

竞争风险数据分析中机器学习改进的逆概率删失加权

付佳琪, 侯文*

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2025年12月27日; 录用日期: 2026年1月21日; 发布日期: 2026年1月29日

摘要

在竞争风险数据分析中, Fine-Gray比例风险模型结合逆概率删失加权(IPCW)是常用的方法, 但传统IPCW权重在处理删失时可能产生不稳定估计。为克服这一局限, 本文引入一种机器学习增强的逆概率加权方法, 把机器学习预测的目标事件概率作为分子纳入权重构建, 并将该权重嵌入IPCW的估计方程中。最后采用Sandwich方差估计量进行统计推断。为验证该方案的可行性与稳健性, 本文选取了几种主流机器学习算法来生成权重中的预测概率, 基于R包中公开的数据进行实例分析, 与传统IPCW与DML相比, 本方法得到了有效且稳定的估计。并通过敏感性分析证实了结果的稳健性。结果表明, 本方法在竞争风险数据分析中展现出了一定的应用潜力。

关键词

竞争风险, 逆概率删失加权, Fine-Gray模型, 因果推断, 机器学习

Improvement of Inverse Probability of Censoring Weighting in Competing Risk Data Analysis by Machine Learning

Jiaqi Fu, Wen Hou*

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: December 27, 2025; accepted: January 21, 2026; published: January 29, 2026

Abstract

In the analysis of competing risks data, the Fine-Gray proportional hazards model combined with

*通讯作者。

Inverse Probability of Censoring Weighting (IPCW) is a commonly used method. However, traditional IPCW weights may lead to unstable estimates when handling censoring. To overcome this limitation, this paper introduces a machine learning-enhanced inverse probability weighting method, which incorporates the target event probability predicted by machine learning as the numerator into weight construction and embeds the resulting weights into the estimating equations of IPCW. Finally, the Sandwich variance estimator is adopted for statistical inference. To verify the feasibility and robustness of the proposed method, several mainstream machine learning algorithms are selected to generate the predicted probabilities in the weights, and a case analysis is conducted based on public data from R packages. Compared with traditional IPCW and DML, the proposed method yields valid and stable estimates. Moreover, sensitivity analysis confirms the robustness of the results. The findings indicate that this method exhibits certain application potential in the analysis of competing risks data.

Keywords

Competing Risks, Inverse Probability of Censoring Weighting (IPCW), Fine-Gray Model, Causal Inference, Machine Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在医学、公共医学与社会科学等领域的随访研究中, 研究者大多关注生存数据中某个单一的、明确结局的事件的发生风险或发生时间。然而, 在实际的随访过程中, 研究对象往往同时暴露在多种互斥的结局事件风险下, 即个体可能经历多种类型的事件, 且一种事件的发生会永久阻碍其他事件的发生。此时就产生了统计分析中一个复杂且关键的问题: “竞争风险”(Competing Risks)。例如, 在骨髓移植治疗随访数据中, “移植相关死亡”就构成了“疾病复发”的竞争事件。此时如果使用传统的分析方法(如 Kaplan-Meier 法或标准 Cox 模型), 将非目标事件视为删失(Censoring)来计算生存概率, 就会高估目标事件的累积发生率, 从而掩盖真实的治疗风险。因此, 为了得到无偏且具有可靠指导意义的结论, 就需要发展并且完善竞争风险数据的分析方法。

Fine 和 Gray (1999)提出了子分布比例风险模型(Subdistribution Proportional Hazards Model), 通常被称为 Fine-Gray 模型[1]。并且该方法首次在竞争风险框架下系统引入了逆概率删失加权(IPCW)的思想。在竞争风险数据中, 研究者分析的核心指标是累计发生函数(Cumulative Incidence Function, CIF), 即在时间点 t 之前发生 k 类特定事件的概率, 记为 $F_k(t) = P(T \leq t, \delta = k)$ 。为了对 CIF 直接建模, Fine 和 Gray (1999)引入了子分布风险函数(Subdistribution Hazard Function)的概念, 它是指 t 时刻发生目标事件的瞬时风险。规定风险过程 $Y_i^*(t) = 1\{T_i \geq t \text{ 或 } (T_i \leq t, \delta_i \neq k)\}$, 即在风险集中除仍在随访的个体外, 也包括已发生竞争事件的个体。Fine-Gray 模型直接对子分布风险进行建模, 其形式如下:

$$\lambda_k(t; \mathbf{Z}_i) = \lambda_{0k}(t) \exp(\beta^T \mathbf{Z}_i),$$

其中 $\lambda_{0k}(t)$ 是基线子分布风险。该模型用回归系数 β 量化了协变量对 CIF 的影响, 具有清晰的临床意义。

而 Fine-Gray 模型在估计子分布风险时, 为了校正右删失可能带来的偏倚, 默认采用了 IPCW 方法。它利用删失时间的 Kaplan-Meier 估计 $\hat{G}(t)$ 来为风险集内的个体赋予权重, 其中 $\hat{G}(t)$ 是删失生存函数(Censoring Survival Function)。加权估计方程如下:

$$U^{IPCW}(\beta) = \sum_{i=1}^n \int_0^{\tau} w_i(t) \{Z_i - \bar{Z}(t, \beta)\} dN_i^{(k)}(t) = 0,$$

其中

$$\bar{Z}^{IPCW}(t, \beta) = \frac{\sum_{j=1}^n w_j(t) Z_j \exp(\beta^T Z_j) Y_j^*(t)}{\sum_{j=1}^n w_j(t) \exp(\beta^T Z_j) Y_j^*(t)}.$$

协变量效应 $\hat{\beta}$ 的标准误由稳健 sandwich 方法获得, R 包默认实现。由于其理论的完备性和结果的直观性, Fine-Gray 模型被视为竞争风险数据分析领域的“金标准”。本文将 Fine-Gray 模型的估计结果作为与我们提出的 ML-IPCW 框架进行比较的基准。

此后, Robins, Hernán 与 Brumback (2000) 在因果推断的边际结构模型框架下, 进一步完善了 IPCW 的理论基础[2]。但这些方法均依赖于一个关键假设: 用于计算权重的删失模型必须被正确设定。随着数据复杂性增加, 传统方法面临模型误设问题时, 机器学习在各种预测任务中展现出了优越的预测性能。这就催生了一个极具前景的研究方向: 将机器学习的强大预测能力与 IPCW 方法相结合。

目前, 该方向的探索已形成多种不同的技术路径。其一, 利用机器学习提升权重估计的稳健性。例如, Lee, Lessler 与 Stuart (2010) 系统比较了逻辑回归与多种机器学习算法在估计倾向得分方面的表现[3]。研究表明, 用机器学习替代传统参数模型来估计权重能显著减少偏差。这一核心思想为解决 IPCW 对删失机制误设敏感的问题提供了关键思路。其二, 将 IPCW 作为组件嵌入机器学习模型, 使其能够处理删失数据。Kvamme, Borgan 与 Scheel (2019) 将 IPCW 权重整合进神经网络的损失函数, 即将 IPCW 权重乘以每个样本的损失, 使标准的前馈网络能够直接进行生存预测[4]。第三, 在统一的半参数框架下进行高阶模型整合。这是当前因果推断领域最为严谨和前沿的分析范式, 以双机器学习(Double Machine Learning, DML)和靶向最大似然估计(Targeted Maximum Likelihood Estimation, TMLE)为代表方法。其中, Chernozhukov 等人(2018)提出的 DML 框架通过理论革新, 为在因果推断中安全地使用任意复杂的机器学习模型提供了统计保障[5]。在本研究中, 我们将 DML 框架应用于竞争风险数据, 旨在通过与 ML-IPCW 及 Fine-Gray 模型的系统对比, 评估不同方法在控制倾向得分所蕴含的混杂信息后的效应估计稳健性。此外, Stensrud 等人(2022)在竞争风险环境下, 将靶向最大似然估计(Targeted Maximum Likelihood Estimation, TMLE)与机器学习结合以估计“可分离效应”。TMLE 首先使用超级学习者(一个多种机器学习算法的加权组合)来初始估计结局和删失机制, 然后通过一个基于影响函数的靶向步骤进行一步更新, 以得到双稳健、半参有效的估计量[6]。这种方法在理论最复杂的场景下, 将机器学习、IPCW 和半参数理论完美地整合在一个框架内, 代表了该领域的最高标准。

尽管现有研究在机器学习与 IPCW 方法结合的上已取得显著进展, 但主要集中于两个方向: 一是利用机器学习优化 IPCW 流程中的“前半程”, 即改进删失概率的估计; 二是在复杂的统一架构下进行整合。但很少有研究考虑利用机器学习结果来稳定 IPCW 估计本身的方差分布。因此, 本文选择在 IPCW 估计方程的稳定化环节引入机器学习预测概率, 以改善权重分布的极端性。基于这一思路, 提出一个机器学习增强的逆概率删失加权(ML-IPCW)统一框架。该框架的核心在于: 将随机森林、极限梯度提升、支持向量机与多层感知机五种主流机器学习算法所预测的目标事件概率作为分子纳入权重构建, 并将该权重嵌入 IPCW 估计方程中。本文基于 R 包中的骨髓移植随访数据(EBMT4)进行实例分析, 以验证所提出方法在实际竞争风险数据中的有效性与实用性。

2. 用于估计稳定化因子的机器学习方法介绍

2.1. ML-IPCW 权重与估计方程

传统的 IPCW 权重对删失概率的倒数进行加权, 用以修正删失带来的偏倚。然而, 该方法在部分个

体的条件删失概率趋近于 0 时会得到极大的权重, 模型的估计结果就会因这一部分的极端权重丧失稳健性[7]。基于此, 本文提出使用机器学习模型来得到一个因子构建机器学习增强的 IPCW 权重(ML-Enhanced IPCW Weight):

$$w_i^{ML}(t) = \frac{\hat{p}_i}{\hat{G}(t_i | \mathbf{Z}_i)},$$

其中分母 $\hat{G}(t_i | \mathbf{Z}_i) = \hat{P}(C > t_i | \mathbf{Z}_i)$ 由拟合 Cox 比例风险模型得到, 用于修正删失。而分子 \hat{p}_i 即为我们提出的机器学习增强因子。它是个体 i 发生目标事件的条件概率: $\hat{p}_i = \hat{P}(\delta_i = 1 | \mathbf{Z}_i)$ 。该值仅依赖于个体的基线协变量, 用以压缩权重, 控制整体方差。

本文用回归系数 β 度量协变量对目标事件累计发生率的影响。在借鉴 Cox 模型的偏似然思想并整合 Fine-Gray 模型的特殊风险集和我们提出的 ML 增强权重后, 给出加权的估计方程(Weighted Estimating Equation)来得到估计值。对于目标事件, 其估计方程为:

$$U^{ML}(\beta) = \sum_{i=1}^n \int_0^{\tau} w_i^{ML}(t) \{ \mathbf{Z}_i - \bar{\mathbf{Z}}^{ML}(t, \beta) \} dN_i^{(k)}(t),$$

其中

$$\bar{\mathbf{Z}}^{ML}(t, \beta) = \frac{\sum_{j=1}^n w_j^{ML}(t) \mathbf{Z}_j \exp(\beta^T \mathbf{Z}_j) Y_j^*(t)}{\sum_{j=1}^n w_j^{ML}(t) \exp(\beta^T \mathbf{Z}_j) Y_j^*(t)},$$

由于该方程没有解析解, 本文通过数值迭代算法找到使最接近零的解, 作为最终估计值 $\hat{\beta}$ 。鉴于模型的半参数特性与权重使用, 本文采用 Sandwich 方差估计量计算的标准误, 其形式为: $\hat{A}^{-1} \hat{B} (\hat{A}^{-1})^T$, 其中 $\hat{A} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial U_i^{ML}(\beta)}{\partial \beta} \bigg|_{\beta=\hat{\beta}}$, $B = \frac{1}{n} \sum_{i=1}^n U_i^{ML}(\hat{\beta}) U_i^{ML}(\hat{\beta})^T$ 。

2.2. 用于权重构建的机器学习方法介绍

本研究提出一种新的权重构建方案, 其核心是利用机器学习模型预测的目标事件概率作为权重的分子。为检验该方案的普适性, 我们选取了 4 种在复杂数据模式识别中表现卓越的算法进行实现与比较。所有模型的预测目标均为二分类事件: 是否发生目标事件($\delta=1$)。

2.2.1. 随机森林(Random Forest, RF)

随机森林是一种强大的集成学习算法, 它通过构建大量决策树并聚合其预测结果来工作。并通过样本随机化(Bagging)和特征随机化来确保每棵树都具有差异性, 从而有效避免过拟合[8][9]。对于分类任务, 最终的预测概率由所有树的预测结果平均得到。

具体实现上, 本研究利用 R 中的 randomForest 包构建了一个分类模型。该模型的响应变量被定义为目标事件的发生状态, 并纳入所有基线协变量作为预测因子。为确保模型的稳定收敛, 我们将决策树的数量设定为 300。同时考虑到数据中事件类型的非平衡性, 通过 classwt 参数为复发类别赋予了更高的权重来提高模型对少数类别事件的识别能力。模型训练完毕后, 通过聚合所有树的类别投票来生成概率预测。数据中被预测为目标事件的概率 $\hat{p}(\mathbf{x})$ 由森林中所有 B 棵决策树的预测结果通过简单平均(即投票比例)得到:

$$\hat{p}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B 1(T_b(\mathbf{x}) = 1),$$

其中 $T_b(\mathbf{x})$ 表示第 b 棵树对样本 \mathbf{x} 的预测类别, $1(T_b(\mathbf{x})=1)$ 是一个指示函数, 当第 b 棵树的预测类别为目标事件时取值为 1, 否则取值为 0。

2.2.2. 极限梯度提升(Extreme Gradient Boosting, XGBoost)

XGBoost 是一种高效的梯度提升决策树算法。它通过加法模型构建一系列决策树, 每一棵树的学习目标都是修正前一棵树所产生的残差。XGBoost 的核心优势在于其在目标函数中显式地加入了正则化项, 并运用二阶泰勒展开进行近似, 从而在保证模型复杂度的同时实现了更快的收敛速度[10]。在第 t 次迭代时, 目标是最小化以下正则化目标函数:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t),$$

其中 l 是损失函数, $\hat{y}_i^{(t-1)}$ 是前 $t-1$ 棵树的预测值, f_t 是第 t 棵树, Ω 是惩罚模型复杂度的正则化项。为了高效优化, XGBoost 使用二阶泰勒展开将目标函数近似为:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t),$$

其中 $g_i = \partial_{y_i} l(y_i, \hat{y}_i^{(t-1)})$ 和 $h_i = \partial_{y_i}^2 l(y_i, \hat{y}_i^{(t-1)})$ 分别为损失函数的一阶和二阶梯度。

本文利用 xgboost 包, 将特征矩阵和标签转换为优化的数据结构。模型设定为二分类逻辑回归, 并采用如下关键超参数进行训练: 学习率为 0.05, 最大树深度为 4, 随机采样率 0.8。模型经过 300 轮迭代后, 输出一个原始预测分数值 $F(\mathbf{x}) = \sum_{k=1}^K f_k(\mathbf{x})$, 其中 f_k 为每棵决策树的贡献。该分数值通过 Sigmoid 函数转换为目标事件的预测概率 $\hat{p}(\mathbf{x})$:

$$\hat{p}(\mathbf{x}) = \sigma(F(\mathbf{x})) = \frac{1}{1 + e^{-F(\mathbf{x})}}.$$

2.2.3. 支持向量机 (Support Vector Machine, SVM)

支持向量机(SVM)是一种基于统计学习理论的强大分类器, 其核心思想是寻找一个能将两类样本最优分离的超平面, 并最大化分类间隔, 这使其具有良好的泛化能力[11][12]。对于线性不可分的问题, SVM 通过核函数 $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ 将输入数据隐式映射到高维特征空间, 从而在该空间中构造一个最大间隔超平面。其中, 本文选用的径向基核函数 $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ 因其强大的局部拟合能力而被广泛使用。SVM 的优化问题可以表述为:

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad \text{s.t.} \quad y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, n,$$

其中, C 是惩罚参数, 用于平衡分类间隔与分类错误, ξ_i 是松弛变量。

本研究使用 e1071 包实现 SVM 的预测。考虑到目标事件在数据中可能占比较少, 我们设置了类别权重以提高模型对少数类的关注度。采用径向基核, 并通过网格搜索 $C \in \{0.5, 1, 2, 4\}$ 和 $\gamma \in \left\{ \frac{1}{p}, \frac{2}{p}, \frac{4}{p} \right\}$ 在的范围内, 以留出集上的 AUC 为评价标准, 选择最优超参数组合。最后, 利用 Platt 缩放法将 SVM 的原始决策函数输出 $f(\mathbf{x})$ 校准为后验概率估计:

$$\hat{p}(\mathbf{x}) = P(y=1|\mathbf{x}) \approx \frac{1}{1 + \exp(A \cdot f(\mathbf{x}) + B)},$$

其中参数 A 和 B 通过在验证集上最小化交叉熵损失来拟合。

2.2.4. 多层感知机 (Multilayer Perceptron, MLP)

多层感知机(MLP)是深度学习领域最基础的前馈神经网络模型。通过多层非线性映射, MLP 能够以任意精度逼近任意连续函数, 具有通用逼近能力和强大的特征表达潜力[13]。一个标准的 MLP 由输入层、至少一个隐藏层和输出层构成。信息从前向后传播, 第 l 层的输出由下式计算:

$$\mathbf{a}^{(l)} = \sigma^{(l)}(\mathbf{z}^{(l)}) = \sigma^{(l)}(\mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}),$$

$\mathbf{W}^{(l)}$ 和 $\mathbf{b}^{(l)}$ 分别为该层的权重矩阵和偏置向量, $\sigma^{(l)}$ 为非线性激活函数。模型通过反向传播(backpropagation)与梯度下降(gradient descent)最小化损失函数(如交叉熵), 以学习最优参数。

在本文的实现中, 我们使用 `nnet` 包构建了一个单隐藏层的 MLP。网络结构包含 6 个隐藏层节点, 并采用 Sigmoid 激活函数。为防止过拟合, 我们设置了权重衰减系数为 $5\text{e-}4$, 这相当于在损失函数中加入了 L_2 正则化项 $\frac{\lambda}{2}\|\mathbf{W}\|^2$ 。输出层同样使用 Sigmoid 激活函数, 使其输出值直接介于 0 到 1 之间, 作为目标事件的预测概率 $\hat{p}(\mathbf{x})$ 。其计算公式为:

$$\hat{p}(\mathbf{x}) = \sigma(z) = \frac{1}{1 + e^{-z}},$$

其中 z 是输出层节点的加权输入。模型使用最大迭代次数 500 进行训练, 以确保充分收敛。

2.3. 双机器学习(DML)框架

为验证 ML-IPCW 框架在因果推断方面的稳健性, 本研究同时采用双机器学习(DML)作为理论对照。DML 是一种基于半参数理论的因果推断方法, 其核心在于通过交叉拟合(cross-fitting)与 Neyman 正交化技术, 构造对第一阶段机器学习模型估计误差稳健的估计量, 从而获得具有渐近正态性的因果效应估计并构建有效的置信区间。

在本研究的实证分析中, 对于每个待分析的风险因素(如预防性治疗、移植年份、年龄分层、供者匹配情况), 我们将其编码为二元处理变量 D , 假设处理变量为 X , 结局变量 Y 为复发事件指示变量。先将样本随机分割为 K 份, 对于每一份数据 k , 使用其余 $K-1$ 份数据训练两个随机森林模型: 处理模型 $\hat{g}^{-k}(X) = E[D|X]$ 与结局模型 $\hat{m}^{-k}(X) = E[Y|X]$, 在第 k 份数据上计算正交化残差: $\tilde{D}_i = D_i - \hat{g}^{-k}(X_i)$ 与 $\tilde{Y}_i = Y_i - \hat{m}^{-k}(X_i)$ 。 θ 由最小二乘回归 $\tilde{Y} = \theta\tilde{D} + \varepsilon$ 得到。遍历所有 K 份数据并将估计量平均, 最终得到处理效应估计量 $\hat{\theta}_{DML}$ 及其方差。我们对 6 个主要风险因素分别重复此流程, 得到各自的因果效应估计。

该方法的理论保障在于其构造的得分函数 $\psi = (W; \theta, \eta)$, 其中观测数据 $W = (Y, D, X)$, nuisance 参数 $\eta = (g, m)$ 。若在真实参数 (θ_0, η_0) 处满足正交条件: $\partial_{\eta} E[\psi = (W; \theta_0, \eta_0)] = 0$, 那么即使 η 的估计收敛速度较慢, 仍然可以得到 θ 的 \sqrt{n} -相合估计, 并构建有效的渐近置信区间。该方法具有双重稳健性: 只要 g_0 与 m_0 中有一个被正确设定, θ 的估计就具有一致性。通过这种多变量、多处理的 DML 分析, 我们能够为每个风险因素提供独立于其他混杂的因果效应估计, 从而与 ML-IPCW 及 Fine-Gray 模型的结果形成系统、多维度的对比, 深入评估不同方法在竞争风险数据中的表现差异。

3. 实例分析

3.1. 数据来源介绍与变量设定

本研究采用的数据集采用 R 软件包 `mstate` 中的 EBMT4 数据集, 该数据集包含接受异基因骨髓移植患者的随访资料。经过数据清洗和质量控制后, 最终纳入分析的患者共 2279 例, 其中复发事件 370 例 (16.8%), 死亡事件 533 例 (24.2%), 删失 1376 例 (59.0%)。设定感兴趣事件为移植后复发, 死亡为竞争事

件。并选取预防性治疗(rx)、移植年份(year)、年龄分层(agecl)及供者匹配情况(match)作为协变量。具体研究指标变量见表 1。为验证所提出的机器学习结合逆概率删失加权方法的有效性, 本文分别构建了随机森林(RF-IPCW)、极端梯度提升(XGBoost-IPCW)、支持向量机(SVM-IPCW)和多层感知机(MLP-IPCW)四种机器学习模型, 并以传统 Fine-Gray 模型与双机器学习作为参照进行比较分析。

Table 1. Variable definitions and meanings
表 1. 变量定义与意义

研究变量	定义	分类
time	事件时间	随访时间(Months/Days/Years)
status	事件类型	1: 目标事件(复发) 2: 竞争事件(死亡) 0: 删失
rx	预防性治疗	no (基线参照) yes
year	移植年份	<1990 (基线参照) 1990~1994 1995~1998
agecl	年龄分层	<20 岁(基线参照) 20~40 岁 >40 岁
match	供者匹配情况	full match (基线参照) gender mismatch

3.2. 主要结果

表 2 展示了各方法的回归系数估计结果。在所有 ML-IPCW 的方法中, 预防性治疗(rxyes)均显示出显著的保护效应, 其风险比(HR)在 1.733~2.331 之间, 这表示相对于对照组, 接受预防性治疗的复发风险约为未接受者的 1.7~2.3 倍, 即预防性治疗使复发风险增加了约 71%~131%。这一结果在统计学上高度显著($p < 0.001$)。传统 Fine-Gray 模型估计的预防性治疗效应相对较小($HR = 1.277, p = 0.047$), 且其他协变量如移植年份、年龄的效应估计与机器学习方法存在明显差异。机器学习方法一致显示较晚的移植年份(1990~1994、1995~1998)和较高的年龄(20~40 岁、>40 岁)与复发风险增加相关, Fine-Gray 模型未检出显著效应, 可能由于删失机制建模偏误导致的估计效率下降, 而 ML-IPCW 通过引入稳定化权重缓解了此问题。DML 得到的预防性治疗效应估计值($HR = 1.039, p = 0.050$)接近零效应且处于统计学显著性边界, 与其他方法相比明显更保守, 这可能是由于 DML 通过正交化过程更有效地控制了由倾向性得分捕捉的混杂因素, 减少了模型设定偏误。

为了进一步在视觉上验证表 2 中各模型(包括 Fine-Gray 模型)的实际风险分层能力, 本研究进行了可视化验证。对于每一种模型, 我们首先计算出其为每位患者预测的风险分数(ML-IPCW 的方法使用复发概率 $\hat{p}(\mathbf{x})$, Fine-Gray 模型使用线性预测值 $X\beta$)。然后根据该分数的中位数, 将患者群体分为“高风险”组和“低风险”组, 并分别在高低风险组内, 采用 Aalen-Johansen 非参数方法估计累积发生率函数(CIF)。最后分别针对两组患者, 利用竞争风险模型计算目标事件(即复发)的累积发生率及其 95%置信区间。

Table 2. Comparison of effect estimates of recurrence risk factors across different methods
表 2. 对比各方法对复发风险因素的效应估计

权重	变量	HR	95% CI (HR)	<i>p</i> 值
随机森林	预防性治疗(yes vs no)	2.310	(1.762~3.029)	<0.001
	移植年份(1990~1994 vs <1990)	5.492	(2.127~14.182)	<0.001
	移植年份(1995~1998 vs <1990)	4.810	(1.843~12.551)	0.001
	年龄分层(20~40 岁 vs <20 岁)	6.754	(1.358~33.598)	0.020
	年龄分层(>40 岁 vs <20 岁)	5.861	(1.177~29.196)	0.031
	供者匹配(gender mismatch vs full match)	1.726	(1.321~2.255)	<0.001
极限梯度提升	预防性治疗(yes vs no)	1.714	(1.315~2.234)	<0.001
	移植年份(1990~1994 vs <1990)	5.517	(2.155~14.125)	<0.001
	移植年份(1995~1998 vs <1990)	5.839	(2.251~15.146)	<0.001
	年龄分层(20~40 岁 vs <20 岁)	8.261	(1.651~41.340)	0.010
	年龄分层(>40 岁 vs <20 岁)	6.542	(1.308~32.728)	0.022
	供者匹配(gender mismatch vs full match)	1.580	(1.206~2.072)	<0.001
支持向量机	预防性治疗(yes vs no)	2.278	(1.738~2.986)	<0.001
	移植年份(1990~1994 vs <1990)	6.104	(2.362~15.778)	<0.001
	移植年份(1995~1998 vs <1990)	4.993	(1.919~12.993)	<0.001
	年龄分层(20~40 岁 vs <20 岁)	6.458	(1.298~32.135)	0.023
	年龄分层(>40 岁 vs <20 岁)	5.125	(1.028~25.549)	0.046
	供者匹配(gender mismatch vs full match)	1.534	(1.175~2.002)	0.002
多层感知机	预防性治疗(yes vs no)	1.718	(1.316~2.243)	<0.001
	移植年份(1990~1994 vs <1990)	5.466	(2.135~13.994)	<0.001
	移植年份(1995~1998 vs <1990)	5.883	(2.265~15.280)	<0.001
	年龄分层(20~40 岁 vs <20 岁)	8.325	(1.663~41.672)	0.010
	年龄分层(>40 岁 vs <20 岁)	6.509	(1.302~32.543)	0.023
	供者匹配(gender mismatch vs full match)	1.598	(1.218~2.097)	<0.001
Fine~Gray 模型	预防性治疗(yes vs no)	1.277	(1.003~1.626)	0.047
	移植年份(1990~1994 vs <1990)	1.153	(0.887~1.499)	0.290
	移植年份(1995~1998 vs <1990)	1.051	(0.790~1.398)	0.730
	年龄分层(20~40 岁 vs <20 岁)	0.892	(0.697~1.141)	0.360
	年龄分层(>40 岁 vs <20 岁)	0.845	(0.623~1.146)	0.280
	供者匹配(gender mismatch vs full match)	1.037	(0.817~1.315)	0.770
双机器学习	预防性治疗(yes vs no)	1.035	(0.996~1.076)	0.080
	移植年份(1990~1994 vs <1990)	1.023	(0.982~1.064)	0.273
	移植年份(1995~1998 vs <1990)	0.978	(0.939~1.019)	0.286
	年龄分层(20~40 岁 vs <20 岁)	0.985	(0.949~1.023)	0.442
	年龄分层(>40 岁 vs <20 岁)	0.967	(0.924~1.011)	0.141
	供者匹配(gender mismatch vs full match)	1.003	(0.968~1.039)	0.886

图 1 展示了基于上述分层计算的 CIF 曲线。结果显示, 在双机器学习(DML)方法中, 高风险组与低风险组的 CIF 曲线几乎完全重合, 其 95%置信区间存在显著重叠, 表明 DML 方法在本研究的竞争风险数据中风险分层能力有限。而在 RF-IPCW、XGBoost-IPCW 和 MLP-IPCW 三种机器学习模型中, “高风险”组(红色曲线)的累积复发率均显著高于“低风险”组(蓝色曲线), 且两组的 95%置信区间清晰分离, 具备更好的风险区分能力。

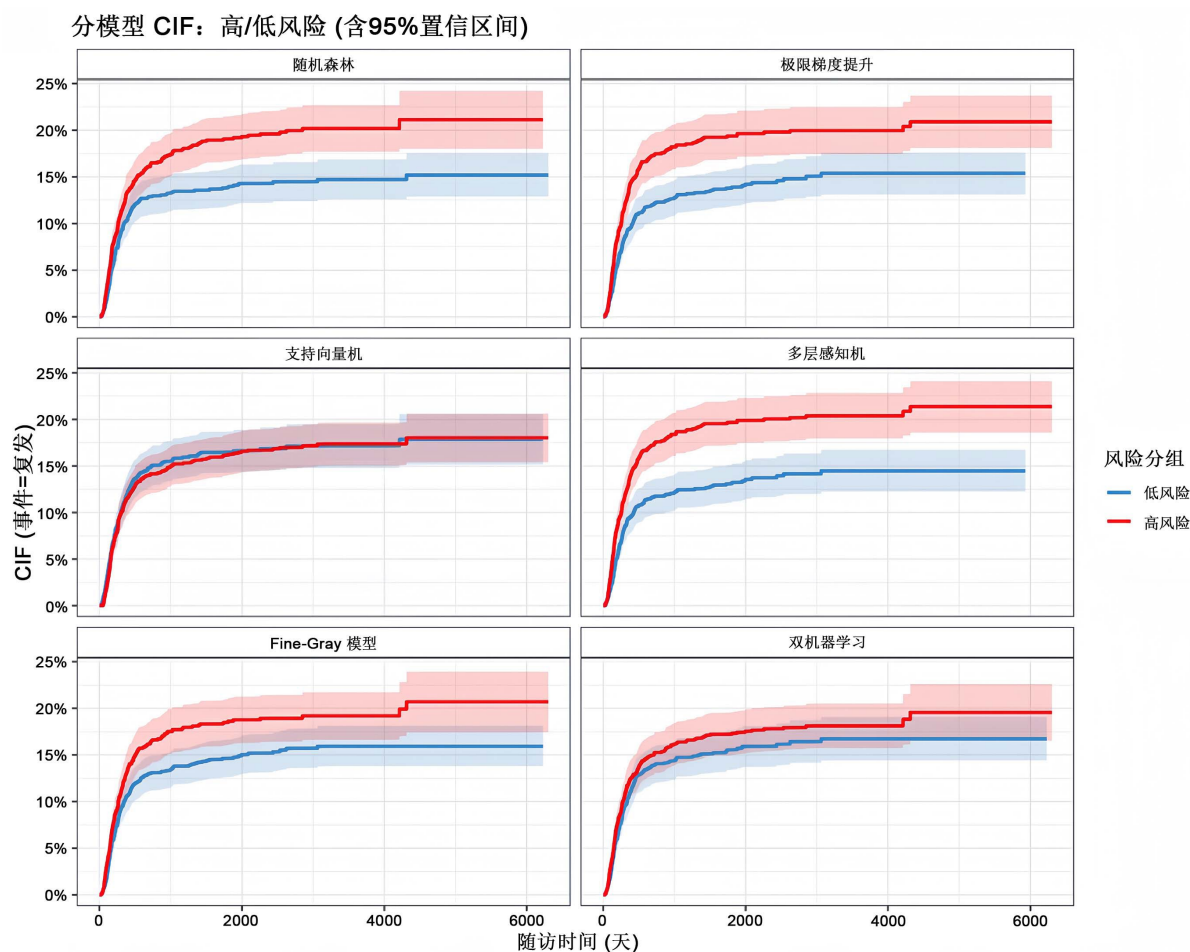


Figure 1. Comparison of CIF curves stratified by the median risk of each model

图 1. 按各模型风险中位数分层的 CIF 曲线对比

该结果直观印证了 ML-IPCW 方法及 Fine-Gray 模型在本研究数据中均具备稳健且有效的风险分层能力, 验证了表 2 中各模型风险估计的有效性。

3.3. 稳健性分析

为进一步验证结果的稳健性, 我们采用倾向得分调整方法进行敏感性分析。以治疗变量(rx)作为因变量, 以移植年份、年龄分层和供者匹配情况作为自变量构建倾向得分模型(表 3), 并将倾向得分作为新的协变量纳入分析。

图 2 展示了治疗组与对照组的倾向得分分布情况。如图所示, 治疗组的倾向得分均值(29.9%)显著高于对照组(22.3%), 两组分布存在部分重叠, 标准化均数差为 0.639 (95% CI: 0.541~0.737)。根据 Cohen 标

准(SMD 小于 0.1 为平衡极好, 0.1~0.2 为平衡良好, 0.2~0.5 为平衡一般, 大于 0.5 为平衡较差), 该 SMD 值说明治疗组与对照组在可观测协变量上存在明显差异, 反映了实际临床中治疗选择的非随机性。不过其分布的充分重叠仍满足倾向得分调整的基本前提。

Table 3. Logistic regression results of the propensity score model
表 3. 倾向得分模型逻辑回归结果

变量	系数	OR 值	95% CI (OR)	p 值
截距	-0.772	0.462	(0.357~0.594)	<0.001
移植年份(1990~1994 vs <1990)	-1.132	0.322	(0.255~0.407)	<0.001
移植年份(1995~1998 vs <1990)	-1.494	0.224	(0.172~0.292)	<0.001
年龄分层(20~40 岁 vs <20 岁)	0.521	1.684	(1.300~2.196)	<0.001
年龄分层(>40 岁 vs <20 岁)	0.862	2.367	(1.740~3.235)	<0.001
供者匹配(gender mismatch vs full match)	-0.041	0.960	(0.757~1.213)	0.735

注: 模型拟合指标: 样本量为 2279, 似然比检验 $\chi^2(5)=166.7$ ($p<0.001$), Nagelkerke $R^2=0.106$, C 统计量为 0.627。

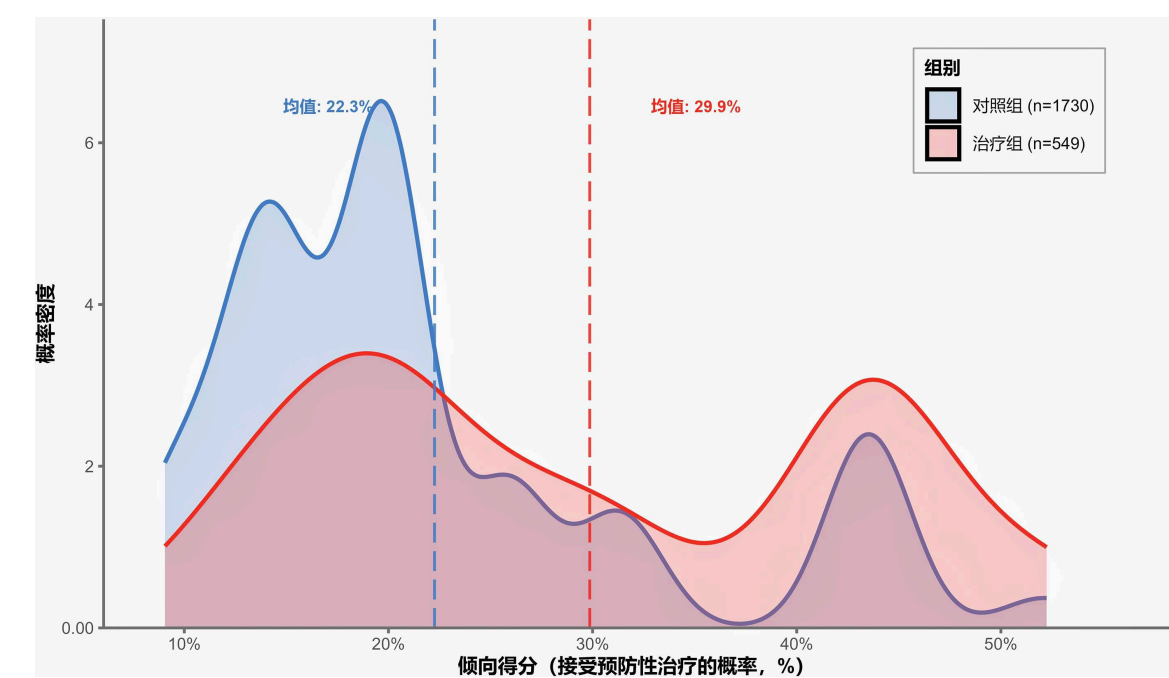


Figure 2. Distribution of propensity scores in the treatment and control groups
图 2. 治疗组与对照组的倾向得分分布

表 4 显示了倾向得分调整后的效应估计, 倾向得分调整后各机器学习方法得到的预防性治疗效应估计与主分析基本一致, HR 值在 2.218~2.961 之间, 均保持统计学显著性($p<0.001$)。倾向得分本身在所有机器学习方法中也显示出显著效应($p<0.001$), 表明其在模型中的重要性。相比之下, Fine-Gray 模型在倾向得分调整后得到的预防性治疗效应估计仍相对保守($HR=1.274, p=0.048$)。而 DML 的估计最为保守($HR=1.039, p=0.050$), 且其置信区间上限接近无效值 1, 提示经正交化调整后, 预防性治疗的效应进一步趋近于零。

Table 4. Comparison of effect estimates after propensity score adjustment across different methods
表 4. 对比各方法倾向得分调整后的效应估计

权重	变量	HR	95% CI (HR)	<i>p</i> 值
随机森林(rx + ps)	预防性治疗(yes vs no)	2.961	(2.223~3.945)	<0.001
	倾向性得分(PS)	0.291	(0.192~0.442)	<0.001
极限梯度提升(rx + ps)	预防性治疗(yes vs no)	2.416	(1.804~3.237)	<0.001
	倾向性得分(PS)	0.300	(0.205~0.439)	<0.001
支持向量机(rx + ps)	预防性治疗(yes vs no)	2.885	(2.160~3.852)	<0.001
	倾向性得分(PS)	0.281	(0.187~0.423)	<0.001
多层感知机(rx + ps)	预防性治疗(yes vs no)	2.218	(1.666~2.953)	<0.001
	倾向性得分(PS)	0.312	(0.212~0.459)	<0.001
Fine-Gray 模型(rx + ps)	预防性治疗(yes vs no)	1.274	(1.002~1.620)	0.048
	倾向性得分(PS)	0.919	(0.771~1.095)	0.340
双机器学习(rx + ps)	预防性治疗(yes vs no)	1.039	(1.000~1.081)	0.050

4. 结论

本研究提出了机器学习增强的逆概率删失加权(ML-IPCW)框架, 以提升竞争风险数据分析的统计功效。实例分析结果表明, ML-IPCW 方法揭示了更强、更显著的治疗保护效应, 并识别了多个传统模型未发现的关键风险因素。同时将 DML 作为对照框架, 经正交化调整后提供了更为保守的效应估计。本框架验证了机器学习在权重稳定化中的巨大潜力, 为竞争风险分析提供了一种更有效且实用的工具。未来研究可聚焦于构建双重稳健(Double Robust)的 ML-IPCW 估计量, 并进一步整合 DML 的正交化技术, 以获取更准确、更稳健的因果效应推断。

基金项目

本项目由 2022 年度辽宁省研究生教育教学改革研究项目(项目编号: LNYJG2022395)资助。

参考文献

- [1] Fine, J.P. and Gray, R.J. (1999) A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, **94**, 496-509. <https://doi.org/10.1080/01621459.1999.10474144>
- [2] Robins, J.M., Hernán, M.Á. and Brumback, B. (2000) Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, **11**, 550-560. <https://doi.org/10.1097/00001648-200009000-00011>
- [3] Lee, B.K., Lessler, J. and Stuart, E.A. (2009) Improving Propensity Score Weighting Using Machine Learning. *Statistics in Medicine*, **29**, 337-346. <https://doi.org/10.1002/sim.3782>
- [4] Kvamme, H., Borgan, Ø. and Scheel, I. (2019) Time-to-Event Prediction with Neural Networks and Cox Regression. *Journal of Machine Learning Research*, **20**, 1-30.
- [5] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., *et al.* (2018) Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal*, **21**, C1-C68. <https://doi.org/10.1111/ectj.12097>
- [6] Stensrud, M.J., Young, J.G., Didelez, V., Robins, J.M. and Hernán, M.A. (2020) Separable Effects for Causal Inference in the Presence of Competing Events. *Journal of the American Statistical Association*, **117**, 175-183. <https://doi.org/10.1080/01621459.2020.1765783>
- [7] Cole, S.R. and Hernan, M.A. (2008) Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*, **168**, 656-664. <https://doi.org/10.1093/aje/kwn164>

-
- [8] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
 - [9] Wager, S. and Athey, S. (2018) Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. *Journal of the American Statistical Association*, **113**, 1228-1242. <https://doi.org/10.1080/01621459.2017.1319839>
 - [10] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
 - [11] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1023/a:1022627411411>
 - [12] Platt, J. (1999) Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in Large Margin Classifiers*, **10**, 61-74.
 - [13] Hornik, K., Stinchcombe, M. and White, H. (1989) Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*, **2**, 359-366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)