Longitudinal Multinomial Relative Entropy-Based Discrete Relative Risk Models for Integrated Prediction of Competing Risk

Qinmengge Li, Kevin He

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109
liqinmg@umich.edu, kevinhe@umich.edu

Abstract

Prognosis prediction is a pivotal aspect of survival analysis, particularly when considering competing risks. The contemporary landscape is enriched with an abundance of biobank data encompassing diverse risk factors like genetics, transcriptomics, and electronic health records, fueling efforts to enhance prognostic predictions. However, the resulting predictive models suffer from rare event rates, limited sample sizes, high dimensionality, and low signal-to-noise ratios. To address these challenges and amplify predictive capabilities, the integration of historical prediction models has emerged as a promising approach. Yet, prevailing integration methods often rest upon the premise of comparable underlying distributions across disparate data sources—a presumption that frequently diverges from reality. Disregarding the inherent heterogeneity among these information sources can inadvertently introduce substantial bias, underscoring the urgency of integrated competing risk analyses that systematically accommodate cohort heterogeneity. In response, we propose an original solution: a longitudinal multinomial relative entropy-based integration framework. This methodology incorporates the established prediction models from the literature, yielding refined prognostic predictions for newly acquired datasets.

Introduction

In clinical studies, the presence of competing risks is a common scenario. Previous research for predicting competing outcomes has predominantly focused on data elements from single data sources. As a result, these analyses often encounter issues such as rare event rates, small sample sizes, high dimensionality and low signal-to-noise ratios. Therefore, the practice of incorporating a local dataset with established prediction models from published sources has gained traction within statistical studies.

However, the conventional integration approaches (Huang, Qin, and Tsai 2016) typically assume a similarity in underlying distributions across multiple data sources, an assumption that is often invalid. Recently, Chen et al. (2021) relaxed the homogeneity assumption and proposed an adaptive double empirical likelihood estimator by introducing a penalty term on the potential discrepancy among data sources. However, this method only applies to external subgroup survival rates at certain time points (e.g. 5-year survival probabilities for patients older than age 65) defined based on a few categorical variables, precluding its application to more general scenarios. The major challenge revolves around effectively addressing the inherent heterogeneity among data sources. Adding to the complexity, data elements in established prediction models are often restricted to those that already exist and, hence, its covariate space may be different from the local cohort. Sometimes even the outcome information is different. For instance, in clinical studies, local data might offer comprehensive insights into competing risks, whereas historical prediction models might merely rely on aggregated failure data without the finer details.

To address these challenges, we propose a longitudinal multinomial relative entropy-based discrete competing risk integration procedure (RE) to robustly integrate published prediction models and newly collected time-to-event data to improve prediction performance, accounting for challenges including heterogeneity, data sharing, and privacy constraints. The proposed method controls the relative weight of the historical information, identifying the most compatible ones and diminishing the weights of less relevant ones. Moreover, the proposed method is flexible to incorporate incomplete historical information; e.g. the new data contains competing risk information but the historical models only have overall failure information.

Method

Discrete Competing Risk Model

Let T_i and C_i denote the underlying death time and censoring time for subject i in the local data, $i = 1, \ldots, n$ where n is the local sample size. Thus, the observed survival time is $X_i = \min\{T_i, C_i\}$. Let $\Delta_i = 1, \ldots, J$ be the index of causes of death where J is the total number of causes. Let $t_1, \ldots, t_K$ be the distinct failure times and $k = 1, \ldots, K$ be the corresponding index. Let $\boldsymbol{Z}_i$ be a p-dimensional covariate vector for the i-th patient. We assume that, conditional on $\boldsymbol{Z}_i$, T_i is independently censored by C_i.

To model the competing risk, let

$$\lambda_j(t_k; \boldsymbol{Z}_i) = P(T_i = t_k, \Delta_i = j | T_i \geq t_k, \boldsymbol{Z}_i)$$

be the discrete cause-specific hazard (Kalbfleisch and Prentice 2002) at time t_k for the i-th patient with covariate $\boldsymbol{Z}_i$.

Note that the overall hazard is the summation of the cause-specific hazards, i.e.

$$\lambda(t_k; \boldsymbol{Z}_i) = P(T_i = t_k | T_i \geq t_k, \boldsymbol{Z}_i) = \sum_{j=1}^{J} \lambda_j(t_k; \boldsymbol{Z}_i).$$

Let $\mathcal{D}_k$ denote the set of labels associated with individuals failing at time t_k. The set of labels associated with individuals censored at time t_k is denoted as $\mathcal{C}_k$, and the at risk set at time t_k is denoted as $\mathcal{R}_k$. The likelihood function is given by,

$$L = \prod_{k=1}^{K} \prod_{j=1}^{J} \left\{ \prod_{i=1}^{n} f_j(t_k; \boldsymbol{Z}_i)^{\delta_{ij}(t_k)} \right\} \prod_{i \in \mathcal{C}_k} S(t_k; \boldsymbol{Z}_i), \quad (1)$$

where $\delta_{ij}(t_k) = I(T_i = t_k, \Delta_i = j)$ is the cause-specific death indicator at time t_k,

$$S(t_k; \boldsymbol{Z}_i) = P(T_i > t_k | \boldsymbol{Z}_i) = \prod_{l=1}^{k} \{1 - \lambda(t_l; \boldsymbol{Z}_i)\}$$

is the survival function, and

$$\begin{aligned} f_j(t_k; \boldsymbol{Z}_i) &= P(T_i = t_k, \Delta_i = j | \boldsymbol{Z}_i) \\ &= \lambda_j(t_k; \boldsymbol{Z}_i) \prod_{l=1}^{k-1} \{1 - \lambda(t_l; \boldsymbol{Z}_i)\} \end{aligned}$$

is the sub-density function of subject i at time t_k. Note that both the survival function and the sub-density function, and hence, the likelihood function in (1) can be fully represented by the cause-specific hazard functions. That is,

$$\begin{aligned} L = \prod_{k=1}^{K} \Bigg[& \prod_{i \in \mathcal{R}_k - \mathcal{D}_k} \{1 - \lambda(t_k; \boldsymbol{Z}_i)\} \\ & \prod_{j=1}^{J} \left\{ \prod_{i=1}^{n} \lambda_j(t_k; \boldsymbol{Z}_i)^{\delta_{ij}(t_k)} \right\} \Bigg]. \end{aligned} \quad (2)$$

Historical Prediction Model

Suppose we have a historical prediction model that returns predicted cause-specific hazards

$$\tilde{\lambda}_j(t_k; \mathbf{Z}_i) = \tilde{P}(T_i = t_k, \Delta_i = j | T_i \geq t_k, \mathbf{Z}_i)$$

for each subject in the local data, where $\tilde{P}$ is a failure time distribution corresponding to the historical model. Then for each of the subjects in the local data, a sequence of historical predicted hazards across multiple time points can be calculated. Note that the method can incorporate various types of historical prediction models, including Cox models, discrete competing risk models, or various types of machine learning algorithms, as long as the historical model provides predicted cause-specific hazards.

Longitudinal Multinomial-Based Relative Entropy

To integrate information across multiple data sources with competing risks, we utilize the following two facts: First, the likelihood of the discrete cause-specific hazards model, as shown in (2), can be fully represented by the discrete cause-specific hazard functions, which are conditional probabilities; second, with the multiple failure types, the discrete cause-specific hazards model can be viewed as a sequence of longitudinal multinomial trials. These motivate us to define a longitudinal multinomial-based relative entropy to measure the disparity between the two probability distributions corresponding to the historical model $\tilde{P}$ and the discrete competing risk model in the local data P,

$$\begin{aligned} \mathrm{d}(\tilde{P} \| P; t_k, \boldsymbol{Z}_i) = & \sum_{j=1}^{J} \tilde{\lambda}_j(t_k, \mathbf{Z}_i) \log \left\{ \frac{\tilde{\lambda}_j(t_k, \mathbf{Z}_i)}{\lambda_j(t_k, \mathbf{Z}_i)} \right\} \\ & + \{1 - \tilde{\lambda}(t_k, \mathbf{Z}_i)\} \log \left\{ \frac{1 - \tilde{\lambda}(t_k, \mathbf{Z}_i)}{1 - \lambda(t_k, \mathbf{Z}_i)} \right\}, \end{aligned} \quad (3)$$

where $\tilde{\lambda}_j(t_k; \mathbf{Z}_i)$ is the predicted cause-specific hazard based on the historical model. The accumulated relative entropy in the sequence of longitudinal multinomial outcomes over subjects in the local data is then given by

$$\mathrm{RE}(\tilde{P} \| P) = \sum_{i=1}^{n} \sum_{k=1}^{K} Y_i(t_k) \mathrm{d}(\tilde{P} \parallel P; t_k, \boldsymbol{Z}_i), \quad (4)$$

where $Y_i(t_k) = I(X_i \geq t_k)$ is the at-risk indicator at time t_k.

To balance the trade-off between the historical prediction model and the local competing risk data, we maximize the log-likelihood of the local data, while keeping the accumulative relative entropy, $\mathrm{RE}(\tilde{P} \| P)$, small. This can be achieved by maximizing a penalized log-likelihood,

$$\ell_\eta = \ell - \eta \, \mathrm{RE}(\tilde{P} \| P),$$

where $\ell = \log L$ is the log-likelihood on the local data, with L defined in (2), η is a tuning parameter weighing the relative importance of the historical model to the local data. In the extreme case of $\eta = 0$, ℓ_η is reduced to the log-likelihood of the local data. When $\eta = \infty$, the resulting model is equivalent to the historical model. Thus, the relative entropy-based integration controls the relative weight of the historical model, assigning larger weights to more comparable ones and diminishing the weights of less relevant ones. The tuning parameter η can be determined by cross-validation to minimize the predictive deviance (Burnham and Anderson 2002; Tutz and Schmid 2016), which is the negative log-likelihood evaluated on the validation data.

Estimation

In this section, we will show that the proposed method is computationally efficient for both low-dimensional and high-dimensional problems, and can be easily extended to various machine learning methods. Specifically, consider the following link function for longitudinal multinomial trials

$$\log \left\{ \frac{\lambda_j(t_k; \boldsymbol{Z}_i)}{1 - \sum_{j^*=1}^{J} \lambda_{j^*}(t_k; \boldsymbol{Z}_i)} \right\} = \gamma_{kj} + \boldsymbol{Z}_i^\top \boldsymbol{\beta}_j,$$

where γ_{kj} is the link-transformed baseline hazard for cause j at time k, and $\boldsymbol{\beta}_j$ denotes a vector of cause-specific coefficient associated with $\boldsymbol{Z}_i$. Let $\boldsymbol{\theta}_j = (\boldsymbol{\gamma}_j^\top, \boldsymbol{\beta}_j^\top)^\top$ be the parameters for cause j, where $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{Kj})^\top$. The corresponding log-likelihood is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k) \left[\sum_{j=1}^{J} \left\{ \delta_{ij}(t_k)(\gamma_{kj} + \boldsymbol{Z}_i^\top \boldsymbol{\beta}_j) \right\} - \log \left\{ 1 + \sum_{j=1}^{J} \exp(\gamma_{kj} + \boldsymbol{Z}_i^\top \boldsymbol{\beta}_j) \right\} \right], \quad (5)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_J^\top)^\top$.

Proposition 1 *Ignoring constant terms not involving $\boldsymbol{\theta}$, the penalized log-likelihood*

$$\ell_\eta(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \eta\, RE(\tilde{P} \| P) \quad (6)$$

is proportional to the following objective function

$$\ell_\eta(\boldsymbol{\theta}) \propto \sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k) \left[\sum_{j=1}^{J} \left\{ \tilde{\delta}_{ij}(t_k)(\boldsymbol{\gamma}_{kj} + \boldsymbol{Z}_i^\top \boldsymbol{\beta}_j) \right\} - \log \left\{ 1 + \sum_{j=1}^{J} \exp(\boldsymbol{\gamma}_{kj} + \boldsymbol{Z}_i^\top \boldsymbol{\beta}_j) \right\} \right], \quad (7)$$

where

$$\tilde{\delta}_{ij}(t_k) = \frac{\delta_{ij}(t_k) + \eta \tilde{\lambda}_j(t_k; \boldsymbol{Z}_i)}{1 + \eta}$$

is the weighted average of the observed cause-specific death indicator and the predicted cause-specific hazards.

Proposition 1 shows that the penalized log-likelihood $\ell_\eta(\boldsymbol{\theta})$ has a similar form, and hence, similar computation cost as the classical discrete competing risk models. Existing estimation procedures, such as Newton-Raphson (Kalbfleisch and Prentice 2002) for low-dimensional settings, Lasso (Hastie, Tibshirani, and Wainwright 2015) for high-dimensional variable selection, XGboost (Chen and Guestrin 2016) for gradient boosting, and Nnet-Survival (Gensheimer and Narasimhan 2019) for deep neural networks can be easily generalized with the proposed method. For example, with a L_1 penalty, a penalized objective function for the integrated Lasso is given by

$$\ell_{\eta,\lambda}(\boldsymbol{\theta}) \quad = \quad \ell_\eta(\boldsymbol{\theta}) - \lambda \|\boldsymbol{\beta}\|_1, \quad (8)$$

where $\|\boldsymbol{\beta}\|_1$ is the L_1 norm of $\boldsymbol{\beta}$, and λ is a tuning parameter that determines the amount of penalization. Since the penalized objective function in (8) has a similar form as the classical Lasso for multinomial outcomes (Hastie, Tibshirani, and Wainwright 2015), the parameter estimation can be easily obtained by the commonly used coordinate descent-based procedures.

Asymptotic Properties

Let $\hat{\boldsymbol{\theta}}(\eta)$ and $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood estimators of the proposed integrated procedure and un-integrated procedure using the local data only. It can be shown that a range $(0, \eta^*]$ exists, on which the asymptotic mean squared error (aMSE) of $\hat{\boldsymbol{\theta}}(\eta)$ is strictly less than that of $\hat{\boldsymbol{\theta}}$. Note that when $\eta = 0$, $\hat{\boldsymbol{\theta}}(\eta)$ reduces to $\hat{\boldsymbol{\theta}}$. When the historical and the new cohorts are homogeneous, $\hat{\boldsymbol{\theta}}(\eta)$ is an unbiased estimator and the variance of $\hat{\boldsymbol{\theta}}(\eta)$ decreases as η increases; hence, aMSE$(\hat{\boldsymbol{\theta}}(\eta)) <$ aMSE$(\hat{\boldsymbol{\theta}})$ for all $\eta > 0$. When the historical and the new cohorts are heterogeneous, for $\eta \in [0, \eta^*]$, ∂aMSE$\{\hat{\boldsymbol{\theta}}(\eta)\}/\partial\eta < 0$ and, hence, aMSE$(\hat{\boldsymbol{\theta}}(\eta)) <$ aMSE$(\hat{\boldsymbol{\theta}})$. Thus, even integrating heterogeneous historical information increases the bias of $\hat{\boldsymbol{\theta}}(\eta)$; due to the bias and variance trade-off, the aMSE of $\hat{\boldsymbol{\theta}}(\eta)$ can still be improved with an appropriate range of integration weight η. Note that the conjectured properties accommodate the case where the newly collected data contain more variables than used in the historical model.

Connection with Existing Methods

Relative entropy, also known as the Kullback-Leibler (KL) divergence (Kullback and Leibler 1951), has been proposed to integrate historical information for binary and ordinal outcomes (Liu and Shum 2003; Schapire et al. 2005; Jiang, He, and Zhang 2016). However, these methods are not applicable for censored time-to-event data, as a censored outcome only provides partial information that the unobserved failure time is greater than the observed censoring time. Recently, (Wang et al. 2021) proposed a discrete hazard-based relative entropy for longitudinal binary outcomes, based on the fact that the likelihood of the discrete failure time models can be fully represented by the discrete hazard functions and, hence, both failed and censored information can be incorporated. This motivates us to extend the discrete hazard-based relative entropy to competing risk settings, measure the discrepancy between the published prediction model and the local data with competing risks.

Extensions

In what follows, we provide two extensions of the proposed integration procedure to incorporate the scenarios with incomplete outcome information: (1) integrate a local competing risk data and a historical overall mortality model without detailed competing risk information; and (2) integrate a local overall mortality data without detailed competing risk information and a historical competing risk model.

Integration with Incomplete Historical Information

Utilizing the fact that the historical (overall) failure time can be viewed as a longitudinal binary outcome, and that the overall hazard function is the summation of the cause-specific hazards, we define the following longitudinal

Bernoulli relative entropy as follows:

$$
\begin{aligned}
&d(\tilde{P}\|P; t_k, Z_i) \\
=&\tilde{\lambda}(t_k, \mathbf{Z}_i) \log\left\{\frac{\tilde{\lambda}(t_k, \mathbf{Z}_i)}{\sum_{j=1}^{J}\lambda_j(t_k, \mathbf{Z}_i)}\right\} \\
&+\{1-\tilde{\lambda}(t_k, \mathbf{Z}_i)\}\log\left\{\frac{1-\tilde{\lambda}(t_k, \mathbf{Z}_i)}{1-\sum_{j=1}^{J}\lambda_j(t_k, \mathbf{Z}_i)}\right\}, \quad (9)
\end{aligned}
$$

which provides a metric to control the relative weight of the information between the historical model based on overall mortality only and the local competing risk data.

Proposition 2 *Ignoring constant terms not involving $\boldsymbol{\theta}$, the penalized log-likelihood $\ell_\eta(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \eta\ RE(\tilde{P}\|P)$, where the accumulative longitudinal Bernoulli relative entropy $RE(\tilde{P}\|P)$ is defined similarly as* (3)*, is proportional to the following objective function*

$$
\begin{aligned}
&\ell_\eta(\boldsymbol{\theta}) \\
&\propto \sum_{i=1}^{n}\sum_{k=1}^{K}\frac{Y_i(t_k)}{1+\eta}\left[\sum_{j=1}^{J}\left\{\delta_{ij}(t_k)(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right. \\
&\left.+\eta\tilde{\delta}_i(t_k)\log\left\{\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right] \\
&-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\log\left\{1+\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}, \quad (10)
\end{aligned}
$$

where

$$\tilde{\delta}_i(t_k) = \tilde{P}(T_i = t_k | T_i \geq t_k, \mathbf{Z}_i) = \tilde{\lambda}(t_k; \boldsymbol{Z}_i)$$

is the predicted event indicator for patient i who is at risk at time t_k.

Integration with Incomplete Local Information

Similarly, we can also integrate a local data with overall mortality information and a historical competing risk model, by utilizing the fact that $\tilde{\lambda}(t_k; \boldsymbol{Z}_i) = \sum_{j=1}^{J}\tilde{\lambda}_j(t_k; \boldsymbol{Z}_i)$ and collapsing all the cause-specific hazards in historical model to the overall hazards. We then modify the longitudinal Bernoulli relative entropy as follows:

$$
\begin{aligned}
&d(\tilde{P}\|P; t_k, Z_i) \\
&=\left\{\sum_{j=1}^{J}\tilde{\lambda}_j(t_k, \mathbf{Z}_i)\right\}\log\left\{\frac{\sum_{j=1}^{J}\tilde{\lambda}_j(t_k, \mathbf{Z}_i)}{\lambda(t_k, \mathbf{Z}_i)}\right\} \\
&+\left\{1-\sum_{j=1}^{J}\tilde{\lambda}_j(t_k, \mathbf{Z}_i)\right\}\log\left\{\frac{1-\sum_{j=1}^{J}\tilde{\lambda}_j(t_k, \mathbf{Z}_i)}{1-\lambda(t_k, \mathbf{Z}_i)}\right\}, \quad (11)
\end{aligned}
$$

which provides a metric to control the relative weight of the information between the historical competing risk model and the local data based on overall mortality only.

Combining the log-likelihood of the discrete overall hazard model on the local data

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\left\{\delta_i(t_k)(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^\top\boldsymbol{\beta})\right\}\right. \\
\left.-\log\left\{1+\exp(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^\top\boldsymbol{\beta})\right\}\right] \quad (12)
\end{aligned}
$$

and the accumulative longitudinal Bernoulli relative entropy, Proposition 3 shows that the resulting penalized log-likelihood has a similar form and computational cost as the classical discrete relative risk model.

Proposition 3 *Ignoring terms not involving $\boldsymbol{\theta}$, the penalized log-likelihood $\ell_\eta(\boldsymbol{\theta})$ is proportional to the following objective function*

$$
\begin{aligned}
\ell_\eta(\boldsymbol{\theta}) \propto \sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\frac{\delta_i(t_k)+\eta\sum_{j=1}^{J}\tilde{\lambda}_j(t_k, \mathbf{Z}_i)}{1+\eta}\right. \\
\left.(\boldsymbol{\gamma}_k+Z_i^\top\boldsymbol{\beta})-\log\left\{1+\exp(\boldsymbol{\gamma}_k+Z_i^\top\boldsymbol{\beta})\right\}\right.
\end{aligned}
$$

Simulation

To evaluate the effectiveness of the proposed integration framework based on longitudinal multinomial relative entropy for addressing discrete competing risk problems, we perform simulation experiments across both low-dimensional and high-dimensional scenarios.

Low-Dimensional Settings

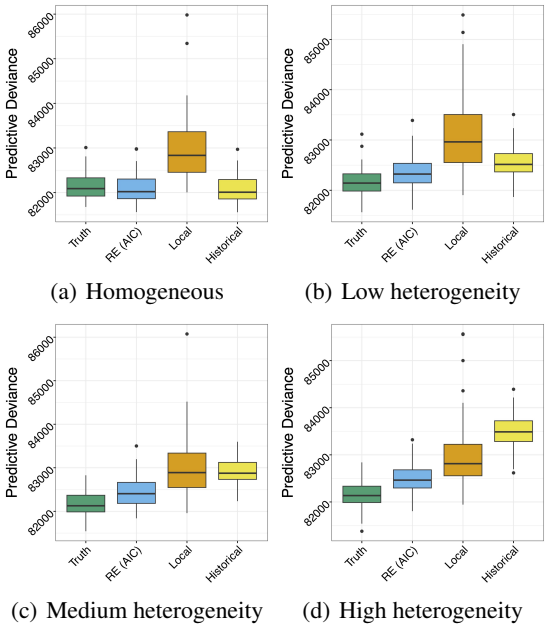
In the low-dimensional setting, a total of four predictors are drawn from independent identically distributed normal distributions with mean zero and unit variance. Two causes of competing failure times are generated from model (5) with 10 distinct failure time points and both causes are independently censored with uniform distribution across the 10 time points. The link-transformed baseline hazards are set as $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = (-2.3, -2.2, -2.08, -1.95, -1.8, -1.61, -1.39, -1.1, -0.7, 0.01)$. The covariate effects for local data are set as $\boldsymbol{\beta}_1 = (-0.05, -0.2, 0.05, 0.2)$, $\boldsymbol{\beta}_2 = (-0.15, 0.1, 0.15, -0.1)$. Two datasets are then drawn from this data generation procedure: the local data with a sample size of 1000 and the testing data with a sample size of 20000. To introduce varying degrees of heterogeneity, we manipulate the covariate effects during historical data generation, then draw 10,000 historical samples from each setting. Specifically:

Homogeneous setting: The historical data are also drawn with the same covariate effects as local data.

Low heterogeneity setting: The covariate effects are adjusted to $\boldsymbol{\beta}_1 = (-0.15, -0.1, 0.15, 0.1)$, $\boldsymbol{\beta}_2 = (-0.25, 0.01, 0.25, -0.01)$.

Medium heterogeneity setting: The covariate effects are adjusted to $\boldsymbol{\beta}_1 = (-0.2, -0.05, 0.2, 0.05)$, $\boldsymbol{\beta}_2 = (-0.3, 0.25, 0.3, -0.25)$.

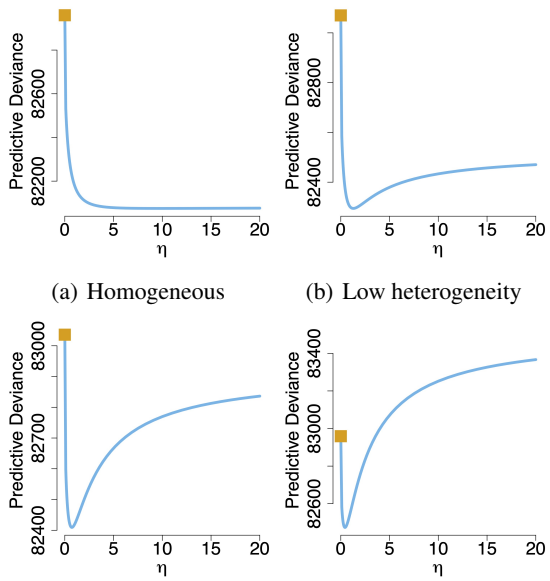
High heterogeneity setting: The covariate effects are adjusted to $\boldsymbol{\beta}_1 = (-0.25, -0.01, 0.25, 0.01)$, $\boldsymbol{\beta}_2 = (-0.35, 0.3, 0.35, -0.3)$.



(a) Homogeneous (b) Low heterogeneity

(c) Medium heterogeneity (d) High heterogeneity

Figure 1: Boxplots of low dimensional simulation results: Boxplots depict model performances assessed through predictive deviances, which represent testing data log-likelihoods evaluated on different models across 100 simulation repetitions. Lower deviance indicates superior prediction. Labels on the x-axis: "Truth" - ground truth model; "RE (AIC)" - proposed integrated model with AIC-selected η; "Local" - model on Local dataset; "Historical" - model on historical dataset.



(a) Homogeneous (b) Low heterogeneity

(c) Medium heterogeneity (d) High heterogeneity

Figure 2: Low dimensional simulation results η curve: the orange square represents the mean predictive deviance of the local model across 100 repeats. The blue line represents the mean predictive deviance of the RE model with varying η across 100 repeats.

The integration approach we propose, labeled as "RE (AIC)," incorporates the use of AIC for fine-tuning the parameter η. This method is juxtaposed against several benchmark models for comparison purposes:

Ground truth model (Truth): The ideal model that corresponds to the actual underlying data generating process;

Local Model (Local): A model fitted exclusively on the local dataset;

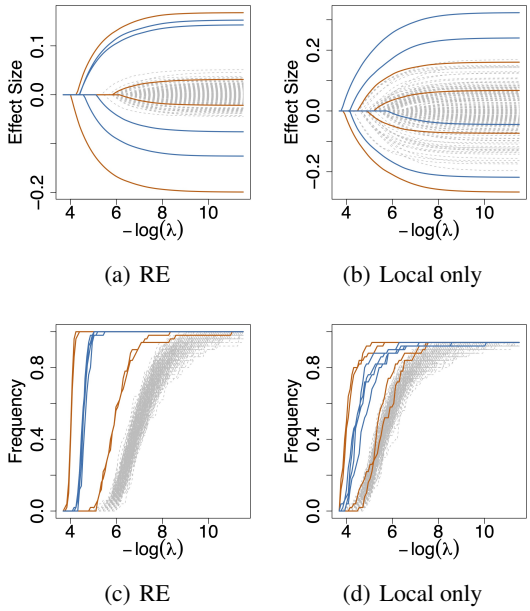
Historical Model (Historical): A model fitted solely on the historical dataset. The performance of these models is assessed using predictive deviance, measured on the testing dataset. Predictive deviance represents the negative log-likelihood of the testing dataset evaluated on the various model estimates. To ensure robustness, the simulation process is iterated 100 times.

In Figure 1, the subfigures labeled as (a), (b), (c), and (d) correspond respectively to the homogeneous, low heterogeneity, medium heterogeneity, and high heterogeneity settings outlined earlier. Our observations indicate that across all these settings, the proposed "RE" model consistently outperforms both the "Local" and "Historical" models in terms of prediction accuracy, as evident from the smaller values of predictive deviance achieved.It's important to acknowledge that in real-world scenarios, the direct sharing of sensitive individual-level historical data is often restricted due to concerns regarding patient privacy. Instead, only summary statistics are typically made available, a concept referred to as the DataSHIELD constraint (Cai, Liu, and Xia 2022). Notably, our proposed method effectively incorporates information from these published summary statistics. This integration not only enhances prediction performance but also ensures the safeguarding of patient privacy.

In Figure 2, panels (a), (b), (c), and (d) correspond to the aforementioned four simulation settings. They illustrate the mean RE predictive deviance across 100 simulation repetitions in relation to the tuning parameter η. As discussed, when $\eta = 0$, the RE model solely relies on local data (depicted as orange squares), while as η approaches infinity, the RE model becomes dependent solely on historical information. These four plots validate the theoretical notion that a suitable range of η exists, ensuring that the RE model consistently outperforms the use of either local or historical information alone. When combined with the insights from Figure 1, we demonstrate that through the application of AIC criteria, the selection of η within this range is achievable. This selection guarantees the aforementioned property, strengthening the argument for the effectiveness of the RE model.

Hign-Dimensional Settings

In the context of the high-dimensional setting, data generation follows a process similar to that of the low-dimensional scenario. However, here we employ 100 predictors, among which four possess non-zero effects of equivalent magnitude to those in the low-dimensional case. Our focus lies in inves-



(a) RE (b) Local only

(c) RE (d) Local only

Figure 3: High dimensional simulation results: (a) and (b) showcase the solution paths of the proposed RE model and Lasso regression fitted on local data only on one of the simulation repeats. The non-zero effects of two different causes are presented as orange and blue lines corresponding to cause 1 and cause 2, and zero effects are the grey dashed lines, depicting the effect size change with regard to the Lasso tunning parameter λ. (c) and (d) showcase the frequency plots of the proposed RE model and Lasso regression fitted on local data only across all simulation repeats. The non-zero effects of two different causes are presented as orange and blue lines corresponding to cause 1 and cause 2, and zero effects are the grey dashed lines, depicting the variable selection frequency with regard to the Lasso tunning parameter λ across the 100 simulation repeats.

tigating a homogeneous simulation setting, where the effect sizes are consistent between historical and local data.

Figure 3, specifically panels (a) and (b), showcases the solution path results from one of the simulation repeats. The proposed high-dimensional RE model, defined as per equation (8), effectively enhances the task of variable selection by integrating historical information. Notably, compared to the Lasso model employing solely local data, our RE model better distinguishes informative signals with non-zero coefficients. The improvement in distinguishing these significant signals contributes to the overall performance enhancement.

To further gauge the variable selection capability of our proposed approach, Figure 3, namely panels (c) and (d), employs stability selection (Meinshausen and Bühlmann 2010). This method serves as an efficient technique to rank the predictive importance of predictors. Specifically, our procedure involves bootstrapping the simulated data 100 times. For each iteration of resampling, both our proposed high-dimensional RE model and the Lasso, utilizing only local data, are applied. The selection frequencies for both techniques are computed as empirical probabilities, quantifying the likelihood of each variable being selected. The outcomes reveal a notable trend. When compared to the Lasso using exclusively local data, our proposed RE model demonstrates an increased likelihood of selecting informative variables. This outcome further solidifies the superiority of our RE model in terms of variable selection in the high-dimensional context.

Discussion

To integrate published survival models with newly collected competing risk data, we propose a longitudinal multinomial relative entropy-based integration procedure. The proposed method is flexible in the following aspects: (1) it controls the relative weight of the historical information, identifying the most compatible ones and diminishing the weights of less relevant ones; (2) it requires less sensitive summary statistics to be obtained from historical data sources, but with no use of any sensitive individual-level historical data; and (3) the proposed procedure has a similar computational cost as the classical discrete competing risk models, and can be easily extended to many machine learning methods, including deep neural networks, high-dimensional variable selection, and gradient boosting. Finally, we note that the predicted cause-specific hazards from the proposed method can be easily combined to obtain the cumulative incidence functions (Lee, Feuer, and Fine 2018) for competing risks.

References

Burnham, K. P.; and Anderson, D. R. 2002. *Model Selection And Multimodel Inference: A Practical Information-Theoretic Approach*. Springer.

Cai, T.; Liu, M.; and Xia, Y. 2022. Individual data protected integrative regression analysis of high-dimensional heterogeneous data. *Journal of the American Statistical Association*, 117(540): 2105–2119.

Chen, T.; and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, 785–794. New York, NY, USA: ACM. ISBN 978-1-4503-4232-2.

Chen, Z.; Ning, J.; Shen, Y.; and Qin, J. 2021. Combining primary cohort data with external aggregate information without assuming comparability. *Biometrics*, 77: 1024–1036.

Gensheimer, M. F.; and Narasimhan, B. 2019. A scalable discrete-time survival model for neural networks. *PeerJ*, 7: e6257.

Hastie, T.; Tibshirani, R.; and Wainwright, M. 2015. *Statistical Learning With Sparsity*. Chapman & Hall CRC Monographs On Statistics And Applied Probability.

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.

Jiang, Y.; He, Y.; and Zhang, H. 2016. Variable selection with prior information for generalized linear models via the prior LASSO method. *Journal of the American Statistical Association*, 111(513): 355–376.

Kalbfleisch, J.; and Prentice, R. 2002. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Second edition.

Kullback, S.; and Leibler, R. A. 1951. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1): 79–86.

Lee, M.; Feuer, E. J.; and Fine, J. P. 2018. On the analysis of discrete time competing risks data. *Biometrics*, 74(4): 1468–1481.

Liu, C.; and Shum, H.-Y. 2003. Kullback-leibler boosting. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, I–I. IEEE.

Matas, A.; Smith, J.; Skeans, M.; Lamb, K.; Gustafson, S.; Samana, C.; Stewart, D.; Snyder, J.; Israni, A.; and Kasiske, B. 2013. OPTN/SRTR 2011 annual data report: kidney. *American Journal of Transplantation*, 13: 11–46.

Meinshausen, N.; and Bühlmann, P. 2010. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4): 417–473.

Schapire, R. E.; Rochery, M.; Rahim, M.; and Gupta, N. 2005. Boosting with prior knowledge for call classification. *IEEE Transactions On Speech And Audio Processing*, 13(2): 174–181.

Tutz, G.; and Schmid, M. 2016. *Modeling Discrete Time-to-Event Data*. Cham: Springer International Publishing.

Wang, D.; Ye, W.; Sung, R.; Jiang, H.; Taylor, J. M. G.; Ly, L.; and He, K. 2021. Kullback-Leibler-Based Discrete Failure Time Models for Integration of Published Prediction Models with New Time-To-Event Dataset. *arXiv:2101.02354*.

Appendix

Proof of Proposition 1

Plug (3) into (4) we have,

$$
\begin{aligned}
&\mathrm{RE}(\tilde{P}\|P) \\
&= \sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k)\left[\sum_{j=1}^{J}\tilde{\lambda}_j(t_k,\mathbf{Z}_i)\log\left\{\frac{\tilde{\lambda}_j(t_k,\mathbf{Z}_i)}{\lambda_j(t_k,\mathbf{Z}_i)}\right\}\right. \\
&\left.+\left\{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)\right\}\log\left\{\frac{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)}{1-\lambda(t_k,\mathbf{Z}_i)}\right\}\right] \\
&= -\sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k)\left[\sum_{j=1}^{J}\tilde{\lambda}_j(t_k,\mathbf{Z}_i)\log\left\{\lambda_j(t_k,\mathbf{Z}_i)\right\}\right. \\
&\left.+\left\{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)\right\}\log\left\{1-\lambda(t_k,\mathbf{Z}_i)\right\}\right]+C \\
&= -\sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k)\left[\sum_{j=1}^{J}\tilde{\lambda}_j(t_k,\mathbf{Z}_i)\log\left\{\frac{\lambda_j(t_k,\mathbf{Z}_i)}{1-\lambda(t_k,\mathbf{Z}_i)}\right\}\right. \\
&\left.+\log\left\{1-\lambda(t_k,\mathbf{Z}_i)\right\}\right]+C \\
&= -\sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k)\left[\sum_{j=1}^{J}\left\{\tilde{\lambda}_j(t_k,\mathbf{Z}_i)(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right. \\
&\left.-\log\left\{1+\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right]+C, \qquad \text{(S1)}
\end{aligned}
$$

where C is constant with respect to the model parameters.

Comparing S1 to (5), we find they share the same form. Plug S1 and (5) into (6) we have

$$
\begin{aligned}
&\ell_\eta(\boldsymbol{\theta}) \\
&= \ell(\boldsymbol{\theta}) - \eta\,\mathrm{RE}(\tilde{P}\|P) \\
&= \sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k)\left[\sum_{j=1}^{J}\left\{\delta_{ij}(t_k)(\gamma_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right. \\
&\left.-\log\left\{1+\sum_{j=1}^{J}\exp(\gamma_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right] \\
&+\eta\sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k)\left[\sum_{j=1}^{J}\left\{\tilde{\lambda}_j(t_k,\mathbf{Z}_i)(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right. \\
&\left.-\log\left\{1+\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right]+C \\
&\propto \sum_{i=1}^{n}\sum_{k=1}^{K} Y_i(t_k)\left[\sum_{j=1}^{J}\left\{\tilde{\delta}_{ij}(t_k)(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right. \\
&\left.-\log\left\{1+\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^\top\boldsymbol{\beta}_j)\right\}\right]+C/(1+\eta).\square
\end{aligned}
$$

Proof of Proposition 2

Plug (9) into (4) we have,

$$
\begin{aligned}
&\mathrm{RE}(\tilde{P}\|P)\\
&=\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\left\{\frac{\tilde{\lambda}(t_k,\mathbf{Z}_i)}{\sum_{j=1}^{J}\lambda_j(t_k,\mathbf{Z}_i)}\right\}\right.\\
&\left.+\{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)\}\log\left\{\frac{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)}{1-\sum_{j=1}^{J}\lambda_j(t_k,\mathbf{Z}_i)}\right\}\right]\\
&=-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\left\{\sum_{j=1}^{J}\lambda_j(t_k,\mathbf{Z}_i)\right\}\right.\\
&\left.+\{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)\}\log\left\{1-\sum_{j=1}^{J}\lambda_j(t_k,\mathbf{Z}_i)\right\}\right]+C\\
&=-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\left\{\frac{\sum_{j=1}^{J}\lambda_j(t_k,\mathbf{Z}_i)}{1-\sum_{j=1}^{J}\lambda_j(t_k,\mathbf{Z}_i)}\right\}\right.\\
&\left.+\log\left\{1-\sum_{j=1}^{J}\lambda_j(t_k,\mathbf{Z}_i)\right\}\right]+C\\
&=-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\left\{\sum_{j=1}^{J}exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right.\\
&\left.-\log\left\{1+\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right]+C, \qquad \text{(S2)}
\end{aligned}
$$

where C is constant with respect to the model parameters.

Plug S2 and (5) into (6) we have

$$
\begin{aligned}
&\ell_\eta(\boldsymbol{\theta})\\
&=\ell(\boldsymbol{\theta})-\eta\,\mathrm{RE}(\tilde{P}\|P)\\
&=\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\sum_{j=1}^{J}\left\{\delta_{ij}(t_k)(\gamma_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right.\\
&\left.-\log\left\{1+\sum_{j=1}^{J}\exp(\gamma_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right]\\
&+\eta\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\left\{\sum_{j=1}^{J}exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right.\\
&\left.-\log\left\{1+\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right]+C\\
&\propto\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{Y_i(t_k)}{1+\eta}\left[\sum_{j=1}^{J}\left\{\delta_{ij}(t_k)(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right.\\
&\left.+\eta\tilde{\delta}_i(t_k)\log\left\{\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}\right]+C/(1+\eta)\\
&-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\log\left\{1+\sum_{j=1}^{J}\exp(\boldsymbol{\gamma}_{kj}+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_j)\right\}.\square
\end{aligned}
$$

Proof of Proposition 3

Plug (11) into (4) we have,

$$
\begin{aligned}
\mathrm{RE}(\tilde{P}\|P)&=\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\left\{\frac{\tilde{\lambda}(t_k,\mathbf{Z}_i)}{\lambda(t_k,\mathbf{Z}_i)}\right\}\right.\\
&\left.+\{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)\}\log\left\{\frac{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)}{1-\lambda(t_k,\mathbf{Z}_i)}\right\}\right]\\
&=-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\{\lambda(t_k,\mathbf{Z}_i)\}\right.\\
&\left.+\{1-\tilde{\lambda}(t_k,\mathbf{Z}_i)\}\log\{1-\lambda(t_k,\mathbf{Z}_i)\}\right]+C\\
&=-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)\log\left\{\frac{\lambda(t_k,\mathbf{Z}_i)}{1-\lambda(t_k,\mathbf{Z}_i)}\right\}\right.\\
&\left.+\log\{1-\lambda(t_k,\mathbf{Z}_i)\}\right]+C\\
&=-\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta})\right.\\
&\left.-\log\left\{1+\exp(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta})\right\}\right]+C, \qquad \text{(S3)}
\end{aligned}
$$

where C is constant with respect to the model parameters.

Plug S3 and failure-time log likelihood, (12), into (6) we have

$$
\begin{aligned}
&\ell_\eta(\boldsymbol{\theta})\\
&=\ell(\boldsymbol{\theta})-\eta\,\mathrm{RE}(\tilde{P}\|P)\\
&=\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\left\{\delta_i(t_k)(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta})\right\}\right.\\
&\left.-\log\left\{1+\exp(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta})\right\}\right]\\
&+\eta\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\tilde{\lambda}(t_k,\mathbf{Z}_i)(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta})\right.\\
&\left.-\log\left\{1+\exp(\boldsymbol{\gamma}_k+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta})\right\}\right]+C\\
&\propto\sum_{i=1}^{n}\sum_{k=1}^{K}Y_i(t_k)\left[\frac{\delta_i(t_k)+\eta\sum_{j=1}^{J}\tilde{\lambda}_j(t_k,\mathbf{Z}_i)}{1+\eta}(\boldsymbol{\gamma}_k+Z_i^{\top}\boldsymbol{\beta})\right.\\
&\left.-\log\left\{1+\exp(\boldsymbol{\gamma}_k+Z_i^{\top}\boldsymbol{\beta})\right\}\right]+C/(1+\eta).\square
\end{aligned}
$$

Real Data Analysis

We implemented the proposed RE data integration approach on the kidney transplantation data from the Organ Procurement and Transplantation Network (OPTN) (Matas et al. 2013). We conducted a comprehensive investigation involving 31,243 European ancestry and 16,339 African ancestry individuals who underwent kidney transplantation surgery between January 2007 and January 2012. Over the course of one year following the surgery, we tracked the occurrence of graft failure or patient death as competing risk events. Among patients of European ancestry, we observed a total of 1,187 deaths and 1,350 graft failures. For patients of African ancestry, there were 583 deaths and 925 graft failures. In the original dataset, failure times were recorded with a daily resolution. However, acknowledging the practical challenges faced by clinicians in assessing the exact failure time, we

decided to utilize a weekly resolution by grouping the failure time into 53 discrete weeks.

We included various covariates, encompassing both donor and recipient demographic characteristics (age, gender, race, BMI). Additionally, we collected donor-specific factors such as hypertension status, frozen organ time, and whether the donated kidney met the criteria for being of low quality (Expanded Criteria Donor [ECD]). Among recipient-specific factors, we considered diabetes status, the duration since the diagnosis of end-stage renal disease (ESRD), and whether the recipient had previously received a kidney transplant.

Our objective is to construct a competing risk model specifically taylored to African ancestry patients, utilizing insights gained from European ancestry patients. We randomly divided the cohort of 16,339 African ancestry patients into two subsets: a training set consisting of 10,000 patients and a testing set with 6,339 patients. To assess the performance of the RE model under different types of prior information, we generated three distinct sets of European prior information:

- (a) Donor and recipients' covariates with competing risk outcomes: This first type of prior contains comprehensive information extracted from European patients, encompassing both donor and recipient covariates with competing risk outcomes. It is the most informative prior among the three.
- (b) Recipients' covariates with competing risk outcomes: The second type of prior includes partial information from European patients, specifically focusing on recipient covariates with competing risk outcomes. It is less informative than the first type of prior as it excludes donor-specific factors.
- (c) Donor and recipients' covariates with failure-time outcomes: The third type of prior also contains partial information from European patients, incorporating both donor and recipient covariates but with failure-time outcomes rather than competing risks. It is less informative than the first type of prior, which includes complete information on competing risk outcomes.

It's important to note that for the African local data, we employed the complete information available, including both donor and recipient covariates with competing risk outcomes, across all three prior settings.

Prior Constructed on Donor and Recipients' Covariates with Competing Risk Outcomes

To construct the prior information in the first setting based on the European patients, we regressed both donor and recipients' covariates on the competing risk outcome by the model given in equation (5) and predicted the cause-specific hazard for African patients in the training set. Then the RE models as given in equation (7) were fitted on the African training set with varying η values. The performance of these RE models was evaluated on the African testing set using two key metrics: Predictive Deviance (PD) and Cause-Specific Concordance Index (C-index). The results of this analysis are presented in Figure 4 and Figure 5. From the results, we found that by leveraging additional European information, the RE model consistently exhibited improved performance, whether evaluated on PD or C-index, compared to using the African training data alone. Additionally, our analysis revealed the presence of a suitable range of η values, where the RE model also outperformed the model using European data alone based on PD or graft failure C-index.

Prior Constructed on Recipients' Covariates with Competing Risk Outcomes

To construct the prior information based on the European patients, we performed the regression that considered only recipients' covariates in relation to the competing risk outcome, using the model specified in equation (5). This allowed us to predict the cause-specific hazard for African patients. Then the RE models as described in equation (7) were fitted on the African training set with varying η values. The RE model performance was evaluated on the African testing set based on PD and C-index as shown in Figure 6 and Figure 7, respectively. From the results, even though prior information only encompassed partial covariates, leveraging it with a proper η still improves the model performance, whether evaluated on PD or C-index, compared to using the African data alone. Furthermore, we demonstrated that the RE model consistently outperformed the model using European data alone evaluated based on graft failure C-index. When considering PD, the RE model outperformed the European data-based model when an appropriate η value was selected.

Prior Constructed on Donor and Recipients' Covariates with Failure-Time Outcomes

To construct the prior information in the third setting based on the European patients, we regressed both donor and recipients' covariates on the failure-time outcome by the model given in equation (12) and predicted the overall hazard for African patients in the training set. Then the RE models as given in equation (10) were fitted on the African training set with varying η values. The RE model performance was evaluated on the African testing set based on PD and C-index as shown in Figure 8 and Figure 9, respectively. From the results, even though the European prior only contained the overall failure-time information leveraging it consistently improved the model performance, either evaluated on PD or C-index, compared to using the African training data alone.

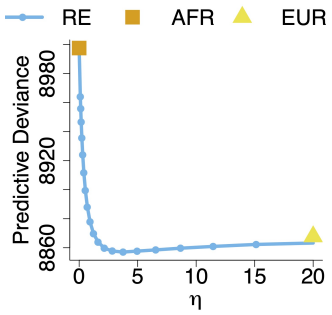


Figure 4: Model predictive deviance with European prior information constructed based on donor and recipients' covariates with competing risk outcomes. RE model performances with varying η are presented in light blue line with round dots. African model performance is presented in orange square. European model performance is presented in yellow triangle.

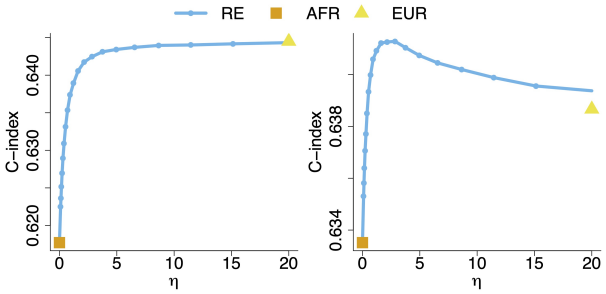


Figure 5: Model C-index with European prior information constructed based on donor and recipients' covariates with competing risk outcomes. The left plot showcases the C-index based on death and the right plot showcases the C-index based on graft failure. RE model performances with varying η are presented in light blue line with round dots. African model performance is presented in orange square. European model performance is presented in yellow triangle.

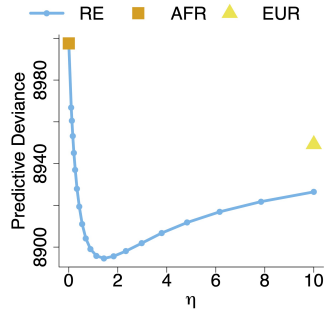


Figure 6: Predictive deviance with European prior information constructed based on recipients' covariates with competing risk outcomes. RE model performances with varying η are presented in light blue line with round dots. African model performance is presented in orange square. European model performance is presented in yellow triangle.

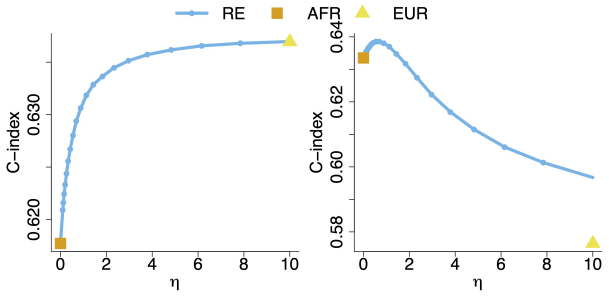


Figure 7: Model C-index with European prior information constructed based on recipients' covariates with competing risk outcomes. The left plot showcases the C-index based on death and the right plot showcases the C-index based on graft failure. RE model performances with varying η are presented in light blue line with round dots. African model performance is presented in orange square. European model performance is presented in yellow triangle.

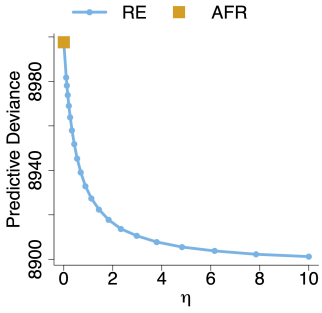


Figure 8: Model predictive deviance with European prior information constructed based on donor and recipients' covariates with failure-time outcomes. RE model performances with varying η are presented in light blue line with round dots. African model performance is presented in orange square.

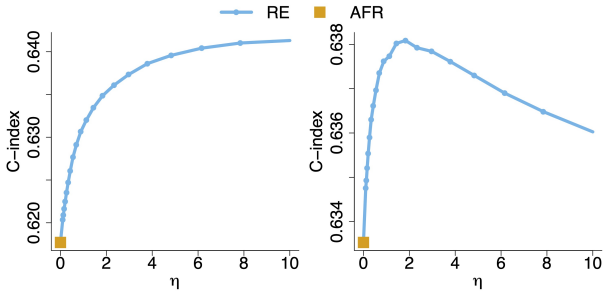


Figure 9: Model C-index with European prior information constructed based on donor and recipients' covariates with failure-time outcomes. The left plot showcases the C-index based on death and the right plot showcases the C-index based on graft failure. RE model performances with varying η are presented in light blue line with round dots. African model performance is presented in orange square.