient information of other Race group and the cumulative baseline hazards at those time points are mostly borrowed from the source cohort by Trans-Cox.

Second, some covariates have similar effect sizes, while others have quite different effect sizes on survival between the two cohorts. The Cox model based on the MDA cohort has similar effect sizes for age and grade, but has a much smaller effect for the race being Black compared to the effect size estimated using the NCDB cohort. Rueth et al. (2014) reported that the Black population tended to be treated not according to the IBC treatment guidelines compared to the White IBC patients, thus increasing the mortality risk using NCDB. However, breast cancer patients treated at MD Anderson generally received their treatment according to the guidelines regardless of their racial background (Shen et al., 2007). As a result, racial status plays a less influential role in the MDA cohort than in the NCDB cohort.

Third, we observe that Trans-Cox can effectively transfer knowledge on the regression coefficients from the NCDB cohort to the MDA cohort for similar effects and simultaneously reduce information borrowing for inconsistent effects. In the upper panel of Table 1, the estimated standard errors of grade and trimodality are substantially reduced, resulting in more precise estimators compared to the Cox model using the MDA cohort alone. Different from other subtypes of breast cancer, all IBC started as Stage III since they involve the skin. The MDA analytic cohort only focuses only on Stage III IBC patients with a fine category of B versus C. Then the stage (Stage IIIB vs Stage IIIC) has a non-significant effect on overall survival by using the MDA cohort only or borrowing information from NCDB with Trans-Cox. The standard errors of the parameters and the 95% confidence intervals of the Trans-Cox method are obtained using bootstrap.

We also estimate the survival curves for patients with or without receiving trimodality and present the results in Figure 5. The left panel shows the estimated survival curve for Black patients with age 50, grade III, stage IIIC, and receiving trimodality. We find that patients receiving trimodality treatment at MD Anderson have better survival outcomes compared to IBC patients with the same baseline characteristics in NCDB even after Trans-Cox borrows information from NCDB. This suggests that MD Anderson may provide better care than the average medical institution in the United States. The right panel contains the survival curves for the patient with the same characteristics (Black, age 50, grade III, stage IIIC) but without receiving trimodality treatment. It is interesting that patients did not have any survival benefit even if they were treated at MD Anderson Cancer Center without using trimodality.

Lastly, we evaluate the prediction performance of fitted models using the concordance index (C-index) (Steck et al., 2007), Uno's C-index (Uno et al., 2011), area under the receiver operating characteristic curve (AUC), and error of estimated personalized risk prediction (Uno et al., 2007). The error of estimated personalized risk prediction at time L is defined as $\sum_{i=1}^{N} \widehat{w}_i \left\{ \widehat{p}_i(L) - 1(Y_i \geq L) \right\} / \sum_{i=1}^{N} \widehat{w}_i$ where $\widehat{w}_i = \frac{1(Y_i \leq L)\delta_i}{\widehat{G}(Y_i)} + \frac{1(Y_i \geq L)}{\widehat{G}(L)}$ and $\widehat{G}$ is the Kaplan-Meier estimation function of the censored time $\{C_i\}$. We evaluate the personalized risk prediction at time points $L = \{20, 30, \cdots, 110\}$. We use the bootstrap-based method to evaluate the survival risk prediction for the MDA data and compare with the true observations (Steyerberg et al., 2001). Figure 6 shows that the Trans-Cox model outperforms the Cox model using the target data only ("Cox-Tonly" in Figure 6) in terms of estimation precision, while both methods exhibit similar predictive performance, as measured by Uno's C index. The Cox model using the combined cohort ("Cox-Both"), the stratified Cox model ("Cox-Stratified"), and the Cox model using the source cohort, NCDB, alone ("Cox-Sonly") have similarly poor performance. Overall, the Trans-Cox and Cox-Tonly models have the best prediction accuracy. In comparison, the other three models (Cox-Sonly, Cox-Both, Cox-Stratified) have lower concordances and much higher prediction errors.

Discussion

Motivated by the challenges in borrowing information from a source cohort to improve the time-varying risk assessment for a target cohort, this article develops a transfer learning-based method that adaptively determines the degree of information transferring from the source cohort. Previous works that combine multiple data sources often need to consider the covariate distribution shift between the study cohorts, e.g., when estimating marginal treatment effects or generalizing research findings to a larger population (Colnet et al., 2020; Wu and Yang, 2021). Our problem focuses on individual risk assessment and does not need to address such covariate distribution shifts. The proposed method can naturally address the heterogeneity between the two cohorts by imposing L-1 penalties. Much of the literature has a developed adaptive Lasso for the Cox model to allow feature selection (Tibshirani, 1997; Zhang and Lu, 2007). Different from these methods, our Trans-Cox model imposes the Lasso penalty on the discrepancy in regression coefficients and baseline hazard functions between the two resources. Our model considers the penalties from both sides in a unified framework to allow simultaneous control of the information sharing for the covariate effects and time-varying baseline hazards under the Cox model.

Bayesian methods may appear to be a natural modeling strategy to borrowing information from the source cohort to improve the estimation in the target cohort. However, there are several challenges to solve the current problem using a standard Bayesian approach. First, when the ratio of sample sizes of the two cohorts is large, the posterior distribution would be dominated by the source cohort if directly fitting the two cohorts with Bayesian methods, although the risk estimation for the target cohort is the purpose. To solve this, a few tuning parameters may need to be involved and carefully selected to balance the sample sizes and heterogeneity between the two cohorts. Second, besides the non-parametric component under the Cox model, the dimension of η and ξ increases with the sample size as well, which also increases the computational cost for a Bayesian approach.

A fundamental assumption of many previous works that borrow information from the source population is the comparability between the two cohorts (Chatterjee et al., 2016; Li et al., 2022). Our method relaxes this assumption by accommodating situations when the two cohorts can be heterogeneous. Such heterogeneity commonly exists in practice due to various types of selection biases, demographic differences, regulatory restrictions, etc. As shown in our results, when the risk models of the two cohorts are different, directly combining the two datasets can result in biased findings. This estimation bias can be substantial when the sample size of the source cohort is much larger than the target cohort.

In contrast, Trans-Cox demonstrates robust estimations even when there are high levels of cohort heterogeneity. This highlights the significance of our research question and the proposed methodology.

Moreover, many recent studies have discussed the challenges of sharing individual-level data in multi-site studies due to feasibility, privacy, and other concerns (Maro et al., 2009; Li et al., 2019; Toh et al., 2011). The existing transfer learning and information borrowing methods usually adapt their proposals to incorporate the summary statistics from the source cohort in the analysis, replacing the need of individual-level data for the source population (Chatterjee et al., 2016; Li et al., 2022). But none of the work is directly applicable to time-to-event data or allows time varying differences in baseline hazards of the two cohorts. We also recognize the importance of this issue and allow Trans-Cox to directly incorporate summary statistics of estimated coefficients and cumulative baseline hazards as the information from the source cohort. This advantage facilitates the use of information from additional sources with the guarantee of protecting patient privacy.

It is worth noting that the primary purpose of using the L-1 penalty in our method is not to control sparsity, but to regulate the amount of information borrowed from the source cohort to the target cohort. The L-1 penalty has been successfully used in the literature to control information discrepancy (Ding et al., 2023; Chen et al., 2021). We employ the L-1 penalty to control the distance in the cumulative baseline hazard component (ξ) due to the high dimension of the failure time points. We also apply it to the covariate coefficient component ($\boldsymbol{\eta}$) for consistency, although other types of penalty could also be used for such a purpose. When high-dimensional covariates are present, feature selection must be incorporated into our proposed method. One simple solution is to add an additional L-1 penalty component on $\boldsymbol{\beta}$, but the implementation details are beyond the scope of this work.

The proposed method can be extended in several ways. First, the current framework only considers point estimators from the source cohort. When the source cohort has a large sample size, for example in the NCDB data, the estimation variations are negligible. However, the estimation uncertainties can be non-negligible when a smaller or more diverse source cohort is considered. Incorporating such uncertainties in the method may better inform the level of knowledge sharing between the cohorts and further improve accuracy.

Second, we focus on the situation where one source cohort is available. Although the current method can be directly extended to multiple source situations by combining all the source data as a single source cohort, this naive extension may over-simplify the complexity of this problem. For example, when there are different directions of the covariate effects and large variations in sample sizes among multiple sources, it is unclear how to achieve the balance between the estimation accuracy and the model flexibility. Li et al. (2020) discussed approaches to identify informative auxiliary cohorts and aggregate these cohorts to improve transfer learning in the high-dimensional linear regression setting. Similar approaches can be considered here to allow for incorporating the information from multiple source populations.

Third, similar to several existing methods (Dahabreh et al., 2020), we assume the interested covariates are available for both cohorts. This could be a limitation in practice as medical institutions or clinical trials sometimes capture different covariates from the population-level data registries (Taylor et al., 2022). Further extensions can be made to allow for incorporating a subset of regression coefficients to improve individualized risk assessment. Finally, it is generally believed that a standard bootstrap may not work well in the regression problem using penalization (Chatterjee and Lahiri, 2011). Although our simulation study demonstrates a reasonable performance for the bootstrap-based inference, the evaluation is restricted by the simulation settings. We acknowledge that the theoretical justification for the inference procedure is beyond the scope of this work and worthy of future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank Dr. Naoto Ueno and his team for sharing IBC patient cohort data, and Dr. Isabelle Bedrosain and the team for collaboration on the NCDB project, which motivated this research. The authors also gratefully acknowledge the constructive comments from the editor, associate editor, and two anonymous reviewers in improving the manuscript.

Funding

This work is supported by the National Institutes of Health (R01CA269696 and R03CA270725).

References

Antonelli J, Zigler C, and Dominici F (2017). Guided bayesian imputation to adjust for confounding when combining heterogeneous data sources in comparative effectiveness research. Biostatistics 18 (3), 553–568. [PubMed: 28334230]

Bilimoria KY, Stewart AK, Winchester DP, and Ko CY (2008). The National Cancer Data Base: a powerful initiative to improve cancer care in the united states. Annals of surgical oncology 15 (3), 683–690. [PubMed: 18183467]

Breslow NE (1972). Contribution to discussion of paper by DR Cox. J. Roy. Statist. Soc., Ser. B 34, 216–217.

Burnham KP and Anderson DR (2004). Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research 33 (2), 261–304.

Cai TT and Wei H (2021). Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. The Annals of Statistics 49 (1), 100–128.

Carlé A, Pedersen IB, Perrild H, Ovesen L, Jørgensen T, and Laurberg P (2013). High age predicts low referral of hyperthyroid patients to specialized hospital departments: evidence for referral bias. Thyroid 23 (12), 1518–1524. [PubMed: 23745710]

Carvalho AL, Nishimoto IN, Califano JA, and Kowalski LP (2005). Trends in incidence and prognosis for head and neck cancer in the united states: a site-specific analysis of the seer database. International journal of cancer 114 (5), 806–816. [PubMed: 15609302]

Chatterjee N, Chen Y-H, Maas P, and Carroll RJ (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. Journal of the American Statistical Association 111 (513), 107–117. [PubMed: 27570323]

Chen Z, Ning J, Shen Y, and Qin J (2021). Combining primary cohort data with external aggregate information without assuming comparability. Biometrics 77 (3), 1024–1036. [PubMed: 32827153]

Colnet B, Mayer I, Chen G, Dieng A, Li R, Varoquaux G, Vert J-P, Josse J, and Yang S (2020). Causal inference methods for combining randomized trials and observational studies: a review. arXiv preprint arXiv:2011.08047

Cox DR (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) 34 (2), 187–202.

Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, and Hernan MA (2020). Extending inferences from a randomized trial to a new target population. Statistics in medicine 39 (14), 1999–2014. [PubMed: 32253789]

Dillon JV, Langmore I, Tran D, Brevdo E, Vasudevan S, Moore D, Patton B, Alemi A, Hoffman M, and Saurous RA (2017). Tensorflow distributions. arXiv preprint arXiv:1711.10604

Ding J, Li J, Han Y, McKeague IW, and Wang X (2023). Fitting additive risk models using auxiliary information. Statistics in Medicine

Dürr O, Sick B, and Murina E (2020). Probabilistic deep learning: With python, keras and tensorflow probability Manning Publications.

Huang C-Y, Qin J, and Tsai H-T (2016). Efficient estimation of the Cox model with auxiliary subgroup survival information. Journal of the American Statistical Association 111 (514), 787–799. [PubMed: 27990035]

Jaiyesimi IA, Buzdar AU, and Hortobagyi G (1992). Inflammatory breast cancer: a review. Journal of clinical oncology 10 (6), 1014–1024. [PubMed: 1588366]

Jeong S and Namkoong H (2020). Robust causal inference under covariate shift via worst-case subpopulation treatment effects. In Conference on Learning Theory, pp. 2079–2084. PMLR.

Jiao FF, Fung CSC, Wan EYF, Chan AKC, McGhee SM, Kwok RLP, and Lam CLK (2018). Five-year cost-effectiveness of the multidisciplinary risk assessment and management programme–diabetes mellitus (RAMP-DM). Diabetes Care 41 (2), 250–257. [PubMed: 29246949]

Karr AF, Fulp WJ, Vera F, Young SS, Lin X, and Reiter JP (2007). Secure, privacy-preserving analysis of distributed databases. Technometrics 49 (3), 335–345.

Kumar A, Guss ZD, Courtney PT, Nalawade V, Sheridan P, Sarkar RR, Banegas MP, Rose BS, Xu R, and Murphy JD (2020). Evaluation of the use of cancer registry data for comparative effectiveness research. JAMA Network Open 3 (7), e2011985–e2011985. [PubMed: 32729921]

Li D, Lu W, Shu D, Toh S, and Wang R (2022). Distributed Cox proportional hazards regression using summary-level information. Biostatistics

Li J, Panucci G, Moeny D, Liu W, Maro JC, Toh S, and Huang T-Y (2018). Association of risk for venous thromboembolism with use of low-dose extended-and continuous-cycle combined oral contraceptives: a safety study using the sentinel distributed database. JAMA internal medicine 178 (11), 1482–1488. [PubMed: 30285041]

Li S, Cai T, and Duan R (2021). Targeting underrepresented populations in precision medicine: A federated transfer learning approach. arXiv preprint arXiv:2108.12112

Li S, Cai TT, and Li H (2020). Transfer learning for high-dimensional linear regression: Prediction, estimation, and minimax optimality. arXiv preprint arXiv:2006.10593

Li S, Cai TT, and Li H (2022). Transfer learning in large-scale gaussian graphical models with false discovery rate control. Journal of the American Statistical Association, 1–13. [PubMed: 35757777]

Li Z, Roberts K, Jiang X, and Long Q (2019). Distributed learning from multiple EHR databases: contextual embedding models for medical events. Journal of biomedical informatics 92, 103138. [PubMed: 30825539]

Liauw SL, Benda RK, Morris CG, and Mendenhall NP (2004). Inflammatory breast carcinoma: outcomes with trimodality therapy for nonmetastatic disease. Cancer 100 (5), 920–928. [PubMed: 14983486]

Lin D-Y and Zeng D (2010). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. Biometrika 97 (2), 321–332. [PubMed: 23049122]

Liu D, Zheng Y, Prentice RL, and Hsu L (2014). Estimating risk with time-to-event data: An application to the women's health initiative. Journal of the American Statistical Association 109 (506), 514–524. [PubMed: 25018574]

Maro JC, Platt R, Holmes JH, Strom BL, Hennessy S, Lazarus R, and Brown JS (2009). Design of a national distributed health data network. Annals of internal medicine 151 (5), 341–344. [PubMed: 19638403]

Masuda H, Brewer T, Liu D, Iwamoto T, Shen Y, Hsu L, Willey J, Gonzalez-Angulo A, Chavez-MacGregor M, Fouad T, et al. (2014). Long-term treatment efficacy in primary inflammatory breast cancer by hormonal receptor- and HER2-defined subtypes. Annals of oncology 25 (2), 384–391. [PubMed: 24351399]

Neath AA and Cavanaugh JE (2012). The bayesian information criterion: background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics 4 (2), 199–203.

Platt J and Kardia S (2015). Public trust in health information sharing: implications for biobanking and electronic health record systems. Journal of personalized medicine 5 (1), 3–21. [PubMed: 25654300]

Raval MV, Bilimoria KY, Stewart AK, Bentrem DJ, and Ko CY (2009). Using the NCDB for cancer care improvement: an introduction to available quality assessment tools. Journal of surgical oncology 99 (8), 488–490. [PubMed: 19466738]

Rueth NM, Lin HY, Bedrosian I, Shaitelman SF, Ueno NT, Shen Y, and Babiera G (2014). Underuse of trimodality treatment affects survival for patients with inflammatory breast cancer: an analysis of treatment and survival trends from the national cancer database. Journal of Clinical Oncology 32 (19), 2018. [PubMed: 24888808]

Shen Y, Dong W, Esteva FJ, Kau S-W, Theriault RL, and Bevers TB (2007). Are there racial differences in breast cancer treatments and clinical outcomes for women treated at MD Anderson Cancer Center? Breast cancer research and treatment 102 (3), 347–356. [PubMed: 17028980]

Steck H, Krishnapuram B, Dehing-Oberije C, Lambin P, and Raykar VC (2007). On ranking in survival analysis: Bounds on the concordance index. Advances in neural information processing systems 20.

Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans M, Vergouwe Y, and Habbema JDF (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. Journal of clinical epidemiology 54 (8), 774–781. [PubMed: 11470385]

Sugiyama M, Krauledat M, and Müller K-R (2007). Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research 8 (5).

Taylor JM, Choi K, and Han P (2022). Data integration: Exploiting ratios of parameter estimates from a reduced external model. Biometrika

Tian Y and Feng Y (2022). Transfer learning under high-dimensional generalized linear models. Journal of the American Statistical Association (just-accepted), 1–30. [PubMed: 35757777]

Tibshirani R (1997). The lasso method for variable selection in the Cox model. Statistics in medicine 16 (4), 385–395. [PubMed: 9044528]

Toh S (2020). Analytic and data sharing options in real-world multidatabase studies of comparative effectiveness and safety of medical products. Clinical Pharmacology & Therapeutics 107 (4), 834–842. [PubMed: 31869442]

Toh S, Platt R, Steiner J, and Brown J (2011). Comparative-effectiveness research in distributed health data networks. Clinical Pharmacology & Therapeutics 90 (6), 883–887. [PubMed: 22030567]

Uno H, Cai T, Pencina MJ, D'Agostino RB, and Wei L-J (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in medicine 30 (10), 1105–1117. [PubMed: 21484848]

Uno H, Cai T, Tian L, and Wei L-J (2007). Evaluating prediction rules for t-year survivors with censored regression models. Journal of the American Statistical Association 102 (478), 527–537.

Van Uden D, Van Laarhoven H, Westenberg A, de Wilt J, and Blanken-Peeters C (2015). Inflammatory breast cancer: an overview. Critical reviews in oncology/hematology 93 (2), 116–126. [PubMed: 25459672]

Wu L and Yang S (2021). Transfer learning of individualized treatment rules from experimental to real-world data. arXiv preprint arXiv:2108.08415

Zhang HH and Lu W (2007). Adaptive lasso for Cox's proportional hazards model. Biometrika 94 (3), 691–703.

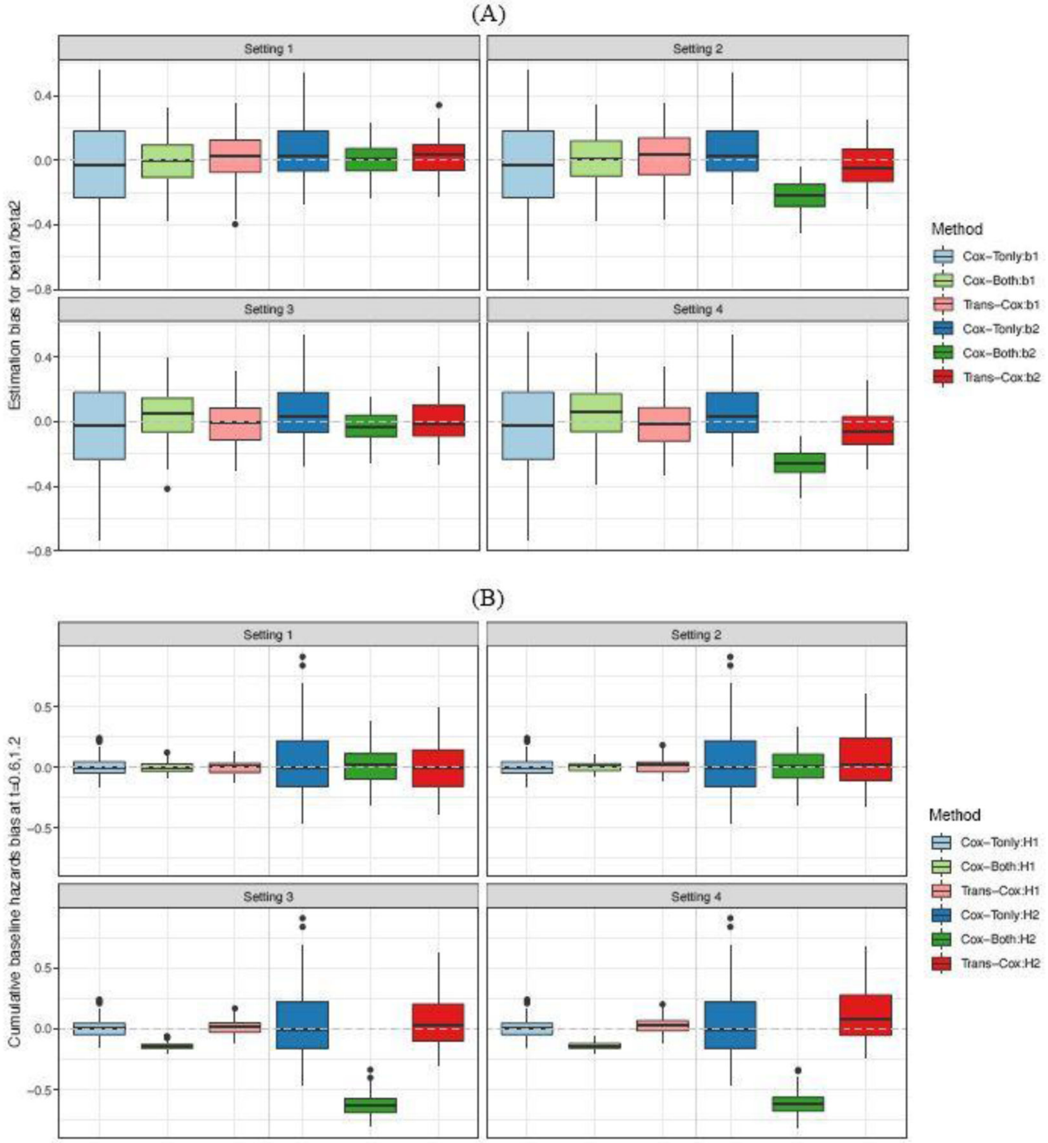


Fig. 1.

Estimation biases for coefficients $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$ (Panel A), as well as cumulative baseline hazards at times 0.6 and 1.2 (Panel B) over 100 Monte Carlo simulations. $H_1 = \widehat{H}_0(0.6)$ and $H_2 = \widehat{H}_0(1.2)$. The dotted gray line shows the place where bias equals zero.

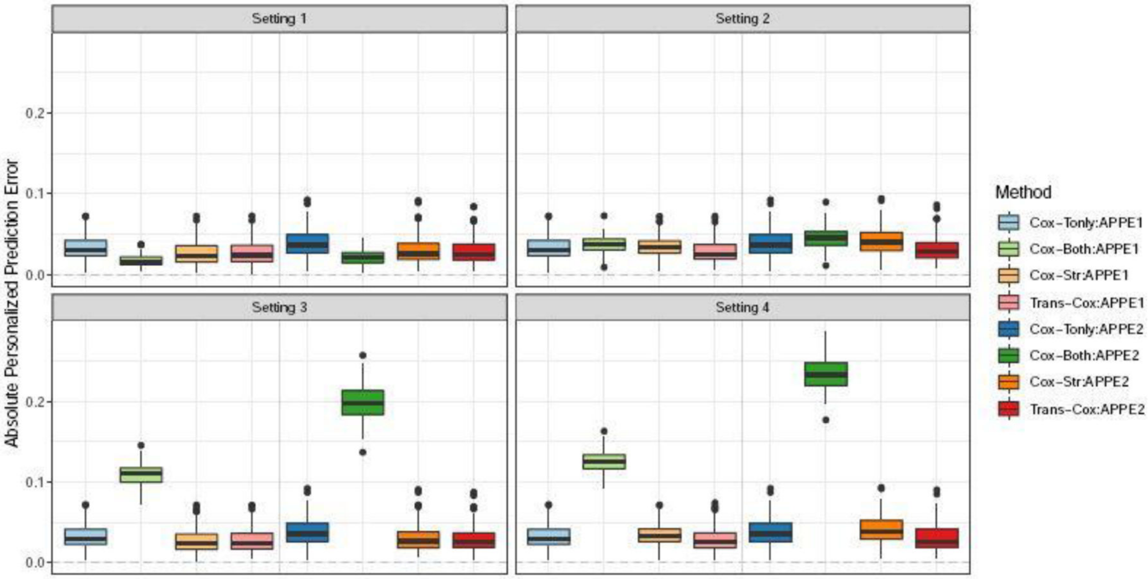


Fig. 2.

Boxplots for absolute personalized prediction error of Trans-Cox and other existing methods in different simulation settings at time 0.6 (APPE1) and 1.2 (APPE2). The results are summarized over 100 Monte Carlo datasets.

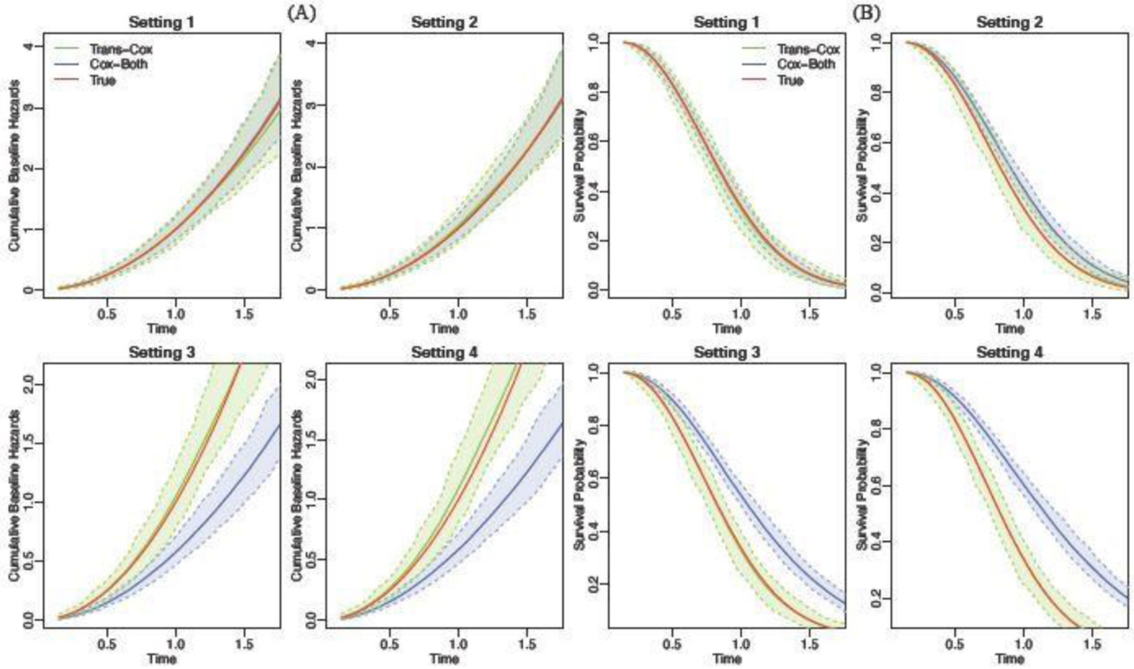
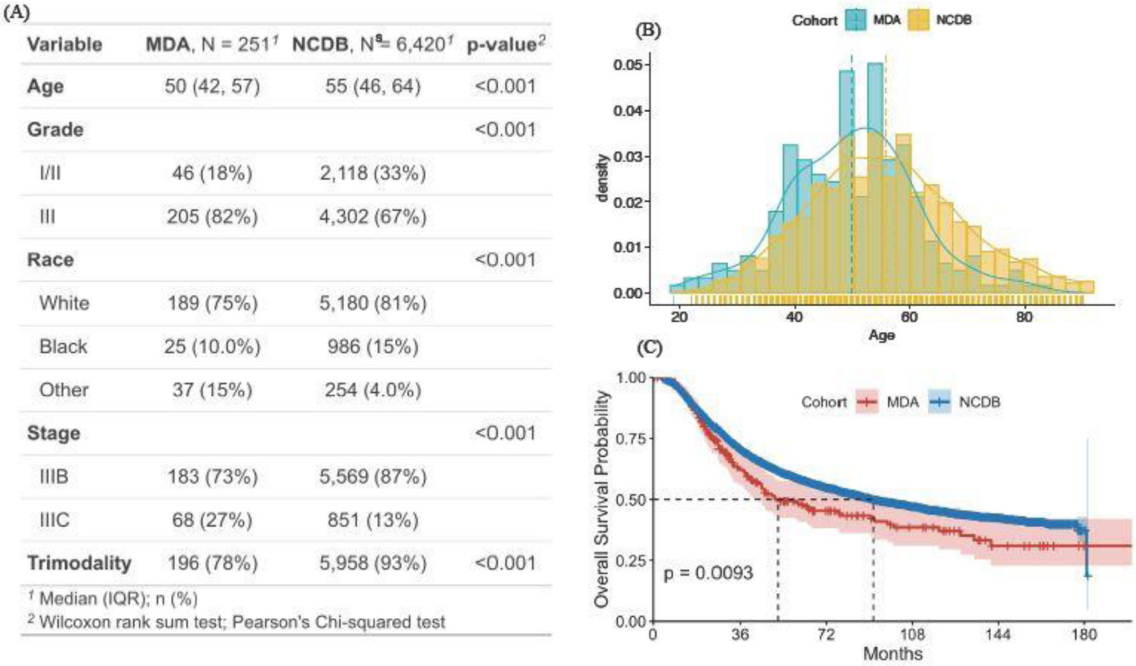


Fig. 3.

The estimated cumulative baseline hazard (Panel A) and the survival curves (Panel B) for Trans-Cox (green) and Cox-Both (blue) in comparison to the true curves (red) over 100 Monte Carlo experiments. For survival curves, the covariates are fixed at $\boldsymbol{X} = (0.5, 1, 0.1, 0.5, 0.5)$.



(A)

Variable	**MDA**, N = 251[1]	**NCDB**, N = 6,420[1]	**p-value**[2]
Age	50 (42, 57)	55 (46, 64)	<0.001
Grade			<0.001
I/II	46 (18%)	2,118 (33%)	
III	205 (82%)	4,302 (67%)	
Race			<0.001
White	189 (75%)	5,180 (81%)	
Black	25 (10.0%)	986 (15%)	
Other	37 (15%)	254 (4.0%)	
Stage			<0.001
IIIB	183 (73%)	5,569 (87%)	
IIIC	68 (27%)	851 (13%)	
Trimodality	196 (78%)	5,958 (93%)	<0.001

[1] Median (IQR); n (%)
[2] Wilcoxon rank sum test; Pearson's Chi-squared test

Fig. 4.

Summary of the patients' characteristics of the MDA and NCDB cohorts.

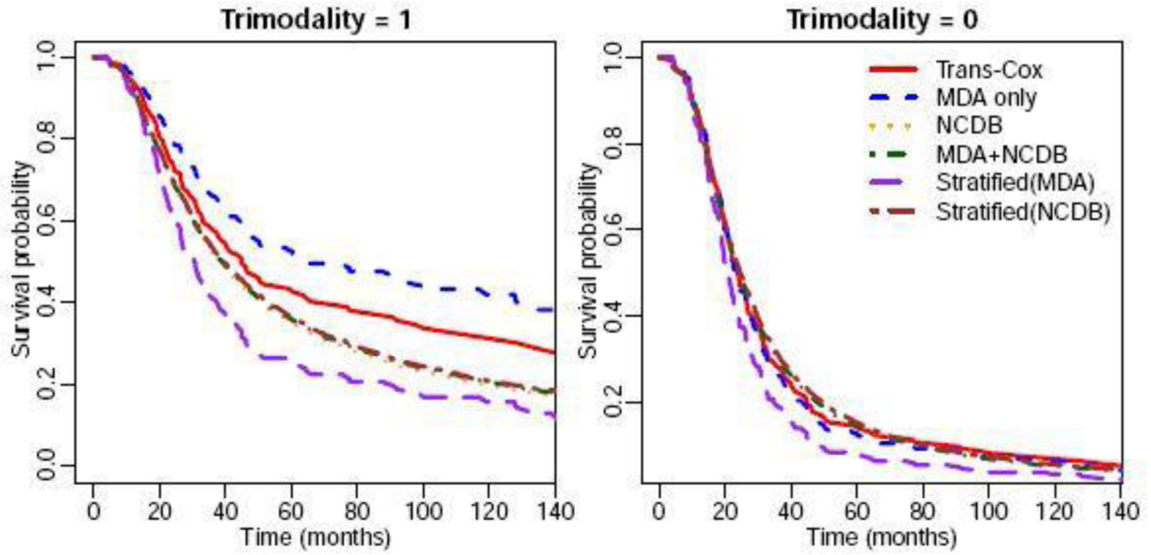


Fig. 5.

The estimated survival curves using different methods for patients with (left panel) or without (right panel) receiving trimodality treatment. Other characteristics were specified as: age 50 years old, grade III, Black race, and stage IIIC.

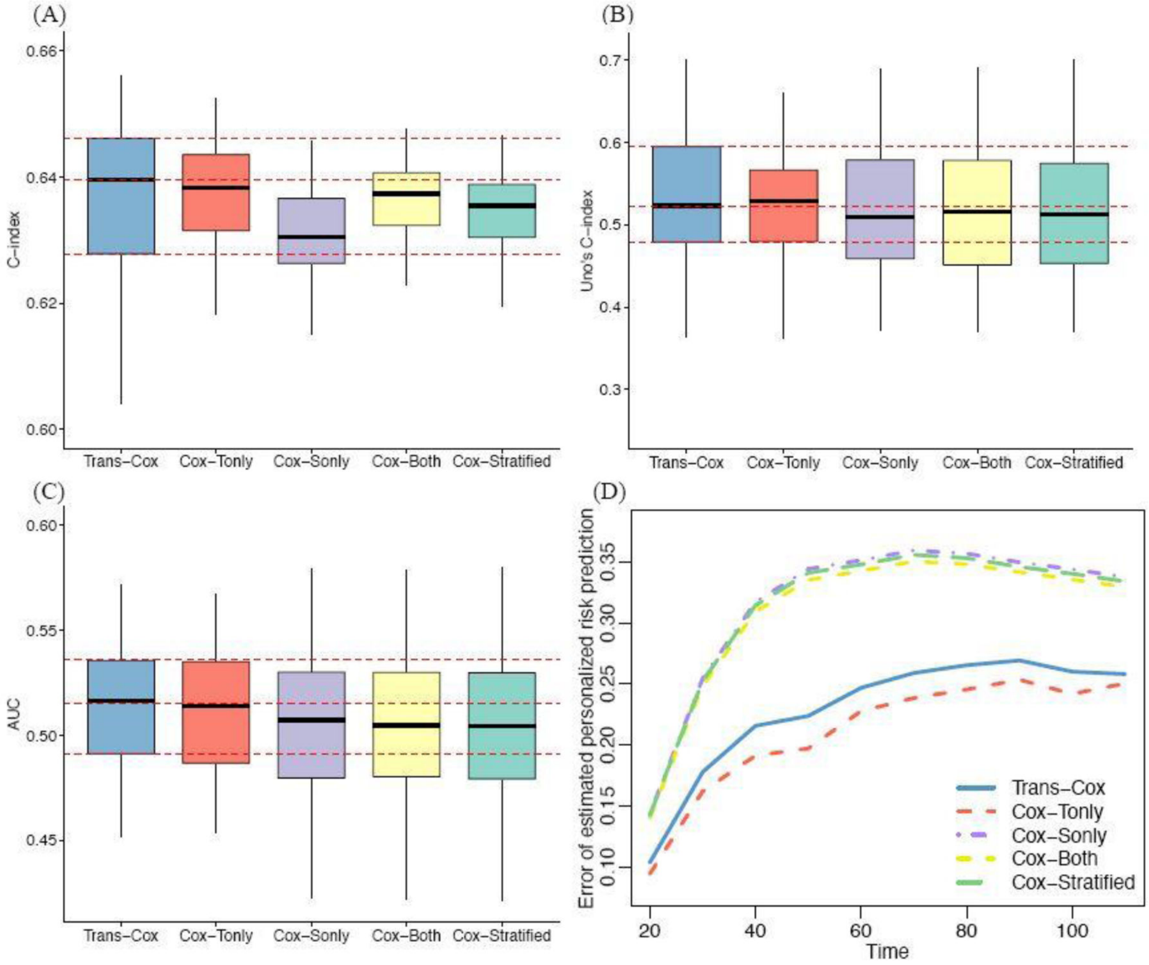


Fig. 6.

Evaluation of the prediction performance of four methods using C-index (Panel A), Uno's C-index (B), AUC (C), and personalized risk prediction error (D).

Table 1

Analysis results of the MDA and NCDB cohorts.

$\hat{\beta}$					
	Trans-Cox	**Cox(MDA)**	**Cox(NCDB)**	**Cox(Combined)** [†]	**Cox(Stratified)**
Normalized Age	0.091	0.074	0.208	0.197	0.199
Grade:III vs I/II	0.616	0.388	0.347	0.353	0.348
Race:Black vs White	0.272	0.163	0.429	0.412	0.416
Race:Other vs White	−0.128	−0.336	−0.113	−0.107	−0.130
Stage:IIIC vs IIIB	−0.059	−0.243	0.386	0.368	0.357
Trimodality:Yes vs. No	−0.838	−1.174	−0.589	−0.636	−0.627
SE [*]					
	Trans-Cox	Cox(MDA)	Cox(NCDB)	Cox(Combined)	Cox(Stratified)
Normalized Age	0.121	0.109	0.021	0.021	0.021
Grade:III vs I/II	0.197	0.269	0.044	0.043	0.044
Race:Black vs White	0.265	0.325	0.051	0.051	0.051
Race:Other vs White	0.261	0.295	0.118	0.109	0.109
Stage:IIIC vs IIIB	0.179	0.244	0.055	0.053	0.054
Trimodality:Yes vs No	0.207	0.230	0.069	0.065	0.065
p-value					
$\hat{\beta}$					
	Trans-Cox	Cox(MDA)	Cox(NCDB)	Cox(Combined)	Cox(Stratified)
Normalized age	0.453	0.498	<0.001	<0.001	<0.001
Grade:III vs I/II	0.002	0.150	<0.001	<0.001	<0.001
Race:Black vs White	0.305	0.616	<0.001	<0.001	<0.001
Race:Other vs White	0.623	0.295	0.118	0.109	0.19
Stage:IIIC vs IIIB	0.753	0.318	<0.001	<0.001	<0.001
Trimodality:Yes vs No	<0.001	<0.001	<0.001	<0.001	<0.001

[*] Boostrap-based standard error estimation with 1000 bootstrap resamplings.

[†] Cox(Combined) is the Cox regression model using the combined data of MDA and NCDB.