arXiv:2602.05022v1 [stat.ME] 4 Feb 2026

双重变量重要性匹配：估计事件概率和发生时间的不同因果效应

Yuqi Li[1] Quinn Lanners[1] Matthew M. Engelhard[1]

[1]杜克大学生物统计与生物信息学系

摘要

在许多临床情境中，估计治疗对生存时间数据的影响，不仅会受到混杂、删失和异质性的影响，还会因存在一个永远不会发生感兴趣事件的治愈亚群而变得复杂。在此类情境下，治疗可能分别对（1）治愈概率和（2）未治愈个体的事件发生时间产生不同影响。标准生存分析和因果推断方法通常不会区分已治愈与未治愈个体，因此难以揭示治疗对治愈概率和事件发生时间的不同作用机制。为应对这些挑战，我们提出一种基于匹配的框架，通过构建不同的匹配组，分别估计治愈概率和事件发生时间的异质性治疗效应（HTE）。我们使用混合治愈模型识别这两个估计目标的特征重要性，并据此为高维空间中的匹配构建加权距离度量。在匹配组内，Kaplan–Meier估计量用于估计治愈概率和预期事件发生时间，并据此推导个体层面处理效应。我们在等尺度约束下，为估计量一致性和距离度量最优性提供理论保证。此外，我们还将估计误差分解为删失、模型拟合和不可约噪声所造成的部分。模拟和真实世界数据分析表明，我们的方法能够在事件发生时间情境中提供易于解释且稳健的异质性处理效应估计值。

1 引言

治疗对临床结局的影响通常可分为短期影响和长期影响。在孕产妇健康领域，分娩后开具预防性阿奇霉素可能会延缓感染，但最终无法预防感染（Tita等，2016）。在心理健康领域，心理治疗可能降低心理健康问题的发生率，但矛盾的是，也可能加快对已存在问题的识别（Jorm等，2017）。在早期黑色素瘤中，免疫治疗可能使部分患者获得长期缓解（即治愈），但对其他患者而言，可能只是延缓复发（Michielin等，2020；Boutros等，2023）。在这些情况以及许多其他情况下，我们可能希望区分治疗对以下两方面的影响：（a）感兴趣事件发生的长期概率；（b）如果该事件发生，其短期发生时间。例如，在向患者提供预后信息以帮助其作出治疗决策，或探究潜在治疗机制时，明确区分这些量至关重要。

在生存分析文献中，可以通过假设存在一个治愈亚群 来对短期效应与长期（甚至无限期）效应之间的区别进行建模；在该亚群中，感兴趣事件永远不会发生（Farewell，1982）。

因此，我们可以区分治疗对以下方面的影响：(a) 治愈概率（即长期效应），以及 (b) 未治愈患者的事件发生时间分布（即短期效应）。如图1所示，从概念上讲，治疗能够且通常确实会带来长期

获益，即使在事件发生时，治疗也可能缩短预期事件发生时间。然而，传统生存分析框架及其相关因果方法通常会混淆这两种效应，将所有删失个体都视为最终可能发生事件。忽视这一区别会导致对治疗效果的理解不够准确，进而妨碍我们理解治疗机制并选择最佳临床行动。

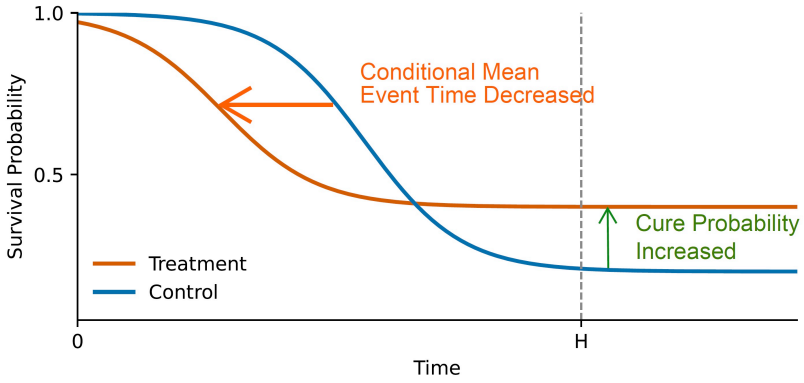

图1：假设的生存曲线：治疗提高了治愈概率，却缩短了条件平均事件时间。

在本研究中，我们旨在估计治疗对事件发生时间结局的这两个临床上不同组成部分的异质性治疗效应（HTE）。为明确界定这些效应，我们设定一个具有临床意义的时间范围 H，在此期间可能发生相关事件。在 H 之前未发生事件的患者被视为已治愈（即未发生事件）。H 之后发生的事件可以忽略，因为它们超出了与原始病症及其相关治疗有关的有限时间窗口。在这一框架下，我们定义两项治疗效应估计目标：（i）治愈概率；（ii）发生事件者的期望事件时间。

随后，我们提出一种基于匹配的方法，用于估计这两类异质性治疗效应。匹配方法因概念简单、透明，在因果推断中得到广泛应用，因为它们与随机研究的设计十分相似（Stuart，2010）。这些优势在临床研究等高风险应用中尤为重要，因为此类研究中常见生存时间数据。在我们的研究设定中，我们针对两个估计目标分别构建匹配组，并认识到事件概率和事件发生时间可能取决于不同的协变量。一项关键挑战在于，观察性数据通常具有高维特征，且包含许多无关或含噪特征。近期研究在保留可解释性的同时应对了这一问题，具体做法是使用辅助结局模型学习距离度量，并提高与结局关联最密切的协变量的权重（Lanners等人，2023；Parikh等人，2022）。我们将这一方法调整用于当前设定，通过学习两种不同的距离度量，分别针对与治愈状态和事件发生时间最相关的协变量进行匹配。我们拟合一个混合治愈模型（Farewell，1982），其似然联合纳入有关治愈机制和事件时间分布的信息。随后，我们提取针对各自关注效应的特征重要性，并据此构建两种专门用于匹配的距离度量。

利用这些距离度量，我们实施双重匹配程序来估计异质性治疗效应。对于每位个体，我们根据从混合治愈模型中学习到的两种不同距离度量，分别构建两个匹配组。每种距离度量都近似于其对应目标估计量的最优协变量加权，从而提高匹配质量并减少估计误差。在每个治疗组内，我们使用 Kaplan-Meier估计量估计时间 H 时的生存概率（可解释为治愈概率），以及在

事件发生于 H 之前这一条件下的平均事件时间。匹配后的治疗组与对照组之间的差异可得出易于解释的异质性治疗效应估计。我们还表明，在标准假设下，所提出的估计量对这两种因果效应均具有一致性。

我们的贡献包括：(1) 提出一种新颖的匹配框架，分别估计治愈概率和事件发生时间上的异质性治疗效应；(2) 基于辅助混合治愈模型中的协变量重要性，构建用于匹配的治愈特异性和时间特异性距离度量；(3) 证明在等尺度约束下，这些距离度量对于异质性处理效应估计是最优的；(4) 证明在标准假设下，匹配组内的 Kaplan-Meier估计量能够一致地估计个体特异性生存曲线；(5) 在多种模拟设置和一项真实案例研究中验证我们的方法；(6) 将异质性处理效应估计误差分解为删失、模型拟合和不可约噪声所导致的部分，从而深入揭示估计不确定性的来源和局限。

2 文献综述

随着大规模数据集日益普及，人们对异质性治疗效应估计的关注不断增加（Wager和Athey，2018；Knaus等人，2021），相应地也出现了众多回归、加权和机器学习方法（Yao等人，2021）。然而，由于这些方法依赖无法检验的假设，近期研究开始强调可解释的方法，以便领域专家对其进行审查（Shikalgar等人，2025；Parikh等人，2023）。在这方面，匹配方法尤其具有吸引力，因为其概念简单，有利于将专家判断纳入分析流程（Stuart，2010；Yu等人，2021a）。

尽管倾向评分匹配（Rosenbaum和Rubin，1983年）和预后评分匹配（Hansen，2008）仍是应用最广泛的方法，但为提升治疗效应估计的可解释性和准确性，越来越多近似精确匹配（AME）框架应运而生（Yu等人，2021b；Diamond和Sekhon，2013；Wang等人，2021；Dieng等人，2019年；Morucci等人，2020年；Katta等人，2024）。与我们采用的方法最接近的是学习拉伸后匹配（MALTS）（Parikh等人，2022）和变量重要性匹配（VIM）；这两种方法都使用拉伸距离度量，对重要变量赋予更高权重。我们的方法将VIM扩展到包含无事件亚群的事件发生时间情境。

基于生存时间数据进行因果推断的匹配方法 在生存分析情境中，匹配方法尤其具有吸引力，因为删失和随访时间不等会使估计更加复杂。在此背景下，治疗效应可以针对多个估计目标进行定义，例如生存曲线、风险比，或受限均值生存时间、中位生存时间等标量汇总指标（Bewick等人，2004）。研究者已开发出匹配方法的扩展形式，以估计上述各类目标量：生存曲线（Austin，2014年；Xu等人，2023）、风险比（Austin和Stuart，2015；Lee等人，2020年），以及受限均值生存时间或中位生存时间（Lin等人，2023；Yang等人，2019年；Jiang等人，2024）。这些方法包括倾向评分匹配等广泛使用的框架（Austin，2014年；Lu等人，2018；Xu等人，2023；Lin等人，2023），以及充分核匹配（Jiang等人，2024）、最优全匹配（Austin和Stuart，2015；Lee等人，2020年）和粗化精确匹配（Yang等人，2019年）等较新的创新方法。

大多数针对生存数据的匹配方法都侧重于总体层面的估计目标。为估计条件处理效应，人们开发了基于建模的方法，例如 Cox 回归、AFT BART 和目标最大似然估计（Ashouri and Henderson, 2024; Zhu and Gallego, 2020; Austin, 2014）。尽管这些方法能够估计条件处理效应，但容易受到模型设定错误的影响，且透明度低于基于匹配的方法，因此更难审查，这限制了它们在高风险情境中的可信度。

存在治愈亚群的生存分析 混合治愈模型和促发时间治愈模型用于处理包含治愈亚群的生存时间数据，二者对治愈群体的表征方式不同（Amico and Van Keilegom, 2018; Boag, 1949; Berkson and Gage, 1952; Yakovlev et al., 1996; Chen et al., 1999; Tsodikov, 1998）。数十年来，人们一直在研究这些模型；近期研究则通过半参数方法和机器学习方法放宽了参数假设（Farewell, 1982; Felizzi et al., 2021; Engelhard and Henao, 2022; Musta et al., 2022）。我们聚焦于混合治愈框架，因为它与我们关注的估计目标更为直接契合。

关于使用含有治愈亚群的生存时间数据进行因果推断，现有研究有限。Gao和Zheng（2017）使用促发时间治愈模型，估计存在不依从情况的随机对照试验中的治疗效应。与我们的目标最为接近的是Sun和Song（2025）以及Wang等人（2024）的混合治愈方法。Sun和Song（2025）提出将贝叶斯因果森林与AFT混合治愈模型结合起来的方法，而Wang等人（2024）则为这一任务提出了半参数估计量。尽管这些基于建模的方法效果良好，但与基于匹配的方法相比，其结果透明度较低，也更难向领域专家阐释。明确考虑治愈亚群并同时关注治愈概率与未治愈者生存情况的基于匹配的方法，仍未得到充分探索。

3 设置与估计目标

考虑一个数据集 $\mathcal{D}$，由 n 名受试者组成，标记为 $i = 1$、.....、n。对于每名受试者 i，令 Z_i 表示其处理状态，其中 $Z_i = 1$ 表示积极治疗，$Z_i = 0$ 表示对照。每名受试者都有一个事件时间 T_i，并受到 C_i 的右删失。我们还观测到一个由 $\boldsymbol{X}_i \in \mathcal{X} \subseteq \mathbb{R}^d$ 表示的 d 治疗前协变量向量，这些协变量在受试者间独立同分布。

为定义因果目标估计量，我们采用潜在结果框架（Rubin，1980）。在该框架下，每位受试者有两个潜在失败时间（$T_i(0)$ $T_i(1)$）和两个潜在删失时间（$C_i(0)$ $C_i(1)$），分别对应两种治疗。由于每位受试者只能观察到一个潜在结局，因此个体层面处理效应无法识别。相反，我们关注这些事件发生时间结局的异质性治疗效应（HTE）。

我们将治疗 z 下的潜在生存概率函数定义为 $S_z(t|\boldsymbol{x}) = \Pr(T(z) > t|\boldsymbol{X} = \boldsymbol{x})$，$\forall \boldsymbol{x} \in \mathcal{X}$。一个常见的因果目标估计量是时间 t时生存率的条件异质性处理效应：

$$\tau(t|\boldsymbol{x}) = S_1(t|\boldsymbol{x}) - S_0(t|\boldsymbol{x}), \quad 0 \le t \le t_{\max}, \tag{1}$$

其中，$t_{\max}$ 为最大随访时间。该估计目标很容易解释为两种治疗组之间的因果风险差。然而，由于它是时间 t的函数，估计 $\tau(t|\boldsymbol{x})$颇具挑战，尤其是在 t 值较大、事件较少发生的情况下。

为克服这些困难，我们设定预先指定的时间阈值 H，并仅关注在 H 之前发生事件的事件时间。这种方法符合许多临床研究中的常见做法，因为 (a) 药物及其他治疗仅在有限时间内有效，且 (b) 无限期随访并不实际。由于可能存在删失，且设定了时间范围 H，我们只能观察到事件时间 $Y_i = \min\{T_i, C_i, H\}$ 以及事件指示变量$\delta_i = \mathbb{I}(T_i < C_i,\ T_i < H)$。

在此设置下，我们定义一个潜变量 $E_i(z) = \mathbb{I}(T_i(z) < H)$，其中 $E_i(z) = 0$ 表示个体已治愈，而 $E_i(z) = 1$ 表示个体未治愈 （二者均指接受治疗z的情况）。由此，我们可以定义以下两个关注的估计目标：（i）治愈**概率**上的 HTE：

$$\pi(\boldsymbol{x}) = \Pr(E_i(1) = 0 \mid \boldsymbol{X} = \boldsymbol{x}) - \Pr(E_i(0) = 0 \mid \boldsymbol{X} = \boldsymbol{x}), \tag{2}$$

以及（ii）条件平均事件时间（CMET）上的 HTE,

$$\begin{aligned}\Delta(\boldsymbol{x}) = \mathbb{E}[T_i(1) \mid T_i(1) < H, \boldsymbol{X} = \boldsymbol{x}] \\ - \mathbb{E}[T_i(0) \mid T_i(0) < H, \boldsymbol{X} = \boldsymbol{x}],\end{aligned} \tag{3}$$

即以非治愈者为条件。下文集中估计 $\pi(\boldsymbol{x})$ 和 $\Delta(\boldsymbol{x})$。

3.1 假设与识别

我们采用稳定单元处理值假设（SUTVA）（Rubin，1980），该假设由一致性和无干扰假设构成。此外，我们假设以下用于识别的标准条件：无混杂性、正值性和非信息性删失（Imbens 和 Rubin，2015）。在这些假设下，以下命题表明，$\pi(x)$ 和 $\Delta(x)$ 均可通过条件生存概率函数 $S_z(t \mid x) = \Pr(T > t \mid Z = z,\ \boldsymbol{X} = \boldsymbol{x})$ 识别。假设和推导的细节见补充材料。

命题1 如果一致性、无混杂性、正值性和非信息性删失假设成立，则这两个估计目标可表示为

$$\pi(\boldsymbol{x}) = S_1(H \mid \boldsymbol{x}) - S_0(H \mid \boldsymbol{x}), \tag{4}$$

$$\begin{aligned}\Delta(\boldsymbol{x}) = \frac{\int_0^H S_1(t|\boldsymbol{x})\,dt - HS_1(H|\boldsymbol{x})}{1 - S_1(H|\boldsymbol{x})} \\ - \frac{\int_0^H S_0(t|\boldsymbol{x})\,dt - HS_0(H|\boldsymbol{x})}{1 - S_0(H|\boldsymbol{x})}.\end{aligned} \tag{5}$$

命题2表明，估计 $\pi(\boldsymbol{x})$ 和 $\Delta(\boldsymbol{x})$ 的关键在于 $z \in \{0, 1\}$ 和 $t \in [0, H]$ 的生存概率函数 $S_z(t \mid \boldsymbol{x})$。这一特征使我们能够通过对匹配组应用 Kaplan–Meier估计量，对目标估计量进行非参数估计，具体方法见第 4.3 节。

3.2 混合治愈模型

混合治愈模型由两个部分组成，分别用于建模 (a) 治愈概率和 (b) 事件时间分布。指定这两个模型的形式后，我们可以通过最大似然估计来估计模型参数。鉴于我们只能观测到时间 Y和事件指示变量 δ，因此在混合治愈模型中，我们考虑密度 $p(y,\ \delta|\boldsymbol{x})$，该密度是治愈状态与事件时间信息的混合，分别对应未治愈个体与已治愈个体。密度和生存函数可由混合治愈模型的两个部分表示，详细推导见补充材料。

在预先指定的时间范围 H下，由混合治愈方法建模的生存函数可以写为:

$$\begin{aligned}S(t) &= \Pr(cure) + [1 - \Pr(cure)]S_N(t) \\ &= \Pr(E = 0) + [1 - \Pr(E = 0)]\Pr(T > t|T \le H) \\ &= \Pr(T > H) + \Pr(T \le H)\Pr(T > t|T \le H) \\ &= \Pr(T > t)\end{aligned} \tag{6}$$

其中，$\Pr(cure)$ 是治愈概率，而在患者未治愈的条件下，$S_N(t)$ 是生存函数。

密度函数可概括如下:

$$
\begin{aligned}
p(y,\delta|\boldsymbol{x}) &= p(y,\delta=1|\boldsymbol{x}) + p(y,\delta=0|\boldsymbol{x}) \\
&= p(E=1|\boldsymbol{x})\mathbb{I}(y<H)\frac{f_T(y|\boldsymbol{x})}{F_T(H|\boldsymbol{x})}S_C(y|\boldsymbol{x}) \\
&+ p(E=1|\boldsymbol{x})\mathbb{I}(y<H)\left(1-\frac{F_T(y|\boldsymbol{x})}{F_T(H|\boldsymbol{x})}\right)f_C(y|\boldsymbol{x}) \\
&+ p(E=0|\boldsymbol{x})\mathbb{I}(y<H)f_C(y|\boldsymbol{x}) \\
&+ p(E=0|\boldsymbol{x})\mathbb{I}(y=H)S_C(y|\boldsymbol{x})
\end{aligned}
\tag{7}
$$

其中， $f_T(y|\boldsymbol{x})$ 和 $F_T(H|\boldsymbol{x})$ 分别是事件时间分布的概率密度函数 和累积分布函数 ， 而 $f_C(y|\boldsymbol{x})$ 和 $S_C(y|\boldsymbol{x})$ 分别是删失时间分布的概率密度函数 和*1 -* 累积分布函数 。在非信息性删失假设下，删失项是干扰成分。

参数模型通常易于拟合，并能得到闭式的 $\pi(\boldsymbol{x})$ 和 $\Delta(\boldsymbol{x})$，但高度依赖于正确的模型设定。相比之下，Kaplan-Meier估计量是一种假设最少的非参数方法，但无法调整协变量。因此，我们提出一个结合参数方法与非参数方法的框架，该框架能降低估计误差，并对模型设定错误具有稳健性。

4 方法论

我们提出一种基于匹配的方法，用于对包含治愈亚群的生存时间数据进行因果推断。与类似的匹配方法（Parikh等人，2022；Lanners等人，2023）一样，我们的方法通过学习距离度量来构建匹配组，以估计治疗效应。我们扩展了这一框架，利用混合治愈模型推导出两种不同的度量，一种针对治愈概率，另一种针对条件平均事件时间（CMET），从而能够针对每个目标估计量进行匹配。我们的方法总结于算法1中，以下三个小节将详细介绍各个步骤。

4.1 拟合混合治愈模型

为便于进行诚实的因果推断，我们首先将数据集划分为训练集（$\mathcal{D}_{\text{tr}}$）和估计集（$\mathcal{D}_{\text{est}}$）（Ratkovic，2019年）。仅使用 $\mathcal{D}_{\text{tr}}$ 中的样本，对每个治疗组$Z \in \{0,1\}$，我们拟合一个包含两个部分的混合治愈模型:

$$
P(\text{cure}|\mathbf{X}) = \sigma(\beta_0 + \mathbf{X}\boldsymbol{\beta}), \quad \sigma(z) = \tfrac{1}{1+e^{-z}}, \tag{8}
$$

$$
\log(T) \sim N(\lambda_0 + \mathbf{X}\boldsymbol{\lambda}, \sigma^2), \tag{9}
$$

其中，方程8指定治愈概率，方程9指定事件时间分布的加速失效时间模型。我们通过在训练集上最大化似然 $\mathcal{L} = \sum_{i=1}^{N} \log\ p(y_i,\ \delta_i|\boldsymbol{x_i})$ 来估计模型参数，所得系数 $\boldsymbol{\beta}$，$\boldsymbol{\lambda}$ 概括了两个结果组成部分的变量重要性。

4.2 学习距离度量

我们使用估计系数来量化特征重要性，并构造加权欧氏距离:

$$
\begin{aligned}
d_{\text{cure}}(\mathbf{x},\mathbf{y}) &= \sqrt{(\mathbf{x}-\mathbf{y})^\top \mathbf{W}_{\text{cure}}(\mathbf{x}-\mathbf{y})},\\
d_{\text{time}}(\mathbf{x},\mathbf{y}) &= \sqrt{(\mathbf{x}-\mathbf{y})^\top \mathbf{W}_{\text{time}}(\mathbf{x}-\mathbf{y})},\\
\mathbf{W}_{\text{cure}} &= \frac{1}{2}\,\mathrm{diag}\big(|\boldsymbol{\beta}_1| + |\boldsymbol{\beta}_0|\big),\\
\mathbf{W}_{\text{time}} &= \frac{1}{2}\,\mathrm{diag}\big(|\boldsymbol{\lambda}_1| + |\boldsymbol{\lambda}_0|\big),
\end{aligned}
\tag{10}
$$

其中， $|\boldsymbol{\beta}_z|$ 和 $|\boldsymbol{\lambda}_z|$ 分别是治疗组 $Z=z$中混合治愈模型的治愈部分和时间部分的估计系数绝对值向量，diag(·)表示对角矩阵。这些权重矩阵会提高与相应结局关联更强的协变量的重要性，从而近似实现对影响最大的特征进行精确匹配。

4.3 匹配与异质性处理效应估计

对于估计集 $\mathcal{D}_{\text{est}}$ 中的每个单位 i ，我们首先分别使用治愈距离度量和时间距离度量，从处理组（M_1）和对照组（M_0）中寻找 K 个最近邻，以确定匹配集。然后，我们基于这些匹配集，使用 Kaplan–Meier 估计量估计异质性处理效应。

根据第 3.1 节，治愈概率上的异质性处理效应是利用通过治愈距离 d_{cure} 识别的匹配组估计的：

$$
\pi(\mathbf{x}_i) = \hat{S}_{M_1}(H) - \hat{S}_{M_0}(H),
\tag{11}
$$

ˆ其中 S 表示 Kaplan–Meier 估计量。

同样，条件平均事件时间上的异质性处理效应是利用通过时间距离 d_{time} 识别的匹配组估计的：

$$
\begin{aligned}
\Delta(\boldsymbol{x}_i) = &\frac{\int_0^H \hat{S}_{M_1}(t)\,dt - H\hat{S}_{M_1}(H)}{1-\hat{S}_{M_1}(H)}\\
&- \frac{\int_0^H \hat{S}_{M_0}(t)\,dt - H\hat{S}_{M_0}(H)}{1-\hat{S}_{M_0}(H)}.
\end{aligned}
\tag{12}
$$

5 理论结果

本节将为我们提出的框架建立理论保证。首先，我们证明，在标准假设下，治愈概率和事件发生时间的匹配估计量均具有一致性。我们还在线性模型假设下刻画了匹配所用的最优距离度量。证明见补充材料。

5.1 匹配估计的一致性

我们首先证明所提出的用于治愈效应和时间效应的匹配估计量的一致性。回顾公式18和19中的目标估计量形式。我们需要证明ˆ$S(H|\boldsymbol{x})$和 $\int_0^H$ˆ$S(t|\boldsymbol{x})\,\mathrm{d}t$的一致性。令 $\{\boldsymbol{X}_{i_1},...,\boldsymbol{X}_{i_K}\} \subset \{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\}$ 表示大小为$\backsim K$的匹配组，该组由根据度量 $d(\cdot,\ \cdot)$选出的 K 个 $\boldsymbol{x}$ 的最近邻构成。令 S_{Kn} 表示基于 $\{(Y_{i_k},\ \delta_{i_k})\}_{k=1}^K$的Kaplan–Meier估计量。

算法1 用于异质性处理效应估计的双变量重要性匹配

要求： 数据 $\mathcal{D}=\{\mathbf{X}_i,\ Y_i,\ \delta_i,\ Z_i\}_{i=1}^N$，时间范围。 H

确保： 治愈和生存时间的异质性处理效应估计值

1: **划分数据**：将 $\mathcal{D}$ 划分为训练集 $\mathcal{D}_{\text{tr}}$ （35%）和估计集 $\mathcal{D}_{\text{est}}$ （65%）2: **训练模型**：

- 分别在 $\mathcal{D}_{\text{tr}}(Z=0)$ 和 $\mathcal{D}_{\text{tr}}(Z=1)$ 上拟合混合治愈模型
- 提取系数 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_0$（治愈模型）和 $\boldsymbol{\lambda}_1$，$\boldsymbol{\lambda}_0$（时间模型）

3: **构建权重矩阵**：

- $\mathbf{W}_{\text{治愈}} \leftarrow \frac{1}{2}$诊断$\big(|\boldsymbol{\beta}_1|+|\boldsymbol{\beta}_0|\big)$
- $\mathbf{W}_{\text{时间}} \leftarrow \frac{1}{2}$diag$\big(|\boldsymbol{\lambda}_1|+|\boldsymbol{\lambda}_0|\big)$

4: **匹配与估计**：

- 对于每个 $i\in\mathcal{D}_{\text{估计}}$：1. 使用 $d_{\text{治愈}}$ 和 $d_{\text{时间}}$ 分别估计治愈效应和时间效应，构建匹配组 M_1，M_0 。2. 利用匹配组的 Kaplan–Meier 估计值，并将其代入公式 36 和 12 中的估计量，计算治愈概率和条件平均事件时间的异质性治疗效应。

5: **返回**：所有估计单位的异质性处理效应估计值。

定义两个**泛函** $F(S):=S(H)$ 和 $G(S):=\int_0^H S(u)\,\mathrm{d}u$。估计量的一致性 $\widehat{\ }$ $\widehat{\ }$ 归结为证明 $F(S_{Kn})\to F(S(\cdot\mid\boldsymbol{x}))$ 和 $G(S_{Kn})\to G(S(\cdot\mid\boldsymbol{x}))$。

定理 1. </style}令 属于 的支撑集，即，对于任意 $\epsilon>0$，$\Pr(d(\boldsymbol{X},\ \boldsymbol{x})<\epsilon)>0$。假设$S(t\mid\boldsymbol{x})$ 关于 $\boldsymbol{x}$ 是局部 Lipschitz 连续的，且对 t一致，即，存在 $\boldsymbol{x}$的一个邻域，记为 $\mathcal{N}_{\boldsymbol{x}}$，以及一个常数 $\tau>H$，使得对所有属于 $\mathcal{N}_{\boldsymbol{x}}$ 的 $\boldsymbol{x}'$，$\boldsymbol{x}''$ 以及 $t\in[0,\tau]$，$|S(t\mid\boldsymbol{x}')-S(t\mid\boldsymbol{x}'')|\le L\cdot d(\boldsymbol{x}',\ \boldsymbol{x}'')$，其中某个有限常数为 L。那么，

$$F(\widehat{S}_{Kn})\xrightarrow{p}F(S(\cdot\mid\boldsymbol{x})),\quad G(\widehat{S}_{Kn})\xrightarrow{p}G(S(\cdot\mid\boldsymbol{x})),$$

当 K，$n\to\infty$ 且 $K=o(n)$ 时，其中 $\xrightarrow{p}$ 表示依概率收敛。

由 S_{Kn} 的一致性已得到确立，以下推论可直接得出。

推论 1. 估计量中定义的两个量 36 和 12，它们是 $\widehat{\ }$ $\widehat{\ }$ $F(S_{Kn})$ 和 $G(S_{Kn})$ 的函数，也是一致的。

5.2 最优距离度量

用于匹配的距离度量应根据特征对所关注估计目标的影响进行自适应加权，从而将局部扰动下的最坏情况差异降至最低。我们通过限制加权欧氏距离下的 Lipschitz 常数对此进行形式化。假设采用线性混合治愈模型，我们推导出用于基于匹配的治愈效应和时间效应估计的最优权重。

设置。 对于对角权重矩阵 $W=\text{diag}(w_1,\ \ldots,\ w_p)$，其中 $w_j\ge 0$ 和 $\sum_{j=1}^p w_j=c>0$，定义加权范数 $\|\boldsymbol{v}\|_W=(\sum_{j=1}^p w_j v_j^2)^{1/2}$ 以及距离 $d_W(\boldsymbol{x},\ \boldsymbol{x}')=\|\boldsymbol{x}-\boldsymbol{x}'\|_W$。

则相应的对偶范数为 $\|\boldsymbol{g}\|_{W^{-1}} = (\sum_{j=1}^{p} g_j^2/w_j)^{1/2}$，约定当 $w_j = 0$ 和 $g_j = 0$时， $g_j^2/w_j =$为 0。对于目标函数 $f(\cdot)$，其在 d_W 下的 Lipschitz 常数为：

$$L(W) = \sup_{x \neq x'} \frac{|f(\boldsymbol{x}) - f(\boldsymbol{x}')|}{d_W(\boldsymbol{x}, \boldsymbol{x}')}. \tag{13}$$

定理 2a（最优治愈权重）。对于治愈概率 $S(H \mid \boldsymbol{x},\ z) = \sigma(\beta_{z0} + \boldsymbol{\beta}_z^\top \boldsymbol{x})$，Lipschitz 常数满足

$$L_{\text{cure}}(W) \leq \tfrac{1}{4} \|\boldsymbol{\beta}_z\|_{W^{-1}}. \tag{14}$$

在等尺度约束 $\sum_{j=1}^{p} w_j = c$ 下，使偏差界最小化的治愈权重矩阵为

$$\begin{aligned} W_{\mathrm{c},z} &= \operatorname{diag}\left(\frac{c\,|\beta_{z1}|}{\sum_{k=1}^{p} |\beta_{zk}|}, \ldots, \frac{c\,|\beta_{zp}|}{\sum_{k=1}^{p} |\beta_{zk}|} \right) \\ &= \operatorname{diag}|\boldsymbol{\beta}_z|. \end{aligned} \tag{15}$$

定理 2b（最优时间权重）。设 $\eta(\boldsymbol{x}) = \lambda_0 + \boldsymbol{\lambda}_z^\top \boldsymbol{x}$ 并假设 $(\log T \mid \boldsymbol{X} = \boldsymbol{x}) \sim \mathcal{N}(\eta(\boldsymbol{x})\ \sigma^2)$。将未治愈者期望事件时间记为 $m(\boldsymbol{x}) = \mathbb{E}[T \mid T < H, \boldsymbol{X} = \boldsymbol{x}]$。那么存在一个有限常数 C_{time} 使得

$$L_{\text{time}}(W) \leq C_{\text{time}} \|\boldsymbol{\lambda}_z\|_{W^{-1}}. \tag{16}$$

在等尺度约束 $\sum_{j=1}^{p} w_j = c$ 下，使偏差界最小化的时间权重矩阵为

$$W_{\mathrm{t},z} = \operatorname{diag}|\boldsymbol{\lambda}_z|. \tag{17}$$

在实践中，必须对治疗组和对照组统一定义距离度量，因为匹配需要采用共同规则。尽管平均权重 10 在各治疗组内并非单独最优，但取平均提供了一种自然的折中。由此得到的偏差界仍受到控制（通过各治疗组中的最大值），且一致性得以保留。

6 模拟实验

6.1 设置

我们使用治愈模型和事件时间模型模拟数据，其中治疗会影响治愈概率和事件发生时间。治疗效应可能存在异质性，并随协变量 $\boldsymbol{X} \in \mathbb{R}^p$而变化；这些协变量包括从 Bern(0.5) 和 N(0, 1) 分别抽取的 $p_{\text{bin}} \leq p$ 二元特征和 $p_{\text{cont}} \leq p$ 连续特征。治疗分配遵循逻辑斯蒂模型： $P(Z = 1 \mid \boldsymbol{X})$ =expit($\boldsymbol{X}^\top \beta_Z$)。针对每个治疗组分别对治愈状态建模： $P(E = 1 \mid Z = z, \boldsymbol{X})$ = expit($\boldsymbol{X}^\top \beta_{E_z}$)。对于未治愈个体（$E = 1$），对数事件时间服从：$(\log\ T \mid Z = z, \boldsymbol{X}) \sim$ N($\boldsymbol{X}^\top \beta_{T_z}$, 1)，并在有限时间范围 H处进行右截断。独立的右删失时间从 $[0, 1.5H]$中均匀抽取，从而产生非平凡的删失率。观测时间 $Y = \min\{T, C, H\}$，事件指示变量为 $\delta = \mathbb{1}\{T < \min(C, H)\}$。包含 20,000 个样本的模拟数据集被划分为训练集（7,000 个）和估计集（13,000 个）。

使用治疗组特异系数（β_{E_1}, β_{E_0}）和（β_{T_1}, β_{T_0}），可以探索各种治疗效应修饰因子设置。我们考虑四种不同的模拟设置，其中治疗效应修饰因子分别存在于：**1.** 仅治愈模型中；**2.** 仅时间模型中；**3.**两个模型中，但对应不相交协变量；**4.** 两个模型中，且对应重叠协变量。更多详情请参阅补充材料。

表 1：四种设置下治愈概率 (×100) 和条件平均事件时间的异质性处理效应估计的平均绝对误差 (±标准差) ($K = 50$)

方法	仅治愈		仅时间		两者+独立		两者+重叠	
	Cure	Time	Cure	Time	Cure	Time	Cure	Time
预言机	6.6 ± 0.2	12.3 ± 1.3	6.3 ± 0.3	22.0 ± 1.8	6.7 ± 0.4	16.0 ± 1.9	6.8 ± 0.3	18.2 ± 1.5
部分预言机	7.7 ± 0.3	26.8 ± 1.5	7.6 ± 0.2	32.6 ± 1.5	8.0 ± 0.3	31.3 ± 1.8	8.3 ± 0.3	33.4 ± 1.5
MCM KNN	7.9 ± 0.3	26.8 ± 1.0	7.6 ± 0.2	33.3 ± 1.5	8.3 ± 0.3	31.4 ± 1.3	8.7 ± 0.3	38.5 ± 1.3
MCM KNN 组合	8.0 ± 0.3	27.6 ± 1.1	7.8 ± 0.2	39.4 ± 1.6	8.7 ± 0.3	33.4 ± 1.2	8.9 ± 0.3	45.9 ± 1.5
特征选择 KNN	8.2 ± 0.3	28.0 ± 1.1	8.0 ± 0.2	43.2 ± 1.8	9.4 ± 0.5	38.1 ± 2.5	9.4 ± 0.4	41.5 ± 2.6
欧几里得近邻算法	9.9 ± 0.3	29.8 ± 1.1	8.3 ± 0.2	48.7 ± 1.6	10.5 ± 0.4	52.1 ± 1.1	10.8 ± 0.4	55.6 ± 1.3
倾向评分 K 近邻	17.2 ± 0.2	37.1 ± 1.3	9.6 ± 0.4	91.9 ± 1.2	19.2 ± 0.2	70.3 ± 0.9	19.2 ± 0.2	88.7 ± 0.9
预后评分 K 近邻	13.6 ± 0.5	38.3 ± 1.8	8.4 ± 0.4	92.9 ± 1.6	14.5 ± 0.5	60.4 ± 2.1	15.4 ± 0.4	68.9 ± 2.4
Cox模型（无匹配）	5.2 ± 0.6	36.2 ± 2.2	8.2 ± 0.7	78.7 ± 3.1	7.7 ± 0.7	58.0 ± 1.3	8.6 ± 0.7	67.5 ± 1.5

6.2 方法比较

我们将本方法与若干匹配类和基于模型的估计量进行比较。匹配类估计量包括：使用欧氏距离的K近邻（KNN）匹配（Prijs等人，2007）、倾向评分匹配（Austin，2014；Lu等人，2018）、基于Cox模型的预后评分匹配（Anand等人，1999）、基于结果的特征选择匹配（Kroupa等人，2019），以及在倾向得分尺度上进行的最优完全匹配（Austin和Stuart，2015；Lee等人，2020）。作为非匹配基线，我们还在各治疗组内分别拟合Cox比例风险模型（Cox，1972；Kalbfleisch和Prentice，2002），以估计组别特异的生存函数，从而估计异质性治疗效应。

作为基准，我们纳入了**两个预言机估计器**：1. 完全预言机可获取真实模型系数作为权重、治愈状态以及两种治疗组下的事件时间，因此能够在不受删失影响的情况下进行最优匹配和异质性处理效应估计。其误差仅反映伯努利分布和事件时间分布变异所带来的内在随机性。2. 部分预言机使用真实模型参数进行匹配，但依赖观测到的、可能经过删失的结局。它反映了在已知协变量重要性的条件下可能达到的最佳表现，从而隔离出模型拟合误差。将预言机估计器与所提出的方法进行比较，可以将误差分解为不可约结果变异，以及模型拟合和删失带来的估计误差。

对于我们的方法，除了分别使用基于治愈的权重（针对治愈概率的异质性处理效应）或基于时间的权重（针对条件平均事件时间的异质性处理效应）进行匹配外，我们还评估了对两种权重取平均的组合版本。我们使用 300 次模拟中相对于真实效应的平均绝对误差（MAE）评估异质性处理效应估计的准确度，并通过标准差和个体层面误差的分布来评估稳健性。

6.3 结果

表 1 报告了在四种模拟设置下，对治愈概率和条件平均事件时间（CMET）的异质性治疗效应进行估计时的平均绝对误差。我们的方法所得平均绝对误差始终略高于完全预言机和部分预言机基线。例如，完全预言机的平均绝对误差约为 6.5，而我们的方法通常在 7.5 到 8.5 之间。差距较小，这表明大部分残余误差（由预言机的平均绝对误差反映）是由数据变异导致的不可约误差，而非估计误差。我们的方法与部分预言机之间的差异较小，反映了模型拟合过程中的估计误差；这进一步表明，我们学得的距离度量与最优匹配结构十分接近。鉴于事件发生时间数据存在删失和未知治愈状态，这些结果表明，我们的方法在这些情形下的表现非常接近最优。

与其他基于匹配的方法相比，我们的方法在所有设定下都显著降低了误差。标准欧几里得匹配无法区分相关协变量，

而倾向评分匹配和预后评分匹配虽然能够处理混杂，但在估计这两种效应时表现不佳。在治疗效应修饰因子同时影响治愈分量和时间分量的设定下，这种改进最为明显。在这类情况下，基于特征选择的匹配的平均绝对误差会显著增加。我们还评估了一种组合权重匹配方法，该方法对两个最优权重取平均。由于组合权重并不直接针对任一目标估计量，因此这种方法的设计与通用匹配策略相似。然而，它的表现始终略优于这些基线方法，凸显了学习权重的优势，即使这些权重并未明确针对相应的估计目标。我们还将 Cox比例风险模型作为非匹配基线纳入评估。当治疗效应修饰因子仅影响事件概率时，Cox表现良好；但当修饰因子影响事件发生时间时，Cox无法捕捉异质性，并产生显著更高的误差，尤其是在估计时间效应时。这些结果表明，当 Cox模型的假设与底层数据相符时，其表现良好；但当异质性同时涉及事件概率和事件发生时间时，其局限性便显现出来。

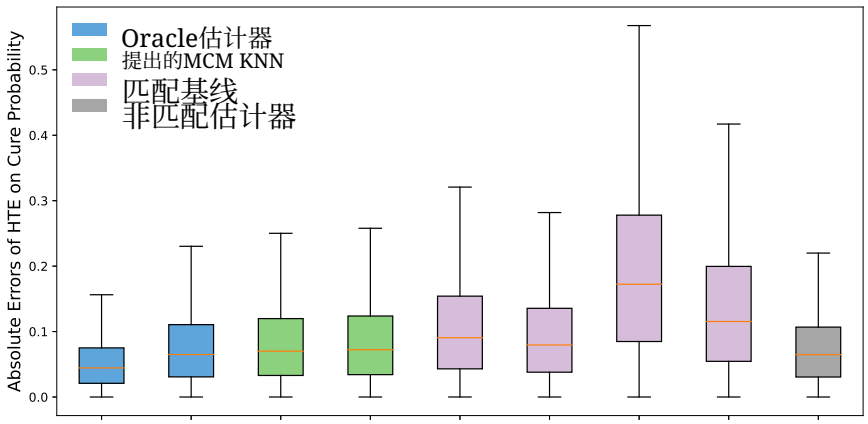

图2：治愈概率的绝对异质处理效应估计误差。

图2展示了单次模拟运行中，在估计集上计算的个体水平异质处理效应治愈概率估计绝对误差的分布。在此设置中，治愈分量和时间分量均包含效应修饰因子，且协变量彼此不相交。正如预期，Oracle估计器的误差和变异性最低。我们提出的方法（基于MCM的匹配）紧随两个Oracle基线之后。与对结果不加区分但无偏的欧几里得匹配相比，我们的方法平均误差和方差更低。其表现与基于特征选择的匹配以及未进行匹配的Cox模型相当，不过如表1所示，后两种方法在不同模拟设置下的表现波动更为明显。

在KNN匹配中选择 K 需要在偏差与方差之间权衡。较小的 K 可得到精确但变异较大的估计；较大的 K 可提高稳定性，但存在匹配不佳的风险。我们在实验中使用 $K = 50$ 来平衡这些因素。在实践中， K 应依据队列规模、特征变异性和领域知识来确定，因为标准验证并不可行。更多实验结果见补充材料。

7 真实世界数据实验

为说明本方法的实际价值——在假设可能不成立且真实模型未知的情况下——我们分析了一个接受异基因造血干细胞移植的急性淋巴细胞白血病患者多中心队列（Ma 等，2021）。所关注的治疗比较单倍型相合移植（单倍型相合造血干细胞移植）与同胞全相合供者移植（同胞全相合供者移植）。共纳入1157例患者，随访时间最长达4193天。

7.1 研究设计与设置

在记录的四种结局中，我们重点关注**无白血病生存期（LFS）**，它反映了复发和死亡情况，可作为长期治愈的替代指标。我们将 $H = 10$ 年（3650天）时的状态定义为治愈。治疗组分布不均衡，治疗组有911名患者，对照组有246名患者（约占78%和22%）。删失率分别为18.1%和25.2%。基线协变量包括人口统计学特征、供者-受者特征、微小残留病状态（MRD）、HLA不匹配和完全缓解状态，共计8–9个关键预测变量。我们使用35%的数据训练混合治愈模型，使用65%的数据进行异质性治疗效应估计，并在匹配中使用 $K = 35$ 个邻居。

7.2 异质性治疗效应估计结果

如图3所示，不同匹配方法得到的估计异质性治疗效应分布差异显著。所提出的方法（MCM）得到的估计异质性治疗效应分布集中且对称，均值约为0.05，四分位距（IQR）与两组之间治愈概率的经验差异相符。基于欧氏距离的匹配得到的分布对称但离散程度很高，而倾向评分（PS）匹配产生的估计值高度集中、变异性有限。预后评分匹配得到的分布呈双峰且部分为负，这可能是由于预后模型与治愈相关结局不匹配所致。基于特征选择的匹配产生的估计值偏斜且不稳定。

对治愈概率的估计异质性治疗效应进行两两相关性分析后发现，分离权重MCM和组合权重MCM变体之间呈较强正相关（0.799），而分离权重MCM与欧几里得匹配之间呈正相关，但相关性中等（0.346）。这表明各方法得出的估计方向一致，但结果存在差异。

总体而言，我们的结果表明，与同胞全相合供者移植相比，单倍型相合造血干细胞移植平均可将白血病无生存期提高约5%。这一改善具有临床意义，但幅度较比较方法所得估计更为保守。这种差异可能源于我们的方法旨在分别估计治愈概率和生存时间的不同效应，而比较方法并非如此。因此，比较方法可能会将治疗对白血病无生存期概率的效应与对短期生存时间的效应混为一谈；后者同样重要，但在临床意义上有所不同。尽管在这一真实世界情境中无法获得真实异质性治疗效应，但我们的理论和模拟结果强烈表明，由于数据集中存在大量治愈人群，我们的方法很可能更适用于该数据集。

8 结论

我们提出了一种新颖的匹配框架，利用基于混合治愈模型得出的协变量特异性距离，分别估计治愈概率和事件发生时间上的异质性治疗效应。我们的方法具有坚实的理论基础，通过匹配的 Kaplan-Meier 曲线实现一致估计，并在各类模拟中表现良好。然而，该方法仍需在多样化的真实世界

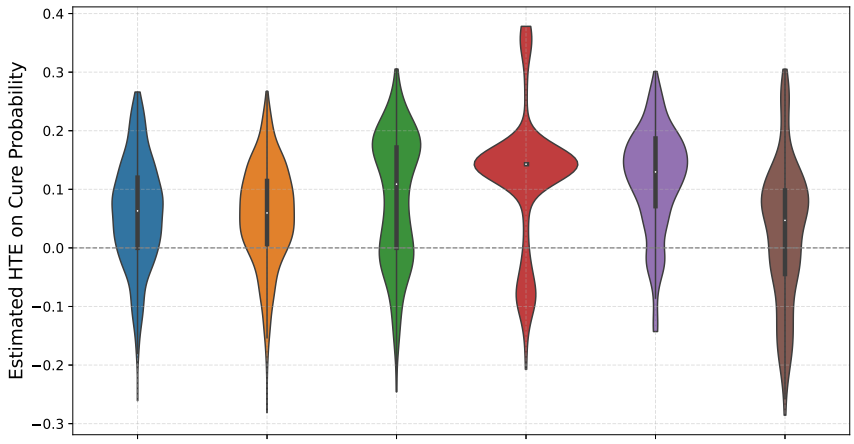

图 3：无白血病生存期（LFS）下治愈概率估计的异质性治疗效应分布。

数据集上进一步验证；尽管纳入了参数化和非参数化元素，它仍可能对模型设定错误敏感。

参考文献

Amico, M. and Van Keilegom, I. (2018). 生存分析中的治愈模型。《统计学及其应用年度评论》，5(1):311–342。Anand, S. S., Smith, A. E., Hamilton, P., Anand, J., Hughes, J. G., and Bartels, P. H. (1999). 结直肠癌智能预后系统评估。《医学中的人工智能》，15(2):193–214。Ashouri, M. and Henderson, N. C. (2024). 用于受限平均生存时间推断的广义贝叶斯加性回归树。*arXiv*预印本 *arXiv:2402.17920*。Austin, P. C. (2014). 倾向评分方法在生存结局或事件发生时间结局中的应用：报告与随机试验所用指标类似的效应指标。《医学统计学》，33(7):1242–1258。Austin, P. C. and Stuart, E. A. (2015). 生存结局的最优全匹配：一种值得更广泛应用的方法。《医学统计学》，34(30):3949–3967。Berkson, J. and Gage, R. P. (1952). 治疗后癌症患者的生存曲线。《美国统计学会会刊》，47(259):501–515。Bewick, V., Cheek, L., and Ball, J. (2004). 统计学综述12：生存分析。《重症监护》，8:1–6。Boag, J. W. (1949). 癌症治疗治愈患者比例的最大似然估计。《皇家统计学会会刊：*B*辑（方法论）》，11(1):15–53。

Boutros, C., Belkadi-Sadou, D., Marchand, A., Roy, S., Routier, E., and Robert, C. (2023). 治愈了吗？免疫治疗应答者的长期结局：聚焦黑色素瘤。《*Current Oncology Reports*》，25(9):989–996。Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). 一种用于含生存比例的生存数据的新贝叶斯模型。《美国统计学会会刊》，94(447):909–919。Cox, D. R. (1972). 回归模型与生命表。《皇家统计学会会刊：*B*辑（方法论）》，34(2):187–220。Diamond, A. and Sekhon, J. S. (2013). 用于估计因果效应的遗传匹配：一种实现观察性研究平衡的通用多变量匹配方法。《经济学与统计学评论》，95(3):932–945。Dieng, A., Liu, Y., Roy, S., Rudin, C., and Volfovsky, A. (2019). 用于因果推断的可解释近似精确匹配。载于第*22*届人工智能与统计国际会议，第2445–2453页。PMLR。Engelhard, M. and Henao, R. (2022). 在失效时间数据神经混合治愈模型中区分“是否发生”与“何时发生”。载于人工智能与统计国际会议，第9571–9581页。PMLR。Farewell, V. T. (1982). 混合模型在长期存活者生存数据分析中的应用。《生物统计学》，第1041–1046页。Felizzi, F., Paracha, N., P¨ohlmann, J., and Ray, J. (2021). 肿瘤学中的混合治愈模型：教程与实践指南。《*PharmacoEconomics-open*》，5:143–155。Gao, X. and Zheng, M. (2017). 对含治愈比例且存在不依从情况的生存数据随机试验中的因果效应进行估计。《统计学通讯：理论与方法》，46(8):4065–4087。Hansen, B. B. (2008). 倾向评分的预后类比量。《生物统计学》，95(2):481–488。Imbens, G. W. and Rubin, D. B. (2015).《统计学、社会科学与生物医学中的因果推断：导论》。剑桥大学出版社，纽约，纽约州。Jiang, F., Zhao, G., Rodriguez-Monguio, R., and Ma, Y. (2024). 高维混杂因素下生存分析中的因果效应估计。《生物统计学》，80(4):ujae110。Jorm, A. F., Patten, S. B., Brugha, T. S., and Mojtabai, R. (2017). 增加治疗服务是否降低了常见精神障碍的患病率？来自四个国家的证据综述。《世界精神病学》，16(1):90–99。Kalbfleisch, J. D. and Prentice, R. L. (2002).《失效时间数据的统计分析》。约翰·威立父子出版公司，第2版。Katta, S., Parikh, H., Rudin, C., and Volfovsky, A. (2024). 用于分析可穿戴设备、传感器和分布数据的可解释因果推断。载于人工智能与统计国际会议，第3340–3348页。PMLR。Knaus, M. C., Lechner, M., and Strittmatter, A. (2021). 异质性因果效应的机器学习估计：经验蒙特卡洛证据。《计量经济学杂志》，24(1):134–161。

Kroupa, P., Morton, C., Le Calvez, K., and Williams, M.（2019）。评估k近邻算法在一系列模拟数据集上的简单、可解释的生存时间预测效果。载于*2019*年*IEEE*第*32*届计算机医学系统国际研讨会（*CBMS*），第367–372页。IEEE。Lanners, Q., Parikh, H., Volfovsky, A., Rudin, C., and Page, D.（2023）。用于因果推断的变量重要性匹配。载于人工智能中的不确定性，第1174–1184页。PMLR。Lee, H.-J., Wong, J. B., Jia, B., Qi, X., and DeLong, E. R.（2020）。在真实世界登记数据库中运用因果推断方法评估生存差异，并与随机临床试验中的结果进行比较。《医学统计学》，39(22):3003–3021。Lin, Z., Ni, A., and Lu, B.（2023）。观察性研究中限制平均生存时间边际因果效应的匹配设计。《因果推断杂志》，11(1):20220035。Lu, B., Cai, D., and Tong, X.（2018）。使用倾向评分匹配设计检验观察性生存数据中的因果效应。《医学统计学》，37(11):1846–1858。Ma, R., Xu, L., Zhang, X., Wang, Y., Chen, H., Chen, Y., Wang, F., Han, W., Sun, Y., Yan, C., 等（2021）。一种主要基于微小/可测量残留病灶定量动态变化的综合评分系统，用于预测急性淋巴细胞白血病患者复发。《临床肿瘤学杂志》，39(15 suppl)。Michielin, O., Atkins, M. B., Koon, H. B., Dummer, R., and Ascierto, P. A.（2020）。长期生存结果对转移性黑色素瘤治疗的影响不断演变。《癌症免疫治疗杂志》，8(2):e000948。Morucci, M., Orlandi, V., Roy, S., Rudin, C., and Volfovsky, A.（2020）。用于可解释个体化治疗效应估计的自适应超盒匹配。载于人工智能不确定性会议，第1089–1098页。PMLR。Musta, E., Patilea, V., and Van Keilegom, I.（2022）。半参数混合治愈模型的两步估计程序。*arXiv*预印本 *arXiv:2207.08237*。Parikh, H., Hoffman, K., Sun, H., Zafar, S. F., Ge, W., Jing, J., Liu, L., Sun, J., Struck, A., Volfovsky, A., 等（2023）。美国危重症患者癫痫样活动对出院结局的影响：一项回顾性横断面研究。《柳叶刀·数字健康》，5(8):e495–e502。Parikh, H., Rudin, C., and Volfovsky, A.（2022）。MALTS：学习拉伸后匹配。《机器学习研究杂志》，23(240):1–42。Prijs, M., Peelen, L., Bresser, P., and Peek, N.（2007）。一种预测生存时间的最近邻方法及其在慢性呼吸系统疾病中的应用。载于欧洲医学人工智能会议，第77–86页。Springer。Ratkovic, M.（2019）。重振回归：通过机器学习实现诚实且有效的因果推断。Rosenbaum, P. R. and Rubin, D. B.（1983）。倾向评分在因果效应观察性研究中的核心作用。《生物统计学》，70(1):41–55。

Rubin, D. B.（1980）。实验数据的随机化分析：关于Fisher随机化检验的评论。《美国统计学会会刊》，75(371):591–593。Shikalgar, S. 等（2025）。一种用于因果推断的两阶段可解释匹配框架。*arXiv*预印本 *arXiv:2504.09635*。Stuart, E. A.（2010）。因果推断的匹配方法：综述与展望。《统计科学》，25(1):1。Stute, W.（1993）。存在协变量时随机删失情形下的一致估计。《统计年刊》，21(1):146–156。Sun, R. and Song, X.（2025）。一种用于估计异质性治疗效应的基于树的贝叶斯加速失效时间治愈模型。《贝叶斯分析》，20(2):345–373。Tita, A. T., Szychowski, J. M., Boggess, K., Saade, G., Longo, S., Clark, E., Esplin, S., Cleary, K., Wapner, R., Letson, K., 等（2016）。剖宫产的辅助性阿奇霉素预防用药。《新英格兰医学杂志》，375(13):1231–1241。Tsodikov, A.（1998）。考虑长期存活者的比例风险模型。《生物统计学》，第1508–1516页。Wager, S. and Athey, S.（2018）。使用随机森林估计和推断异质性治疗效应。《美国统计学会会刊》，113(523):1228–1242。Wang, T., Morucci, M., Awan, M. U., Liu, Y., Roy, S., Rudin, C., and Volfovsky, A.（2021）。FLAME：一种快速、大规模、近乎精确匹配的因果推断方法。《机器学习研究杂志》，22(31):1–41。Wang, Y., Deng, Y., and Zhou, X.-H.（2024）。针对含治愈亚群的生存时间数据的因果推断。《生物统计学》，80(2):ujae028。Xu, M., Xi, Z., Zhao, Q., Yang, W., Tan, J., Yi, P., Zhou, J., and Huang, T.（2023）。甲状腺乳头状癌中侵袭性甲状腺外侵犯与生存的因果推断：倾向评分匹配与加权分析。《内分泌学前沿》，14:1149826。Yakovlev, A. Y., Tsodikov, A. D., and Asselain, B.（1996）。肿瘤潜伏期的随机模型及其生物统计学应用，第1卷。世界科学出版社。Yang, L.-P., Wang, Z.-X., He, M.-M., Jin, Y., Ren, C., Wang, Z.-Q., Wang, F.-H., Li, Y.-H., Wang, F., and Xu, R.-H.（2019）。姑息性胃切除术和/或转移灶切除术对伴有同时性转移的胃癌患者的生存获益：一项基于人群的研究，采用倾向评分匹配和粗化精确匹配。《癌症杂志》，10(3):602。Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., and Zhang, A.（2021）。因果推断综述。《*ACM* 数据知识发现汇刊（*TKDD*）》，15(5):1–46。Yu, R., Small, D. S., Harding, D., Aveldanes, J., and Rosenbaum, P. R.（2021a）。整合定量与定性研究的观察性研究最优匹配。《统计学与公共政策》，8(1):42–52。Yu, R., Small, D. S., Harding, D., Aveldanes, J., and Rosenbaum, P. R.（2021b）。整合定量与定性研究的观察性研究最优匹配。《统计学与公共政策》，8(1):42–52。

Zhu, J. 和 Gallego, B.（2020年）。观察性生存分析中异质性治疗效应的目标估计。《生物医学信息学杂志》，107:103474。

补充材料

A 缺失证明

A.1 命题1的证明

本节介绍假设并给出**命题1**的详细证明。

命题2 如果一致性、无混杂性、正值性和非信息性删失假设成立，则这两个目标估计量可以表示为

$$\pi(\boldsymbol{x}) = S_1(H \mid \boldsymbol{x}) - S_0(H \mid \boldsymbol{x}), \tag{18}$$

$$\begin{aligned}\Delta(\boldsymbol{x}) = &\frac{\int_0^H S_1(t|\boldsymbol{x})\,dt - HS_1(H|\boldsymbol{x})}{1 - S_1(H|\boldsymbol{x})} \\ &- \frac{\int_0^H S_0(t|\boldsymbol{x})\,dt - HS_0(H|\boldsymbol{x})}{1 - S_0(H|\boldsymbol{x})}.\end{aligned} \tag{19}$$

证明。 假设我们有以下假设:

假设1（一致性）对所有 i,

$$T_i = Z_i T_i(1) + (1 - Z_i) T_i(0) \tag{20}$$

$$C_i = Z_i C_i(1) + (1 - Z_i) C_i(0). \tag{21}$$

假设2（无混杂性）

$$\{T_i(1), T_i(0), C_i(1), C_i(0)\} \perp\!\!\!\perp Z_i \mid \boldsymbol{X}_i. \tag{22}$$

假设3（正值性）对所有 x,

$$0 < \Pr(Z_i = 1 \mid \boldsymbol{X}_i = \boldsymbol{x}) < 1 \tag{23}$$

假设4（非信息性删失）

$$\{T_i(1), T_i(0)\} \perp\!\!\!\perp \{C_i(1), C_i(0)\} \mid \boldsymbol{X}_i \tag{24}$$

命题1的证明如下:

$$\begin{aligned}\pi(\boldsymbol{x}) &= \Pr(E_i(1) = 0 \mid \boldsymbol{X} = \boldsymbol{x}) - \Pr(E_i(0) = 0 \mid \boldsymbol{X} = \boldsymbol{x}) \\ &= \Pr(T(1) > H \mid \boldsymbol{X} = \boldsymbol{x}) - \Pr(T(0) > H \mid \boldsymbol{X} = \boldsymbol{x}) \\ &= \Pr(T(1) > H \mid Z = 1, \boldsymbol{X} = \boldsymbol{x}) - \Pr(T(0) > H \mid Z = 0, \boldsymbol{X} = \boldsymbol{x}) \\ &= \Pr(T > H \mid Z = 1, \boldsymbol{X} = \boldsymbol{x}) - \Pr(T > H \mid Z = 0, \boldsymbol{X} = \boldsymbol{x}) \\ &= S_1(H \mid \boldsymbol{x}) - S_0(H \mid \boldsymbol{x}),\end{aligned} \tag{25}$$

由于我们可以推导出下式:

$$
\begin{aligned}
E[T_i(z) \mid T_i(z) < H, \boldsymbol{X} = \boldsymbol{x}] &= \int t \cdot p(T = t \mid T < H, Z = z, \boldsymbol{X} = \boldsymbol{x})\, dt \\
&= \int t \cdot \frac{p(T = t, T < H \mid Z = z, \boldsymbol{X} = \boldsymbol{x})}{p(T < H \mid Z = z, \boldsymbol{X} = \boldsymbol{x})}\, dt \\
&= \int t \cdot \frac{p(T = t \mid Z = z, \boldsymbol{X} = \boldsymbol{x}) \cdot p(T < H \mid T = t, Z = z, \boldsymbol{X} = \boldsymbol{x})}{p(T < H \mid Z = z, \boldsymbol{X} = \boldsymbol{x})}\, dt \\
&= \frac{\int_0^H t \cdot p(T = t \mid Z = z, \boldsymbol{X} = \boldsymbol{x})\, dt}{\int_0^H p(T = t \mid Z = z, \boldsymbol{X} = \boldsymbol{x})\, dt} \\
&= \frac{\int_0^H t\, dS_z(t \mid \boldsymbol{x})}{\int_0^H dS_z(t \mid \boldsymbol{x})} \\
&= \frac{-H S_z(H \mid \boldsymbol{x}) + \int_0^H S_z(t \mid \boldsymbol{x})\, dt}{1 - S_z(H \mid \boldsymbol{x})}
\end{aligned}
\tag{26}
$$

因此，第二个目标估计量可表示为:

$$
\begin{aligned}
\Delta(\boldsymbol{x}) &= \mathbb{E}[T_i(1) \mid T_i(1) < H, \boldsymbol{X} = \boldsymbol{x}] - \mathbb{E}[T_i(0) \mid T_i(0) < H, \boldsymbol{X} = \boldsymbol{x}] \\
&= \frac{\int_0^H S_1(t|\boldsymbol{x})\, dt - H S_1(H|\boldsymbol{x})}{1 - S_1(H|\boldsymbol{x})} - \frac{\int_0^H S_0(t|\boldsymbol{x})\, dt - H S_0(H|\boldsymbol{x})}{1 - S_0(H|\boldsymbol{x})}
\end{aligned}
\tag{27}
$$

A.2 混合治愈模型

给定预先指定的时间范围 H，采用混合治愈方法建模的生存函数可写为:

$$
\begin{aligned}
S(t) &= \Pr(cure) + [1 - \Pr(cure)] S_N(t) \\
&= \Pr(E = 0) + [1 - \Pr(E = 0)] \Pr(T > t | T \le H) \\
&= \Pr(T > H) + \Pr(T \le H) \Pr(T > t | T \le H) \\
&= \Pr(T > t)
\end{aligned}
\tag{28}
$$

其中，$\Pr(cure)$ 表示治愈概率，而 $S_N(t)$ 表示患者未治愈条件下的生存函数。

似然。令 T 表示真实事件时间，C 表示删失时间，$Y =$ 表示在固定时间范围 H 下的观测时间，即 $\min(T, C, H)$，并令事件指示变量为 $\delta = \mathbb{I}(T < C, T < H)$。协变量记为 $\boldsymbol{x}$，潜在治愈指示变量记为 $E = \mathbb{I}(T < H)$，其中 $E = 1$ 表示未治愈个体，$E = 0$ 表示治愈个体。

记:

- $f_T(t \mid \boldsymbol{x})$：在给定协变量$\boldsymbol{x}$的条件下，事件时间 T 的概率密度函数（PDF）；
- $F_T(t \mid \boldsymbol{x}) = \Pr(T \le t \mid \boldsymbol{x})$：$T$的累积分布函数（CDF）；
- $S_T(t \mid \boldsymbol{x}) = 1 - F_T(t \mid \boldsymbol{x}) = \Pr(T > t \mid \boldsymbol{x})$：$T$的生存函数;
- $f_C(c \mid \boldsymbol{x})$：给定 $\boldsymbol{x}$ 时删失时间 C 的密度函数;

- $S_C(c \mid \boldsymbol{x}) = \Pr(C > c \mid \boldsymbol{x})$: C 的生存函数。

数据 $p(y,\ \delta|\boldsymbol{x})$ 的密度（似然）可概括如下:

$$
\begin{aligned}
p(y,&\delta|\boldsymbol{x}) = p(y, \delta = 1|\boldsymbol{x}) + p(y, \delta = 0|\boldsymbol{x}) \\
&= p(E = 1|\boldsymbol{x})\mathbb{I}(y < H)\frac{f_T(y|\boldsymbol{x})}{F_T(H|\boldsymbol{x})}S_C(y|\boldsymbol{x}) + p(E = 1|\boldsymbol{x})\mathbb{I}(y < H)\left(1 - \frac{F_T(y|\boldsymbol{x})}{F_T(H|\boldsymbol{x})}\right)f_C(y|\boldsymbol{x}) \\
&+ p(E = 0|\boldsymbol{x})\mathbb{I}(y < H)f_C(y|\boldsymbol{x}) + p(E = 0|\boldsymbol{x})\mathbb{I}(y = H)S_C(y|\boldsymbol{x})
\end{aligned} \tag{29}
$$

证明。 对于式 29 中的第一项,

$$
\begin{aligned}
p(y, \delta = 1|\boldsymbol{x}) &= p(y, \delta = 1, E = 1|\boldsymbol{x}) + p(y, \delta = 1, E = 0|\boldsymbol{x}) \\
&= p(E = 1|\boldsymbol{x})p(y, \delta = 1|\boldsymbol{x}, E = 1) + 0 \\
&= p(E = 1|\boldsymbol{x})p(T = y, y < H, y < C|\boldsymbol{x}, T < H) \\
&= p(E = 1|\boldsymbol{x})\mathbb{I}(y < H)p(T = y, y < C|\boldsymbol{x}, T < H) \\
&= p(E = 1|\boldsymbol{x})\mathbb{I}(y < H)p(T = y|T < H, \boldsymbol{x})p(C > y|\boldsymbol{x}, T < H) \\
&= p(E = 1|\boldsymbol{x})\mathbb{I}(y < H)\frac{p(T = y, T < H|\boldsymbol{x})}{p(T < H|\boldsymbol{x})}p(C > y|\boldsymbol{x}) \\
&= p(E = 1|\boldsymbol{x})\mathbb{I}(y < H)\frac{p(T = y|\boldsymbol{x})}{p(T < H|\boldsymbol{x})}p(C > y|\boldsymbol{x}) \\
&= p(E = 1|\boldsymbol{x})\mathbb{I}(y < H)\frac{f_T(y|\boldsymbol{x})}{F_T(H|\boldsymbol{x})}S_C(y|\boldsymbol{x})
\end{aligned} \tag{30}
$$

对于式 29 中的第二项,

$$
\begin{aligned}
p(y, \delta = 0, |\boldsymbol{x}) &= p(y, \delta = 0, E = 1|\boldsymbol{x}) + p(y, \delta = 0, E = 0|\boldsymbol{x}) \\
&= p(E = 1|\boldsymbol{x})p(y, \delta = 0|\boldsymbol{x}, E = 1) + p(E = 0|\boldsymbol{x})p(y, \delta = 0|\boldsymbol{x}, E = 0) \\
&= p(E = 1|\boldsymbol{x})p(y, \delta = 0|\boldsymbol{x}, T < H) + p(E = 0|\boldsymbol{x})p(y, \delta = 0|\boldsymbol{x}, T \geq H)
\end{aligned} \tag{31}
$$

对于式 31 第一项中的 $p(y,\ \delta = 0|\boldsymbol{x},\ T < H)$:

$$
p(y, \delta = 0|\boldsymbol{x}, T < H) = p(C = y, T > y, H > y|\boldsymbol{x}, T < H) + p(H = y, T \geq y, C \geq y|\boldsymbol{x}, T < H) \tag{32}
$$

$$
= \mathbb{I}(y < H)\frac{p(T > y, T < H|\boldsymbol{x})}{p(T < H|\boldsymbol{x})}p(C = y|\boldsymbol{x}) + \mathbb{I}(y = H)p(T \geq H, C \geq H|\boldsymbol{x}, T < H) \tag{33}
$$

$$
= \mathbb{I}(y < H)\left(1 - \frac{F_T(y|\boldsymbol{x})}{F_T(H|\boldsymbol{x})}\right)f_C(y|\boldsymbol{x}) + 0 \tag{34}
$$

对于式 31 第二项中的 $p(y,\ \delta = 0|\boldsymbol{x},\ T \geq H)$:

$$
\begin{aligned}
p(y, \delta = 0|\boldsymbol{x}, T \geq H) &= p(T > y, C = y, H > y \mid \boldsymbol{x}, E = 0) + p(T \geq y, C \geq y, H = y \mid \boldsymbol{x}, E = 0) \\
&= \mathbb{I}(y < H)\frac{p(T > y, T \geq H \mid \boldsymbol{x})}{p(T \geq H \mid \boldsymbol{x})}p(C = y \mid \boldsymbol{x}) + \mathbb{I}(y = H)\frac{p(T \geq y, T \geq H \mid \boldsymbol{x})}{p(T \geq H \mid \boldsymbol{x})}p(C \geq y \mid \boldsymbol{x}) \\
&= \mathbb{I}(y < H)f_C(y \mid \boldsymbol{x}) + \mathbb{I}(y = H)S_C(y \mid \boldsymbol{x})
\end{aligned} \tag{35}
$$

代入 根据上述推导的各项，我们可以得到式（29）所示的似然函数 .

A.3 定理 1 的证明

设置。 设 $S(t \mid \boldsymbol{x}) := \Pr(T > t \mid \boldsymbol{X} = \boldsymbol{x})$为给定协变量$\boldsymbol{x}$时的条件生存曲线。设 $\boldsymbol{X}_1$、……、 $\boldsymbol{X}_n$ 为从分布 P中抽取的独立同分布样本。对于每个 $i \in \{1, 2, \ldots, n\}$，设 T_i为随机事件时间，其生存函数为 $S(t \mid \boldsymbol{X}_i)$；设 C_i 为与 T_i独立的随机删失时间。定义观测时间 $Y_i = \min\{T_i, C_i\}$ 和事件指示变量 $\delta_i = \mathbb{I}(T_i < C_i)$。

考虑一个大小为 m, $\{\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m}\} \subset \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ 的匹配组，用于估计$S(t \mid \boldsymbol{x})$。按照惯例，这些 m 个单位是从所有单位中，根据某个距离度量 $d(\cdot, \cdot)$，选取与 $\boldsymbol{x}$ 距离最近的单位。令 $S_{mn}(t)$ 为 Kaplan-Meier估计量，其采用 $\{(Y_{k_i}, \delta_{k_i})\}_{i=1}^{m}$。

定义两个泛函

$$F(S) := S(H), \quad G(S) := \int_0^H S(u)\,\mathrm{d}u.$$

下面的定理表明，在一定条件下， $F(S_{mn})$ 收敛于 $F(S(\cdot \mid \boldsymbol{x}))$，且 $G(S_{mn})$ 收敛于 $G(S(\cdot \mid \boldsymbol{x}))$。

定理 1. 设 $\boldsymbol{x}$ 属于 $\boldsymbol{X}$ 的支撑集，即，对于任意 $\epsilon > 0$, $P(d(\boldsymbol{X}, \boldsymbol{x}) < \epsilon) > 0$。假设 $S(t \mid \boldsymbol{x})$ 关于 $\boldsymbol{x}$ 是局部 Lipschitz 连续的，且对 t 一致，即，存在 $\boldsymbol{x}$ 的一个邻域，记为 $\mathcal{N}_{\boldsymbol{x}}$，以及一个常数 $\tau > H$，使得对于所有 $\boldsymbol{x}', \boldsymbol{x}''$ 属于 $\mathcal{N}_{\boldsymbol{x}}$ 且 $t \in [0, \tau]$, $|S(t \mid \boldsymbol{x}') - S(t \mid \boldsymbol{x}'')| \le L \cdot d(\boldsymbol{x}', \boldsymbol{x}'')$。那么，

$$F(\widehat{S}_{mn}) \xrightarrow{p} F(S(\cdot \mid \boldsymbol{x})), \quad G(\widehat{S}_{mn}) \xrightarrow{p} G(S(\cdot \mid \boldsymbol{x})),$$

当 $m, n \to \infty$ 满足 $m = o(n)$ 时。

推论 1. 如下所示在第 4.3 节定义的两个估计量是 F 和 G的函数，它们也具有一致性。

$$\pi(\mathbf{x}_i) = \hat{S}_{M_1}(H) - \hat{S}_{M_0}(H), \tag{36}$$

$$\Delta(\boldsymbol{x}_i) = \frac{\int_0^H \hat{S}_{M_1}(t)\,dt - H\hat{S}_{M_1}(H)}{1 - \hat{S}_{M_1}(H)} - \frac{\int_0^H \hat{S}_{M_0}(t)\,dt - H\hat{S}_{M_0}(H)}{1 - \hat{S}_{M_0}(H)} \tag{37}$$

证明. 我们将通过三个引理来证明**定理 1** 。第一个引理表明，随着样本量增加，匹配组会越来越接近 $\boldsymbol{x}$ 。

引理 1. 令 $d_{mn} = \max_i d(\boldsymbol{X}_{k_i}, \boldsymbol{x})$，其中 $\boldsymbol{x}$ 属于 $\boldsymbol{X}$的支撑集。那么， $d_{mn} \xrightarrow{p} 0$.

引理 1 的证明。定义 $d_i = d(\boldsymbol{X}_i, \boldsymbol{x})$。那么， d_{mn}是 d_1、……、 d_n的第 m个次序统计量，记为 $d_{(m)}$。令 $Q(c) = P(d(\boldsymbol{X}, \boldsymbol{x}) \le c)$ 为 $d(\boldsymbol{X}, \boldsymbol{x})$的分布的累积分布函数。对于任意 $\alpha \in (0, 1)$，样本分位数 $d_{(\lfloor \alpha n \rfloor)}$依概率收敛于 α分位数 $Q^{-1}(\alpha)$，因为样本分位数 $d_{(\lfloor \alpha n \rfloor)}$是一个满足估计方程 $\psi_i(c) = \mathbb{I}(d_i \le c) - \alpha$的 M-估计量。因此，当 $n \to \infty$时， $d_{(\lfloor \alpha n \rfloor)} = Q^{-1}(\alpha) + o_p(1)$。由于 $m = o(n)$，对于任意 $\alpha \in (0, 1)$，当 n 充分大时， $m < \alpha n$。因此，当 $n \to \infty$时， $d_{(m)} \le d_{(\lfloor \alpha n \rfloor)} = Q^{-1}(\alpha) + o_p(1)$。由于 α 是任意选取的，并且 $\lim_{\alpha \to 0} Q^{-1}(\alpha) = 0$，这是因为 $\boldsymbol{x}$ 属于 $\boldsymbol{X}$的支撑集，我们有 $d_{(m)} = o_p(1)$，由此得到 $d_{mn} \xrightarrow{p} 0$。

第二个引理表明，Kaplan–Meier估计量与混合生存函数强相合，这一点已由Stute（1993）确立。

引理 2. 定义 $\bar{S}(t) = \frac{1}{m}\sum_{i=1}^{m} S(t \mid \boldsymbol{X}_{k_i})$。则，

$$\sup_{0\le t\le\tau} |\widehat{S}_{mn}(t) - \bar{S}(t)| \xrightarrow{p} 0$$

当 $m,\ n\to\infty$ 且 $m = o(n)$ 时。

¯第三个引理表明，只要协变量$\boldsymbol{X}_{k_i}$ 接近 $\boldsymbol{x}$，混合生存函数 $S(t)$ 就不会偏离 $S(t \mid \boldsymbol{x})$ 太远。

引理 3。 如果 $\boldsymbol{x}$ 属于 $\boldsymbol{X}$ 的支撑集，且 $S(t \mid \boldsymbol{x})$关于 t一致地局部 Lipschitz 连续，那么 $S(t)$ 在 $[0,\tau]$ 上强相合于 $S(t \mid \boldsymbol{x})$。

引理 3 的证明。

$$|\bar{S}(t) - S(t \mid \boldsymbol{x})| \le \frac{1}{m}\sum_{i=1}^{m} |S(t \mid \boldsymbol{X}_{k_i}) - S(t \mid \boldsymbol{x})|.$$

此外，当 $d(\boldsymbol{X}_{k_i},\ \boldsymbol{x}) < \epsilon$ 时，

$$|S(t \mid \boldsymbol{X}_{k_i}) - S(t \mid \boldsymbol{x})| \le L d(\boldsymbol{X}_{k_i}, \boldsymbol{x}).$$

因此，对于任意 $\delta > 0$,

$$\begin{aligned}
P\left(|\bar{S}(t) - S(t \mid \boldsymbol{x})| \ge \delta\right) &\le P\left(\frac{1}{m}\sum_{i=1}^{m} |S(t \mid \boldsymbol{X}_{k_i}) - S(t \mid \boldsymbol{x})| \ge \delta\right) && (38)\\
&\le P\left(\frac{1}{m}\sum_{i=1}^{m} |S(t \mid \boldsymbol{X}_{k_i}) - S(t \mid \boldsymbol{x})| \ge \delta, d_{mn} < \epsilon\right) + P(d_{mn} \ge \epsilon) && (39)\\
&\le P\left(\frac{1}{m}\sum_{i=1}^{m} |S(t \mid \boldsymbol{X}_{k_i}) - S(t \mid \boldsymbol{x})| \ge \delta \mid d_{mn} < \epsilon\right) + P(d_{mn} \ge \epsilon) && (40)\\
&\le P(L d_{mn} \ge \delta \mid d_{mn} < \epsilon) + P(d_{mn} \ge \epsilon) && (41)\\
&\le P(d_{mn} \ge \delta/L) + P(d_{mn} \ge \epsilon). && (42)
\end{aligned}$$

¯由于 d_{mn} 依概率收敛于 0，上述概率也收敛于 0。因此， $S(t)$ 在 $[0,\tau]$ 上强相合于 $S(t \mid \boldsymbol{x})$。

ˆ**主定理的证明** 由这三个引理可知，在定理 1 的条件下，$S_{mn}(t)$ 强相合于 $S(t \mid \boldsymbol{x})$。根据三角不等式，

$$\left|\widehat{S}_{mn}(t) - S(t \mid \boldsymbol{x})\right| \le \left|\widehat{S}_{mn}(t) - \bar{S}(t)\right| + \left|\bar{S}(t) - S(t \mid \boldsymbol{x})\right|.$$

因此，对任意 $\epsilon > 0$,

$$\mathbb{P}\left(\left|\widehat{S}_{mn}(t) - S(t \mid \boldsymbol{x})\right| > \epsilon\right) \le \mathbb{P}\left(\left|\widehat{S}_{mn}(t) - \bar{S}(t)\right| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\left|\bar{S}(t) - S(t \mid \boldsymbol{x})\right| > \frac{\epsilon}{2}\right).$$

由于右侧两项在 $n\to\infty$ 时均收敛于零，我们得出：

$$\mathbb{P}\left(\left|\widehat{S}_{mn}(t) - S(t \mid \boldsymbol{x})\right| > \epsilon\right) \to 0,$$

这意味着

$$\widehat{S}_{mn}(t) \xrightarrow{P} S(t \mid \boldsymbol{x}).$$

作为一个特例， $\widehat{S}_{mn}(H)$ 与 $S(H \mid \boldsymbol{x})$ 一致。这证明了 $F(\widehat{S}_{mn}) \xrightarrow{p} F(S(\cdot \mid \boldsymbol{x}))$. 此外,

$$|G(\widehat{S}_{mn}) - G(S(\cdot \mid \boldsymbol{x}))| = \left| \int_0^H \widehat{S}_{mn}(u) - S(u \mid \boldsymbol{x}) \, \mathrm{d}u \right| \leq H \sup_{t \in [0,H]} |\widehat{S}_{mn}(t) - S(t \mid \boldsymbol{x})| \xrightarrow{p} 0.$$

由此完成了对 $G(\widehat{S}_{mn}) \xrightarrow{p} G(S(\cdot \mid \boldsymbol{x}))$ 的证明。

A.4 定理 2a 和 2b 的证明

设置。 对于对角权重矩阵 $W = \mathrm{diag}(w_1, \ldots, w_p)$，其中 $w_j \geq 0$ 且 $\sum_{j=1}^p w_j = c > 0$，定义加权范数 $\|\boldsymbol{v}\|_W = (\sum_{j=1}^p w_j v_j^2)^{1/2}$ 和距离 $d_W(\boldsymbol{x}, \boldsymbol{x}') = \|\boldsymbol{x} - \boldsymbol{x}'\|_W$。

则相应的对偶范数为 $\|\boldsymbol{g}\|_{W^{-1}} = (\sum_{j=1}^p g_j^2 / w_j)^{1/2}$，约定当 $w_j = 0$ 且 $g_j = 0$ 时 $g_j^2 / w_j =$ 为 0。对于目标函数 $f(\cdot)$，其在 d_W 下的 Lipschitz 常数为:

$$L(W) = \sup_{x \neq x'} \frac{|f(\boldsymbol{x}) - f(\boldsymbol{x}')|}{d_W(\boldsymbol{x}, \boldsymbol{x}')}. \tag{43}$$

定理 2a. 对于治愈概率 $S(H \mid \boldsymbol{x}, z) = \sigma(\beta_{z0} + \boldsymbol{\beta}_z^\top \boldsymbol{x})$，Lipschitz 常数满足

$$L_{\mathrm{cure}}(W) \leq \tfrac{1}{4} \|\boldsymbol{\beta}_z\|_{W^{-1}}. \tag{44}$$

在等尺度约束 $\sum_{j=1}^p w_j = c$ 下，该界由治愈权重矩阵最小化

$$\begin{aligned} W_{\mathrm{c},z} &= \mathrm{diag}\left(\frac{c\,|\beta_{z1}|}{\sum_{k=1}^p |\beta_{zk}|}, \ldots, \frac{c\,|\beta_{zp}|}{\sum_{k=1}^p |\beta_{zk}|} \right) \\ &= \mathrm{diag}|\boldsymbol{\beta}_z|. \end{aligned} \tag{45}$$

定理 2a 的证明。 治愈函数的梯度为

$$\nabla_{\boldsymbol{x}} S(H \mid \boldsymbol{x}, z) = \sigma'(\beta_{z0} + \boldsymbol{\beta}_z^\top \boldsymbol{x}) \, \boldsymbol{\beta}_z, \qquad \sigma'(u) = \sigma(u)\{1 - \sigma(u)\} \leq \tfrac{1}{4}. \tag{46}$$

因此,

$$\|\nabla_{\boldsymbol{x}} S(H \mid \boldsymbol{x}, z)\|_{W^{-1}} \leq \tfrac{1}{4} \, \|\boldsymbol{\beta}_z\|_{W^{-1}} = \tfrac{1}{4} \Big(\sum_{j=1}^p \tfrac{\beta_{zj}^2}{w_j} \Big)^{1/2}. \tag{47}$$

根据对偶范数不等式（中值定理与 H¨older 相结合），

$$|S(H \mid \boldsymbol{x}, z) - S(H \mid \boldsymbol{x}', z)| \leq \sup_{\boldsymbol{x}} \|\nabla_{\boldsymbol{x}} S(H \mid \boldsymbol{x}, z)\|_{W^{-1}} \, d_W(\boldsymbol{x}, \boldsymbol{x}'), \tag{48}$$

由此得到 Lipschitz界

$$L_{\mathrm{cure}}(W) \leq \tfrac{1}{4} \Big(\sum_{j=1}^p \tfrac{\beta_{zj}^2}{w_j} \Big)^{1/2}. \tag{49}$$

T要最小化 $\sum_j \beta_{zj}^2 / w_j$，满足 $w_j \geq 0$ 和 $\sum_j w_j = c$，应用柯西–施瓦茨不等式 z:

$$\Big(\sum_{j=1}^p |\beta_{zj}| \Big)^2 = \Big(\sum_j \sqrt{w_j} \cdot \tfrac{|\beta_{zj}|}{\sqrt{w_j}} \Big)^2 \leq \Big(\sum_j w_j \Big) \Big(\sum_j \tfrac{\beta_{zj}^2}{w_j} \Big) = c \sum_j \tfrac{\beta_{zj}^2}{w_j}. \tag{50}$$

因此 $\sum_j \beta_{zj}^2/w_j \geq (\sum_j |\beta_{zj}|)^2/c$，当且仅当 $w_j \propto |\beta_{zj}|$ 时等号成立。施加 $\sum_j w_j = c$ 可确定比例常数，从而得到唯一最小值点

$$w_j^\star = \frac{c\,|\beta_{zj}|}{\sum_{k=1}^p |\beta_{zk}|}, \qquad j = 1, \ldots, p, \tag{51}$$

从而得到最优治愈权重矩阵 $W_{c,\ z}$ 并完成证明。

定理 2b（最优时间权重）。设 $\eta(\boldsymbol{x}) = \lambda_0 + \boldsymbol{\lambda}_z^\top \boldsymbol{x}$，并假设 $(\log T \mid \boldsymbol{X} = \boldsymbol{x}) \sim \mathcal{N}(\eta(\boldsymbol{x})\ \sigma^2)$。将未治愈者期望事件时间记为 $m(\boldsymbol{x}) = \mathbb{E}[T \mid T < H, \boldsymbol{X} = \boldsymbol{x}]$。则存在有限常数 C_{time}，使得

$$L_{\text{time}}(W) \leq C_{\text{time}} \|\boldsymbol{\lambda}_z\|_{W^{-1}}. \tag{52}$$

在等尺度 约束 $\sum_{j=1}^p w_j = c$ 下，该界由时间权重矩阵使其达到最小

$$W_{\text{t},z} = \text{diag}|\boldsymbol{\lambda}_z|. \tag{53}$$

定理 2b 的证明。

在给定 $T < H$ 的条件下，未治愈者的事件时间服从截断对数正态分布：

$$T \sim \text{LogNormal}(\eta(\boldsymbol{x}), \sigma^2) \quad \text{truncated to } (0, H), \tag{54}$$

其中，$\eta(\boldsymbol{x}) = \lambda_{z0} + \boldsymbol{\lambda}_z^\top \boldsymbol{x}$。

其条件累积分布函数为

$$F_{T|T<H}(t \mid \boldsymbol{x}, z) = \frac{\Phi\Big(\frac{\log t - \eta(\boldsymbol{x})}{\sigma}\Big)}{\Phi\Big(\frac{\log H - \eta(\boldsymbol{x})}{\sigma}\Big)}, \qquad t \in (0, H), \tag{55}$$

其中，$\Phi(\cdot)$ 表示标准正态累积分布函数。

根据定义，未治愈者期望事件时间为

$$m(\boldsymbol{x}, z) = \mathbb{E}[T \mid T < H, \boldsymbol{X} = \boldsymbol{x}, Z = z] = e^{\eta(\boldsymbol{x}) + \sigma^2/2}\, \frac{\Phi\Big(\frac{\log H - \eta(\boldsymbol{x}) - \sigma^2}{\sigma}\Big)}{\Phi\Big(\frac{\log H - \eta(\boldsymbol{x})}{\sigma}\Big)}. \tag{56}$$

因此，$m(\boldsymbol{x},\ z)$ 仅通过 $\eta(\boldsymbol{x}) = \lambda_{z0} + \boldsymbol{\lambda}_z^\top \boldsymbol{x}$ 依赖于 $\boldsymbol{x}$。根据链式法则，

$$\nabla_{\boldsymbol{x}} m(\boldsymbol{x}, z) = \frac{\partial m(\eta)}{\partial \eta}\, \boldsymbol{\lambda}_z. \tag{57}$$

正则性。假设 $\boldsymbol{x}$ 取值于紧集 $\mathcal{X}$，使得 $\eta(\boldsymbol{x}) \in [\eta_{min}, \eta_{max}]$，其中 $\eta_{\min} = \inf_{\boldsymbol{x}\in\mathcal{X}} \eta(\boldsymbol{x})$，$\eta_{\max} = \sup_{\boldsymbol{x}\in\mathcal{X}} \eta(\boldsymbol{x})$。

存在一个有限常数

$$C_{\text{time}} = \sup_{\eta(\boldsymbol{x}) \in [\eta_{min}, \eta_{max}]} \left| \frac{\partial m(\eta)}{\partial \eta} \right| < \infty, \tag{58}$$

在这一正则性条件下，

$$\|\nabla_{\boldsymbol{x}} m(\boldsymbol{x}, z)\|_{W^{-1}} = \left| \tfrac{\partial m(\eta)}{\partial \eta} \right| \|\boldsymbol{\lambda}_z\|_{W^{-1}} \ \leq\ C_{\text{time}} \|\boldsymbol{\lambda}_z\|_{W^{-1}}. \tag{59}$$

根据对偶范数不等式（中值定理结合 H¨older），

$$|m(\boldsymbol{x},z)-m(\boldsymbol{x}',z)| \le \sup_{\boldsymbol{x}} \|\nabla_{\boldsymbol{x}} m(\boldsymbol{x},z)\|_{W^{-1}}\, d_W(\boldsymbol{x},\boldsymbol{x}'), \tag{60}$$

这意味着 Lipschitz界

$$L_{\text{time}}(W) \le C_{\text{time}}\|\boldsymbol{\lambda}_z\|_{W^{-1}} = C_{\text{time}}\Big(\sum_{j=1}^{p}\tfrac{\lambda_{zj}^2}{w_j}\Big)^{1/2}. \tag{61}$$

最后，考虑在等尺度约束 $\sum_{j=1}^{p} w_j = c > 0$ 下最小化该 Lipschitz界。应用柯西–施瓦茨不等式，

$$\Big(\sum_{j=1}^{p}|\lambda_{zj}|\Big)^2 = \Big(\sum_{j=1}^{p}\sqrt{w_j}\cdot\tfrac{|\lambda_{zj}|}{\sqrt{w_j}}\Big)^2 \le \Big(\sum_{j=1}^{p} w_j\Big)\Big(\sum_{j=1}^{p}\tfrac{\lambda_{zj}^2}{w_j}\Big) = c\sum_{j=1}^{p}\tfrac{\lambda_{zj}^2}{w_j}. \tag{62}$$

当且仅当 $w_j \propto |\lambda_{zj}|$时，等号成立。施加 $\sum_{j=1}^{p} w_j = c$ 后，可得唯一最小值点

$$w_j^{\star} = \frac{c\,|\lambda_{zj}|}{\sum_{k=1}^{p}|\lambda_{zk}|}, \qquad j=1,\dots,p, \tag{63}$$

该点达到最小 Lipschitz界，并给出最优时间权重矩阵 $W_{\mathrm{t},\ z}$ =diag$|\boldsymbol{\lambda}_z|$。□

B 附加实验

本节将介绍更多模拟研究，以进一步考察所提出框架的稳健性和可推广性。我们首先概述四种实验设置中使用的真实数据生成机制。随后，我们报告不同匹配组大小 ($K = 100$) 下的补充结果，并探讨涉及其他影响治疗分配和结局的混杂变量的情形。这些实验共同对该方法在不同数据复杂度和建模假设下的性能进行了更全面的评估。

B.1 数据生成机制

事件概率模型。对于治疗组 $z \in \{0,1\}$，在时间界限 H 之前发生事件的概率通过逻辑回归建模:

$$\Pr(E_z = 1 \mid \boldsymbol{X}) = \operatorname{expit}(\boldsymbol{X}^{\top}\boldsymbol{\beta}_{E_z}), \tag{64}$$

其中，expit(x) $= 1/(1+e^{-x})$。相应的治愈概率为 $1-\Pr(E_z = 1 \mid \boldsymbol{X})$。治愈概率上的治疗效应源于 $\boldsymbol{\beta}_{E_1}$ 和 $\boldsymbol{\beta}_{E_0}$之间的差异。因此，治愈概率上的真实异质性治疗效应为

$$\begin{aligned}\mathrm{HTE}_{\text{cure}}(\boldsymbol{X}) &= [1-\Pr(E_1=1\mid\boldsymbol{X})] - [1-\Pr(E_0=1\mid\boldsymbol{X})] \\ &= \Pr(E_0=1\mid\boldsymbol{X}) - \Pr(E_1=1\mid\boldsymbol{X}).\end{aligned} \tag{65}$$

未治愈个体的事件时间模型。 给定 $E_z = 1$，事件时间服从截断对数正态加速失效时间（AFT）模型：

$$\log(T_z) \mid (E_z = 1, \boldsymbol{X}) \sim \mathcal{N}(\boldsymbol{X}^\top \boldsymbol{\lambda}_z, \sigma_z^2), \quad T_z \le H, \tag{66}$$

其中， $H = 800$ 表示行政删失时间界限。截断对数正态分布对应的理论均值为

$$\mathbb{E}[T_z \mid \boldsymbol{X}, T_z \le H] = \exp\Big(\boldsymbol{X}^\top \boldsymbol{\lambda}_z + \tfrac{1}{2}\sigma_z^2\Big) \frac{\Phi\Big(\frac{\log H - \boldsymbol{X}^\top \boldsymbol{\lambda}_z - \sigma_z^2}{\sigma_z}\Big)}{\Phi\Big(\frac{\log H - \boldsymbol{X}^\top \boldsymbol{\lambda}_z}{\sigma_z}\Big)}, \tag{67}$$

其中， $\Phi(\cdot)$为标准正态累积分布函数，且我们设定了 $\sigma_z^2 = 1$。事件发生时间的治疗效应由 $\boldsymbol{\lambda}_1$ 与 $\boldsymbol{\lambda}_0$ 之间的差异所致。因此，期望事件时间的真实异质性治疗效应定义为

$$\mathrm{HTE}_{\mathrm{time}}(\boldsymbol{X}) = \mathbb{E}[T_1 \mid \boldsymbol{X}, T_1 \le H] - \mathbb{E}[T_0 \mid \boldsymbol{X}, T_0 \le H]. \tag{68}$$

观测数据生成。 对于每位受试者 $i = 1, \ldots\ldots, n$，观测生存时间 Y_i和事件指示变量 δ_i 按如下方式生成：

$$T_i = Z_i T_{1i} + (1 - Z_i) T_{0i}, \tag{69}$$
$$E_i = Z_i E_{1i} + (1 - Z_i) E_{0i}, \tag{70}$$
$$C_i \sim \mathrm{Uniform}(0, 1.5H), \tag{71}$$
$$Y_i = \min(T_i, C_i, H), \quad \delta_i = \mathbb{I}(T_i < \min(C_i, H)). \tag{72}$$

治疗分配 Z_i 按如下方式生成：

$$\Pr(Z_i = 1 \mid \boldsymbol{X}_i) = \mathrm{expit}(\boldsymbol{X}_i^\top \boldsymbol{\beta}_Z), \tag{73}$$

其中， $\boldsymbol{\beta}_Z$ 控制治疗对协变量的依赖程度。

B.2 四种模拟情景的参数设置

我们考虑四种模拟情景，以代表治疗对治愈概率和事件发生时间产生效应的不同机制。每种情景都规定治疗组如何改变相应的参数向量 $\boldsymbol{\beta}_{E_z}$ 和 $\boldsymbol{\Lambda}_z$。

协变量向量记为

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_{20})^\top,$$

其中 X_1–X_{10} 是从伯努利(0.5)分布中独立生成的二元变量， X_{11}–X_{20} 是从 $\mathcal{N}(0, 1)$中独立抽取的连续变量。

情景特定参数。 在所有情景中，治疗分配模型均相同：

$$\Pr(Z = 1 \mid \boldsymbol{X}) = \mathrm{expit}(0.5 X_{20}), \tag{74}$$

使得 $\boldsymbol{\beta}_Z = (0, \ldots, 0, 0.5)^\top$。

(1) 仅治愈。 治疗只影响事件概率，而治疗组之间的事件时间模型相同。

$$\Pr(E=1 \mid Z=1, \boldsymbol{X}) = \operatorname{expit}\big(0.7 + 0.5X_{17} + 0.5X_{18} + 0.8X_{19}\big), \tag{75}$$
$$\Pr(E=1 \mid Z=0, \boldsymbol{X}) = \operatorname{expit}\big(1.2 + 0.3X_{1} + 0.5X_{2} + 0.2X_{10}\big), \tag{76}$$
$$\log T \mid (E=1, \boldsymbol{X}) \sim \mathcal{N}\big(3.5 + Z + 0.3X_{2} + 0.5X_{19},\, 1^2\big), \tag{77}$$

其中，共享时间模型意味着 $\lambda_{1,\ 0} - \lambda_{0,\ 0} = 1$ （仅截距偏移，无异质性效应）。

(2) 仅时间。 治疗仅影响事件时间模型；各治疗组的治愈概率保持不变。

$$\Pr(E=1 \mid \boldsymbol{X}) = \operatorname{expit}\big(1 + 0.2X_{1} - 0.1X_{2} - 0.8X_{17} + 0.5X_{18} + 0.4X_{19} - 0.8X_{20}\big), \tag{78}$$
$$\log T \mid (E=1, Z=1, \boldsymbol{X}) \sim \mathcal{N}\big(4.5 + 0.4X_{1} - 0.2X_{2} + 0.3X_{3} + 0.8X_{15} - 0.6X_{16} - 0.5X_{17},\, 1^2\big) \tag{79}$$
$$\log T \mid (E=1, Z=0, \boldsymbol{X}) \sim \mathcal{N}\big(3.5 + 0.2X_{1} + 0.2X_{2} + 0.5X_{3} + 0.2X_{15} - 0.1X_{16} - 0.1X_{17},\, 1^2\big). \tag{80}$$

(3) 两者 + 相互独立。 治疗同时影响治愈和时间成分，但两者使用互不重叠的协变量集。

$$\Pr(E=1 \mid Z=1, \boldsymbol{X}) = \operatorname{expit}\big(0.7 + 0.3X_{17} + 0.7X_{18} + 0.2X_{19} - 0.8X_{20}\big), \tag{81}$$
$$\Pr(E=1 \mid Z=0,\ \boldsymbol{X}) = \text{expit函数}\big(1.5\big), \tag{82}$$
$$\log T \mid (E=1, Z=1, \boldsymbol{X}) \sim \mathcal{N}\big(4.5 + 1.2X_{2} - 0.1X_{3} + 0.3X_{15},\, 1^2\big), \tag{83}$$
$$\log T \mid (E=1, Z=0, \boldsymbol{X}) \sim \mathcal{N}\big(3.5 + 0.3X_{2} + 0.2X_{3} + 0.3X_{15},\, 1^2\big). \tag{84}$$

(4) 两者 + 存在重叠。 治疗同时影响两种机制，且部分协变量同时影响治愈和时间成分。

$$\Pr(E=1 \mid Z=1, \boldsymbol{X}) = \operatorname{expit}\big(0.7 + 0.2X_{3} + 0.3X_{17} + 0.7X_{18} + 0.2X_{19} - 0.8X_{20}\big), \tag{85}$$
$$\Pr(E=1 \mid Z=0, \boldsymbol{X}) = \text{expit函数}\big(1.5 - 0.4X_{3}\big), \tag{86}$$
$$\log T \mid (E=1, Z=1, \boldsymbol{X}) \sim \mathcal{N}\big(4.5 + 1.2X_{2} + 0.3X_{15} + 0.5X_{17} - 0.6X_{18},\, 1^2\big), \tag{87}$$
$$\log T \mid (E=1, Z=0, \boldsymbol{X}) \sim \mathcal{N}\big(3.5 + 0.3X_{2} + 0.4X_{3} + 0.3X_{15} + 0.1X_{17} - 0.2X_{18},\, 1^2\big). \tag{88}$$

B.3 当 $K = 100$ 时的结果

为评估所提出的方法对匹配参数 K 选择的稳健性，除了正文中报告的 $K = 50$ 结果外，我们还使用 $K = 100$ 个最近邻重复了所有模拟实验。与主要实验一致，每个模拟数据集包含 $n = 20,000$ 个样本，其中 35%（7，000 个）用于训练混合治愈模型，65%（13，000 个）用于治疗效应估计。表2报告了四种设置下，治愈概率和条件平均事件时间（CMET）的异质性治疗效应（HTE）估计的平均绝对误差（MAE，±标准差）。

总体而言，结果表明，所提出的方法在匹配组较大时仍然稳定，在所有情景下的估计准确度仅有轻微变化。事实上，当增加

K 能够提升有效匹配质量且不会引入过度异质性时，我们的方法往往能获得略高的估计准确度。由于 Cox 模型不依赖匹配，其结果与正文中的结果保持不变。在实践中， K 的最佳选择可能取决于可用样本量，以及偏差缩减与方差控制之间的权衡。

表2：四种设置下治愈概率（×100）和条件平均事件时间的异质性处理效应估计的平均绝对误差（±标准差）$K = 100$

方法	仅治愈		仅时间		两者+独立		两者+重叠	
	Cure	Time	Cure	Time	Cure	Time	Cure	Time
预言机	5.3 ± 0.4	12.3 ± 0.5	4.6 ± 0.2	22.0 ± 0.8	5.2 ± 0.2	16.0 ± 0.6	5.3 ± 0.2	18.2 ± 0.7
部分预言机	5.9 ± 0.2	19.0 ± 0.7	5.1 ± 0.3	27.7 ± 1.0	5.8 ± 0.3	20.8 ± 0.8	6.0 ± 0.3	24.4 ± 0.9
MCM KNN	6.6 ± 0.3	18.1 ± 0.6	5.4 ± 0.3	27.5 ± 1.0	6.6 ± 0.3	21.2 ± 0.8	6.9 ± 0.3	24.8 ± 1.0
MCM KNN（组合）	6.6 ± 0.3	18.3 ± 0.6	5.6 ± 0.3	34.8 ± 1.2	7.0 ± 0.3	24.4 ± 1.0	7.1 ± 0.4	27.5 ± 1.2
特征选择 KNN	7.0 ± 0.3	19.5 ± 0.7	5.7 ± 0.3	38.9 ± 1.5	8.2 ± 0.4	36.3 ± 1.3	8.4 ± 0.4	36.0 ± 1.4
欧几里得近邻算法	8.6 ± 0.3	22.5 ± 0.8	6.0 ± 0.3	44.9 ± 1.8	9.7 ± 0.5	50.1 ± 1.7	10.1 ± 0.5	51.7 ± 1.9
倾向评分 K 近邻	16.3 ± 0.5	30.7 ± 1.0	7.6 ± 0.4	90.7 ± 2.5	18.5 ± 0.6	67.2 ± 2.0	18.5 ± 0.7	85.6 ± 2.8
预后评分 K 近邻	14.3 ± 0.4	29.9 ± 1.0	6.1 ± 0.3	88.0 ± 2.2	12.7 ± 0.4	56.5 ± 1.8	13.6 ± 0.5	67.1 ± 2.4
Cox模型（未匹配）	5.2 ± 0.6	36.2 ± 2.2	8.2 ± 0.7	78.7 ± 3.1	7.7 ± 0.7	58.0 ± 1.3	8.6 ± 0.7	67.5 ± 1.5

B.4 轻度混杂下的额外实验

为评估对治疗–结局混杂的稳健性，我们使用与第 B.2 节相同的治愈模型和事件时间模型重复了所有模拟，但修改了治疗分配机制，使其依赖于额外的结局相关协变量。这会引入轻度混杂，因为某些预测变量会同时影响治疗和结局。具体而言，对于四种模拟设置中的每一种，治疗分配模型如下：

(1) 仅治愈。

$$\Pr(Z = 1 \mid \boldsymbol{X}) = \operatorname{expit}(0.2X_2 + 0.8X_3 + 0.3X_{19} - 0.5X_{20} + 0.5). \tag{89}$$

(2) 仅时间。

$$\Pr(Z = 1 \mid \boldsymbol{X}) = \operatorname{expit}(0.3X_2 - 0.2X_3 + 0.4X_{16} + 0.5X_{17} + 0.5). \tag{90}$$

(3) 两者 +独立。

$$\Pr(Z = 1 \mid \boldsymbol{X}) = \operatorname{expit}(0.3X_3 - 0.2X_4 + 0.4X_{16} + 0.6X_{17} + 0.5). \tag{91}$$

(4) 两者 +重叠。

$$\Pr(Z = 1 \mid \boldsymbol{X}) = \operatorname{expit}(0.4X_3 + 0.3X_{16} + 0.6X_{17} - 0.5X_{18} + 0.5). \tag{92}$$

结果。 表3报告了在这些轻度混杂治疗模型下，使用 $K = 50$ 最近邻对治愈概率和条件平均事件时间（CMET）进行异质性处理效应（HTE）估计的平均绝对误差（MAE， ±标准差）。

与无混杂基线相比，引入轻度混杂如预期那样使平均绝对误差（MAE）小幅上升，但总体性能排序保持不变。所提出的基于MCM的估计量在全部四种设置中均保持接近部分预言机的估计准确度，表明所学距离度量具有较强稳健性。这些结果表明，该方法能够容忍治疗分配与协变量之间的中等程度依赖，而异质性处理效应估计不会出现实质性下降。

表 3：轻度混杂下异质性处理效应估计的平均绝对误差（±标准差）（$K = 50$ ）

方法	仅治愈		仅时间		两者+独立		两者+重叠	
	Cure	Time	Cure	Time	Cure	Time	Cure	Time
预言机	6.6 ± 0.2	17.4 ± 1.3	6.3 ± 0.2	25.6 ± 1.6	6.6 ± 0.3	20.2 ± 1.2	6.7 ± 0.3	22.5 ± 1.4
部分预言机	7.4 ± 0.4	28.3 ± 1.5	7.4 ± 0.3	35.0 ± 1.9	7.7 ± 0.4	29.7 ± 1.6	7.8 ± 0.4	32.5 ± 1.7
MCM KNN	8.0 ± 0.4	26.8 ± 1.2	7.8 ± 0.4	40.6 ± 2.3	8.2 ± 0.5	33.4 ± 1.9	8.1 ± 0.5	35.5 ± 1.9
MCM KNN（组合）	8.2 ± 0.5	27.2 ± 1.3	7.7 ± 0.4	34.7 ± 2.0	8.2 ± 0.5	30.5 ± 1.7	8.6 ± 0.4	33.4 ± 1.8
特征选择 KNN	8.3 ± 0.3	27.6 ± 1.4	7.8 ± 0.5	40.9 ± 2.2	8.5 ± 0.5	33.3 ± 1.8	8.6 ± 0.5	36.1 ± 1.9
欧几里得近邻算法	9.6 ± 0.6	30.6 ± 1.7	8.4 ± 0.6	48.9 ± 2.5	10.5 ± 0.7	53.1 ± 2.4	11.0 ± 0.8	57.2 ± 2.5
倾向评分 K 近邻	16.5 ± 0.8	34.4 ± 1.6	9.6 ± 0.5	81.1 ± 2.7	19.1 ± 0.8	70.4 ± 2.6	18.8 ± 0.8	66.8 ± 2.3
预后评分 K 近邻	14.0 ± 0.7	36.2 ± 1.8	7.7 ± 0.5	91.1 ± 4.1	15.7 ± 0.7	54.9 ± 2.3	16.2 ± 0.8	63.2 ± 2.4
Cox模型（未匹配）	4.0 ± 0.4	39.4 ± 2.3	9.7 ± 0.6	74.4 ± 3.5	7.4 ± 0.5	57.7 ± 2.6	8.4 ± 0.5	69.1 ± 1.7