arXiv:2602.13533v1 [stat.ME] 14 Feb 2026

删失和缺失数据情况下两个层级终点胜率的估计与统计推断

Yi Liu*[1, 2]，Huiman Barnhart[2,3]，Sean O'Brien[2,3]，Yuliya Lokhnygina[2,3]，和Roland A. Matsouaka[2,3]

[1]北卡罗来纳州立大学统计系，美国北卡罗来纳州罗利[2]杜克临床研究所，美国北卡罗来纳州达勒姆[3]杜克大学生物统计与生物信息学系，美国北卡罗来纳州达勒姆

*通讯作者：Yi Liu，yliu297@ncsu.edu

摘要

胜率（WR）是随机临床试验中比较治疗效果的常用指标，适用于按层级排序的终点。基于计数的方法（如Pocock算法）是估计WR的标准方法。然而，该算法无法妥善处理数据删失或缺失的参与者，可能导致估计值存在偏倚且效率低下，尤其是在治疗组间存在异质性删失或缺失数据时。尽管近期已有针对分层生存时间终点的扩展方法解决了部分局限，但除计算量较大的多重插补方法外，现有方法均无法处理包含存在缺失数据的非生存终点的情形。本文提出一种简单的非参数最大似然估计量（NPMLE），用于估计存在删失和缺失数据的两个层级终点的WR。该方法利用所有观测数据，无需强参数假设，并提供闭式渐近方差估计量。我们通过模拟研究以及基于HEART-FID和ISCHEMIA试验的两个数据示例，展示了该方法的性能。所提出的方法可提供一致估计量、提高估计效率，并且在非信息性删失和随机缺失（MAR）假设下具有稳健性，为现有WR估计方法提供了一种灵活的替代方案。我们提供了用户友好的R软件包WinRS，以便于实施。

关键词随机临床试验，复合终点，非信息性删失，随机缺失，非参数最大似然估计。

1 引言

随机临床试验常使用复合终点，全面评估治疗对目标人群的影响。例如，心血管疾病试验通常会评估死亡、卒中、心肌梗死和生活质量等终点。在试验设计阶段将多个终点纳入层级结构，有助于利用更多信息评估治疗疗效和安全性

（Barnhart等人，2025；Finkelstein和Schoenfeld，1999）。胜率（WR）由Pocock等人（2012）提出，是一种有用且恰当的指标，可用于比较具有层级终点的两个治疗组。

Pocock算法被广泛用于估计胜率（WR）；它与Finkelstein-Schoenfeld方法（Finkelstein和Schoenfeld，1999）相似，并通过广义成对比较（GPC）框架，将Wilcoxon-Mann-Whitney检验的范式扩展到多种不同类型的结局（分类、连续、离散等）（Pocock等人，2012；Buyse等人，2025）。该方法已广泛应用于真实世界研究（Abdalla等人，2016；Anker等人，2009；Ferreira等人，2020；Krittayaphong等人，2024；Redfors等人，2020；Maron等，2020；Monzo等人，2024），并且已有易于使用的软件包可供实施（Cui和Huang，2025；Dong等人，2023；Buyse等人，2025）。然而，当数据包含删失值或缺失值（这在纵向研究和事件时间研究中很常见）时，Pocock估计量可能会产生偏倚（Buyse等人，2025）。

假设我们要比较两种治疗方案（以下简称“治疗”与“安慰剂”）在一系列按预先设定的层级从高到低排序的优先级终点上的表现。我们可以从两个治疗组中各选一名参与者组成一对，并依次比较他们的各项终点。Pocock方法的核心思想是，对参与者所有可能的两两比较中的胜、负和平局进行计数，并估计胜率（WR）。从排名最高的终点开始，我们逐对评估治疗组中的参与者表现是更好（即“胜”）还是更差（即“负”）。如果该对结果为平局，或无法进行比较，则继续比较排名第二的终点。这个过程会持续进行，直到决出胜者；如果遍历层级中的最后一个终点后仍无法明确分出胜负，则将该对判定为平局。评估完所有参与者对后，我们统计胜、负和平局的总数，并将胜数与负数之比作为胜率的估计值（Pocock等人，2012）。遍历整个层级后仍无法判定结果的参与者对记为平局，这些平局不用于胜率计算。

显然，如果终点数据存在删失或缺失，某些配对的实际胜负状态可能无法正确判定；因此，Pocock算法可能导致估计的WR出现偏倚。例如，对于至少有一项终点测量缺失的配对，即使数据可用时实际情况可能并非如此，该配对在该终点上仍被视为平局。类似地，对于存在删失的生存时间终点，胜负取决于配对参与者的共同随访时间；因此，我们可能无法恰当地利用该配对的全部观测数据。特别是，当一名参与者在另一名参与者发生事件之前被删失时，该配对会被视为平局，并在后续终点上进行比较。然而，如果不存在删失，这种平局可能并非真正的平局，而可能是实际的胜利或失败。

正如Li 等人(2024)所展示的，即使在最有利的缺失数据机制下，缺失

完全随机缺失（MCAR），即缺失与观测数据和未观测数据均相互独立（Rubin，1976），Pocock方法也可能产生有偏估计。因此，使用Pocock方法对删失或缺失数据进行胜率的一致估计，可能需要对缺失数据机制和删失机制作出强有力的假设，这些假设可能比标准的完全随机缺失和治疗组间非信息性同质删失假设更强。

Pocock对删失和缺失数据的处理可能导致估计量依赖于删失分布或缺失数据机制，因此可能无法一致地估计胜率（WR）。从估计目标的角度看，这种依赖并不可取，因为胜率应反映固定设计和研究人群下的治疗效应，而非特定试验中观察到的删失或缺失模式。这一观点与ICH E9（R1）增补文件一致，该文件强调，在临床试验中，处理缺失和删失数据必须是估计目标框架不可或缺的一部分（Kahan等人，2024）。

多位作者研究了删失和缺失对胜率估计的影响，并提出了可能的解决方法。Dong等人（2020a）指出，当主要终点被删失时可能产生偏倚，尤其是在治疗具有长期获益的情况下，并强调删失时间和随访时长是偏倚的关键影响因素。在此发现的基础上，Dong等人（2020b）提出了一种逆概率删失加权（IPCW）估计量，可在非信息性删失下对胜率进行一致估计，随后又扩展了该方法，将基线和时变协变量纳入其中，以处理依赖于协变量的删失（Dong等人，2021）。Mao （2024）进一步指出，较高的删失率会增加平局数量，使真实胜率比的恢复更加困难。不过，他们提出的解决方案仅适用于层级式生存时间终点。Wang等人（2025a）采用了不同思路，开发了一种联合插补算法，用于插补缺失的纵向结局和临床结局，然后应用基于计数的方法，根据时间范围估计胜率。尽管该方法能够灵活处理多种结局类型，但计算量可能较大，且对插补模型设定错误较为敏感。除Wang等人（2025a）外，现有方法均聚焦于仅由生存结局构成的层级，尚未涉及层级中还包含非生存终点的情形，例如横断面生活质量（QOL）或反应评分。

本文聚焦于一种具体且具有重要实际意义的情形：两个按层级排序的终点，其中第一个是受删失影响的终末事件时间结局，第二个是可能存在缺失的非生存终点。这种结构常见于优先考虑生存结局，同时辅以患者报告或临床结局的试验。例如，HEART-FID 优先考察 12 个月内的死亡时间，其次考察 12 个月时的 6 分钟步行距离。类似地，ISCHEMIA 考察 4 年内的死亡时间，其次考察第 4 年时的西雅图心绞痛问卷 — 心绞痛频率评分（SAQ7-AF）。尽管我们的方法专为这种结构设计，但该框架也可推广到更广泛的情形。

我们引入一个一维汇总指标，即 $\mathcal{S}$-评分，它既保留两个终点的层级排序，又能进行非参数估计。通过将胜率表示为 $\mathcal{S}$-评分分布的函数，我们得到一个代入估计量，无需强建模假设即可同时处理删失和缺失。无需进行两两比较，我们的方法便可避免在存在删失或缺失数据时施加强参数假设。以这种方式构建问题，使我们能够运用 M 估计的一般理论，进而为推导基于影响函数的方差估计量提供了有原则的依据。因此，我们能够为所提出的胜率估计量推导出一个闭式渐近方差估计量。

通过避免多重两两比较所带来的问题以及指定插补模型的需求，我们提出的 $\mathcal{S}$-评分方法提供了一种简单的非参数最大似然估计量（NPMLE），同时对非信息性删失和随机缺失（MAR）机制均具有稳健性。 $\mathcal{S}$-评分方法基于分布，与 Pocock算法等基于计数的方法有本质区别。为便于推广和使用，我们还开发了一个用户友好的 R软件包`WinRS`，可从以下网址免费获取：`https://github.com/yiliu1998/WinRS`。我们的研究方法框架适用于包含两个优先级终点的随机临床研究：优先级最高的生存终点，以及优先级次之的连续型或有序型终点。该框架可直接应用于第一终点满足独立于协变量的非信息性删失，以及第二终点满足随机缺失机制的情况；在后一种情况下，缺失可能依赖于已观测的基线协变量。

本文其余部分安排如下。在第2节，我们介绍统计学预备知识和所提出的方法。在第3节，我们开展广泛的模拟研究，以考察并展示所提出估计量在有限样本下的表现。在第4节，我们通过两个案例研究展示该框架的实用性：HEART-FID 和 ISCHEMIA 临床试验。最后，在第5节，我们对本文进行总结并提出一些看法。我们感谢使用 ChatGPT-5.2 对语言和语法进行润色；在准备本文过程中，我们未以其他方式使用任何大语言模型。

2 方法论

2.1 符号与设定

考虑一项随机研究，其中有两个治疗组， a 和 b，人数分别为 n_a 和 n_b。我们关注的是具有两个层级终点的情形。第一终点 Y_1是一个终末事件时间结局，存在删失，在区间 $[0, h]$内测量；其中， h 表示研究的有限时间范围，所有 Y_1 发生在 h 之后的参与者均视为行政性删失。

第二终点 Y_2在研究结束时（即时间 h）测量。请注意，实际中，参与者通常会在 h前后一个较短的时间窗口内完成随访，并假设 h 的值相同。因此，视为在时间 h测量了 Y_2 。部分参与者的 Y_2 可能缺失。我们假设 Y_2的取值位于有界范围 $[0,\tau]$内，其中 $\tau > 0$为某个有限值；或者，可以通过适当的单调变换将其转换到 $[0,\tau]$ 范围内，使得 Y_1和 Y_2均为正值，且数值越大表示结局越好。

注意， Y_2 如果终末事件发生在随访之前，则无法测量。因此，如果参与者的 Y_2 在 Y_1 研究期间内被观察到，我们不考虑其$[0,h]$. 假设存在的情况。这类似于竞争风险情景（Gooley 等人，1999；Fine 和 Gray，1999；Putter 等人，2007）或多状态模型（Hougaard，1999）中的情形，在这些情形下，终末事件 Y_1 发生后， Y_2 可能无法明确定义，因为 Y_1 的发生排除了测量、观察或经历 Y_2的可能性。

2.2 胜率比估计目标与传统估计方法

对于在研究期间 $[0,h]$可观察到的生存时间终点 Y_1 ，存在相应的事件时间T_1，其中 $T_1 \in (0, \infty)$，使得

$$Y_1 = \min(T_1, h) + I(T_1 > h),$$

其中， h 表示研究的时间范围， $I(\cdot)$为指示函数。该表述给出了 T_1的粗化版本，代表了事件可能直到研究结束仍未发生的实际情形。为方便起见，如果终末事件发生在时间范围h之后，我们记录 $Y_1 = h+1$。如果删失发生在 h之前， Y_1 可能会被右删失，这是一种缺失数据。

我们关注的估计目标是胜率（WR），其基于预先规定的时间范围或随访时间，根据两个层级终点定义 $h < \infty$。该胜率定义明确，将随访时间纳入其中，并对应于Wang等人（2025a）所考察的“限定时间胜率”估计目标。令 Y_{1z} 和 Y_{2z}分别表示组 $z \in \{a,b\}$的第一终点和第二终点。对于随机抽取的一对参与者（一人来自每个组），若 $Y_{1a} > Y_{1b}$，则组 a 胜；反之，若 $Y_{1a} < Y_{1b}$，则组a 负，以上均依据第一终点判定。当两者在 $Y_{1a} = Y_{1b}$上打平时，则使用第二终点进行比较：若 $Y_{2a} > Y_{2b}$，组 a 胜；若 $Y_{2a} < Y_{2b}$，组 a 负。仅当两者在两个终点上均打平时才视为平局，即（$Y_{1a} = Y_{1b}$ 和 $Y_{2a} = Y_{2b}$）。我们关注的胜率定义为组 a 获胜的概率与组 a失败的概率之比，表示为

$$\theta^{\mathrm{WR}} = \frac{P(Y_{1a} > Y_{1b}) + P(Y_{1a} = Y_{1b}, Y_{2a} > Y_{2b})}{P(Y_{1a} < Y_{1b}) + P(Y_{1a} = Y_{1b}, Y_{2a} < Y_{2b})}. \tag{1}$$

该估计目标基于时间范围 h ，能够提供实际、具有意义且易于解读的治疗效应

并针对两个层级终点。虽然该估计目标取决于时间范围，但不取决于 Y_1 的删失分布，也不取决于 Y_2 的缺失数据机制。

对于既无删失也无缺失的数据，可通过 Pocock 配对比较算法（Pocock等人，2012）获得该 WR 估计目标的一致估计量。简要回顾如下：将两个治疗组中一组的每位参与者分别与另一组的每位参与者配对，由此形成 $n_a \times n_b$ 个可能的不同配对。首先根据层级中的第一个终点对每一对参与者进行比较：如果 a组中的参与者表现优于另一位参与者，则将其判定为“胜者”，而另一位参与者属于 b组 a；随后将该配对判定为 a组的一场“胜利”。如果这对参与者在第一个终点上打平，算法便继续比较层级中的下一个终点，依此类推，直到所有配对均被判定为 a组的胜者、败者或平局。随后通过以下公式估计 WR:

$$\widehat{\theta}^{\mathrm{WR}} = \frac{\text{Total number of wins in group } a}{\text{Total number of loses in group } a}. \tag{2}$$

Pocock 估计量的方差通常基于Bebu和Lachin(2016)提出的 U 统计量框架，从而可以构建 Wald 型置信区间。此外，也可采用其他稳健方差估计方法，例如Matsouaka (2022)提出的方法。

在下一节中，我们针对两个层级终点的设定，提出一种估计胜率的简单方法。

2.3 不存在删失或缺失数据时的估计方法

假设不存在删失或缺失数据。考虑以下评分 $\mathcal{S}$，这是一个整合两个终点信息的新随机变量:

$$\mathcal{S} = Y_1 + I(Y_1 > h)Y_2. \tag{3}$$

如附录 A.1 所示，我们有如下等价关系。

$$\{\mathcal{S}_a > \mathcal{S}_b\} = \{Y_{1a} > Y_{1b}\} \cup \{Y_{2a} > Y_{2b}, Y_{1a} = Y_{1b}\} \quad \text{with probability 1.} \tag{4}$$

这种等价关系依赖于以下事实：在我们的设定中，能够在时间 Y_2 h观察到$Y_1 > h$，意味着 Y_1 是终末事件发生的时间（见第 2.1 节）。因此， $\{Y_{2a} > Y_{2b}, Y_{1a} = Y_{1b} \le h\}$ 不可能发生，其概率为零。

注 1. 对于任意固定时间点 $k \le h$，概率 $P(Y_{2a} > Y_{2b}, Y_{1a} = Y_{1b} = k)$ 和 $P(Y_{2a} <$

Y_{2b}，$Y_{1a} = Y_{1b} = k$) 均为零，这是因为事件时间 T_1 在 $[0, h]$ 上服从连续分布，而 $\{Y_{1a} = Y_{1b} = k\}$ 是一个零概率集。因此，事件 $\{Y_{2a} > Y_{2b}, Y_{1a} = Y_{1b} \leq h\}$ 的概率也为零——无论 Y_1 是否为终末事件。这意味着，在连续时间设定下，可将 Y_1 为终末事件这一条件放宽，从而推出方程 (4)，进而拓宽我们的方法的适用范围。

根据（4）中的等价关系，我们有 $P(\mathcal{S}_a > \mathcal{S}_b) = P(Y_{1a} > Y_{1b}) + P(Y_{2a} > Y_{2b}, Y_{1a} = Y_{1b})$。这有助于将（1）中定义的胜率估计目标重新写为

$$\theta^{\mathrm{WR}} = \frac{P(\mathcal{S}_a > \mathcal{S}_b)}{P(\mathcal{S}_b > \mathcal{S}_a)} = \frac{\int \bar{F}_a(s) dF_b(s)}{\int \bar{F}_b(s) dF_a(s)}, \tag{5}$$

其中，对于 $z \in \{a, b\}$，F_z 和 $\bar{F}_z$ 分别是组 z中 $\mathcal{S}$评分的累积分布函数（CDF）和生存函数。因此，估计胜率可以基于两个治疗组中 $\mathcal{S}$评分的估计分布。本文后续将所提出的方法称为“$\mathcal{S}$评分方法”或“$\mathcal{S}$评分估计量”。

根据公式（4），条件 $\mathcal{S}_a > \mathcal{S}_b$ 可作为组 a“获胜”的判据。因此，基于层级终点 Y_1 和 Y_2 判定获胜或失败，可简化为基于单一终点 $\mathcal{S}$判定获胜或失败。$\mathcal{S}$的支撑集为 $(0, h + \tau + 1]$。注意，$\mathcal{S}$的累积分布函数和生存函数在区间 $[h, h + 1]$ 内有跳跃，因为根据定义，Y_1 从 h 跳跃至 $h + 1$。因此，我们定义，对于所有$s \in [h, h + 1)$，当 $z \in \{a, b\}$时，$F_z(s) = F_z(h)$。以完整数据 $\mathcal{O}^{\mathrm{full}} = (Z, Y_1, Y_2)$ 为基础、以 Z 作为治疗组时，θ^{WR} 的简单非参数最大似然估计量为

$$\widehat{\theta}^{\mathrm{WR}}(\mathcal{O}^{\mathrm{full}}) = \frac{\widehat{P}(\mathcal{S}_a > \mathcal{S}_b)}{\widehat{P}(\mathcal{S}_b > \mathcal{S}_a)} = \frac{\int_0^{h+\tau+1} \{1 - \widehat{F}_a(s)\} d\widehat{F}_b(s)}{\int_0^{h+\tau+1} \{1 - \widehat{F}_b(s)\} d\widehat{F}_a(s)}, \tag{6}$$

其中，$\widehat{F}_a$ 和 $\widehat{F}_b$ 是累积分布函数（Kaplan 和 Meier，1958）的 Kaplan-Meier 估计，将 $\mathcal{S}$ 视为事件时间变量，其中所有参与者均在其观测值 $\mathcal{S}$处发生事件。具体而言，

$$\widehat{F}_z(s) = 1 - \prod_{s_i^{(z)} \leq s} \left(1 - \frac{d_i^{(z)}}{n_i^{(z)}}\right), \tag{7}$$

其中，$s_i^{(z)}$ 表示来自以 i为索引的数据中被视为“事件时间”的 $\mathcal{S}$ 的不同观测值，$d_i^{(z)}$ 表示在时间 $s_i^{(z)}$ 发生的事件数，$n_i^{(z)}$ 表示在组 z中时间 $s_i^{(z)}$ 之前紧邻时点的风险集个体数，其中 $z \in \{a, b\}$。

所提出的估计量（6）与 Pocock估计量（2）的不同之处在于，前者是基于分布的方法，而非基于计数的方法。因此，可以利用M估计理论得到基于影响函数的方差估计量。

迄今为止，我们提出的估计量适用于完全观测数据 $\mathcal{O}^{\text{full}}$。然而在实践中，更现实的情形是： Y_1 在时间范围 h之前发生删失，且 Y_2 在随访时间存在缺失数据。为解决这一问题，Pocock方法通过采用配对参与者的共同随访时间，在 Y_1 中进行成对比较，以处理删失；如果其中一人存在缺失数据，则在 Y_2 中将配对参与者视为平局。这种方法可能产生有偏估计，因为基于共同随访时间（$< h$）在 Y_1 中进行的胜/负/平分类可能因删失而不同于假设两位参与者均观察至时间 h时所得到的分类。此外，在 Y_2中因缺失数据而看似平局的两位参与者，如果其 Y_2 值均被完整观测，也可能并非平局。Pocock方法可能隐含依赖于关于删失和缺失机制的强假设，这些假设可能比针对 Y_1 的标准非信息性删失假设和针对 Y_2 的完全随机缺失（MCAR）假设更强（Li 等人，2024；Mao，2024）。尽管在生存分析中，针对 Y_1 的独立删失假设被广泛接受（Cox，1972；Kaplan 和 Meier，1958），尤其是在随机研究中，但针对 Y_2 的完全随机缺失（MCAR）假设往往并不现实。正如 Mao（2024）所指出的，Pocock方法对生存终点的删失率较为敏感。

在下一节中，我们将所提出的 $\mathcal{S}$-评分方法加以扩展，用以估计我们的设定下 Y_1 中存在删失数据且 Y_2 中存在缺失数据时的胜率。

2.4 存在删失或缺失数据时的估计方法

假设 Y_1 在 h 之前被右删失，其删失随机变量为 C_1 ，并令

$$\widetilde{Y}_1 = \min(Y_1, C_1, h) + I(Y_1 > h, C_1 \geq h) \text{ with } \Delta_1 = I(Y_1 \leq C_1, Y_1 \leq h) + I(Y_1 > h, C_1 \geq h).$$

~ Y_1 的观测值可由(Y_1, Δ_1), 表示，其中 $\Delta_1 = 0$ 表示区间 $[0, h]$内发生右删失。注意， $Y_1 = h + 1$ 表示因 h 的时间范围所致的行政删失，此处不将其视为删失。对于 Y_2中的缺失数据，令 R_2 为非缺失指示变量，即当 Y_2 被观测到时为 $R_2 = 1$，否则为 $R_2 = 0$ 。注意，若终末事件发生在 h之前，则 Y_2 自然缺失。

~ ~ ~我们将完整数据的粗化水平版本记为 $\mathcal{O} = (Z, Y_1, \Delta_1, Y_2, R_2)$，其中 $Y_2 = Y_2 R_2$，并作出以下两个假设。

假设1 （非信息性删失）。 在每个治疗组内，删失均为非信息性的，其含义为 $C_1 \perp\!\!\!\perp (Y_1, Y_2) \mid Z = z$, 其中 $\perp\!\!\!\perp$ 表示“独立性”。

假设2（随机缺失（随机缺失））。 对于非生存终点 Y_2，时间范围 h之后的缺失数据为随机缺失，即 $Y_2 \perp\!\!\!\perp R_2 \mid Y_1 > h, Z = z$。

在生存分析文献中，通常假设删失 C_1 与事件时间 Y_1 相互独立（Kaplan 和 Meier，1958；Cox，1972），以确保可以基于非参数 Kaplan-Meier 估计量识别事件的生存函数。更一般地，在假设 1中，我们允许不同治疗组的删失机制有所不同。在假设 $C_1 \perp\!\!\!\perp Y_2 \mid Z = z$ 1 中纳入该条件是为了确保样本中 $Y_1 > h$ 的部分仍具有代表性，可用于估计超出 $\mathcal{S}$-评分的 h分布。例如，当 $C_1 < h$ 但 $Y_1 > h$时， Y_2 可能存在但未被观测到。如果 Y_2依赖于 C_1，那么在估计 $\mathcal{S}$的 $s \geq h$-评分累积分布函数时就会引入偏倚。

我们还提出随机缺失假设（假设 2），即在治疗组归属条件（条件 $Y_1 > h$）下，只要尚未发生终末事件（条件 $Z = z$），任何缺失数据均为随机缺失。条件 $Y_1 > h$ 表示当 Y_2 时，可以在随访时间测量。

假设2本质上与研究设计中的“单调缺失数据”概念相符，该概念指一种缺失数据模式：某个变量出现缺失意味着所有后续变量（按预先设定的顺序）也都缺失（Tsiatis，2007）。在我们的情境中，当 Y_1 在 h之前被删失（这也是一种缺失）时， Y_2 会自动缺失（根据一种与 Y_1 删失无关的机制，依据假设1）。因此，我们仅讨论参与者中 Y_2 的缺失数据模式，且这些参与者满足 $Y_1 > h$。

根据观测数据，我们将基于完整数据、在式（3）中定义的 $\mathcal{S}$-评分从 $\mathcal{O}^{\text{full}}$ 扩展到 $\mathcal{O}$。我们 ~ ~ ~ ~引入以下粗化的 $\mathcal{S}$，将其视为“右删失”的生存终点。记 $\mathcal{S} = Y_1 + I\ (Y_1 > h)\ Y_2$ ~ ~，并将“事件”指示变量记为 $\Delta_{\mathcal{S}} = I\ (\Delta_1 = 1,\ Y_1 \leq h)\ + I\ (R_2 = 1,\ Y_1 = h + 1)$。注意，以概率~ ~1， $\{\Delta_1 = 1, Y_1 \leq h\}$ 与 $\{R_2 = 1, Y_1 = h + 1\}$ 是互斥事件，因此 $\Delta_{\mathcal{S}} \in \{0, 1\}$。因此，我们的观测数据可改写为 $\mathcal{O}^{\mathcal{S}} = (Z,\ \widetilde{\mathcal{S}},\ \Delta_{\mathcal{S}})$。我们在图 1 中总结了从两个终点到 $\mathcal{S}$-评分的转换。

~注意， $\mathcal{S}$ 是 $\mathcal{S}$ 的观测版本，其中 $\mathcal{S}$ 的“删失”基于：(i) Y_1的右删失状态，以及 (ii) Y_2的缺失。在假设 1 和假设2 下，删失是非信息性的。因此，基于 $\mathcal{O}^{\mathcal{S}}$ 计算的各组 Kaplan-Meier估计量，是 $\mathcal{S}$各组生存函数的无偏估计量。因此，对于存在删失和缺失的数据，我们得到以下简单的 WR 非参数最大似然估计量：

$$\widehat{\theta}^{\text{WR}}(\mathcal{O}^{\mathcal{S}}) = \frac{\widehat{P}(\mathcal{S}_a > \mathcal{S}_b)}{\widehat{P}(\mathcal{S}_b > \mathcal{S}_a)} = \frac{\displaystyle\int_0^{h+\tau+1} \{1 - \widehat{F}_a(s)\} d\widehat{F}_b(s)}{\displaystyle\int_0^{h+\tau+1} \{1 - \widehat{F}_b(s)\} d\widehat{F}_a(s)}, \tag{8}$$

其中， $\widehat{F}_a$ 和 $\widehat{F}_b$ 分别是基于 $\mathcal{O}^{\mathcal{S}}$ 、针对每个组的 $\mathcal{S}$ 的累积分布函数的 Kaplan-Meier 估计。该

公式（8）中的估计量被证明对于 θ^{WR} 在附录 A.2中是一致的。

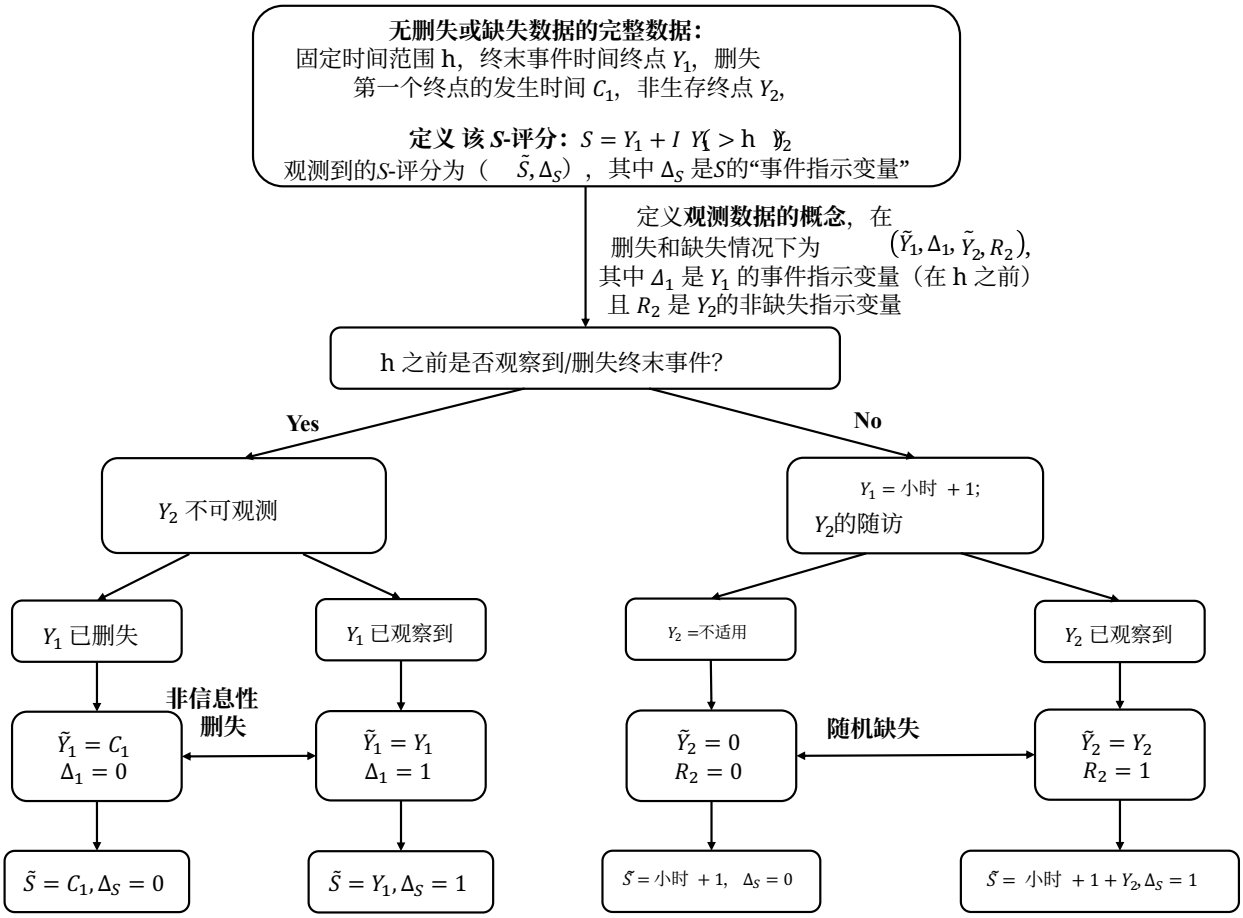


图 1：所提出的 $\mathcal{S}$-评分方法流程示意图。

注释 2. 我们根据假设 $\mathcal{S}$-评分时，隐含地假设了非信息性删失。根据假设 1 和 2，以及 $\mathcal{S}$ 和 $\Delta_{\mathcal{S}}$的定义，我们可以作出这一假设。具体而言，我们可以为 $C_{\mathcal{S}}$的 $\mathcal{S}$-评分定义一个“删失”变量，如下所示：

$$C_{\mathcal{S}} = \begin{cases} C_1, & \text{if } C_1 < h, \\ h + 0.5, & \text{if } C_1 \geq h \text{ and } R_2 = 0, \\ h + \tau + 1 + 0.5, & \text{if } C_1 \geq h \text{ and } R_2 = 1. \end{cases}$$

在假设1 和 2下， $\mathcal{S} \perp\!\!\!\perp C_{\mathcal{S}} \mid Z = z$因此，基于变量 $\tilde{\mathcal{S}} = \min(\mathcal{S}, C_{\mathcal{S}})$ 且满足 $\Delta_{\mathcal{S}} = I(\mathcal{S} < C_{\mathcal{S}})$ 的*Kaplan-Meier*估计量可用于(8)。

2.5 方差估计

为了对所提出的非参数最大似然估计量（8）进行统计推断，我们需要为估计量（6）构造适当的方差估计量。常用的方法是非参数自助法（Efron和Tibshirani，1994年），该方法以有效性和准确性著称。鉴于所提出的估计量（6）是正则且渐近线性（RAL）的，并且相对于 $\mathcal{S}$-评分函数是光滑的（除了生存函数中的某些跳跃点外，以概率1成立），自助法方法能够提供准确的方差估计，这一点由Shao和Tu (2012)加以论证。通过自助法得到的置信区间（置信区间）可以用多种方式构造。例如，可以利用多个自助法样本得到的估计方差来构造Wald型置信区间。或者，也可以将自助法分布的 $\alpha/2$ 和$1-\alpha/2$分位数作为界限来构造置信区间。

然而，自助法的一个显著缺点是计算负担较重，且会随着样本量增加而加重。因此，通常会利用渐近理论，在点估计量之外推导出一个闭式解析方差估计量，以便快速估计不确定性。本文利用影响函数（IF）理论（Van der Vaart，2000；Stefanski 和 Boos，2002年）为（6）提出一个渐近方差估计量。影响函数记为$\varphi(\mathcal{O}^{\mathcal{S}};\ F_a,\ F_b)$，它取决于观测数据以及两个治疗组的 $\mathcal{S}$-评分累积分布函数。（6）的影响函数公式推导见附录A.3。此外，影响函数 $\varphi(\mathcal{O}^{\mathcal{S}};\ F_a,\ F_b)$满足

$$\widehat{\theta}^{\mathrm{WR}}=\widehat{\theta}^{\mathrm{WR}}(\mathcal{O}^{\mathcal{S}})=\theta^{\mathrm{WR}}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(\mathcal{O}_i^{\mathcal{S}};F_a,F_b)+o_p(n^{-1/2}).$$

由此，我们可以利用标准中心极限定理近似渐近方差，得到

$$\widehat{\mathcal{V}}_{\mathrm{IF}}=\widehat{\mathcal{V}}_{\mathrm{IF}}(\widehat{\theta}^{\mathrm{WR}})=\frac{1}{n}\sum_{i=1}^{n}\varphi(\mathcal{O}_i^{\mathcal{S}};\widehat{F}_a,\widehat{F}_b)^2.$$

换言之，我们可以证明以下一致性和渐近正态性（见附录A.3），

$$\sqrt{n/\widehat{\mathcal{V}}_{\mathrm{IF}}}(\widehat{\theta}^{\mathrm{WR}}-\theta^{\mathrm{WR}})\xrightarrow{d}\mathcal{N}(0,1).$$

2.6 纳入基线协变量的扩展

在本节中，我们将假设2 扩展为允许缺失数据机制除了取决于组别指示变量外，还取决于基线协变量。利用基线协变量向量$\mathbf{X}$，我们提出以下假设：

假设3（协变量随机缺失）。 $R_2 \perp\!\!\!\perp Y_2 \mid Y_1 > h,\ Z = z,\ \mathbf{X}$。

假设 3 推广了假设2，因为它允许个体特异性缺失模式。在假设 3下，一致估计需要进行协变量调整。相比之下，若假设2成立，则无需进行协变量调整也能实现一致估计。不过，使用协变量对缺失数据建模可能有助于提高效率（Liuet al.，2025b；Gao et al.，2024）。

在假设1 和3下，可利用M估计理论（Van derVaart，2000）通过求解一组无偏估计方程来估计胜率。具体而言，我们构建一组堆叠的估计方程，其中纳入了用于估计 Y_1 的累积分布函数的Kaplan–Meier估计量、针对 Y_2 缺失机制的治疗特异性回归模型参数，以及针对 Y_2 的协变量调整的累积分布函数估计量。缺失机制和 Y_2 的累积分布函数均仅基于满足$Y_1 > h$ 的数据子集进行估计。关于估计方程、由此得到的协变量调整点估计量及其渐近性质的完整技术细节，见附录A.4。

尽管我们将方程（8）称为非参数最大似然估计量（NPMLE），但更准确地说，协变量调整版本是一种半参数估计量，因为它纳入了一个针对缺失机制的参数模型，该模型由向量 $\boldsymbol{\beta}_z$ （其维度与 $\mathbf{X}$ 相同）进行参数化，用于 $z = a,\ b$本质上，这是一个包含协变量主效应、治疗指示变量，以及基线协变量与治疗指示变量交互效应的模型。与第 2.5 节中非参数最大似然估计量的方差估计量类似，我们为协变量调整半参数估计量推导了基于影响函数的方差估计量（见附录 A.4）。这一协变量调整估计量及其基于影响函数的方差估计量也已在我们的 R 软件包 `WinRS` 中实现。

3 模拟研究

我们进行了模拟研究，将所提出的方法与Pocock等人（2012）和Wang等人（2025a）的方法，以及逆概率删失加权方法（Dong等人（2020b）；该方法仅适用于生存终点）进行比较，并分别简称为“Pocock”“Wang”和“IPCW”。对于第3.1 节所述的数据生成过程，我们模拟了两组蒙特卡洛数据。首先，在每组规模为 $n_{\text{super}} = 10^7$ 的超总体中，为 $z \in \{a, b\}$ 生成完整数据$\mathcal{O}^{\text{full},\ z} = (Z = z,\ Y_1,\ Y_2)$。从合并后的样本池 $\mathcal{O}^{\text{full},\ a} \cup \mathcal{O}^{\text{full},\ b}$中，随机抽取 10^7 对独立配对，每对各包含来自每个组的一名参与者。随后，根据这些配对之间的成对比较，统计组 a 和组 b的获胜次数，从而计算胜率。鉴于样本量大、比较次数多且数据完整（即没有删失或缺失数据），估计真实胜率的不确定性可忽略不计。借助这一大样本，我们可以计算“真实”胜率，并用其评估下文第二组蒙特卡洛数据中用于比较不同估计方法的性能指标。

~ ~其次，我们模拟了2，000个独立的观测数据集 $\mathcal{O}^z = (Z = z,\ Y_1,\ \Delta_1,\ Y_2,\ R_2)$，在 $z \in \{a, b\}$中按1:1的比例随机分配，并采用不同样本量 $(n_a,\ n_b) \in \{(100,\ 100),\ (1000,\ 1000)\}$ 以代表小型和大型试验。对于每次重复模拟，我们使用上述四种方法（Pocock、Wang、IPCW和我们的 $\mathcal{S}$ 评分）估计胜率。

Pocock方法通过`WinRatio`R软件包（Duarte和Ferreira，2022）实现，使用默认的非匹配对算法及其内置的置信区间估计量。

我们遵循Wang等人（2025a）附录 A.1 中关于右删失事件时间和连续型结局进行多重插补的说明，以实现其方法（先插补后计数）。按照其模拟研究框架，我们选择对50个插补数据集进行插补，并将Pocock算法应用于每个插补数据集。

令 $\widehat{\theta}_i^{\mathrm{WR}}$和 $\widehat{\mathcal{V}}_i$分别表示第 i个插补数据集的点估计值和方差估计值，$i = 1,\ \ldots\ldots,\ 50$。相应的点估计值和方差可估计为

$$\bar{\theta}^{\mathrm{WR}} = \frac{1}{50}\sum_{i=1}^{50}\widehat{\theta}_i^{\mathrm{WR}} \quad \text{and} \quad \widehat{\mathcal{V}}_{\mathrm{pooled}} = \widehat{\mathcal{V}}_{\mathrm{within}} + \left(1 + \frac{1}{50}\right)\widehat{\mathcal{V}}_{\mathrm{between}},$$

其中 $\widehat{\mathcal{V}}_{\mathrm{within}} = \dfrac{1}{50}\sum_{i=1}^{50}\widehat{\mathcal{V}}_i, \quad \widehat{\mathcal{V}}_{\mathrm{between}} = \dfrac{1}{49}\sum_{i=1}^{50}\left(\widehat{\theta}_i^{\mathrm{WR}} - \bar{\theta}^{\mathrm{WR}}\right)^2.$

^随后，我们基于 $\mathcal{V}_{\mathrm{pooled}}$构造 Wald 型置信区间。

对于Dong等人（2020b）提出的逆概率删失加权方法，我们注意到，该方法最初是为包含一个或多个删失生存结局的情形开发的。尽管这种情形与我们的研究有所不同，该方法仍在`WINS` R软件包中实现，可用于我们的模拟。具体而言，我们使用`win.ratio()` 函数实现该方法，将第一个终点类型指定为`"tte"`（事件时间），第二个指定为`"continuous"`，并将估计方法设为`"ipcw"`。对于连续型终点，该软件包在存在缺失数据时采用完整病例分析。由于Wang方法和逆概率删失加权方法的计算量都很大，我们仅在正文中展示的情景下应用这两种方法，而未在在线补充材料（网络附录S.1）中展示的情景下应用。

对于我们的 $\mathcal{S}$评分方法，我们使用两种不同的 Wald 型置信区间为 θ^{WR} 构造置信区间：一种基于由影响函数推导出的闭式方差估计量（IF-Wald，第2.5节）；另一种使用 1,000 次非参数自助法重复抽样估计的标准误（BT-Wald）。我们还考虑一种基于百分位数的置信区间，使用自助法估计值的 2.5% 和 97.5% 分位数（BT-QT）。

我们使用以下四项指标评估了WR估计方法在2,000个模拟数据集上的表现：

- 绝对再　　绝对相对偏倚（ARB%）：$(\widehat{\theta}^{\mathrm{WR}} - \theta^{\mathrm{WR}})\%/\theta^{\mathrm{WR}}$ 的均值的绝对值。

- 均方根误差（RMSE）：$(\widehat{\theta}_{\text{胜率}} - \theta_{\text{胜率}})^2$ 均值的平方根。

- 覆盖概率（CP%）：真实的 $\theta_{\text{胜率}}$ 落在构造的95%置信区间内的次数所占百分比。由于蒙特卡罗误差，若CP%处于$95 \pm 1.96 \times \sqrt{95 \times (100-95)/2,\ 000} = [94, 96]$ 则认为其与名义95%水平~~没有显著差异~~。

- 置信区间宽度：量化各方差估计量之间差异的95%置信区间的平均宽度。

3.1 数据生成过程

我们在不同情景下生成数据，这些情景涵盖样本量、 Y_1的右删失模式、 Y_2 的缺失数据模式，以及 $\theta^{\text{WR}} = 1$ 或2的治疗效应，以全面评估所提出的方法。

对于每个 $z \in \{a, b\}$，我们均从伽马分布 $\Gamma(\alpha,\ \lambda)$生成事件时间 $T_{1z} \sim \Gamma(\alpha_{Tz},\ \lambda_{Tz})$和删失时间 $C_{1z} \sim \Gamma(\alpha_{Cz},\ \lambda_{Cz})$，其中 α和 λ分别为形状参数和尺度参数。给定 $h = 90$后，我们将观测时间生成为 $Y_{1z} = \min(T_{1z},\ h) + I(T_{1z} > h)$，并将事件指示变量生成为 $\Delta_{1z} = I(Y_{1z} \le C_{1z},\ Y_{1z} \le h) + I(Y_{1z} > h,\ C_{1z} > h)$。

在 $\tau = \max(Y_2) = 50$的基础上，我们还生成了 $Y_{2z} \sim I(Y_{1z} > h) \cdot \min\{\max(\mathcal{N}(\mu_z,\ \sigma_z^2),\ 0)\ \tau\}$，并将缺失指示变量设为 $R_{2z}^* \sim I(T_{1z} \le h) + I(T_{1z} > h) \cdot$ 伯努利分布(β_z)（因此，非缺失指示变量为 $R_{2z} = 1 - R_{2z}^*$）。也就是说，我们仅为具有 $Y_1 > h$ 的个体生成了 Y_2，因为 Y_1 是终末事件，只有在 h之后才能测量 Y_2，而 Y_2 可能会根据 β_z的取值而缺失。

不同模拟情景所使用的参数值如下：

- 治疗效应：$\theta_{\text{胜率}} = 1$ 对于 $\alpha_{Ta} = \alpha_{Tb} = 2.5$、$\lambda_{Ta} = \lambda_{Tb} = 0.04$、$\mu_a = \mu_b = 10$和 $\sigma_a = \sigma_b = 10$；以及 $\theta_{\text{胜率}} = 2$，当 $\alpha_{Ta} = 2.5$，$\alpha_{Tb} = 4$、$\lambda_{Ta} = 0.04$，$\lambda_{Tb} = 0.10$、$\mu_a = 10$，$\mu_b = 20$、$\sigma_a = 10$和 $\sigma_b = 20$时。

- Y_1 的右删失模式：

  – 无删失数据：代码中的 $C_1 \equiv h + 1$（其值始终大于 Y_1）；– 按治疗组同质删失：低（中等）删失率分别为 $\alpha_{Ca} = \alpha_{Cb} = 1.8$、$\lambda_{Ca} = \lambda_{Cb} = 0.01$（0.02），总体上两组的删失观测约占 20%（40%）；–按治疗组异质性删失：$(\alpha_{Ca}, \alpha_{Cb}) = (3.2, 1.5)$，$(\lambda_{Ca}, \lambda_{Cb}) = (0.04, 0.08)$（低（中等）删失率分别为 (0.02,0.05)，总体上两组的删失观测分别占 20%（40%），但各组的删失率不同。

- Y_2的缺失模式：

  – 无缺失数据：$\beta_a = \beta_b = 0$；

– 完全随机缺失：我们考虑两种情况（β_a，β_b）$=(0.2, 0.2)$ 和 $(0.4,0.4)$，这两种情况分别导致 Y_2 中约有 20% 和 40% 的数据缺失，且两组的缺失数据机制相同；– 随机缺失：我们考虑（β_a，β_b）$=(0.15, 0.25)$ 和 $(0.3,0.5)$，两组总体缺失数据比例分别约为 20% 和 40%。然而，在这些情况下，治疗组和对照组的缺失数据比例不同。也就是说，缺失数据机制取决于治疗分配，并非完全随机。

总体而言，我们的实验考虑了 $2\times5\times5\times2=100$ 种情景，以评估三种竞争性点估计方法（$\mathcal{S}$-评分、Pocock方法、Wang方法和逆概率删失加权）以及三种用于 $\mathcal{S}$-评分方法的置信区间估计量（IF-Wald、BT-Wald和BT-QT）的表现。

3.2 结果

我们仅在表1中展示部分模拟结果，重点关注样本量较大（$n_a=n_b=1000$）、平均删失率和缺失数据率适中（40%）的情景。其余结果均见在线补充材料（网络附录S.1）。

在 Y_1删失的三种情形中，存在缺失的 Y_2 时，除 $\theta^{\mathrm{WR}}=1$ 在完全随机缺失下的情形外，无论是完全随机缺失还是随机缺失，Pocock方法均产生较大的偏倚，且覆盖概率较低。对于 Y_1 和 Y_2 均完整观测的数据（最上方情形），我们的 $\mathcal{S}$-评分方法与Pocock方法产生相同的绝对相对偏倚和均方根误差，方差估计仅有细微差异，这证实了二者在完整数据下点估计的等价性。当 Y_1 在各治疗组间发生删失（无论同质还是异质）时，Pocock方法对 $\theta^{\mathrm{WR}}=1$ 和2的均方根误差均较大；当 $\theta^{\mathrm{WR}}=2$时，覆盖概率较低。此外，当 $\theta^{\mathrm{WR}}=2$ 时，只要 Y_1 存在删失数据，或 Y_2存在缺失数据，Pocock方法就会产生偏倚，而我们的方法在所有情形下均保持一致性并能进行有效推断。

此外，对于为我们的方法构建 95%置信区间的三种方法（IF-Wald、BT-Wald 和 BT-QT），我们观察到它们在覆盖概率和置信区间宽度方面差异极小，这支持了我们提出的基于影响函数的方差估计量的有效性，也表明这三种置信区间方法均可用于量化不确定性。我们观察到，所有基于 $\mathcal{S}$-评分的方法的覆盖概率都非常接近 95% 的名义水平，且均处于$[94,\ 96]$ 范围内，但有三种情况除外（96.40、96.40 和 93.85，表中以粗体标出），这些结果也接近 95%。

我们观察到，当 Y_1 被删失时，Wang 方法可能存在偏倚，且覆盖概率较差；但当 Y_1 未被删失时，无论 Y_2的缺失机制如何，该方法总体上仍较为稳健。这可能是因为插补模型始终假设删失时间服从指数分布（但此处数据是基于伽马分布生成的），同时允许对缺失

表1：部分模拟结果，其中（n_a， n_b）=(1000，1000)。

删失 Y_1的	缺失 Y_2的	方法	$\theta^{\mathrm{WR}}=1$ ARB%	RMSE	CP%	宽度	$\theta^{\mathrm{WR}}=2$ ARB%	RMSE	CP%	宽度
无删失 Y_1	No	$\mathcal{S}$-评分（IF-Wald）	0.37	0.052	95.00	0.20	0.30	0.114	95.00	0.45
		$\mathcal{S}$-评分（BT-Wald）	0.37	0.052	94.50	0.20	0.30	0.114	95.05	0.45
		$\mathcal{S}$-评分（BT-QT）	0.37	0.052	94.00	0.20	0.30	0.114	94.35	0.45
		Pocock	0.37	0.052	94.65	0.20	0.30	0.114	94.90	0.45
		逆概率删失加权*	0.37	0.052	**96.20**	0.20	0.30	0.114	**96.60**	0.47
	MCAR（40%）	$\mathcal{S}$-评分（IF-Wald）	0.10	0.052	95.05	0.20	0.01	0.113	95.45	0.45
		$\mathcal{S}$-评分（BT-Wald）	0.10	0.052	94.80	0.20	0.01	0.113	95.50	0.45
		$\mathcal{S}$-评分（BT-QT）	0.10	0.052	94.50	0.20	0.01	0.113	95.00	0.45
		Pocock	0.11	0.054	95.45	0.21	9.67	0.226	**53.30**	0.41
		Wang	0.11	0.053	94.55	0.20	0.08	0.113	95.30	0.45
		逆概率删失加权*	0.28	0.052	96.00	0.21	9.66	0.227	**57.60**	0.43
	MAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.48	0.051	**96.40**	0.20	0.08	0.110	95.90	0.45
		$\mathcal{S}$-评分 (BT-Wald)	0.48	0.051	**96.40**	0.20	0.08	0.110	95.85	0.45
		$\mathcal{S}$-评分 (BT-QT)	0.48	0.051	95.85	0.20	0.08	0.110	95.70	0.45
		Pocock	14.57	0.152	**15.95**	0.18	16.80	0.362	**8.70**	0.38
		Wang	0.89	0.052	**96.30**	0.20	0.11	0.115	95.90	0.45
		逆概率删失加权*	14.11	0.148	**19.00**	0.18	16.99	0.367	**10.80**	0.39
40% 删失 Y_1，同质的组间	No	$\mathcal{S}$-评分 (IF-Wald)	0.27	0.061	94.15	0.24	0.14	0.131	94.15	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.27	0.061	94.30	0.24	0.14	0.131	94.05	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.27	0.061	94.20	0.23	0.14	0.131	**93.85**	0.49
		Pocock	0.33	0.068	94.55	0.26	2.70	0.152	**92.05**	0.53
		Wang	0.37	0.063	**93.95**	0.24	6.39	0.183	**76.40**	0.47
		逆概率删失加权*	0.32	0.071	**93.60**	0.26	2.77	0.143	94.60	0.54
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.03	0.059	94.70	0.24	0.16	0.126	94.80	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.03	0.059	94.70	0.24	0.16	0.126	95.05	0.50
		$\mathcal{S}$-评分（BT-QT）	0.03	0.059	94.65	0.23	0.16	0.126	94.80	0.49
		Pocock	0.05	0.069	94.80	0.27	7.57	0.204	**77.30**	0.51
		Wang	0.00	0.062	**93.85**	0.23	6.13	0.176	**77.80**	0.47
		逆概率删失加权*	0.27	0.068	94.80	0.27	7.12	0.198	**80.40**	0.52
	MAR（40%）	$\mathcal{S}$-评分（IF-Wald）	0.56	0.059	94.90	0.24	0.22	0.126	94.90	0.50
		$\mathcal{S}$-评分（BT-Wald）	0.56	0.059	95.10	0.24	0.22	0.126	95.35	0.50
		$\mathcal{S}$-评分（BT-QT）	0.56	0.059	94.90	0.24	0.22	0.126	94.85	0.49
		Pocock	8.66	0.107	**73.65**	0.25	11.20	0.265	**55.85**	0.49
		Wang	0.39	0.062	94.10	0.23	6.06	0.175	**77.80**	0.47
		逆概率删失加权*	8.30	0.105	**75.60**	0.25	10.92	0.261	**59.60**	0.50
40% 删失 Y_1，异质组间（20% 对 60%）	No	$\mathcal{S}$-评分（IF-Wald）	0.14	0.060	95.45	0.24	0.13	0.125	94.90	0.50
		$\mathcal{S}$-评分（BT-Wald）	0.14	0.060	95.45	0.24	0.13	0.125	95.25	0.50
		$\mathcal{S}$-评分（BT-QT）	0.14	0.060	95.65	0.24	0.13	0.125	95.35	0.50
		Pocock	7.95	0.102	**76.35**	0.25	5.92	0.180	**84.50**	0.53
		Wang	20.79	0.222	**17.15**	0.29	7.89	0.218	**83.65**	0.55
		逆概率删失加权*	8.05	0.099	**77.20**	0.25	6.34	0.182	**84.60**	0.53
	MCAR（40%）	$\mathcal{S}$-评分（IF-Wald）	0.27	0.061	95.40	0.24	0.30	0.128	95.20	0.50
		$\mathcal{S}$-评分（BT-Wald）	0.27	0.061	95.25	0.24	0.30	0.128	95.40	0.50
		$\mathcal{S}$-评分（BT-QT）	0.27	0.061	94.85	0.24	0.30	0.128	94.40	0.50
		Pocock	2.57	0.073	**92.80**	0.27	7.53	0.205	**77.70**	0.52
		Wang	20.61	0.220	**18.25**	0.29	8.13	0.224	**80.60**	0.56
		逆概率删失加权*	2.67	0.073	**93.00**	0.27	7.76	0.210	**78.20**	0.52
	MAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.75	0.062	95.50	0.24	0.38	0.128	95.30	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.75	0.062	95.55	0.25	0.38	0.128	95.30	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.75	0.062	95.00	0.24	0.38	0.128	94.45	0.50
		Pocock	9.89	0.117	**67.20**	0.25	9.76	0.241	**65.60**	0.51
		Wang	20.84	0.223	**16.85**	0.29	8.18	0.225	**80.40**	0.56
		逆概率删失加权*	9.91	0.115	**64.80**	0.25	9.90	0.243	**62.60**	0.51

区间 [94, 96] 外的覆盖概率以粗体突出显示。*逆概率删失加权方法已在 `WINS` R软件包中实现，用于第一个终点 Y_1，并排除 Y_2中的缺失数据受试者。

治疗组的数据（因此在我们的模拟中对随机缺失具有稳健性）。尽管如此，他们的方法总体上产生的偏倚和均方根误差均大于我们的方法。在某些情景下，Wang 方法的表现也不如 Pocock方法，例如，在删失具有异质性且 $\theta^{\mathrm{WR}} = 1$ 时。当异质删失率为 20%（组 a）和 60%（组 b）且 $\theta^{\mathrm{WR}} = 2$ 时，Wang 方法看起来可能表现更好；但通过一项简单实验，我们发现，调换删失率（组 a 为 60%，组 b 为 20%）会导致偏倚显著增大，覆盖概率也明显变差。一个可能的原因仍是，对事件时间进行错误插补，可能会导致在 h 之前插补的事件过多或过少，从而使胜负计数产生偏倚，并可能导致估计量比 Pocock方法的偏倚更大。在 Y_1 中没有删失、但 Y_2 中存在缺失数据的情景下，Wang 方法优于 Pocock方法，且结果有效。

关于 WINS软件包中实现的逆概率删失加权方法，我们发现，在 Y_1 中没有删失且 Y_2中没有缺失的情况下，该方法得到的绝对相对偏倚和均方根误差与我们的 $\mathcal{S}$评分方法及 Pocock 方法相同，证实了它们在完整数据下的等价性。差异仅体现在方差估计上，逆概率删失加权方法的覆盖概率略高。当 Y_2 被完整观测，且潜在偏倚仅由 Y_1中的删失导致时，逆概率删失加权方法在同质删失下仍保持一致性，但在异质性删失下会产生偏倚，而我们的方法仍保持一致性。更一般地说，当 Y_1 中同时存在删失且 Y_2中存在缺失时，逆概率删失加权方法会产生偏倚，这可能是因为该方法只能应用于 Y_2中没有缺失数据的受试者子集。当真实胜率比为 1 时，逆概率删失加权方法在完全随机缺失和同质删失下表现良好，但对异质性删失和缺失较为敏感。当真实胜率比为 2 时，删失或缺失都会导致绝对相对偏倚和均方根误差大幅升高，并造成覆盖不足。总体而言，这些结果表明，与其他竞争方法相比，我们的方法在所考察的模拟设置中具有稳健性和更优的表现。

此外，尽管我们未展示其他情景（例如样本量较小（$n_a = n_b = 100$）或删失和缺失率较低（20%））的结果，但这些结果可见在线补充材料（网络附录S.1）。我们在此简要总结研究发现。总体而言，Pocock方法的偏倚模式与表1所示一致。我们的方法始终具有较小的偏倚和均方根误差。在较小样本中（$n_a = n_b = 100$），我们提出的 $\mathcal{S}$评分方法的三种置信区间估计量（IF-Wald、BT-Wald和BT-QT）的覆盖率偶尔略微超出 $[94,\ 96]$ 范围（最低为93.55，最高为96.45）；鉴于该 $[94, 96]$ 范围是根据估计覆盖概率的抽样分布正态近似得出的，这可能是蒙特卡洛变异所致。尽管如此，大多数覆盖概率值都落在 $[94,\ 96]$范围内，且接近95%的名义水平。所提出的基于影响函数的方差估计量始终能实现 $[93.65,\ 96.00]$, 的覆盖率，只有两个情景略低于94（93.90和93.65），这证实了其用于统计推断的有效性。与通过自助法方法估计的方差相比，该方法还具有更高的

计算效率，因为它避免了重复重抽样。

最后，我们开展了一项额外实验，以评估在样本量减少时，所提出的基于影响函数的方差估计量的小样本表现。具体而言，我们生成了具有 $n_a = n_b = 50$和 $n_a = n_b = 25$的数据，并评估主模拟中考虑的 $\mathcal{S}$-评分估计量的三种置信区间估计方法的覆盖概率。 $n_a = n_b = 50$时，基于影响函数的方差估计量所得的覆盖概率仍接近名义水平，总体处于 $[94, 96]$ 范围内。然而，我们发现，当样本量减少至 $n_a = n_b = 25$时，基于影响函数的方差估计量通常得到低于名义范围下限的覆盖概率，但所有数值仍高于91.80。 $n_a = n_b = 25$ 的详细结果见在线补充材料（网络附录 S.1.2）。这一表现符合预期，因为基于影响函数的方差估计量依赖大样本理论，在非常小的样本中可能不够准确，体现了计算效率与有限样本精度之间的自然权衡。相比之下，自助法的覆盖概率接近名义水平，且在小样本情境下计算负担并不重。综合来看，这些发现给出了一个实用指南：基于影响函数的方差估计量凭借其计算效率和可靠的覆盖率，适用于中等至大样本量；而在小样本应用中，基于自助法的方差估计更为可取。我们的实证发现表明，如果总样本量为 $n < 100$，建议采用自助法。

4 真实数据分析

我们将所提出的方法应用于两个案例研究。第一个研究是 HEART-FID试验（Mentz 等人，2023），这是一项双盲随机临床试验，旨在评估羧基麦芽糖铁与安慰剂对心力衰竭、射血分数降低和缺铁患者的疗效。第二个研究是 ISCHEMIA试验（Maron等人，2020），该试验将中度或重度缺血患者随机分配至初始侵入性策略（包括血管造影，并在可行时进行血运重建）加药物治疗，或仅采用药物治疗的初始保守策略；若药物治疗失败，则进行血管造影。

我们采用 $\mathcal{S}$-评分方法，并分别在调整和不调整协变量的情况下 Y_2 处理缺失数据，如第2.4 节和2.6节所示。为调整缺失数据，我们使用包含若干预先指定协变量的逻辑回归模型。我们还评估并比较了三种方差估计方法（IF-Wald、BT-Wald 和 BT-QT），这些方法也用于我们的模拟研究。对于基于自助法的方法，我们使用 1,000 个自助法数据重复样本。

4.1 研究 I：HEART-FID试验

HEART-FID试验（Mentz 等人，2023）随机分配并纳入了 3,065 名参与者，其中 1,532 人属于治疗组。主要分析包括三个层级终点：一年内至死亡时间、12 个月内心力衰竭住院次数，以及从基线至第6个月6分钟步行距离的变化。

为展示我们的方法，我们选择了两个层级终点：12个月内的至死亡时间（Y_1）和第12个月的6分钟步行距离（Y_2）。参与者在第12个月（$h = 364$ 天）行政性删失；根据观测到的第12个月最大6分钟步行距离，我们设定 $\tau = 864$。为填补缺失的 Y_2 值，我们考虑了 Mentz 等人（Mentz 等人 (2023)）使用的两个基线协变量：基线时的6分钟步行距离，以及参与者的地理区域（北美洲、亚太地区或欧洲）。

表2：HEART-FID试验数据的汇总统计量。

变量（均值（标准差）或 n （%））		总体 ($n = 3,065$)	对照组 ($n = 1,533$)	治疗组 ($n = 1,532$)
协变量	基线 6 分钟步行距离	274.3 (109.5)	274.7 (109.4)	273.9 (109.7)
	地区			
	北美洲	1,442 (47.0%)	721 (47.0%)	721 (47.1%)
	亚太地区	210 (6.9%)	105 (6.8%)	105 (6.9%)
	欧洲	1,413 (46.1%)	707 (46.1%)	706 (46.1%)
Y_1	$h = 364$ 天前死亡	289 (9.4%)	158 (10.3%)	131 (8.6%)
	h之前右删失	80 (2.6%)	49 (3.2%)	31 (2.0%)
		在具有 $Y_1 > h$的参与者中		
Y_2		总体 ($n = 2,696$)	对照组 ($n = 1,326$)	治疗组 ($n = 1,370$)
	第12个月6分钟步行距离	290.7 (115.8)	293.1 (120.1)	288.5 (111.5)
	缺失 Y_2 数据	419（15.5%）	208（15.7%）	211（15.4%）

对于 Y_2，汇总统计量未纳入缺失数据。SD：标准差； n(%)：人数和百分比。

总体而言，参与者中在 $h = 364$ 天之前被删失的比例不足3%，如表 2所示；约15%的参与者存在 Y_2 值缺失。不过，这些 Y_2 缺失数据主要来自北美洲（21%），高于欧洲或亚太地区（两者均为11%）。此外，存在缺失数据的参与者基线时的6分钟步行距离平均较短。治疗组间在 $h = 364$ 天之前的死亡率以及均值 Y_2 （不含缺失数据）差异无统计学意义（p值 $= 0$分别为0.10和0.35）。以治疗、两个协变量及其交互作用（Z、$\mathbf{X}$和 $Z \cdot \mathbf{X}$）为自变量，对缺失指示变量进行逻辑回归分析，结果显示治疗与基线协变量之间无显著交互作用（p值 $= 0.36$ 和0.72），但协变量主效应显著（p值 < 0.001）。

为评估该缺失模型对随机缺失假设可能违反情况的稳健性，我们计算了 E值（VanderWeele 和 Ding, 2017），结果为 1.63，这表明某个未测量因素

若要完全解释观察到的关联，在控制观测协变量后，其与观测到第二终点的概率以及结局之间的关联强度均需达到至少 1.63 的风险比。

表3 展示了采用我们的方法、协变量调整方法以及 Pocock算法进行分析的结果。考虑少量删失数据和缺失值后，无论是否进行协变量调整，得到的WR估计值都略小于 Pocock算法的结果，且95%置信区间更窄。尽管如此，总体结论保持不变。这些WR估计值也小于原研究主要分析中报告的数值（Mentz 等人，2023），这是因为我们在此仅使用了两个终点，而非原研究中的三个层级终点，并且使用了第12个月而非第6个月的6分钟步行距离。

为了更好地理解估计的胜率与 $\mathcal{S}$评分之间的关系，我们进一步在图 $\mathcal{S}$中绘制了按治疗分层的2评分Kaplan–Meier曲线。如图所示，两种治疗组的Kaplan–Meier曲线非常相似，这与估计的胜率接近 1 相一致。

表3：HEART-FID试验：第12个月时至死亡时间和6分钟步行距离的胜率估计。

方法	WR估计值	95%置信区间	置信区间宽度
$\mathcal{S}$-评分 (IF-Wald)	1.01	(0.92, 1.10)	0.18
$\mathcal{S}$-评分 (BT-Wald)	1.01	(0.92, 1.10)	0.18
$\mathcal{S}$-评分 (BT-QT)	1.01	(0.93, 1.11)	0.18
协变量调整的 $\mathcal{S}$评分（IF-Wald）	1.02	(0.94, 1.11)	0.18
协变量调整的 $\mathcal{S}$评分（BT-Wald）	1.02	(0.93, 1.12)	0.18
协变量调整的 $\mathcal{S}$评分（BT-QT）	1.02	(0.94, 1.12)	0.18
Pocock	1.04	(0.96, 1.14)	0.19

CI：置信区间。

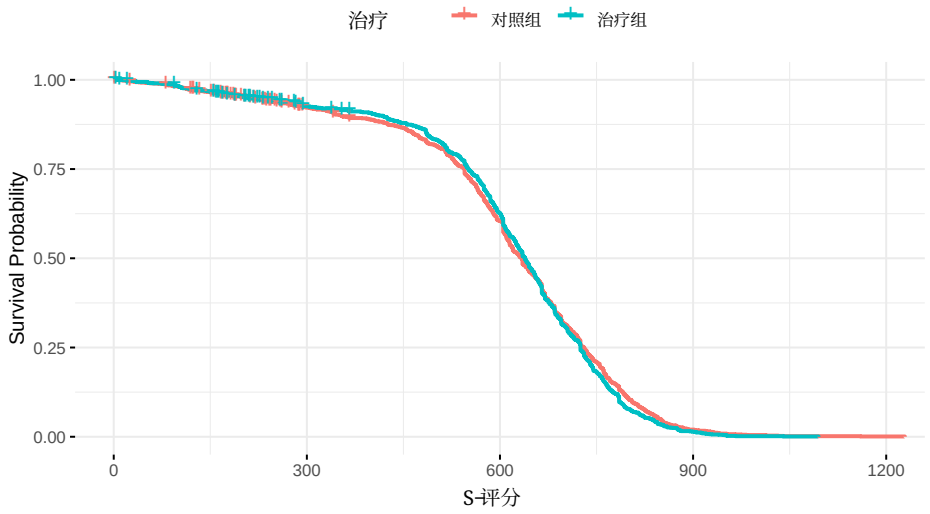


图 2：HEART-FID研究中 $\mathcal{S}$评分的治疗组特异性Kaplan-Meier曲线。

此例展示了真实胜率比与 1 无显著差异的情况，两组的 Y_1 删失率较低， Y_2 数据缺失比例适中。因此，如模拟研究所示，在这种情况下， $\mathcal{S}$评分方法和Pocock 方法都可能具有一致性。不过，我们观察到 $\mathcal{S}$评分方法的置信区间略窄，表明其效率可能有所提高。这一结果与我们的模拟结果一致，即使两种方法都具有一致性，我们的 $\mathcal{S}$评分方法也能持续得到效率更高的估计值。

为评估非信息性删失假设可能遭到违反时结果的敏感性，我们进行了补充的最佳和最坏情形分析。这些分析通过假设治疗组中所有删失参与者在研究结束时均存活，而对照组中所有删失参与者均于次日死亡（最佳情形分析），来设定极端但合理的删失机制；在最坏情形分析中则反转这些假设。在本例中，最佳情形分析得出的胜率比估计值为1.02（95%置信区间： [0.93，1.11]），而最坏情形分析得出的结果为1.01（95%置信区间： [0.92，1.10]）；主要分析的结果为1.01（95%置信区间： [0.92，1.10]）。这些结果表明，最佳、最坏和主要分析的结果差异极小，说明在这种情况下，估计的WR对极端删失假设高度不敏感，也为本例中非信息性删失假设的合理性提供了支撑。

最后，为评估所假设的随机缺失模型可能存在的设定错误，我们在在线补充材料（网络附录 S.2）中，通过改变 $\mathcal{S}$ 的缺失模型，对协变量调整后的 Y_2 -评分估计量进行了补充分析。具体而言，我们考虑了若干种替代的缺失模型设定，并使用 `SuperLearner` R软件包（Van der Laan 等人，2007）实现，包括：(i) 包含协变量间所有二阶交互项的逻辑回归模型（`SL.glm.interaction`）；(ii) 一个

随机森林学习器（SL.randomForest）；以及 (iii) 一个集成学习器，结合了不含交互项的逻辑回归（SL.glm）、含交互项的逻辑回归（SL.glm.interaction）和随机森林（SL.randomForest）。结果表明，所得 WR 估计值与表 3 中报告的结果总体一致，不同缺失模型设定之间没有显著变化。这些发现进一步支持了所提出的估计量对于合理的缺失模型设定错误具有稳健性。

4.2 研究II：ISCHEMIA试验

ISCHEMIA试验共将5,179名参与者随机分组，其中2,588名被分配至治疗组（Maron等人，2020；Spertus等人，2020）。在Maron等人（2020）中，主要终点是心血管原因死亡、心肌梗死，或因不稳定型心绞痛、心力衰竭或复苏后的心脏骤停而住院的复合终点。次要终点合并了心血管原因死亡和心肌梗死。Spertus等人（2020）特别关注了对西雅图心绞痛问卷—心绞痛发作频率（SAQ7-AF）评分这一生活质量终点的分析。

为便于说明，我们选择了两个终点，即至全因死亡的时间 Y_1 （4年内）以及随机化后4年的SAQ7-AF评分 Y_2。我们将 h 固定为1,460天，在该时间点参与者被行政性删失。此外，由于max(Y_2) $= 100$，我们设定 $\tau = 100$。我们还选择了5个协变量，以校正缺失的 Y_2 数据：年龄、性别、糖尿病史、射血分数和估算肾小球滤过率。汇总统计量见下方表4。

表4：ISCHEMIA试验数据的汇总统计量。

变量（均值（标准差）或 n （%））		总体 ($n = 5,179$)	对照组 ($n = 2,591$)	治疗组 ($n = 2,588$)
协变量	Age	64.3（9.6）	64.2（9.7）	64.4（9.5）
	性别 = 女性	1,168 (22.6%)	562 (21.7%)	606 (23.4%)
	糖尿病史	2,164 (41.8%)	1,093 (42.2%)	1,071 (41.4%)
	射血分数	60.1 (8.2)	59.9 (8.3)	60.2 (8.1)
	肾小球滤过率	83.2 (22.6)	83.1 (22.5)	83.2 (22.8)
Y_1	$h = 1$，460 天前死亡	247 (4.8%)	123 (4.7%)	124 (4.8%)
	h前右删失	3,264 (63.0%)	1,627 (62.8%)	1,637 (63.3%)
Y_2		**在具有 $Y_1 > h$ 的参与者中**		
		总体 ($n = 1,688$)	对照组 ($n = 841$)	治疗组 ($n = 827$)
	SAQ7-AF评分	93.7 (13.4)	92.8 (14.1)	94.6 (12.5)
	缺失 Y_2 数据	351 (21.0%)	176 (20.9%)	175 (21.2%)

标准差：标准差； n （%）：人数和百分比。

总体中， 参与者中，约有 5% 死亡，63% 在 $h = 1{,}460$ 天前被删失，而 e

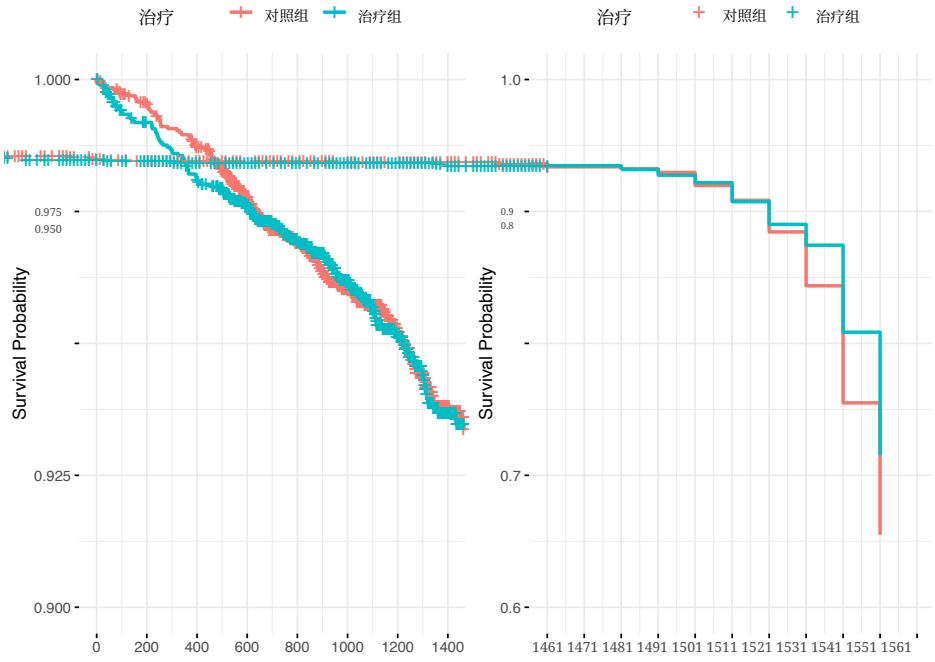
21% 的 SAQ7-AF 评分缺失。总体而言，治疗组间的协变量、全因死亡以及 h 之前 SAQ7-AF 评分被删失的比例在统计学上没有差异。然而，各组的 SAQ7-AF 平均评分存在显著差异（p 值为 $= 0.01$）。采用基线协变量、治疗分配及其交互作用建立的 SAQ7-AF 评分缺失指示变量逻辑回归模型显示，该缺失与年龄、糖尿病史和肾小球滤过率显著相关。

与HEART-FID研究类似，我们评估了该缺失模型的E值（VanderWeele和Ding，2017），结果为1.98，表明该结果对随机缺失假设可能遭到违反具有中等程度的稳健性。这一结果表明，在给定观测协变量的条件下，某个未测量因素与第二终点被观测到的概率及结局的风险比至少均需达到1.98，才能完全解释掉观测到的关联。

表5：ISCHEMIA试验：第48个月至死亡时间和SAQ7-AF评分的胜率估计。

方法	估计值	95%置信区间	置信区间宽度
$\mathcal{S}$-评分（IF-Wald）	1.26	(1.04, 1.49)	0.45
$\mathcal{S}$-评分 (BT-Wald)	1.26	(1.04, 1.48)	0.44
$\mathcal{S}$-评分 (BT-QT)	1.26	(1.06, 1.49)	0.42
协变量校正的 $\mathcal{S}$-评分 (IF-Wald)	1.27	(1.05, 1.49)	0.44
经协变量调整的 $\mathcal{S}$-评分（BT-Wald）	1.27	(1.05, 1.49)	0.44
经协变量调整的 $\mathcal{S}$-评分（BT-QT）	1.27	(1.08, 1.50)	0.42
Pocock	1.00	(0.90, 1.12)	0.22

置信区间：置信区间。



(a) 当$S \leq \mathrm{h} = 1460$时的S-评分　　(b) 当$S > h = 1460$时的S-评分

图3：ISCHEMIA研究中 $\mathcal{S}$-评分的治疗组特异性Kaplan-Meier曲线。

关于估计结果，我们的 $\mathcal{S}$评分方法得出的估计值与Pocock算法显著不同（表5）。未调整和经协变量调整的 $\mathcal{S}$评分方法得出的点估计相近，分别为1.26和1.27，且在0.05水平上具有统计学显著性，而Pocock方法的结果则不具有统计学显著性。调整后与未进行协变量调整的结果相近，这表明调整模型对估计的影响较小，可能是因为各治疗组的缺失数据率没有显著差异。

与HEART-FID研究类似，我们在图 $\mathcal{S}$中绘制了3所示的治疗组特异性Kaplan–Meier曲线。为更清晰地展示生存概率，我们使用阈值 $h = 1460$ （时间范围）将总体曲线分成两个面板（a）和（b）。当 $\mathcal{S} \leq h = 1460$时，所有生存概率均超过0。9，这种划分有助于更清楚地观察。如面板（a）所示，两个治疗组的生存曲线相交：对照组在较早时点的生存率较高，而治疗组随后赶上，并在 $\mathcal{S} \approx 800$之后略高。对于 $\mathcal{S} > 1000$，两条曲线大致重叠。如面板（b）所示，治疗组往往具有较大的 $\mathcal{S}$评分，这与表4 中的生活质量结果一致（治疗组的平均生活质量值较高），并支撑估计的WR大于1的合理性。

我们的结果与Spertus等人（2020仅分析纵向生活质量时的发现一致。我们的模拟结果也支撑这一结论，表明无论治疗组间的缺失或删失是否同质，当真实胜率比不等于1且存在缺失或删失数据时，Pocock方法可能存在偏倚。

在治疗组间，这类缺失或删失是否同质或异质。鉴于本例中两组的删失率和缺失数据占比均为中等至较高（见表4），我们的模拟表明， $\mathcal{S}$-评分方法比Pocock算法能提供更可靠的估计值。

为评估非信息性删失假设可能遭到违反时结果的敏感性，我们按照第4.1节所述进行了最佳和最坏情形分析。最佳情形分析得出的胜率比估计值为 1.68（95%置信区间： $[1.39,\ 1.98]$），而最坏情形分析得出的结果为 0.91（95%置信区间： $[0.75,\ 1.07]$），主要分析结果则为 1.26（95%置信区间： $[1.04,\ 1.49]$）。置信区间采用所提出的基于影响函数的方差估计量构建。估计效应的范围如此之大，反映了这些极端情景下较高的删失比例。需要更多有关删失机制的信息，才能缩小估计效应的范围。因此，研究者应谨慎解读结果，并对非信息性删失假设的合理性作出说明。

最后，与 HEART-FID 分析类似，我们通过改变在线补充材料（网络附录 S.2）中 $\mathcal{S}$的缺失模型，对协变量调整后的 Y_2 评分估计量进行了补充分析。所考察的其他缺失模型设定与 HEART-FID 分析中使用的相同。所得 WR 估计值与表5中报告的结果非常接近，表明不同模型之间没有实质性差异。这些发现进一步支持了所提出的估计量对于合理的缺失模型误设具有稳健性。

5 总结

本文针对涉及两个层级终点的临床试验中胜率估计的一项关键空白展开研究：一个是受右删失影响的至终末事件时间，另一个是存在缺失数据的非生存结局。通过将两个终点中的可用信息映射到单一终点，即 $\mathcal{S}$-评分，我们能够在非信息性删失和随机缺失假设下构建简单的胜率非参数最大似然估计量。这与 Pocock方法采用的基于计数的方法形成对比；后者可能会舍弃观测数据或有价值的信息，以将一对结果判定为胜、负或平，从而可能导致估计值有偏或效率较低。

通过大量模拟，我们证明，该方法能够产生更稳定且高效的胜率估计值，尤其是在治疗效应为胜率 > 1、非生存终点存在较大比例缺失值（在完全随机缺失和随机缺失两种情况下），或生存终点的删失率不可忽略的情景下。模拟结果还凸显了在许多此类情况下，Pocock方法的效率损失。与Wang等人（2025a）的插补方法相比，我们的方法计算高效，并且对模型误设更具稳健性。

此外，第2.5 节提出的渐近方差估计量是一致的，因此能够对胜率估计量进行有效的不确定性量化。然而，在小样本情形下（例如样本量为 $n_a = n_b = 25$的模拟），该估计量可能存在轻微的覆盖不足；因此，对于总样本量小于100的情况，我们建议采用基于自助法的方差估计。总体而言，我们的方法为分析两个层级终点提供了一种实用且稳健的解决方案，无需舍弃任何观测数据，最终能够通过胜率对治疗效应进行更可靠的估计和统计推断。

此外，我们在第 4 节将我们的方法应用于 HEART-FID 和 ISCHEMIA 两项真实临床试验。这两项研究展示了互补的情境，涵盖删失较少与删失较多，以及协变量依赖的缺失等情形。与模拟结果一致，我们的方法有望减轻 Pocock 计数法带来的偏倚。

我们也认识到本研究存在若干有待未来研究解决的局限性。首先，所提出的 $\mathcal{S}$-评分方法专为本文考察的双终点情形设计，目前尚不具备更广泛的适用性。将该方法扩展或调整以适用于更复杂的层级结构，例如多分类终点（Li 等人，2024）或涉及两个以上终点的情景，并非易事。未来，我们计划在一个更灵活的框架下探索基于分布的方法来估计 WR，以处理范围更广的层级终点配置。其次，我们尚未扩展协变量调整，将 Y_1 的删失分布建模为 $\mathbf{X}$ 的函数，类似于第2.6节的结果。我们有意作此取舍；将每个时间点的删失概率建模为协变量（可能随时间变化）的函数更具挑战性，而且需要估计多个额外的干扰参数。第三，我们承认，违反非信息性假设和/或随机缺失假设可能会损害可识别性，并导致 S 分数估计量产生偏倚。针对这些假设开展更正式的敏感性分析（Robins 等人，2000）是对当前框架的重要扩展，我们将在未来研究中加以探索。此外，我们还计划开发双重稳健（甚至多重稳健）估计量（Robins 等人，1994；Bai 等人，2013），结合结局模型与缺失机制模型，从而进一步提高对模型错设的稳健性。

最后，未来研究还可以探索若干潜在的扩展方向。我们方法背后的总体策略或可用于观察性研究（不进行随机化）、多中心临床试验以及目标亚组分析中的胜率分析（Zhuang 等人，2025；Liu et al.，2025a，2024）。此外，将非参数方法扩展到 Wang 的先插补后计数方法（Wang等人，2025a）也值得探讨；例如，可采用结合机器学习工具的样本分割（Van der Laan 等人，2007；Chernozhukov 等人，2018；Wang等人，2025b），以防插补模型设定错误。

数据可用性声明

本文使用的数据不公开，但可向试验赞助方申请获取。

软件

用于实现所提出方法的用户友好型 R 软件包可在作者的 GitHub 页面获取 https://github.com/yiliu1998/WinRS。

致谢

Yi Liu 获得美国国立卫生研究院（NIH）下属国家心肺血液研究所（NHLBI）颁发的 T32HL079896 奖项资助。本文内容完全由作者负责，不一定代表 NIH 的官方观点。杜克临床研究所生物统计学与数据科学研究基金也为本研究提供了支持。

参考文献

Safa Abdalla、Maria E Montez-Rath、Patrick S Parfrey 和 Glenn M Chertow。分析复合结局的胜率方法：在 EVOLVE 试验中的应用。*Contemporary Clinical Trials*，48：119–124，2016。

Stefan D Anker、Josep Comin Colet、Gerasimos Filippatos、Ronnie Willenheimer、Kenneth Dickstein、Helmut Drexler、Thomas F L¨uscher、Boris Bart、Waldemar Banasiak、Joanna Niegowska 等。羧基麦芽糖铁用于心力衰竭和缺铁患者。新英格兰医学杂志，361(25)：2436–2448，2009。

Xiaofei Bai、Anastasios A Tsiatis 和 Sean M O'Brien。观察性研究中采用分层抽样时治疗特异性生存分布的双稳健估计量。*Biometrics*，69(4)：830–839，2013。

HuimanX Barnhart、Yuliya Lokhnygina、Roland A Matsouaka 和 Frank WRockhold。针对具有层级结构的多个生存时间终点，采用胜统计量进行试验设计。*Statistics in Biopharmaceutical Research*，17(2):197–210，2025。

Ionut Bebu 和 John M Lachin。基于优先组分复合结局的胜率分析大样本推断。*Biostatistics*，17(1):178–187，2016。

Marc Buyse、Johan Verbeeck、Everardo D. Saad、Micka¨el De Backer、Vaiva Deltuvaite-Thomas 和 Geer t

Molenberghs，编者。《广义成对比较手册：以患者为中心的分析方法》。Chapman and Hall/CRC，纽约，2025。ISBN 9781003390855。doi: 10.1201/9781003390855。

Victor Chernozhukov、Denis Chetverikov、Mert Demirer、Esther Duflo、Christian Hansen、Whitney Newey 和 James Robins。用于治疗和结构参数的双重／去偏机器学习。《计量经济学杂志》，2018。

David R Cox。回归模型与生命表。《英国皇家统计学会杂志：*B*辑（方法学）》，34(2)：187–202，1972。

Ying Cui 和 Bo Huang。WINS：R wins软件包。综合*R*档案网络（*CRAN*），2025。网址：https://cran.r-project.org/web/packages/WINS/。

Gaohong Dong、Bo Huang、Yu-Wei Chang、Yodit Seifu、James Song 和 David C Hoaglin。胜率：删失和随访时间的影响及其在非比例风险中的应用。《药物统计学》，19(3)：168–177，2020a。

Gaohong Dong、Lu Mao、Bo Huang、Margaret Gamalo-Siebers、Jiuzhou Wang、GuangLei Yu 和 David C Hoaglin。删失逆概率加权（IPCW）调整胜率统计量：独立删失情况下的无偏估计量。《生物制药统计学杂志》，30(5)：882–899，2020b。

Gaohong Dong、Bo Huang、Duolao Wang、JohanVerbeeck、Jiuzhou Wang 和 David C Hoaglin。调整依赖删失下的胜统计量。《药物统计学》，20(3):440–450，2021。

Gaohong Dong、Bo Huang、Johan Verbeeck、Ying Cui、James Song、Margaret Gamalo-Siebers、Duolao Wang、David C Hoaglin、Yodit Seifu、Tobias M¨utze 等。胜统计量（胜率、胜算和净获益）可以相互补充，展示治疗效应对事件时间结局的影响强度。《药物统计学》，22(1):20–33，2023。

K Duarte 和 JP Ferreira。Winratio：针对优先级结局的胜率及其95%置信区间。综合*R*档案网络（*CRAN*），2022。网址：https://cran.r-project.org/web/packages/WinRatio/。

Bradle y Efron 和 Robert J Tibshirani。自助法简介。CRC出版社，1994年。

Jo˜ao Pedro Ferreira、Pardeep S Jhund、K´evin Duarte、Brian L Claggett、Scott D Solomon、Stuart Pocock、Mark C Petrie、FaiezZannad 和 JohnJV McMurray。《心血管试验中胜率的应用》。心力衰竭，8(6):441–450，2020。

Jason P Fine 和 Robert J Gray。竞争风险亚分布的比例风险模型。《美国统计学会杂志》，94(446):496–509，1999。

Dianne M Finkelstein 和 David A Schoenfeld。在临床试验中结合死亡率和纵向测量。《医学统计学》，18(11):1341–1354，1999。Ying Gao、Yi Liu 和 Roland Matsouaka。随机化下调整协变量何时有帮助？对当前实践的比较研究。*BMC*医学研究方法学，24(1):250，2024。

Ted A Gooley、Wendy Leisenring、John Crowley 和 Barry E Storer。在存在竞争风险时估计失败概率：对旧估计量的新表述。《医学统计学》，18(6): 695–706，1999。

Philip Hougaard。多状态模型：综述。*Lifetime Data Analysis*，5:239–264，1999。

BrennanCKahan、JoannaHindley、MarkEdwards、Suzie Cro 和andTimPMorris。估计目标框架：ICH E9（R1）增补文件入门。*BMJ*，384，2024。

Edward L Kaplan 和 Paul Meier。《基于不完整观测的非参数估计》。《美国统计学会杂志》，53(282):457–481，1958。

Rungroj Krittayaphong、Sukrit Treewaree 和 Gregory YH Lip。心房颤动患者整体护理优化护理路径的组成部分：基于cool-af登记研究的胜率分析。*Europace*，26(9):euae237，2024。

Heng Li、Wei-Chen Chen、Nelson Lu、Rong Tang 和 Yu Zhao。缺失数据情形下胜算比参数的难以捉摸。*Therapeutic Innovation & Regulatory Science*，58(3):431–432，2024。

Yi Liu、Alexander W Levis、Sharon-Lise Normand 和 Larry Han。分布偏移下的多源保形推断。*Proceedings of Machine Learning Research*，235:31344–31382，2024。

Yi Liu、Alexander W Levis、Ke Zhu、Shu Yang、Peter B Gilbert 和 Larry Han。分布偏移下因果生存分析的定向数据融合。*arXiv* 预印本 *arXiv:2501.18798*，2025a。

Yi Liu、Ke Zhu、Larry Han 和 Shu Yang。Coadvise：随机对照试验中的协变量调整、变量选择与缺失数据插补。*arXiv* 预印本 *arXiv:2501.08945*，2025b。

Lu Mao。界定胜率的估计目标：将真实效应与删失区分开来。临床试验，21(5): 584–594，2024。

David J Maron、Judith S Hochman、Harmony R Reynolds、Sripal Bangalore、Sean M O'Brien、William E Boden、Bernard R Chaitman、Roxy Senior、Jose L´opez-Send´on、Karen P Alexander 等。稳定型冠状动脉疾病的初始侵入性或保守治疗策略。新英格兰医学杂志，382(15):1395–1407，2020。

Roland A Matsouaka。匹配胜统计量的稳健统计推断。《医学研究统计方法》，31(8):1423–1438，2022。

Robert J Mentz、Jyotsna Garg、Frank W Rockhold、Javed Butler、Carmine G De Pasquale、Justin A Ezekowitz、Gregory D Lewis、Eileen O'Meara、Piotr Ponikowski、Richard W Troughton 等。羧基麦芽糖铁治疗缺铁性心力衰竭。新英格兰医学杂志，389(11):975–986，2023。

Luca Monzo、Bruno Levy、Kevin Duarte、Guillaume Baudry、Alain Combes、Alexandre Ouattara、Cl´ement Delmas、Antoine Kimmoun 和 Nicolas Girerd。在重症医学试验中使用胜率分析。《美国呼吸与重症医学杂志》，209(7):798–804，2024。

Stuart J Pocock、Cono A Ariti、Timothy J Collier 和 Duolao Wang。胜率：一种基于临床优先级分析临床试验复合终点的新方法。《欧洲心脏杂志》，33(2):176–182，2012。

Hein Putter、Marta Fiocco 和 Ronald B Geskus。生物统计学教程：竞争风险与多状态模型。《医学统计学》，26(11):2389–2430，2007。

Bj¨orn Redfors、John Gregson、Aaron Crowley、Thomas McAndrew、Ori Ben-Yehuda、Gregg W Stone 和 Stuart J Pocock。复合终点的胜率方法：基于既往经验的实践指南。《欧洲心脏杂志》，41(46):4391–4399，2020。

James M Robins、Andrea Rotnitzky 和 Lue Ping Zhao。当某些回归变量并非总能观测到时回归系数的估计。《美国统计学会杂志》，89(427):846–866，1994年。

James M. Robins、Andrea Rotnitzky 和 Daniel O. Scharfstein。《缺失数据与因果推断模型中的选择偏倚和未测量混杂的敏感性分析》。载于 M. Elizabeth Halloran 和 Donald Berry 编，《流行病学、环境与临床试验中的统计模型》，第 1–94 页，纽约，纽约州，2000 年。施普林格纽约。ISBN 978-1-4612-1284-3。

Donald B Rubin. 推断与缺失数据。*Biometrika*，63(3):581–592，1976。Jun Shao 和 Dongsheng Tu。刀切法与自助法。Springer Science & Business Media，2012。

John A Spertus、Philip G Jones、David J Maron、Sean M O'Brien、Harmony R Reynolds、Yves Rosenberg、Gregg WStone、FrankE Harrell Jr、William E Boden、William S Weintraub 等。冠状动脉疾病患者接受侵入性或保守治疗后的健康状况结局。新英格兰医学杂志，382(15):1408– 1419，2020。

Leonard A Stefanski 和 Dennis D Boos。M估计微积分。美国统计学家，56(1)：29–38，2002年。

Anastasios Tsiatis。《半参数理论与缺失数据》。Springer Science & Business Media，2007。Mark J Van der Laan、Eric C Polley 和 Alan E Hubbard。超级学习器。《统计学在遗传学与分子生物学中的应用》，6(1):1–23，2007。Aad W Van der Vaart。《渐近统计学》，第 3 卷。剑桥大学出版社，2000。Tyler J VanderWeele 和 Peng Ding。观察性研究中的敏感性分析：引入 E值。《内科学年鉴》，167(4):268–274，2017。Tuo Wang、Ying Li 和 Yongming Qu。限制时间获胜比：从估计目标到估计。《Statistics in Biopharmaceutical Research》，17(1):136–148，2025a。Yiming Wang、Yi Liu 和 Shu Yang。加权平均治疗效应的速率双重稳健估计。《因果推断杂志》，13(1):20240073，2025b。Haotian Zhuang、Xiaofei Wang 和 Stephen L George。多区域随机临床试验中治疗效应异质性的评估。《Statistics in Biopharmaceutical Research》，17(3):315–322，2025。

A 附录：技术性证明

A.1 方程的证明 (4)

首先，让我们证明 $Y_{1b} < Y_{1a}$ 或 $Y_{1b} = Y_{1a}$，$Y_{2b} < Y_{2a}$ 以概率 1 推出 $\mathcal{S}_b < \mathcal{S}_a$。

若 $Y_{1b} < Y_{1a} \le h$，则 $\mathcal{S}_b = Y_{1b} < Y_{1a} = \mathcal{S}_a$。若 $Y_{1b} < Y_{1a} = h+1$，则有 $\mathcal{S}_b = Y_{1b} < Y_{1a} + Y_{2a} = \mathcal{S}_a$，因为 $Y_{2a} > 0$。现在，若 $Y_{1b} = Y_{1a} = h+1$ 且 $Y_{2b} < Y_{2a}$，则 $\mathcal{S}_b = h + Y_{2b} < h + Y_{2a} = \mathcal{S}_a$。最后，事件 $\{Y_{1b} = Y_{1a} \le h, Y_{2b} < Y_{2a}\}$ 的概率为零，因为给定 Y_1 是一个终末事件，而且我们不考虑 Y_2 在时间 h 之前假设存在的情况，因此可以不讨论这种情形。

现在，证明 $\mathcal{S}_b < \mathcal{S}_a$ 蕴含 $Y_{1b} < Y_{1a}$ 或 $Y_{1b} = Y_{1a}$，$Y_{2b} < Y_{2a}$。

首先，若 $\mathcal{S}_a \ge h+1$，则唯一的可能性是 $Y_{1a} = h+1$，且存在 $Y_{2a} \ge 0$ 使得$\mathcal{S}_a = h+1+Y_{2a}$。在这种情况下，若 $\mathcal{S}_b \ge h+1$ 也成立，则存在 $Y_{2b} \ge 0$ 使得 $h+Y_{2b} = \mathcal{S}_b < \mathcal{S}_a = h+Y_{2a}$，这表明 $Y_{1b} = Y_{1a} = h+1$ 以及 $Y_{2b} < Y_{2a}$。然而，若 $\mathcal{S}_b \le h$，则 $Y_{1b} \le h$，这意味着$Y_{1b} < Y_{1a} = h+1$。其次，若 $\mathcal{S}_a \le h$，则 $\mathcal{S}_b \le h$。由于在这种情况下，我们只能有 $\mathcal{S}_a = Y_{1a}$ 和$\mathcal{S}_b = Y_{1b}$。

因此，$Y_{1b} < Y_{1a}$。

证明完成。

A.2 估计量的一致性 $\widehat{\theta}^{\mathrm{WR}}(\mathcal{O}^{\mathcal{S}})$ 在 (8)

我们采用第 2 节中的所有记号。为证明所提出的非参数最大似然估计量（8）对 θ^{WR}具有一致性，我们首先注意到，对于 $z \in \{a, b\}$，在假设1 和2下，无需证明即可得出：

$$\sup_{s \in [0, h+\tau+1]} \left| \widehat{F}_z(s) - F_z(s) \right| \xrightarrow{p} 0, \quad \text{as } n \to \infty. \tag{A.1}$$

接下来，我们证明第2.4节中的“代入估计量”（6 ）的一致性（将上述^累积分布函数的Kaplan-Meier估计量 F_z 代入相应项）。我们首先考虑（6）的分子。将目标泛函定义为

$$\phi(F_a, F_b) = \int_0^{h+\tau+1} \{1 - F_a(s)\}\, dF_b(s),$$

并注意到，代入估计量可表示为

$$\widehat{\phi} = \phi(\widehat{F}_a, \widehat{F}_b) = \int_0^{h+\tau+1} \{1 - \widehat{F}_a(s)\}\, d\widehat{F}_b(s).$$

我们的目标是证明 $\widehat{\phi} \xrightarrow{p} \phi(F_a, F_b)$。为此，考虑差值 $\left| \widehat{\phi} - \phi(F_a, F_b) \right|$. 通过加上并

减去一个中间项，我们得到

$$\begin{aligned}&\left|\int_0^{h+\tau+1}\{1-\widehat{F}_a(s)\}\,d\widehat{F}_b(s)-\int_0^{h+\tau+1}\{1-F_a(s)\}\,dF_b(s)\right|\\&\le\left|\int_0^{h+\tau+1}\{1-\widehat{F}_a(s)\}\,d\widehat{F}_b(s)-\int_0^{h+\tau+1}\{1-\widehat{F}_a(s)\}\,dF_b(s)\right|\\&\quad+\left|\int_0^{h+\tau+1}\{1-\widehat{F}_a(s)\}\,dF_b(s)-\int_0^{h+\tau+1}\{1-F_a(s)\}\,dF_b(s)\right|.\end{aligned}$$

^对于第一项，注意到由于对所有 t均有 $0\le 1-F_a(s)\le 1$，因此

$$\left|\int_0^{h+\tau+1}\{1-\widehat{F}_a(s)\}\,d\widehat{F}_b(s)-\int_0^{h+\tau+1}\{1-\widehat{F}_a(s)\}\,dF_b(s)\right|\le\sup_{s\in[0,h+\tau+1]}\left|\widehat{F}_b(s)-F_b(s)\right|$$

By ^由 $F_b(s)$的一致一致性，右端依概率收敛于零。

对于第二项，我们有

$$\left|\int_0^{h+\tau+1}\{1-\widehat{F}_a(s)\}\,dF_b(s)-\int_0^{h+\tau+1}\{1-F_a(s)\}\,dF_b(s)\right|=\left|\int_0^{h+\tau+1}\big(\widehat{F}_a(s)-F_a(s)\big)\,dF_b(s)\right|.$$

由于 $dF_b(s)$是 $[0,h+\tau+1]$上的有限测度，因此

$$\left|\int_0^{h+\tau+1}\big(\widehat{F}_a(s)-F_a(s)\big)\,dF_b(s)\right|\le\sup_{s\in[0,h+\tau+1]}\left|\widehat{F}_a(s)-F_a(s)\right|\Big(F_b(h+\tau+1)-F_b(0)\Big).$$

^由 $F_a(s)$的一致一致性以及 $F_b(h+\tau+1)-F_b(0)$的有限性，该项亦依概率收敛于零。

综合这些结果，我们得出

$$\left|\widehat{\phi}-\phi(F_a,F_b)\right|\xrightarrow{p}0,$$

由此可得 $\int_0^{h+\tau+1}\{1-\widehat{F}_a\ s\ \}\,d\widehat{F}_b\ s\ \xrightarrow{p}P\ \mathcal{S}_a>\mathcal{S}_b$ $\int_0^{h+\tau+1}\{1-(\ )$ $\widehat{F}_b\ s\ \}\,d\widehat{F}_a\ s\ \xrightarrow{p}P\ \mathcal{S}_a<\mathcal{S}_b$ $\widehat{\theta}^{\mathrm{WR}}\xrightarrow{p}\theta^{\mathrm{WR}}\quad n\to\infty$

( ) ( )。类似地，可以证明 ( ) ( ) ( )。因此，根据斯卢茨基定理，当 时，。

最后，如果已知生存时间是离散的，并假设 $F_a(s)$和 $F_b(s)$函数是某些概率质量函数（pmf）的累积分布函数，则可在上述证明中将积分符号 $\int$替换为 $\sum$。具体而言，不失一般性地，假设时间取值为 $\{0,1,\dots,h+\tau+1\}$，我们

可以将上述代入估计量重写为

$$\widehat{\phi} = \phi(\widehat{F}_a, \widehat{F}_b) = \sum_{s=0}^{h+\tau+1} \{1 - \widehat{F}_a(s)\}\{\widehat{F}_b(s) - \widehat{F}_b(s-1)\}.$$

证明的其余部分类似。

A.3 基于导出影响函数的渐近方差估计

回顾一下，由 $\mathcal{S}$-评分定义的胜率表示为

$$\theta^{\mathrm{WR}} = \frac{P(\mathcal{S}_a > \mathcal{S}_b)}{P(\mathcal{S}_a < \mathcal{S}_b)} = \frac{\int \bar{F}_a(s) dF_b(s)}{\int \bar{F}_b(s) dF_a(s)}.$$

为简洁起见，我们记

$$N = \int \bar{F}_a(s) dF_b(s), \quad D = \int \bar{F}_b(s) dF_a(s), \tag{A.2}$$

因此，胜率为 $\theta^{\mathrm{WR}} = N/D$。

我们在假设 $\mathcal{S}$-评分取值于有限支撑集上的条件下，推导胜率估计量的方差。该假设在实践中并不构成限制，因为观测到的结局必然具有有限精度，因此属于一个有限但可能非常大的集合。将 $\mathcal{S}$ 视为有限网格上的离散随机变量，可以避免测度论上的复杂性，同时准确反映试验数据的可观测特性。这一表述为应用 M 估计理论推导估计量的渐近方差提供了严格的基础。

不失一般性，对于组 z，我们假设所提出的 $\mathcal{S}$-评分变量 $\mathcal{S}_z$ 取值于离散值集合 $\{1, 2 \ldots, M\}$，对应的概率质量为 p_1^z，p_2^z、.....、p_M^z。这里 $M = h + \tau + 1$，以便于后续推导中的记号表示。这意味着我们在假设 Y_1 和 Y_2 变量均为离散变量的基础上推导方差，尽管这两者都可以是连续变量。这是因为假设支撑集有限，便可以用有限个参数对问题进行参数化，从而使用 M估计理论构造估计量并推导渐近方差。在实践中，当观测到的生存时间或非生存变量为连续变量时，我们的推导和结果仍然适用，只需将样本中的有限个观测值视为参数总数即可。

生存函数 $\bar{F}_s^z = P(\mathcal{S}_z > s)$。令 $p_0 = 0$ 和 $F_M^z = 0$，

其中 $z \in \{a, b\}$。根据这些新的定义，上述（A.2）可改写为

$$N = \sum_{s=1}^{M} (\bar{F}_{s-1}^{b} - \bar{F}_{s}^{b}) \bar{F}_{s}^{a}, \quad D = \sum_{s=1}^{M} (\bar{F}_{s-1}^{a} - \bar{F}_{s}^{a}) \bar{F}_{s}^{b}.$$

此外，我们考虑基于条件生存概率的参数化形式，其中令

$$\bar{F}_{s}^{z} = \prod_{j=1}^{s} q_{j}^{z}, \quad s = 1, \ldots, M-1.$$

其中 $\bar{F}_{0}^{z} = 1$ 和 $\bar{F}_{M}^{z} = 0$，对两者均适用 $z \in \{a, b\}$。这里，$q_{j}^{z} = P(\mathcal{S}_z > j \mid \mathcal{S}_z \geq j-1) = F_{j}^{z} / F_{j-1}^{z}$ 是给定参与者存活超过 $j-1$ 时，其在时间 j 的条件生存概率，对两者均适用$z \in \{a, b\}$。

然后，条件生存概率 $\boldsymbol{q} = (q_1^a$、……、q_{M-1}^a，q_1^b、……、$q_{M-1}^b)$ 可通过以下估计方程估计：

$$\frac{1}{n} \sum_{i=1}^{n} \psi_{s}^{z}(\mathcal{O}_{i}^{\mathcal{S}}) = 0,$$

其中，对于一份观测数据 $\mathcal{O}^{\mathcal{S}}$，我们记

$$\psi_{s}^{z} = \psi_{s}^{z}(\mathcal{O}^{\mathcal{S}}) = I(Z = z) E_{s} (A_{s} - \widehat{q}_{s}^{z}),$$

其中 $s \in \{1, \ldots, M-1\}$ 和 $z \in \{a, b\}$，其中解 $\widehat{q}_{s}^{z}$ 对应于 $\mathcal{S}_z = s$时刻的 q_{s}^{z} 的 Kaplan-Meier估计量。E_s 是表示个体是否处于 $\mathcal{S} = s$ 时刻风险集中的指示变量，而 A_s 是表示个体是否存活超过时间 $\mathcal{S} = s$ 的指示变量。我们将这些估计方程堆叠成向量

$$\Psi(\boldsymbol{q}) = \left(\psi_{1}^{a}, \ldots, \psi_{M-1}^{a}, \psi_{1}^{b}, \ldots, \psi_{M-1}^{b}\right)'.$$

令 $B(\boldsymbol{q}) = -\partial \Psi(\boldsymbol{q}) / \partial \boldsymbol{q}$，这是一个对角矩阵，因为上述向量中的每个方程都仅取决于 $\boldsymbol{q}$ 中的一个元素，即 q_{s}^{z}。对于组 z，关于 q_{s}^{z} 的导数为

$$\frac{\partial \psi_{s}^{z}}{\partial q_{s}^{z}} = -I(Z = z) E_{s}.$$

因此，$B(\boldsymbol{q})$ 的（总体）对角线元素为 $\mathbb{E}\{I(Z = z) E_{s}\}$。因此，根据 Van der Vaart(2000)，这些 q 的影响函数（IF）为

$$\eta = \eta(\mathcal{O}^{\mathcal{S}}) = B(\boldsymbol{q})^{-1} \psi,$$

其中 $\mathcal{O}^{\mathcal{S}}$ 是第 2.4 节中定义的观测数据。实际应用中，可通过将期望替换为样本均值来估计这些量。即

$$\widehat{\eta}(\mathcal{O}^{\mathcal{S}}) = \frac{I(Z=z)E_s(A_s - \widehat{q}_s^z)}{\widehat{\mathbb{E}}\{I(Z=z)E_s\}}.$$

接下来，为了获得非参数最大似然估计量的影响函数（6），针对 θ^{WR}，我们还需要求出以下导数 e:

$$C = \left(\frac{\partial \theta^{\mathrm{WR}}}{\partial q_1^a}, \ldots, \frac{\partial \theta^{\mathrm{WR}}}{\partial q_{M-1}^a}, \frac{\partial \theta^{\mathrm{WR}}}{\partial q_1^b}, \ldots, \frac{\partial \theta^{\mathrm{WR}}}{\partial q_{M-1}^b} \right).$$

对于每个元素 $\frac{\partial \theta^{\mathrm{WR}}}{\partial q_t^z}$，有 $\frac{\partial \theta^{\mathrm{WR}}}{\partial q_s^z} = \sum_{k=s}^{M-1} \frac{\partial \theta^{\mathrm{WR}}}{\partial \bar{F}_k^z} \frac{\partial \bar{F}_k^z}{\partial q_s^z}$（根据链式法则）。

注意，所有 $\bar{F}_k^z$（其中 $k \geq s$）对于给定的 $z \in \{a, b\}$ 均依赖于 q_s^z。我们按如下方式计算 $\frac{\partial F_k^z}{\partial q_s^z}$。由于，$\log \bar{F}_k^z = \sum_{j=1}^{k} \log q_j^z$，对 q_s^z 求导（$s \leq k$），得到 $\partial \log \bar{F}_k^z / \partial q_s^z = \frac{1}{q_s^z}$. 因此，根据链式法则，

$$\frac{\partial \bar{F}_k^z}{\partial q_s^z} = \bar{F}_k^z \frac{1}{q_s^z} = \frac{\bar{F}_k^z}{q_s^z}, \quad t \leq k.$$

对于 $s > k$，导数为零，因为在时间 k 之后，$\bar{F}_k^a$ 和 $\bar{F}_k^b$ 不再依赖于 q_s^a 或 q_s^b。

此外，对于 $\frac{\partial \theta^{\mathrm{WR}}}{\partial \bar{F}_s^z}$，应用商法则，

$$\frac{\partial \theta^{\mathrm{WR}}}{\partial \bar{F}_s^a} = \frac{D(\bar{F}_{s-1}^b - \bar{F}_s^b) - N(\bar{F}_s^b - \bar{F}_{s+1}^b)}{D^2} \quad \text{and} \quad \frac{\partial \theta^{\mathrm{WR}}}{\partial \bar{F}_s^b} = \frac{D(\bar{F}_s^a - \bar{F}_{s+1}^a) - N(\bar{F}_{s-1}^a - \bar{F}_s^a)}{D^2},$$

其中 N 和 D 分别是式（A.2）中定义的分子和分母。

最后，根据德尔塔方法，非参数最大似然估计量（6）的影响函数为

$$\begin{aligned}
\varphi(\mathcal{O}^{\mathcal{S}}) &= C\eta(\mathcal{O}^{\mathcal{S}}) \\
&= \sum_{t=1}^{M-1} \frac{1}{q_s^a} \sum_{k=s}^{M-1} \bar{F}_k^a D^{-2} \left[D(\bar{F}_{k-1}^b - \bar{F}_k^b) - N(\bar{F}_k^b - \bar{F}_{k+1}^b) \right] \frac{I(Z=a)E_s(A_s - q_s^a)}{\mathbb{E}\{I(Z=a)E_s\}} \\
&\quad + \sum_{t=1}^{M-1} \frac{1}{q_s^b} \sum_{k=s}^{M-1} \bar{F}_k^b D^{-2} \left[D(\bar{F}_k^a - \bar{F}_{k+1}^a) - N(\bar{F}_{k-1}^a - \bar{F}_k^a) \right] \frac{I(Z=b)E_s(A_s - q_s^b)}{\mathbb{E}\{I(Z=b)E_s\}}.
\end{aligned}$$

为估计 $\varphi(\mathcal{O}^{\mathcal{S}})$，可以代入估计得到的生存函数和条件生存函数，并以样本均值替代期望，从而得到影响函数的估计值 $\widehat{\varphi}(\mathcal{O}^{\mathcal{S}})$。因此，基于估计值的方差

影响函数可表示为

$$\widehat{\mathcal{V}}_{\mathrm{IF}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\varphi}(\mathcal{O}_i^{\mathcal{S}})^2.$$

在此处考虑的有限支撑情形下，根据标准的 M 估计理论，非参数最大似然估计量是正则且渐近线性的（RAL）。

估计量 $\widehat{\theta}^{\mathrm{WR}}$ 是一致的， $n^{-1}\widehat{\mathcal{V}}_{\mathrm{IF}}$ 是其渐近方差的估计值，即

$$\sqrt{n/\widehat{\mathcal{V}}_{\mathrm{IF}}}(\widehat{\theta}^{\mathrm{WR}} - \theta^{\mathrm{WR}}) \xrightarrow{d} \mathcal{N}(0,1).$$

A.4 协变量调整估计量的推导与性质

考虑 Y_1 和 Y_2 均为离散变量，并且不失一般性，假设 $\mathcal{S}$-评分的支撑集为 $\{0,1,\ldots,h,h+1,\ldots,h+\tau+1\}$，其中 $h<\infty$ 为正数（时间范围），且 $\tau=\max(Y_2)<\infty$。$F_z(s)$、 $F_{1z}(s)$ 和 $F_{2z}(s)$ 分别表示治疗组$Z=z$ 中 $\mathcal{S}$、 Y_1 和 Y_2 在 s 处的累积分布函数，其中 $z\in\{a,b\}$。

对于任意 $s\le h$，我们有 $F_z(s)=F_{1z}(s)$，因此可以直接应用 Kaplan-Meier估计量。对于$s>h$，注意到 $\mathcal{S}=h+1+Y_2=h+1+(s-h-1)=s$，这对应于 $Y_1>h$和 $Y_2=s-h-1$这一事件。因此，$\mathcal{S}$ 在 s 处的累积分布函数为 $F_z(s)=F_{1z}(h)+1-F_{1z}(h)F_{2z}(s-h-1)$。相应地，我们的目标是使用 Kaplan-Meier估计量估计 $s\in\{1,\ldots,h\}$ 的 $F_{1z}(s)$，并使用协变量调整估计量估计 $s\in\{1,\ldots,\tau\}$ 的 $F_{2z}(s)$。

为此，我们重点估计参数向量（$F_{1z}(1)\ldots,F_{1z}(h)$ $\boldsymbol{\beta}_z, F_{2z}(1)\ldots,F_{2z}(\tau)$），其中 $\boldsymbol{\beta}_z$ 对 R_{2z}的缺失模型进行参数化。具体而言，我们将 $\pi_z(\mathbf{X})=P(R_{2z}=1\mid \mathbf{X})$定义为在给定协变量X条件下观测到 Y_2的倾向评分。在假设3下， $\pi_z(\mathbf{X})$作为汇总评分，表示参与者在给定其基线协变量的情况下观测到 Y_2的概率。由于 $\pi_z(\mathbf{X})$通常未知，我们指定一个参数回归模型（通常为逻辑回归），其形式为$\pi_z(\mathbf{X};\boldsymbol{\beta}_z)=\{1+\exp(-\mathbf{X}'\boldsymbol{\beta}_z)\}^{-1}$，其中 $\boldsymbol{\beta}_z$ 通常与X具有相同的维度。上述参数向量的无偏

估计方程组定义为

$$\frac{1}{n}\sum_{i=1}^{n} I(Z_i = z) \begin{pmatrix} \Delta_{1i}\{I(Y_{1i} \le 1) - F_{1z}(1)\} \\ \vdots \\ \Delta_{1i}\{I(Y_{1i} \le h) - F_{1z}(h)\} \\ \phi_z(\boldsymbol{\beta}_z; R_{2i}, \mathbf{X}_i) \\ R_{2i}\pi_z(\mathbf{X}_i; \boldsymbol{\beta}_z)^{-1}\{I(Y_{2i} \le 1) - F_{2z}(1)\} \\ \vdots \\ R_{2i}\pi_z(\mathbf{X}_i; \boldsymbol{\beta}_z)^{-1}\{I(Y_{2i} \le \tau) - F_{2z}(\tau)\} \end{pmatrix} = 0, \tag{A.3}$$

其中，前 h 项定义了 Y_{1z} 的累积分布函数的 Kaplan-Meier估计量，$\phi_z(\boldsymbol{\beta}_z; R_2, \mathbf{X})$ 是针对 $\boldsymbol{\beta}_z$ 的估计方程（例如逻辑回归模型的得分方程），最后的 τ 项定义了 Y_2 的累积分布函数的逆概率加权估计量。

令 $(\widehat{F}_{1z}(1), \ldots, \widehat{F}_{1z}(h), \widehat{\boldsymbol{\beta}}_z, \widehat{F}_{2z}(1), \ldots, \widehat{F}_{2z}(\tau))$ 表示上述估计方程组的解。于是，所得的 $\mathcal{S}$-评分的累积分布函数估计量为

$$\widehat{F}_z(s) = I(s \le h) \cdot \widehat{F}_{1z}(s) + I(s > h) \cdot [\widehat{F}_{1z}(h) + \{1 - \widehat{F}_{1z}(h)\}\widehat{F}_{2z}(s - h - 1)], \tag{A.4}$$

对于 $s \in \{0, 1, \ldots, h, h+1, \ldots, h+\tau+1\}$ 满足 $\widehat{F}_z(0) = 0$。将该协变量调整估计量代入（8），即可得到胜率的协变量调整估计量，该估计量在第2.6节中称为半参数估计量。

到目前为止，我们已推导出在生存时间和非生存终点均取离散值的情形下的协变量调整估计量。如果至少一个终点是连续的，则式（A.3）中的相应 M估计方程将变为无限维，从而带来理论和计算方面的挑战。一种实用的解决办法是将连续结局粗化为离散类别，例如将精确时间四舍五入到最接近的天或月。不过，将协变量调整估计量及其渐近方差估计量推广到完全连续的情形，仍有待未来研究解决。

A.4.1 协变量调整估计量的一致性

为证明一致性，我们采用与附录A.2中类似的论证。一旦证明协变量调整后的累积分布函数估计量收敛（即 $\widehat{F}_{1z}(s) \xrightarrow{p} F_{1z}(s)$ 对于 $s \in [0, h]$ 以及 $\widehat{F}_{2z}(s) \xrightarrow{p} F_{2z}(s)$ 对于 $s \in [0, \tau]$，且对两者均成立 $z \in \{a, b\}$），我们便可证明协变量调整后的胜率估计量收敛于真实胜率比。

我们假设倾向得分模型设定正确（依据假设 3），即估计参数收敛于概率极限 $\widehat{\boldsymbol{\beta}}_z \xrightarrow{p} \boldsymbol{\beta}_z^*$，且 $\pi_z(\mathbf{X}) = \pi_z(\mathbf{X}; \boldsymbol{\beta}_z^*)$ 是真实的倾向评分，对于 $z \in \{a, b\}$。

$\widehat{\ }$在第 $F_{1z}(s)$ 上的一致收敛性 $s \in [0, h]$ 已在第 A.2节中证明；我们只需关注$\curlywedge F_{2z}(s)$ 对任意 $s \in [0, \tau]$。注意，利用估计方程（A.3），我们有

$$\widehat{F}_{2z}(s) = \frac{n^{-1}\sum_{i=1}^{n} I(Z_i = z) R_{2i} \pi_z(\mathbf{X}_i; \widehat{\boldsymbol{\beta}}_z)^{-1} I(Y_{2i} \le s)}{n^{-1}\sum_{i=1}^{n} I(Z_i = z) R_{2i} \pi_z(\mathbf{X}_i; \widehat{\boldsymbol{\beta}}_z)^{-1}} := \frac{I_1}{I_2}.$$

分子 $I_1 = \frac{1}{n}\sum_{i=1}^{n} I(Z_i = z) \frac{R_{2i}}{\pi_z}(\mathbf{X}; \widehat{\boldsymbol{\beta}}_z) I(Y_{2i} \le s) \xrightarrow{p} \mathbb{E}\left\{ I(Z = z) \frac{R_2}{\pi_z}(\mathbf{X})\, I(Y_2 \le s) \right\}$，其中

$$\begin{aligned}
\mathbb{E}\left\{ I(Z = z) \frac{R_2}{\pi_z(\mathbf{X})} I(Y_2 \le s) \right\} &= \mathbb{E}\left[\mathbb{E}\left\{ I(Z = z) \frac{R_2}{\pi_z(\mathbf{X})} I(Y_2 \le s) \,\middle|\, \mathbf{X}, Z = z \right\} \right] \\
&= \mathbb{E}\left[\frac{I(Z = z)}{\pi_z(\mathbf{X})} \mathbb{E}\left\{ R_2 I(Y_2 \le s) \mid \mathbf{X}, Z = z \right\} \right] \\
&= \mathbb{E}\left[\frac{I(Z = z)}{\pi_z(\mathbf{X})} \mathbb{E}\{ I(Y_2 \le s) \mid \mathbf{X}, Z = z \} \mathbb{P}(R_2 = 1 \mid \mathbf{X}, Z = z) \right] \quad \text{(by Assumption 3)} \\
&= \mathbb{E}\left[\frac{I(Z = z)}{\pi_z(\mathbf{X})} \pi_z(\mathbf{X}) \mathbb{E}\{ I(Y_2 \le s) \mid \mathbf{X}, Z = z \} \right] = \mathbb{P}(Z = z, Y_2 \le s).
\end{aligned}$$

此外，$I_2 = \frac{1}{n}\sum_{i=1}^{n} I(Z_i = z) \frac{R_{2i}}{\pi_z}(\mathbf{X}; \widehat{\boldsymbol{\beta}}_z) \xrightarrow{p} \mathbb{E}\left\{ I(Z = z) \frac{R_2}{\pi_z}(\mathbf{X}) \right\}$ 其中

$$\begin{aligned}
\mathbb{E}\left\{ I(Z = z) \frac{R_2}{\pi_z(\mathbf{X})} \right\} &= \mathbb{E}\left[\mathbb{E}\left\{ I(Z = z) \frac{R_2}{\pi_z(\mathbf{X})} \,\middle|\, \mathbf{X}, Z = z \right\} \right] = \mathbb{E}\left[\frac{I(Z = z)}{\pi_z(\mathbf{X})} \mathbb{E}\left\{ R_2 \mid \mathbf{X}, Z = z \right\} \right] \\
&= \mathbb{E}\left[\frac{I(Z = z)}{\pi_z(\mathbf{X})} \mathbb{P}(R_2 = 1 \mid \mathbf{X}, Z = z) \right] = \mathbb{E}\left[\frac{I(Z = z)}{\pi_z(\mathbf{X})} \pi_z(\mathbf{X}) \right] = \mathbb{P}(Z = z).
\end{aligned}$$

因此，

$$\widehat{F}_{2z}(s) \xrightarrow{p} \frac{\mathbb{P}(Z = z, Y_2 \le s)}{\mathbb{P}(Z = z)} = \mathbb{P}(Y_2 \le s \mid Z = z) = F_{2z}(s).$$

由于对任意 $s \in [0, \tau]$，在 $0 < \tau < \infty$的情况下，证明过程都相同，因此我们可以构造 $\curlywedge F_{2z}(s)$ 在 s上的一致一致性。随后，估计量（A.4）在 $\mathcal{S} \in [0, h+\tau+1]$ 上的一致性即可立即得到。

A.4.2 协变量调整估计量的方差估计

除了证明估计量的一致性之外，我们还推导了估计量的渐近方差（A.4），推导过程与附录 A.3中的类似。我们还考虑 $\mathcal{S}$-评分 $\mathcal{S}_z$ 组 z 取值于

属于集合 $\{1,\ldots,h,h+1,\ldots,h+\tau+1\}$，其中 h和 τ是一些正整数。对于 $\mathcal{S}_z \le h$，我们有$\mathcal{S}_z = Y_{1z}$。在 h之后， $\mathcal{S}_z = 1+h+Y_{2z}$，并且我们对 h之后的取值使用逆概率加权权重。

由于倾向评分由 $\boldsymbol{\beta}_z$ 参数化，我们将参数向量视为条件生存概率，并将 $(\boldsymbol{\beta}_a', \boldsymbol{\beta}_b')'$ 定义为 $\boldsymbol{\theta} = (\boldsymbol{\beta}_a', q_1^a, \ldots, q_{M-1}^a, \boldsymbol{\beta}_b', q_1^b, \ldots, q_{M-1}^b)'$，其中 $M = h+\tau+1$。参数向量 $\boldsymbol{\theta}$ 可以通过求解以下方程联合估计

$$\frac{1}{n}\sum_{i=1}^{n}\phi_z(\mathcal{O}^{\mathcal{S}}) = 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\psi_s^z(\mathcal{O}_i^{\mathcal{S}}) = 0,$$

对于一份观测数据 $\mathcal{O}^{\mathcal{S}}$，我们有

$$\begin{aligned}
\phi_z &= \phi_z(\mathcal{O}^{\mathcal{S}}) = I(Z=z)\phi_z(\widehat{\boldsymbol{\beta}}_z; R_2, \mathbf{X}) = 0,\\
\psi_s^z &= \psi_s^z(\mathcal{O}^{\mathcal{S}}) = I(Z=z)E_s(A_s - \widehat{q}_s^z) = 0 \quad \text{for } s \le h, \quad \text{and}\\
\psi_s^z &= \psi_s^z(\mathcal{O}^{\mathcal{S}}) = I(Z=z)\pi_z(\mathbf{X};\widehat{\boldsymbol{\beta}}_z)^{-1}E_s(A_s - \widehat{q}_s^z) = 0 \quad \text{for } s > h,
\end{aligned}$$

作为示例，我们考虑这样一种特殊情形：使用逻辑回归估计倾向评分（这是实践中常见的选择）。最后，我们将说明，倾向得分模型的选择不会影响方差公式的表达式，因为倾向评分在 θ^{WR} 估计目标的定义中不起作用，且它仅是一个干扰参数。

令 $\pi_z(\mathbf{X};\boldsymbol{\beta}_z) = \{1+\exp(-\mathbf{X}'\boldsymbol{\beta}_z)\}^{-1}$ 为通过逻辑回归估计的倾向评分，并定义 $\phi_z(\boldsymbol{\beta}_z;\ R_2, \mathbf{X}) = \{R_2 - \pi_z(\mathbf{X};\ \boldsymbol{\beta}_z)\}\mathbf{X},$，其中 $z \in \{a,b\}$。我们将这些估计方程堆叠成向量

$$\Psi(\boldsymbol{q}) = \left(\phi_a, \psi_1^a, \ldots, \psi_{M-1}^a, \phi_b, \psi_1^b, \ldots, \psi_{M-1}^b\right)'.$$

令 $B(\boldsymbol{\theta}) = -\partial\Psi(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$，该矩阵在多个分块上为对角矩阵，因为上述向量中的每个方程仅取决于 $\boldsymbol{\theta}$ 中的一个唯一元素。对于组 z，其导数计算如下：

$$\begin{aligned}
\frac{\partial\phi_z}{\partial\boldsymbol{\beta}_z} &= -\pi_z(\mathbf{X};\boldsymbol{\beta}_z)\{1-\pi_z(\mathbf{X};\boldsymbol{\beta}_z)\}\mathbf{X}\mathbf{X}' \quad \text{(specific to logistic regression)},\\
\frac{\partial\psi_s^z}{\partial q_s^z} &= -I(Z=z)E_s \quad \text{for } s \le h, \quad \text{and}\\
\frac{\partial\psi_s^z}{\partial q_s^z} &= -I(Z=z)\pi_z(\mathbf{X};\boldsymbol{\beta}_z)^{-1}E_s \quad \text{for } t > h.
\end{aligned}$$

根据 Van der Vaart(2000)，影响函数 $\boldsymbol{\theta}$ 为

$$\eta = \eta(\mathcal{O}^{\mathcal{S}}) = B(\boldsymbol{\theta})^{-1}\Psi(\boldsymbol{\theta}),$$

其中 $\mathcal{O}^{\mathcal{S}}$ 是包含协变量$\mathbf{X}$的观测数据。

接下来，为了求得 A.4 中协变量调整估计量的影响函数，我们需要进一步求出以下导数： θ^{WR}

$$C=\left(\frac{\partial\theta^{\mathrm{WR}}}{\partial\boldsymbol{\beta}_a'},\frac{\partial\theta^{\mathrm{WR}}}{\partial q_1^a},\ldots,\frac{\partial\theta^{\mathrm{WR}}}{\partial q_{M-1}^a},\frac{\partial\theta^{\mathrm{WR}}}{\partial\boldsymbol{\beta}_b'},\frac{\partial\theta^{\mathrm{WR}}}{\partial q_1^b},\ldots,\frac{\partial\theta^{\mathrm{WR}}}{\partial q_{M-1}^b}\right)'.$$

注意到

$$\frac{\partial\theta^{\mathrm{WR}}}{\partial\boldsymbol{\beta}_z}=\mathbf{0},$$

这是因为胜率的定义不涉及倾向评分。倾向评分只是我们在估计过程中推导出的一个中间干扰函数。

其他推导与 A.3 节中的推导完全相同。最后，协变量调整估计量（A.4）的影响函数为

$$\begin{aligned}\varphi(\mathcal{O}^{\mathcal{S}})&=C\eta(\mathcal{O}^{\mathcal{S}})\\&=\sum_{s=1}^{M-1}\frac{1}{q_s^a}\sum_{k=s}^{M-1}\bar{F}_k^aD^{-2}\Big[D(\bar{F}_{k-1}^b-\bar{F}_k^b)-N(\bar{F}_k^b-\bar{F}_{k+1}^b)\Big]\frac{I(Z=a)\{I(s\le h)+I(s>h)\pi_a(\mathbf{X};\boldsymbol{\beta}_a)^{-1}\}E_s(A_s-q_s^a)}{\mathbb{E}[I(Z=a)\{I(s\le h)+I(s>h)\pi_a(\mathbf{X};\boldsymbol{\beta}_a)^{-1}\}E_s]}\\&\quad+\sum_{s=1}^{M-1}\frac{1}{q_s^b}\sum_{k=s}^{M-1}\bar{F}_k^bD^{-2}\Big[D(\bar{F}_k^a-\bar{F}_{k+1}^a)-N(\bar{F}_{k-1}^a-\bar{F}_k^a)\Big]\frac{I(Z=b)\{I(s\le h)+I(s>h)\pi_b(\mathbf{X};\boldsymbol{\beta}_b)^{-1}\}E_s(A_s-q_s^a)}{\mathbb{E}[I(Z=b)\{I(s\le h)+I(s>h)\pi_b(\mathbf{X};\boldsymbol{\beta}_b)^{-1}\}E_s]},\end{aligned}$$

其中，符号 N、 D、 $\bar{F}_k^a$、 $\bar{F}_k^b$在第 A.3 节中定义。因此，我们可以用估计得到的影响函数的均方来一致地估计渐近方差。

在线补充材料

S.1 完整模拟结果

S.1.1 完整模拟设置的表格

表 S.1.1：模拟结果： Y_1无删失。

			$\theta^{\mathrm{WR}}=1$				$\theta^{\mathrm{WR}}=2$			
(n_a, n_b)	缺失Y_2的数据	方法	ARB%	RMSE	CP%	宽度	ARB%	RMSE	CP%	宽度
(100, 100) 小	No	$\mathcal{S}$-评分 (IF-Wald)	1.78	0.166	95.55	0.66	2.69	0.390	94.85	1.48
		$\mathcal{S}$-评分 (BT-Wald)	1.78	0.166	96.25	0.68	2.69	0.390	95.60	1.55
		$\mathcal{S}$-评分 (BT-QT)	1.78	0.166	95.10	0.67	2.69	0.390	**93.55**	1.54
		Pocock	1.78	0.166	95.40	0.67	2.69	0.390	95.10	1.51
	MCAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	1.88	0.168	94.95	0.66	1.68	0.369	95.00	1.46
		$\mathcal{S}$-评分 (BT-Wald)	1.88	0.168	95.50	0.68	1.68	0.369	95.80	1.54
		$\mathcal{S}$-评分 (BT-QT)	1.88	0.168	94.65	0.67	1.68	0.369	94.70	1.52
		Pocock	1.93	0.171	95.35	0.68	3.15	0.358	94.25	1.43
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	1.65	0.170	94.75	0.66	2.06	0.391	94.25	1.46
		$\mathcal{S}$-评分 (BT-Wald)	1.65	0.170	95.40	0.68	2.06	0.391	95.30	1.54
		$\mathcal{S}$-评分 (BT-QT)	1.65	0.170	94.45	0.67	2.06	0.391	**93.95**	1.52
		Pocock	1.76	0.175	95.25	0.70	7.72	0.392	**90.25**	1.37
	MAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	1.78	0.171	95.20	0.66	2.13	0.392	94.90	1.47
		$\mathcal{S}$-评分 (BT-Wald)	1.78	0.171	95.40	0.68	2.13	0.392	95.80	1.54
		$\mathcal{S}$-评分 (BT-QT)	1.78	0.171	94.00	0.67	2.13	0.392	94.95	1.53
		Pocock	5.82	0.169	**93.30**	0.63	6.39	0.382	**92.60**	1.38
	MAR （40%）	$\mathcal{S}$-评分（IF-Wald）	2.20	0.173	94.85	0.66	2.19	0.373	95.45	1.47
		$\mathcal{S}$-评分（BT-Wald）	2.20	0.173	95.30	0.68	2.19	0.373	96.40	1.54
		$\mathcal{S}$-评分（BT-QT）	2.20	0.173	**93.60**	0.67	2.19	0.373	94.95	1.53
		Pocock	13.12	0.200	**85.80**	0.60	14.93	0.441	**84.05**	1.27
(1000, 1000) 大	No	$\mathcal{S}$-评分（IF-Wald）	0.37	0.052	95.00	0.20	0.30	0.114	95.00	0.45
		$\mathcal{S}$-评分 (BT-Wald)	0.37	0.052	94.50	0.20	0.30	0.114	95.05	0.45
		$\mathcal{S}$-评分 (BT-QT)	0.37	0.052	94.00	0.20	0.30	0.114	94.35	0.45
		Pocock	0.37	0.052	94.65	0.20	0.30	0.114	94.90	0.45
	MCAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.04	0.053	94.70	0.20	0.06	0.113	95.10	0.45
		$\mathcal{S}$-评分 (BT-Wald)	0.04	0.053	94.65	0.20	0.06	0.113	95.05	0.45
		$\mathcal{S}$-评分 (BT-QT)	0.04	0.053	94.25	0.20	0.06	0.113	94.95	0.45
		Pocock	0.03	0.054	94.40	0.21	4.64	0.145	**86.15**	0.43
	MCAR （40%）	$\mathcal{S}$-评分 (IF-Wald)	0.10	0.052	95.05	0.20	0.01	0.113	95.45	0.45
		$\mathcal{S}$-评分 (BT-Wald)	0.10	0.052	94.80	0.20	0.01	0.113	95.50	0.45
		$\mathcal{S}$-评分（BT-QT）	0.10	0.052	94.50	0.20	0.01	0.113	95.00	0.45
		Pocock	0.11	0.054	95.45	0.21	9.67	0.226	**53.30**	0.41
	MAR （20%）	$\mathcal{S}$-评分（IF-Wald）	0.51	0.054	94.05	0.20	0.05	0.113	95.25	0.45
		$\mathcal{S}$-评分（BT-Wald）	0.51	0.054	94.20	0.20	0.05	0.113	95.25	0.45
		$\mathcal{S}$-评分（BT-QT）	0.51	0.054	93.70	0.20	0.05	0.113	95.00	0.45
		Pocock	6.69	0.084	**73.10**	0.19	8.26	0.201	**64.20**	0.42
	MAR （40%）	$\mathcal{S}$-评分（IF-Wald）	0.48	0.051	**96.40**	0.20	0.08	0.110	95.90	0.45
		$\mathcal{S}$-评分（BT-Wald）	0.48	0.051	**96.40**	0.20	0.08	0.110	95.85	0.45
		$\mathcal{S}$-评分（BT-QT）	0.48	0.051	95.85	0.20	0.08	0.110	95.70	0.45
		Pocock	14.57	0.152	**15.95**	0.18	16.80	0.362	**8.70**	0.38

区间 [94, 96]之外的覆盖概率以粗体突出显示。

表 S.1.2：两个治疗组中的低（约 20%）同质删失模拟结果。

$(n_a,\ n_b)$	缺失数据 Y_2	方法	$\theta^{\mathrm{WR}}=1$ ARB%	RMSE	CP%	宽度	$\theta^{\mathrm{WR}}=2$ ARB%	RMSE	CP%	宽度
(100, 100) 小	No	$\mathcal{S}$-评分（IF-Wald）	1.70	0.180	95.20	0.68	2.34	0.396	94.90	1.51
		$\mathcal{S}$-评分（BT-Wald）	1.70	0.180	95.85	0.71	2.34	0.396	95.80	1.60
		$\mathcal{S}$-评分（BT-QT）	1.70	0.180	94.65	0.70	2.34	0.396	94.40	1.58
		Pocock	1.89	0.189	95.00	0.73	2.24	0.411	94.95	1.60
	MCAR (20%)	$\mathcal{S}$-评分（IF-Wald）	1.52	0.177	94.35	0.68	1.54	0.387	94.95	1.50
		$\mathcal{S}$-评分 (BT-Wald)	1.52	0.177	94.90	0.71	1.54	0.387	95.45	1.58
		$\mathcal{S}$-评分 (BT-QT)	1.52	0.177	94.35	0.70	1.54	0.387	94.60	1.56
		Pocock	1.74	0.189	95.05	0.74	2.49	0.394	**93.70**	1.53
	MCAR （40%）	$\mathcal{S}$-评分 (IF-Wald)	1.64	0.178	94.85	0.69	1.90	0.381	95.75	1.50
		$\mathcal{S}$-评分（BT-Wald）	1.64	0.178	95.40	0.71	1.90	0.381	95.95	1.58
		$\mathcal{S}$-评分（BT-QT）	1.64	0.178	95.15	0.70	1.90	0.381	95.40	1.56
		Pocock	2.36	0.195	95.35	0.76	6.11	0.390	**92.95**	1.49
	MAR (20%)	$\mathcal{S}$-评分（IF-Wald）	1.49	0.175	94.65	0.68	1.97	0.398	94.65	1.50
		$\mathcal{S}$-评分 (BT-Wald)	1.49	0.175	95.35	0.71	1.97	0.398	95.35	1.59
		$\mathcal{S}$-评分 (BT-QT)	1.49	0.175	95.25	0.70	1.97	0.398	94.40	1.57
		Pocock	4.69	0.181	**93.90**	0.70	4.93	0.402	**93.30**	1.50
	MAR （40%）	$\mathcal{S}$-评分 (IF-Wald)	2.38	0.176	95.70	0.69	1.43	0.390	94.65	1.49
		$\mathcal{S}$-评分 (BT-Wald)	2.38	0.176	**96.20**	0.72	1.43	0.390	95.30	1.58
		$\mathcal{S}$-评分 (BT-QT)	2.38	0.176	95.25	0.71	1.43	0.390	94.80	1.56
		Pocock	10.95	0.199	**89.20**	0.66	12.44	0.435	**87.30**	1.39
(1000, 1000) 大	No	$\mathcal{S}$-评分 (IF-Wald)	0.18	0.054	94.15	0.21	0.16	0.113	95.60	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.18	0.054	94.45	0.21	0.16	0.113	95.90	0.46
		$\mathcal{S}$-评分 (BT-QT)	0.18	0.054	**93.80**	0.21	0.16	0.113	95.40	0.46
		Pocock	0.18	0.056	94.90	0.22	0.15	0.117	95.95	0.48
	MCAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.09	0.053	95.15	0.21	0.25	0.119	94.75	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.09	0.053	95.15	0.21	0.25	0.119	94.95	0.46
		$\mathcal{S}$-评分 (BT-QT)	0.09	0.053	94.55	0.21	0.25	0.119	94.45	0.46
		Pocock	0.13	0.058	94.80	0.23	3.84	0.143	**88.75**	0.46
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.18	0.054	95.15	0.21	0.06	0.119	94.40	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.18	0.054	95.15	0.21	0.06	0.119	94.40	0.46
		$\mathcal{S}$-评分 (BT-QT)	0.18	0.054	94.55	0.21	0.06	0.119	94.35	0.46
		Pocock	0.24	0.059	94.45	0.23	7.90	0.201	**69.90**	0.45
	MAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.51	0.055	94.85	0.21	0.01	0.118	94.45	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.51	0.055	94.85	0.21	0.01	0.118	94.70	0.46
		$\mathcal{S}$-评分 (BT-QT)	0.51	0.055	94.00	0.21	0.01	0.118	94.30	0.46
		Pocock	5.72	0.079	**81.70**	0.21	6.97	0.185	**75.80**	0.45
	MAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.76	0.052	96.00	0.21	0.00	0.117	95.10	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.76	0.052	95.95	0.21	0.00	0.117	95.10	0.46
		$\mathcal{S}$-评分 (BT-QT)	0.76	0.052	95.85	0.21	0.00	0.117	95.25	0.46
		Pocock	12.23	0.132	**36.95**	0.20	13.77	0.305	**29.15**	0.42

区间 [94, 96]之外的覆盖概率以粗体突出显示。

表 S.1.3：模拟结果：两个治疗组中度（约 40%）且同质删失。

$(n_a,\ n_b)$	缺失 Y_2的数据	方法	$\theta^{\mathrm{WR}}=1$ ARB%	RMSE	CP%	宽度	$\theta^{\mathrm{WR}}=2$ ARB%	RMSE	CP%	宽度
(100, 100) 小	No	$\mathcal{S}$-评分（IF-Wald）	1.90	0.199	95.10	0.77	2.42	0.424	95.25	1.63
		$\mathcal{S}$-评分（BT-Wald）	1.90	0.199	**96.15**	0.80	2.42	0.424	**96.40**	1.74
		$\mathcal{S}$-评分（BT-QT）	1.90	0.199	95.00	0.79	2.42	0.424	95.25	1.71
		Pocock	2.28	0.224	94.95	0.88	0.43	0.465	94.85	1.82
	MCAR (20%)	$\mathcal{S}$-评分（IF-Wald）	2.54	0.199	95.30	0.77	1.73	0.422	94.45	1.61
		$\mathcal{S}$-评分 (BT-Wald)	2.54	0.199	95.70	0.81	1.73	0.422	95.65	1.73
		$\mathcal{S}$-评分 (BT-QT)	2.54	0.199	95.20	0.80	1.73	0.422	95.00	1.70
		Pocock	3.26	0.227	95.25	0.90	3.18	0.448	**93.70**	1.75
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	1.40	0.200	**93.90**	0.77	2.50	0.428	94.75	1.63
		$\mathcal{S}$-评分 (BT-Wald)	1.40	0.200	94.95	0.80	2.50	0.428	95.90	1.74
		$\mathcal{S}$-评分 (BT-QT)	1.40	0.200	94.65	0.79	2.50	0.428	94.80	1.72
		Pocock	2.25	0.228	95.25	0.91	4.78	0.455	**93.15**	1.73
	MAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	3.13	0.204	95.10	0.78	2.24	0.425	**93.65**	1.62
		$\mathcal{S}$-评分 (BT-Wald)	3.13	0.204	95.85	0.82	2.24	0.425	95.05	1.74
		$\mathcal{S}$-评分 (BT-QT)	3.13	0.204	94.65	0.81	2.24	0.425	94.30	1.71
		Pocock	1.06	0.219	94.25	0.87	4.42	0.451	**92.60**	1.73
	MAR （40%）	$\mathcal{S}$-评分 (IF-Wald)	2.82	0.198	95.30	0.77	2.60	0.433	94.95	1.63
		$\mathcal{S}$-评分 (BT-Wald)	2.82	0.198	96.00	0.82	2.60	0.433	95.85	1.74
		$\mathcal{S}$-评分（BT-QT）	2.82	0.198	94.90	0.81	2.60	0.433	94.50	1.72
		Pocock	6.37	0.213	94.35	0.83	8.61	0.465	**90.70**	1.67
(1000, 1000) 大	No	$\mathcal{S}$-评分（IF-Wald）	0.27	0.061	94.15	0.24	0.14	0.131	94.15	0.50
		$\mathcal{S}$-评分（BT-Wald）	0.27	0.061	94.30	0.24	0.14	0.131	94.05	0.50
		$\mathcal{S}$-评分（BT-QT）	0.27	0.061	94.20	0.23	0.14	0.131	**93.85**	0.49
		Pocock	0.33	0.068	94.55	0.26	2.70	0.152	**92.05**	0.53
	MCAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.37	0.061	95.30	0.24	0.16	0.129	94.70	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.37	0.061	95.75	0.24	0.16	0.129	94.95	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.37	0.061	94.70	0.23	0.16	0.129	94.85	0.49
		Pocock	0.42	0.069	95.00	0.27	5.11	0.172	**87.05**	0.52
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.03	0.059	94.70	0.24	0.16	0.126	94.80	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.03	0.059	94.70	0.24	0.16	0.126	95.05	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.03	0.059	94.65	0.23	0.16	0.126	94.80	0.49
		Pocock	0.05	0.069	94.80	0.27	7.57	0.204	**77.30**	0.51
	MAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.66	0.062	95.45	0.24	0.27	0.128	94.65	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.66	0.062	95.60	0.24	0.27	0.128	94.85	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.66	0.062	94.70	0.24	0.27	0.128	94.80	0.49
		Pocock	3.86	0.077	**90.80**	0.26	6.86	0.194	**80.10**	0.51
	MAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.56	0.059	94.90	0.24	0.22	0.126	94.90	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.56	0.059	95.10	0.24	0.22	0.126	95.35	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.56	0.059	94.90	0.24	0.22	0.126	94.85	0.49
		Pocock	8.66	0.107	**73.65**	0.25	11.20	0.265	**55.85**	0.49

区间 [94, 96]之外的覆盖概率以粗体突出显示。

表 S.1.4：模拟 结果：两组中较低（约 20%）且具有异质性的删失。

$(n_a,\ n_b)$	缺失数据 Y_2	方法	$\theta^{\mathrm{WR}}=1$ ARB%	RMSE	CP%	宽度	$\theta^{\mathrm{WR}}=2$ ARB%	RMSE	CP%	宽度
（100，100）小	No	$\mathcal{S}$-评分（IF-Wald）	1.81	0.179	95.05	0.69	1.91	0.383	94.85	1.51
		$\mathcal{S}$-评分（BT-Wald）	1.81	0.179	95.55	0.72	1.91	0.383	95.80	1.60
		$\mathcal{S}$-评分（BT-QT）	1.81	0.179	94.20	0.71	1.91	0.383	95.40	1.58
		Pocock	4.46	0.183	**93.80**	0.70	0.36	0.392	95.40	1.59
	MCAR（20%）	$\mathcal{S}$-评分（IF-Wald）	0.93	0.179	**93.65**	0.69	1.98	0.386	95.35	1.51
		$\mathcal{S}$-评分（BT-Wald）	0.93	0.179	94.50	0.72	1.98	0.386	**96.30**	1.60
		$\mathcal{S}$-评分（BT-QT）	0.93	0.179	94.20	0.71	1.98	0.386	95.60	1.58
		Pocock	3.13	0.189	**93.75**	0.73	2.35	0.395	95.05	1.57
	MCAR（40%）	$\mathcal{S}$-评分（IF-Wald）	1.67	0.178	95.05	0.70	2.50	0.399	95.00	1.52
		$\mathcal{S}$-评分（BT-Wald）	1.67	0.178	95.15	0.72	2.50	0.399	95.65	1.61
		$\mathcal{S}$-评分（BT-QT）	1.67	0.178	95.00	0.71	2.50	0.399	94.65	1.59
		Pocock	1.03	0.194	94.95	0.77	4.37	0.400	**93.15**	1.54
	MAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	2.09	0.182	94.60	0.70	2.76	0.398	95.05	1.53
		$\mathcal{S}$-评分 (BT-Wald)	2.09	0.182	95.10	0.72	2.76	0.398	**96.15**	1.62
		$\mathcal{S}$-评分 (BT-QT)	2.09	0.182	94.20	0.71	2.76	0.398	94.40	1.60
		Pocock	7.46	0.192	**92.25**	0.69	4.02	0.397	**93.80**	1.54
	MAR（40%）	$\mathcal{S}$-评分 (IF-Wald)	2.70	0.188	94.90	0.70	2.75	0.392	95.45	1.53
		$\mathcal{S}$-评分（BT-Wald）	2.70	0.188	95.50	0.73	2.75	0.392	**96.45**	1.62
		$\mathcal{S}$-评分（BT-QT）	2.70	0.188	94.45	0.72	2.75	0.392	94.65	1.60
		Pocock	10.40	0.202	**89.50**	0.68	8.94	0.409	**90.70**	1.47
(1000, 1000) 大	No	$\mathcal{S}$-评分（IF-Wald）	0.25	0.056	95.30	0.21	0.03	0.118	94.60	0.46
		$\mathcal{S}$-评分（BT-Wald）	0.25	0.056	95.25	0.22	0.03	0.118	94.80	0.47
		$\mathcal{S}$-评分 (BT-QT)	0.25	0.056	94.60	0.21	0.03	0.118	94.35	0.46
		Pocock	6.00	0.082	**79.35**	0.21	2.35	0.131	**92.35**	0.48
	MCAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.25	0.055	95.20	0.21	0.05	0.119	95.00	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.25	0.055	95.35	0.22	0.05	0.119	94.85	0.47
		$\mathcal{S}$-评分 (BT-QT)	0.25	0.055	94.55	0.21	0.05	0.119	94.80	0.46
		Pocock	3.72	0.068	**91.15**	0.22	4.58	0.152	**86.85**	0.47
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.17	0.055	94.70	0.21	0.05	0.119	94.25	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.17	0.055	94.60	0.22	0.05	0.119	94.25	0.47
		$\mathcal{S}$-评分（BT-QT）	0.17	0.055	94.20	0.21	0.05	0.119	94.05	0.46
		Pocock	1.04	0.061	94.10	0.23	6.94	0.186	**76.30**	0.46
	MAR（20%）	$\mathcal{S}$-评分（IF-Wald）	0.58	0.056	95.10	0.22	0.32	0.120	94.75	0.47
		$\mathcal{S}$-评分（BT-Wald）	0.58	0.056	95.20	0.22	0.32	0.120	94.70	0.47
		$\mathcal{S}$-评分（BT-QT）	0.58	0.056	94.25	0.21	0.32	0.120	94.10	0.46
		Pocock	8.86	0.104	**63.60**	0.21	6.46	0.179	**78.85**	0.46
	MAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.67	0.057	94.70	0.22	0.15	0.119	95.15	0.46
		$\mathcal{S}$-评分 (BT-Wald)	0.67	0.057	94.90	0.22	0.15	0.119	94.90	0.47
		$\mathcal{S}$-评分 (BT-QT)	0.67	0.057	94.50	0.22	0.15	0.119	94.35	0.46
		Pocock	12.21	0.133	**39.95**	0.21	11.39	0.262	**45.65**	0.44

超出区间 [94, 96] 的覆盖概率以粗体突出显示。

表 S.1.5：模拟结果：两组均存在中等程度（约 40%）的异质性删失。

$(n_a,\ n_b)$	缺失 Y_2的数据	方法	$\theta^{\mathrm{WR}}=1$ ARB%	RMSE	CP%	宽度	$\theta^{\mathrm{WR}}=2$ ARB%	RMSE	CP%	宽度
(100, 100) 小	No	$\mathcal{S}$-评分 (IF-Wald)	2.07	0.205	94.60	0.79	2.80	0.441	94.45	1.65
		$\mathcal{S}$-评分 (BT-Wald)	2.07	0.205	95.70	0.83	2.80	0.441	95.20	1.77
		$\mathcal{S}$-评分 (BT-QT)	2.07	0.205	94.35	0.82	2.80	0.441	**93.75**	1.74
		Pocock	5.55	0.218	**93.40**	0.84	2.69	0.468	**92.60**	1.80
	MCAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	2.68	0.206	95.30	0.79	2.76	0.438	94.45	1.65
		$\mathcal{S}$-评分 (BT-Wald)	2.68	0.206	**96.45**	0.84	2.76	0.438	95.75	1.77
		$\mathcal{S}$-评分 (BT-QT)	2.68	0.206	95.45	0.83	2.76	0.438	94.60	1.74
		Pocock	2.95	0.219	95.25	0.87	3.55	0.469	**93.25**	1.79
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	1.96	0.208	94.25	0.79	1.83	0.426	**93.60**	1.63
		$\mathcal{S}$-评分 (BT-Wald)	1.96	0.208	**96.15**	0.84	1.83	0.426	94.75	1.75
		$\mathcal{S}$-评分 (BT-QT)	1.96	0.208	94.80	0.83	1.83	0.426	94.15	1.72
		Pocock	0.84	0.228	94.40	0.90	5.50	0.458	**92.70**	1.75
	MAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	2.72	0.208	94.85	0.79	2.27	0.446	94.35	1.64
		$\mathcal{S}$-评分 (BT-Wald)	2.72	0.208	95.75	0.85	2.27	0.446	95.75	1.76
		$\mathcal{S}$-评分 (BT-QT)	2.72	0.208	95.05	0.84	2.27	0.446	94.35	1.73
		Pocock	6.57	0.219	**93.65**	0.83	5.34	0.477	**92.50**	1.75
	MAR （40%）	$\mathcal{S}$-评分 (IF-Wald)	4.08	0.216	94.45	0.81	2.99	0.445	94.50	1.65
		$\mathcal{S}$-评分 (BT-Wald)	4.08	0.216	95.70	0.87	2.99	0.445	95.75	1.77
		$\mathcal{S}$-评分 (BT-QT)	4.08	0.216	94.30	0.86	2.99	0.445	**93.90**	1.75
		Pocock	8.29	0.224	**92.10**	0.83	7.08	0.467	**91.35**	1.72
(1000, 1000) 大	No	$\mathcal{S}$-评分 (IF-Wald)	0.14	0.060	95.45	0.24	0.13	0.125	94.90	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.14	0.060	95.45	0.24	0.13	0.125	95.25	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.14	0.060	95.65	0.24	0.13	0.125	95.35	0.50
		Pocock	7.95	0.102	**76.35**	0.25	5.92	0.180	**84.50**	0.53
	MCAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.22	0.062	95.10	0.24	0.30	0.127	95.25	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.22	0.062	95.15	0.24	0.30	0.127	95.35	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.22	0.062	94.80	0.24	0.30	0.127	94.35	0.50
		Pocock	5.41	0.085	**86.80**	0.26	6.57	0.189	**82.00**	0.52
	MCAR (40%)	$\mathcal{S}$-评分 (IF-Wald)	0.27	0.061	95.40	0.24	0.30	0.128	95.20	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.27	0.061	95.25	0.24	0.30	0.128	95.40	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.27	0.061	94.85	0.24	0.30	0.128	94.40	0.50
		Pocock	2.57	0.073	**92.80**	0.27	7.53	0.205	**77.70**	0.52
	MAR (20%)	$\mathcal{S}$-评分 (IF-Wald)	0.52	0.062	95.35	0.24	0.32	0.127	95.30	0.50
		$\mathcal{S}$-评分 (BT-Wald)	0.52	0.062	95.45	0.24	0.32	0.127	95.35	0.50
		$\mathcal{S}$-评分 (BT-QT)	0.52	0.062	94.90	0.24	0.32	0.127	94.55	0.50
		Pocock	8.86	0.109	**72.20**	0.25	7.72	0.206	**77.45**	0.52
	MAR （40%）	$\mathcal{S}$-评分（IF-Wald）	0.75	0.062	95.50	0.24	0.38	0.128	95.30	0.50
		$\mathcal{S}$-评分（BT-Wald）	0.75	0.062	95.55	0.25	0.38	0.128	95.30	0.50
		$\mathcal{S}$-评分（BT-QT）	0.75	0.062	95.00	0.24	0.38	0.128	94.45	0.50
		Pocock	9.89	0.117	**67.20**	0.25	9.76	0.241	**65.60**	0.51

区间 [94, 96]之外的覆盖概率以粗体突出显示。

表 S.1.6：主文各模拟情景下，基于影响函数和自助法的 $\mathcal{S}$-评分置信区间（置信区间）估计量的覆盖概率。 $n_a = n_b = 25$。超出 $[94, 96]$范围的覆盖概率值以粗体标出。

Y_1的删失	Y_2的缺失	$\theta^{\mathrm{WR}} = 1$			$\theta^{\mathrm{WR}} = 2$		
		IF-Wald	BT-Wald	BT-QT	IF-Wald	BT-Wald	BT-QT
无删失 Y_1	No	**92.95**	94.40	94.20	**92.80**	94.80	**93.85**
	完全随机缺失（40%）	**92.10**	94.50	94.10	**93.55**	95.50	**93.55**
	随机缺失（40%）	**93.85**	95.20	94.10	**93.25**	95.60	95.10
40% 删失 Y_1, 组间同质	No	**93.60**	94.70	95.40	**91.90**	94.60	94.25
	完全随机缺失（40%）	**92.75**	95.50	94.80	**91.80**	95.60	94.55
	随机缺失（40%）	94.40	**96.40**	95.70	**92.85**	95.70	94.90
40%删失 Y_1, 组间异质	No	**92.05**	95.35	95.20	**92.80**	94.75	95.10
	完全随机缺失（40%）	**93.75**	**96.40**	94.80	**93.55**	95.10	94.10
	随机缺失（40%）	**93.40**	96.00	95.20	**92.50**	95.80	95.10

S.2 附加数据分析结果

S.2.1 HEART-FID研究中变化随机缺失模型的敏感性分析

表 S.2.1：HEART-FID试验：在第二终点采用不同缺失模型的情况下，对死亡时间和第12个月6分钟步行距离的胜率估计进行的其他敏感性分析。

缺失模型	方法	估计值	95%置信区间	置信区间宽度
逻辑回归	协变量校正的 $\mathcal{S}$-评分（IF-Wald）	1.02	(0.94, 1.11)	0.18
	协变量校正的 $\mathcal{S}$-评分（BT-Wald）	1.02	(0.93, 1.12)	0.18
	协变量校正的 $\mathcal{S}$-评分（BT-QT）	1.02	(0.94, 1.12)	0.18
逻辑回归（交互）	协变量调整的 $\mathcal{S}$-评分（IF-Wald）	1.02	(0.94, 1.11)	0.18
	协变量调整的 $\mathcal{S}$-评分（BT-Wald）	1.02	(0.94, 1.11)	0.18
	协变量校正的 $\mathcal{S}$评分（BT-QT）	1.02	(0.94, 1.11)	0.17
随机森林	协变量校正的 $\mathcal{S}$评分（IF-Wald）	1.02	(0.93, 1.11)	0.18
	协变量校正的 $\mathcal{S}$评分（BT-Wald）	1.02	(0.93, 1.10)	0.18
	协变量调整的 $\mathcal{S}$-评分（BT-QT）	1.02	(0.93, 1.11)	0.18
集成学习	协变量调整的 $\mathcal{S}$-评分（IF-Wald）	1.02	(0.93, 1.11)	0.18
	协变量校正的 $\mathcal{S}$评分（BT-Wald）	1.02	(0.93, 1.11)	0.17
	协变量校正的 $\mathcal{S}$评分（BT-QT）	1.02	(0.94, 1.11)	0.17

CI：置信区间。逻辑回归（交互）：包含协变量间所有两两交互项的逻辑回归模型。集成学习：使用`SuperLearner` R软件包的集成模型，包含逻辑回归、逻辑回归（交互）和随机森林。

S.2.2 敏感性分析：ISCHEMIA研究中不同的MAR模型

表S.2.2：ISCHEMIA试验：在第二终点的不同缺失机制模型下，针对死亡时间及第48个月SAQ7-AF评分的胜率估计的额外敏感性分析。

缺失模型	方法	估计值	95%置信区间	置信区间宽度
逻辑回归	协变量调整的 $\mathcal{S}$-评分（IF-Wald）	1.27	(1.05, 1.49)	0.44
	协变量调整的 $\mathcal{S}$-评分（BT-Wald）	1.27	(1.05, 1.49)	0.44
	协变量调整的 $\mathcal{S}$-评分（BT-QT）	1.27	(1.08, 1.50)	0.42
逻辑回归（交互）	协变量调整的 $\mathcal{S}$-评分（IF-Wald）	1.28	(1.05, 1.50)	0.45
	协变量调整的 $\mathcal{S}$-评分（BT-Wald）	1.28	(1.05, 1.40)	0.45
	协变量调整的 $\mathcal{S}$-评分（BT-QT）	1.28	(1.08, 1.53)	0.45
随机森林	协变量调整的 $\mathcal{S}$-评分（IF-Wald）	1.27	(1.05, 1.49)	0.44
	协变量调整的 $\mathcal{S}$-评分（BT-Wald）	1.27	(1.06, 1.48)	0.42
	协变量调整的 $\mathcal{S}$-评分（BT-QT）	1.27	(1.09, 1.51)	0.43
集成学习	经协变量调整的 $\mathcal{S}$-评分（IF-Wald）	1.27	(1.05, 1.49)	0.44
	经协变量调整的 $\mathcal{S}$-评分（BT-Wald）	1.27	(1.06, 1.48)	0.43
	经协变量调整的 $\mathcal{S}$-评分（BT-QT）	1.27	(1.08, 1.50)	0.42

CI：置信区间。逻辑回归（交互）：包含协变量间所有两两交互项的逻辑回归模型。集成学习：由`SuperLearner` R软件包构建、包含逻辑回归、逻辑回归（交互）和随机森林的集成模型。