《生物统计学》000，1–??
000 0000
DOI：000

终末事件存在时复发事件治疗效应的模型辅助因果推断

Yiyuan Huang[1], **Ling Zhou**[2],**Min Zhang**[3],和 **Peter X.K. Song**[1,*]

[1]密歇根大学公共卫生学院生物统计学系，美国密歇根州安娜堡，MI 48109[2]西南财经大学统计研究中心，中国四川成都[3]清华大学万科公共卫生学院，中国北京 *电子邮箱：pxsong@umich.edu

arXiv:2601.01245v1 [stat.AP] 3 Jan 2026

摘要：本文旨在评估终末期心力衰竭管理中接受机械循环支持（MCS）设备治疗的患者所获益处。此类研究中，在存在终末事件的情况下，针对复发事件风险治疗效应的假设检验面临挑战。基于累积频数的现有方法会不合理地对生存时间较长的患者不利，因为他们往往会经历更多复发事件。存活期间（WA）检验为解决这一生存时长偏倚问题提供了方案，并且在复发事件率随时间保持恒定时表现良好。然而，若违反这一恒定发生率假设，WA检验可能导致I类错误膨胀，并对治疗效应估计不准确。为弥补这一方法学空白，我们在针对复发事件和终末事件的可分离治疗效应因果推断框架下，提出比例率边际结构模型辅助检验（PR-MSMaT）。通过模拟研究，我们证明，在复发事件率随时间变化时，PR-MSMaT能够适当控制I类错误，同时获得与WA检验相当的检验效能。我们使用PR-MSMaT比较了机械循环支持设备类型与胃肠道出血术后风险之间的关系，研究对象为参加机械循环辅助支持跨机构登记项目的患者。关键词：竞争风险；假设检验；边际结构模型；心脏移植。

本文已提交，等待在《生物统计学》上发表。

1. 引言

心脏移植仍是终末期心力衰竭患者的金标准，但由于供体器官严重短缺，其应用受到限制；美国每年仅进行约3,000例成人心脏移植。因此，长期机械循环支持已成为当前临床实践中心力衰竭管理的一种选择。自2006年以来，机械循环支持跨机构登记处（INTERMACS）每年报告的心室辅助装置植入手术超过2,500例。这些数据最初推动了我们的方法学开发，并在本文中进行了分析。实际上，接受连续流左心室辅助装置（LVAD）治疗的患者占植入病例的大多数，而双心室辅助装置（BiVAD）则较少使用，仅用于右心室功能严重障碍的患者。尽管这些设备显著改善了生存状况和生活质量，但也带来了沉重的不良事件负担，包括出血、感染、卒中和器械故障。其中，胃肠道出血（GIB）是最常见且临床影响显著的并发症之一，往往需要患者反复住院、接受输血和内镜操作。已知复发性胃肠道出血与死亡率升高和生活质量下降相关（Hammer 等，2024）。一个核心临床问题是：对于临床特征相当且预期生存时间相同的患者，选择植入双心室辅助装置还是左心室辅助装置，是否会导致复发性胃肠道出血事件风险不同?

尽管这一问题在临床上十分重要，但由于两个主要原因，目前仍未得到解决。首先，单中心研究中器械使用者的样本量通常不足以支持检验效能足够的统计比较。其次，随机临床试验在实施上困难重重，伦理方面也较为复杂；由于不确定不同器械的比较风险，患者和临床医生可能不愿意随机分配器械类型。因此，迄今为止，现有文献中的已发表证据尚不足以确定这两种器械的获益。一些研究表明，与延迟进行LVAD至BiVAD转换相比，早期计划性使用BiVAD支持可改善部分临床结局（Fitzpatrick et al., 2009）；

而另一些研究则报告，BiVAD接受者的出血和感染发生率较高（Cleveland et al., 2011）。INTERMACS等大规模登记数据或可成为有望得出更具说服力结论的数据来源。值得注意的是，INTERMACS汇总了全国各地医院的患者级数据，初步发现BiVAD患者的平均生存率较低（Kirklin et al., 2015）。需要指出的是，既往分析无法区分出血差异究竟反映了真实的器械效应，还是由患者生存差异所致；后者是主要的混杂因素。两个器械组之间差异性的生存分布，使分析和结果解释变得复杂。这促使我们开发严谨且易于解释的统计方法，从大型电子健康记录（EHR）数据库中获取有效的数据证据，以应对上述挑战。

我们的研究主要关注复发性胃肠道出血的风险。它是术后常见的并发症，会显著增加发病率、成本和资源利用。由于术后分析随访期间可能发生多次胃肠道出血，因此复发事件分析是研究这一结局的首选方法（Andersen and Gill，1982；Pepe and Cai，1993；Lawless and Nadeau，1995）。使用这一分析方法的一项主要挑战是存在死亡等终末事件，这些事件会阻止后续复发，并形成竞争风险情形。若简单地将死亡视为一种删失，则所得推断以存活的受试者为条件，这可能会使不同生存状况患者所使用的设备之间的比较产生偏倚，也缺乏因果可解释性。

为解决复发事件分析中的生存偏倚，已有多种统计方法被提出。Ghosh–Lin估计量（Ghosh和Lin，2000）在估计复发事件边际平均事件数时纳入死亡，但由于存活时间更长的组会累积更多事件，该方法可能会对这些组不利。为缓解这一问题，人们提出了存活期间（WA）估量目标，将事件平均数除以限制平均生存时间进行标准化，从而衡量

每单位存活时间内的事件发生率（Mao，2023年；Schmidli等人，2021年；Wei等人，2023年）。尽管这一WA检验取得了重要进展，但它隐含地假设复发事件发生率随时间保持恒定。实际上，事件强度往往会随时间变化，例如术后早期上升，随后逐渐减弱；当基线发生率随时间变化，或器械组之间的生存分布存在显著差异时，WA检验容易导致I类错误膨胀。这些局限表明，需要开发能够同时适应随时间变化的复发事件发生率以及治疗组间差异性生存分布的新方法。

上述问题与我们的研究动机实例高度相关。首先，我们注意到，心室辅助装置的类型可能通过两条不同路径影响胃肠道出血复发：一是直接影响复发性出血过程（例如通过剪切应力和抗凝要求），二是通过生存产生间接影响，因为存活时间较长的患者有更长的风险暴露时间。厘清这两条路径对于准确推断不同装置对应的出血风险至关重要。基于这一临床认识，我们采用由Stensrud等人（2022年）提出、并由Janvin等人（2023年）扩展至复发事件–终末事件情境的可分离效应框架。该框架将治疗（即装置）分解为两个假设性组成部分，分别作用于复发事件和终末事件路径；详见图2，第3节中有详细说明。简而言之，这种分解得出一个易于解释的估计目标，用于评估治疗对复发事件的直接效应，且不局限于存活者，同时还能适应随时间变化的基线发生率。本文的贡献在于提出一种新的统计推断方法，以丰富现有研究；现有研究主要关注可分离效应框架下的估计，而对正式假设检验的开发有限。

我们所发展的统计推断专注于一种模型辅助得分型假设检验，称为 PR-MSMaT，旨在检验终末事件存在时对复发事件的直接治疗效应。我们的方法基于可分离效应框架，生成一个具有清晰因果机制解释的单一时间积分检验。如论文所示，

PR-MSMaT 在不同基线事件率和差异性生存情况下均能良好地控制 I类错误，从而克服基于 WA 的检验的局限性。

本文结构安排如下。第2节介绍WA范式下的一种假设检验程序，并通过模拟实验展示基于WA的检验的局限性。第3节介绍可分离效应框架（SEF）（Stensrud等，2022年；Janvin等，2023年），该框架用于构建假设检验程序PR-MSMaT。第4节提出比例率边际结构模型辅助检验（PR-MSMaT），并讨论其解析性质。第5节重点介绍模拟实验，比较所提出的PR-MSMaT与基于WA的检验在各种情景下的I类错误和统计功效。第6节展示所提出方法在INTERMACS数据集中的应用，用于检验心室辅助装置之间胃肠道出血负担的差异，最后在第7节给出总结性评论。

2. 基于存活期间（WA）的检验：重新审视

数十年来，在存在终末事件的情况下分析复发事件的方法已有长足发展。Nelson–Aalen（NA）估计量是最早的方法之一，该方法将死亡视为复发过程的删失事件。这一假设实际上允许个体在死亡后继续累积事件，因此在死亡率不可忽略的情形下，会系统性地高估复发事件负担。为解决这一问题，Ghosh和Lin（2000）（GL）考虑到存活时间D之后不会再发生事件，提出了一种基于平均频数函数的非参数框架，$\mu(t) = \mathbb{E}\{N^*(t)\}$，其中$N^*(t)$表示截至时间$t$的复发事件累积数。Ghosh和Lin（2000）还提出了相应的假设检验程序。然而，GL检验未能恰当考虑接受器械植入患者的实际情况：延长生存时间的器械类型可能仅仅因为患者存活更久而增加复发事件计数，而不一定是因为器械提高了发病率。

为弥补这一局限，Mao（2023年）考虑了一种基于存活期间（WA）估计量的方法，该估计量以生存时间为基准，对复发事件累积数取平均。具体而言，在预先指定的时间 τ，存活期间损失率（WALR）定义为

$$\ell(\tau) = \frac{\mathbb{E}\{N^*(\tau)\}}{\mathbb{E}(D \wedge \tau)}.$$

该量表示在区间 $[0, \tau]$ 内每单位生存时间的复发事件期望数，因此可以校正NA方法和GL方法中固有的生存诱导偏倚。Mao（2023年）进一步推导了基于WALR的治疗比较推断，其中还考虑了一个修正后的估计目标 $L(t) = \ell(t)t$，以避免 $\ell(t)$ 在t较小时表现不稳定；该估计目标称为生存补全累积损失函数。

值得指出的是，只有当真实的基线复发事件率随时间保持恒定时，基于WA估计目标的检验才能对复发事件进行公平比较。可以说，这一条件在许多实际研究中可能并不成立。例如，在我们的示例中，很难认为患者的胃肠道出血风险轨迹会随时间保持恒定。在这种情况下，现有的基于WA的检验可能失效，并导致I类错误膨胀。

为说明这一潜在陷阱，我们开展了一项模拟实验，模拟复发事件率随时间变化的随机试验。为简化分析，我们假设复发事件相互独立，且复发事件与终末事件相互独立。基线复发事件率设为一个随时间变化的阶梯函数：起始为每年2次事件，此后在最长5年的随访期间，每1.2年增长至原来的 exp(0.5) 倍。通过两种复发事件率比（RR $= 1.00,\ 0.75$）和三种终末事件风险比（HR $= 1.0,\ 0.5,\ 0.2$）设定治疗效应，得到六种情景。每种情景均以$n = 500$ 名受试者进行模拟，并完成100轮蒙特卡洛重复模拟；使用现有的R软件包 *WA* （Mao，2023年）估计治疗组和对照组的WA函数（生存校正的累积损失函数）。图1展示了100轮重复模拟所得WA函数的平均值。从中我们发现：

(i) 在治疗对复发事件没有影响（RR $= 1$），但会影响生存（HR $\neq 1$）的情景中，例如RR $= 1.0$ 且HR $= 0.5$的情景，两个治疗组的WA估计值似乎存在显著差异。因此，基于此类WA估计值进行的假设检验可能无法良好地控制I类错误。

(ii) 在治疗确实可能同时影响复发事件和终末事件的情形下，例如RR $= 0.75$ 和HR $= 0.2$的场景，估计的WA曲线在第4年左右交叉（即并非真正可分离），这意味着基于WA的检验在第4年左右检验治疗效应时可能会损失统计功效。

这类数值证据表明，当基线事件率随时间变化时，如果治疗同时通过复发过程和生存发挥作用，WA方法在假设检验中可能会导致I类错误率膨胀并降低统计功效。因此，需要一种新的推断方法，能够明确考虑治疗、复发事件和终末事件之间的关联路径。

[图1见此处。]

3. 可分离效应框架（SEF）

为了更好地评估治疗对复发事件的影响，可以考虑两条路径：如图2所示，一条是直接效应，即治疗不通过生存而直接影响复发过程，例如我们研究示例中的胃肠道出血；另一条是间接效应，即治疗通过生存影响复发过程，因为存活时间较长的人往往会经历更多复发事件。主要目标是在评估中区分这两条路径并单独考察直接效应，同时排除间接的生存介导效应（Stensrud等，2022年）。

[图2见此处。]

可分离效应框架（SEF）最初由 Stensrud 等人（2022）针对存在竞争事件的二元结局提出，随后 Janvin 等人（2023）将其扩展至复发

事件情形。此后还提出了其他几种扩展方法（Stensrud 等人，2021年、2023年）。采用这一框架，我们旨在为 SEF 范式中的直接效应开发一种假设检验程序。本文主要考虑随机试验情形，重点关注治疗、复发过程与终末事件之间的关系，而忽略其他所有混杂因素。如果治疗并非随机分配，则可采用匹配或 IPTW 加权等统计方法平衡治疗组，以应用这一 SEF 方法。

令 $A \in \{0,1\}$ 表示正在研究的主要治疗。考虑将随访时间窗口离散划分为 $0 = t_0 < t_1 < \cdots < t_K = \tau$，其中 τ 为最大随访时间。对于区间单元 $(t_{k-1}, t_k]$，令 D_k 表示该时间区间内的死亡指标，Y_k 表示截至 t_k 的复发事件累积数。不失一般性，我们定义假设的有向无环图（DAG）中在时间区间 $(t_{k-1}, t_k]$ 内的拓扑顺序，如图3a所示，其中 D_k 先于 Y_k 发生。为方便起见，对于任意过程 X_k，我们用 $\Delta X_k := X_k - X_{k-1}$ 表示其变化，用 $\bar{X}_k = \{X_j\}_{j=0}^{k}$ 表示其历史。

在 SEF 范式中，治疗 A 被认为可分解为两个不同的治疗成分：A_Y，它只影响复发事件的发生；以及 A_D，它只影响终末事件。因此，接受 $A_Y = A_D = a$ 可能会产生与接受 $A = a$、$a \in \{0,1\}$ 相同的结果。图3b展示了这两个成分在 SEF 框架中的组织方式。尽管在实践中，我们并不总能确定由两个治疗成分构成的确切分解，但从概念上讲，我们可以借助这一假设框架，对其机制路径进行因果推断。

[[图3见此处。]]

形式化地，将反事实复发结局定义为 $Y_{a_Y, a_D k}$，表示受试者接受 $A_Y = a_Y$ 和 $A_D = a_D$ 时，截至 t_k 发生的复发事件累计数。同样，我们定义

反事实死亡指标$D^{a_Y,a^D k}$。反事实均值函数定义为

$$\mu_k^{a_Y,a_D} = E\left[Y_k^{a_Y,a_D}\right].$$

因此，在固定a_D时，对比 $\mu^{a_Y=1,\ a^D k}$ 与 $\mu^{a_Y=0,\ a^D k}$ 可体现**可分离直接效应**；而在固定a_Y 时，对比 $\mu^{a_Y,\ a^D=1\ k}$ 与 $\mu^{a_Y,\ a^D=0\ k}$ 则可体现**可分离间接效应** (Janvin等，2023年)。如图2所示，这些定义从概念上将总体治疗效应划分为不同且易于解释的作用路径。

为识别可　分离效应，我们采用因果推断的以下标准假设：

(i) 一致性：若$A = a$，则$D_K = D_K^a$，$Y_K = Y_{aK}$；

(ii) 可交换性：$(Y_{aK}\ D_K^a) \perp A$；

（iii）正值性：$f_{D_{k+1},Y_k}(0,y_k) > 0$意味着$P(A = a|D_{k+1} = 0, Y_k = y_k) > 0$，对所有$a \in \{0,1\}, k \in \{0,\cdots,K\}$均成立。

此外，对于 SEF，以下条件成立：

（iv）可忽略组件条件：

$$Y_{k+1} \perp A_D | A_Y, \bar{D}_{k+1}, \bar{Y}_k;$$

$$D_{k+1} \perp A_Y | A_D, \bar{Y}_k, \bar{D}_k, \text{for all } k \in \{0,\cdots,K\}.$$

（iv）中的条件确保分解后的A_Y和A_D处于互不重叠的路径上　.

在识别条件下，反事实均值函数的 G 公式为

$$\mu_k^{a_Y,a_D} = \sum_{t=1}^{k} E\left[\frac{I(A = a_Y)}{\pi_A(a_Y)} \prod_{j=0}^{t} \frac{\pi_{D_j}^{a_D}}{\pi_{D_j}^{a_Y}} \Delta Y_t\right],\ k \in \{0,\cdots,K\},$$

where $\pi_A(\bullet) = P(A = \bullet)$以及 $\pi_{D_j}^{\bullet} = f_{\bullet j}(D j)$，其中$f_{\bullet j}(\circ) = P(D j = \circ | Y_{j-1}, D_{j-1}, A = \bullet)$.

对于每个时间区间，均值函数的变化估计为

$$\Delta\hat{\mu}_t^{a_Y,a_D} = \hat{S}_t^{a_Y} \Delta\hat{B}_t^Y, \tag{1}$$

其中，$\hat{S}^{a_Y}{}_t$是组a_Y的生存函数的离散 Kaplan–Meier估计量，并且

$$\Delta\hat{B}_t^Y = \frac{\sum_{i=1}^n \hat{W}_{i,t}^D(a_Y,a_D) I(A_i = a_Y) Z_{i,t} \Delta Y_{i,t}}{\sum_{j=1}^n I(A_j = a_Y) Z_{j,t}},$$

其中 $Z_{i,t}$ 为风险指示变量，并且

$$\hat{W}_{i,t}^{D}(a_Y, a_D) = \prod_{q=1}^{t} \left(\frac{1 - \Delta\hat{\Lambda}_{i,q}^{D|\mathcal{F}}(a_D)}{1 - \Delta\hat{\Lambda}_{i,q}^{D|\mathcal{F}}(a_Y)} \right)^{1-D_{i,q}} \left(\frac{\Delta\hat{\Lambda}_{i,q}^{D|\mathcal{F}}(a_D)}{\Delta\hat{\Lambda}_{i,q}^{D|\mathcal{F}}(a_Y)} \right)^{D_{i,q}},$$

其中 $\Delta\hat{\Lambda}^{D|\mathcal{F}}{}_{i,q}(a) = \hat{P}(D_{i,q} = 1 | \bar{D}_{i,q-1} = 0, \bar{Y}_{i,q-1}, A_i = a)$。累积估计量为 $\hat{\mu}^{a_Y,a_D k} = \sum_{t=1}^{k} \Delta\hat{\mu}^{a_Y,a^D t}$（Janvin等，2023年）。

加权过程 $\hat{W}_{Di,t}(a_Y, a_D)$ 旨在模拟一项假想随机试验，在该试验中，A_Y 和 A_D 可以分别实施。具体而言，它调整观测数据分布，使a_D 下的反事实生存过程与a_Y下的反事实生存过程具有可比性，并且仅在$a_D \neq a_Y$时产生效应。直观地说，即使治疗A 是随机分配的，这种通过权重进行的校准也用于消除组间不平衡，使我们能够恢复A_Y 和A_D 独立随机分配时复发事件的分布。对于在反事实a_D 下比在其实际分配下更可能死亡的个体，权重会降低其权重；对于死亡可能性较低的个体，则提高其权重，从而有效地重塑风险集，以反映反事实干预下复发事件的分布。通过这种方式，加权使各组的生存经历保持一致，因此估计的复发事件过程之间的差异可以完全归因于A_Y的直接效应，而不会受到A_D所诱导的差异性生存的偏倚影响。

上述对SEF的综述介绍了一个框架，在该框架中，治疗可被严谨地分解为两个分别仅作用于复发事件和终末事件的组成部分。因此，SEF中的直接效应表示仅归因于A_Y、同时固定A_D 时复发事件过程的对比。通过将这一关键关注路径与间接效应分离，我们可以检验关于复发事件直接治疗效应的假设，同时减轻WA方法在时变复发事件率下所观察到的偏差。

4. 比例率边际结构模型辅助检验（PR-MSMaT）

基于SEF，我们针对复发事件的直接治疗效应提出了一种假设检验。其逻辑与传统生存分析类似：Kaplan–Meier估计量启发了对数秩检验和将治疗作为唯一协变量的Cox比例风险模型；同样，Janvin等（2023）提出的可分离估计量启发了沿用基于得分的检验思路的比例率边际结构模型（PR-MSM）。

令 $\mu^{a_Y,a^D}(t)$ 表示在治疗成分设定为 (a_Y,a_D) 时，截至时间 t 的复发事件反事实边际平均事件数。PR-MSM 的形式为：

$$\mu^{a_Y,a_D}(t) = \mu^{a_Y=0,a_D}(t)\exp\left(\beta^{a_D}a_Y\right), \quad a_D \in \{0,1\}, \tag{2}$$

这表明，对于每个固定的 a_D，不同 a_Y 下的反事实均值函数在时间上相差一个常数乘法因子。因此，可通过求解以下估计方程来估计 β^{a_D} ：

$$\int_0^\tau \sum_{i=1}^{n} Z_i(t)A_i\left\{d\hat{\mu}^{a_Y=1,a_D}(t) - \exp(\beta^{a_D})\,d\hat{\mu}^{a_Y=0,a_D}(t)\right\} = 0, \tag{3}$$

其中，$\hat{\mu}^{a_Y,a^D}(t)$ 是反事实均值函数估计量，而 $Z_i(t)$ 是风险过程。关于式 (3) 的详细推导，请参见附录I。

在 $H_0:\beta^{a_D}=0$ （无直接治疗效应）下，式 (3) 左侧可简化为

$$U_n \;=\; \int_0^\tau \sum_{i=1}^{n} Z_i(t)A_i\left\{d\hat{\mu}^{a_Y=1,a_D}(t) - d\hat{\mu}^{a_Y=0,a_D}(t)\right\}. \tag{4}$$

令 $\widehat{\mathrm{Var}}(U_n)$ 表示 $\mathrm{Var}(U_n)$ 的一致估计量，该估计量可通过基于微分方程的累积风险估计转换得到（Ryalen 等，2018年）。我们的 PR-MSM 辅助检验（PR-MSMaT）统计量定义为标准化得分检验统计量：

$$T_n \;=\; \frac{U_n}{\widehat{\mathrm{Var}}(U_n)^{1/2}}.$$

定理 1：假设 *(i)* $\hat{W}\,Di,t\,(a_Y,a_D)$ 和 $\hat{S}^{a_Y}t$ 一致收敛于其总体参数；*(ii)* 风险过程 $Z_i(t)$ 与 *0* 和 ∞ 保持正距离；且 *(iii)* 标准

鞅林德伯格条件成立。那么，在 H_0： $\beta a^D = 0$ 下，当 $n \to \infty$ 时，检验统计量 $T_n d \longrightarrow N(0,1)$。

定理1表明，所提出的 PR-MSMaT 是用于检验复发事件不存在可分离直接治疗效应这一零假设的有效检验。定理1的证明见附录II。

PR-MSMaT 扩展了可分离效应估计量，提供了一种正式的假设检验，用于概括预先指定随访时间段内的治疗效应。与原有估计量不同，后者得出的是时间特异性对比，可能较难综合；所提出的检验则对直接治疗效应进行总体推断。该检验易于临床实践中的更广泛受众理解；尤其是，它在 SEF 范式下具有因果可解释性，这与实践中广泛使用的 log-rank 检验或 Cox得分检验类似。第5节的模拟研究表明，在复发事件率随时间变化的情形下（这是实践中经常遇到的现实情况，包括我们的研究示例），PR-MSMaT 能够维持适当的 I类错误控制，且检验效能高于存活期间检验。

5. 模拟实验

按照Janvin等（2023）给出的模拟设计，我们评估了在针对治疗A的双臂随机试验中，PR-MSMaT相较于基于WA的检验的有限样本性能，其中治疗组为$A = 1$，对照组为$A = 0$。在这种情况下，所有对比均按治疗组与对照组进行评估。在$K = 1000$个区间内，采用加性强度模型生成结局（$t_{k-1}, t_k]$、$k = 1, \ldots, 1000$。对于每个区间，我们考虑

$$P(\Delta Y_k = 1 \mid A_Y, D_k = 0) = \beta_{Y,0,k} + \beta_{Y,A} A_Y,$$

$$P(D_k = 1 \mid A_D, \bar{Y}_{k-1}) = \beta_{D,0} + \beta_{D,A} A_D + \beta_{Y,D} Y_{k-1},$$

每个区间内至多发生一次复发事件。我们固定 $\beta_{D,A} = -0.5/K$，将 $\beta_{Y,A} = 0$设为零假设情形以评估 I类错误情形，并将 $\beta_{Y,A} = -0.5/K$ 用于检验效能评估。每种配置均通过 $N = 1000$ 轮蒙特卡罗重复进行考察，样本量为 $n = 1000$。

三种基线复发事件率 $\beta_{Y,0,k}$ 依据以下时间模式进行设计：

(a) **恒定：** $\beta_{Y,\,0,\,k} = 2/K$ 对所有k均成立。

(b) **递减：** 从 $3/K$ 开始，每 200 个间隔减少 $0.5/K$ （降至 $1/K$）。(c) **递增：** 从 $1/K$ 开始，每 200 个间隔增加 $0.5/K$ （升至 $3/K$）。

我们考察三类假设：**双侧：** H_0： $\theta = 0$ 对比H_1： $\theta \neq 0$；**左侧：** H_0： $\theta \geqslant 0$对比H_1： $\theta < 0$；**右侧：** H_0： $\theta \leqslant 0$ 对比H_1： $\theta > 0$。参数 θ 表示治疗组与对照组复发事件的差异；对于我们的 PR-MSMaT，该参数对应于 β^{a_D}，而对于基于WA的检验，则对应于WALR比值。

模拟结果汇总于图4。当基线复发事件率恒定时（即情景(a)），我们的 PR-MSMaT 和 WA检验均能维持名义 I类错误率，且检验效能相近。在时间齐性率下，这种一致性证实了两种方法在该情景中的表现均良好。相反，当基线复发事件率随时间变化时，两者的表现便出现分歧。在发生率递减时（即情景(b)），WA在双侧检验和左侧检验中均出现I类错误膨胀。在发生率递增时（即情景(c)），WA在双侧检验和右侧检验中同样出现I类错误膨胀。令人惊讶的是，我们的 PR-MSMaT 在所有情景下都能很好地控制I类错误。在检验效能方面，我们的 PR-MSMaT 在情景(b)中的检验效能与 WA检验相当，在情景(c)中的检验效能甚至高于 WA。需要注意的是，在情景(b)和(c)中，WA均存在I类错误膨胀的缺陷。这表明 PR-MSMaT 能够可靠地应对非平稳基线复发事件过程，且表现明显优于WA检验。

方向性检验结果有助于解释 WA 检验的表现。WA 检验通过用限制平均生存时间对复发事件累积数进行标准化来汇总数据，并据此比较各组别的“存活期间发生率”。这一过程实际上将每位受试者观察到的事件总数视为其观察到的生存期间以大致恒定的单位时间发生率产生的结果

。当真实基线发生率随时间下降时，大多数事件会较早发生，因此较早死亡的受试者，其一生中处于高发生率时期的比例往往更大。在这种情况下，WA 会将较高的“存活期间平均事件发生率”归于这类生存时间较短的受试者，即使他们实际上并不比生存时间更长的受试者更容易发生事件。因此，WA 可能会系统性地将较早死亡视为对照组事件强度较高的证据；这会使估计的治疗效应（治疗组相对于对照组）偏低，并导致左侧 I 类错误率膨胀。同样，当基线发生率随时间上升时，事件会集中在较晚时期发生，因此生存时间较长的受试者仅仅因为在高发生率时期持续处于观察状态，就会累积更多事件。WA 将此视为某一组别的“存活期间发生率”升高，从而使估计的治疗效应偏高，并导致右侧 I 类错误率膨胀。

相比之下，我们的 PR-MSMaT 直接针对治疗对复发事件过程的可分离通路特异性效应，同时固定生存路径。由于它不会将高风险或低风险时段内仅由随访时间长短不均所导致的差异归因于治疗效应，因此 PR-MSMaT 避免了这些方向性偏倚。因此，在保持检验效能具有竞争力的同时，它在所有随时间变化的基线模式下均能维持名义 I 类错误率。

[图4见此处。]

6. 应用

现在，我们分析第1节中介绍的引例数据。在 SEF 中，我们关注的问题是：在生存路径固定的情况下，器械类型是否会对复发性胃肠道出血（GIB）产生显著的直接效应。按照这一设定，复发事件率的任何差异都可归因于器械本身，而非生存差异。我们的数据集包含在2006年4月1日至 $\geqslant$ 19 岁成年患者，他们接受了首次连续流左心室辅助装置植入，可伴有或不伴有后续右心室辅助装置植入（双心室辅助装置），截至

2017年12月31日。我们排除了搏动性装置、单独的右心室辅助装置、全人工心脏以及非首次左心室辅助装置植入。主要结局为复发性胃肠道出血事件，定义为上消化道或下消化道出血，或来源不明的隐匿性出血阳性事件。死亡被视为终末事件（或干扰事件）。为模拟随机试验情境，我们使用植入前的人口统计学、临床、实验室、超声心动图及血流动力学因素进行 1:1 倾向评分匹配。匹配后，基线特征均衡良好（所有单变量检验的 p 值均为 > 0.05 ）。匹配后数据集中基线因素的汇总表见补充材料。经过 1:1 匹配后，分析样本中每组（双心室辅助装置组和左心室辅助装置组）包含 $n = 1{,}051$ 名患者。

我们按器械类型（左心室辅助装置 vs. 双心室辅助装置）总结了匹配队列中的复发性胃肠道出血（胃肠道出血）事件和总生存期。图5展示了胃肠道出血负担及其随访时间的两幅描述性图表。从生存对比中可以清楚地看出，双心室辅助装置组的死亡率显著早于左心室辅助装置组。我们还使用 WA检验和 PR-MSMaT，在 5 个时间范围 $\tau \in \{1, 2, 5, 10, 15\}$（月）内评估了双心室辅助装置与左心室辅助装置之间的治疗对比。图6报告了标准化的 Z值和 p值，两种方法均采用对比双心室辅助装置 –左心室辅助装置，这意味着负的 Z值有利于双心室辅助装置。

值得注意的是，WA 和 PR-MSMaT 实际上针对不同的估量目标，因此它们对图5中的生存不平衡反应不同；图中双心室辅助装置组患者往往更早死亡。WA 将随生存时间完整计算的累积频数除以平均生存时间，这能减少生存者偏倚，但无法将其完全消除。如果事件发生率随时间变化，尤其是在植入后不久较高，那么双心室辅助装置组较早发生死亡意味着其观测时间中有更大比例落在高风险窗口内，这会抬高存活期间发生率，并产生正的 Z，表明双心室辅助装置组的胃肠道出血负担更重，即使该组直接发生事件的倾向并不更高。PR-MSMaT 使用可分离效应加权，使各组的生存路径（A_D）保持一致，并在固定死亡过程的情况下检验对复发事件的直接效应（A_Y

不变。在图6中，随访时间（τ）较长时 PR-MSMaTZ 值为负，这表明在固定生存路径后，双心室辅助装置的直接复发事件负担低于左心室辅助装置（可归因于直接效应的胃肠道出血事件更少）。因此，符号不一致（WA 为正、PR-MSMaT 为负）反映了间接生存路径对双心室辅助装置组高风险早期时段的过度加权。

[图5见此处。]

[图6见此处。]

参与胸外科医师协会（STS）INTERMACS 项目的医院使用标准化数据表收集该人群的纵向数据。本研究使用这些数据已获得 STS参与者用户文件工作组的批准。STS 不要求患者同意向 STS INTERMACS 提交数据，相关要求由各医院的当地政策决定。本文所表达的观点并不代表 STS 的观点。

7. 讨论

我们开发了一种推断方法，称为 PR-MSMaT。这是一种在可分离效应框架（SEF）下构建的得分型检验，用于评估复发事件的直接治疗效应，同时明确固定生存路径。从概念上讲，PR-MSMaT 通过一种实用的得分检验来评估 Stensrud等（2022）和 Janvin等（2023）的可分离效应估计目标，该检验类似于对数秩检验或 Cox得分检验。它针对的是路径特异性对比，而非由复发过程和终末过程共同驱动的机制混合效应；通过汇总整个随访期间的信息，得出一个具有明确定义的因果估计目标、且易于解释的单一汇总统计量。在涵盖多种数据生成机制的大量模拟中，研究表明 PR-MSMaT 能够维持名义 I类错误率，并且对时变基线风险具有稳健性。值得注意的是，在现有的存活期间（WA）检验出现以下情况时，它仍能保持有效性：

系统性 I类错误膨胀。事实上，SEF 能有效地将复发事件动态与差异性生存解耦，并减少因生存时长不同而引发的虚假关联。在我们的实证研究中，我们发现，固定生存路径能够使我们分离出复发性胃肠道出血（GIB）负担中仅由器械选择本身造成的部分，而不受死亡率差异的影响。这一发现具有临床意义；在考虑患者预期生存期的前提下，它有助于临床医生选择器械，以尽量降低出血风险。此外，由于 SEF 在执行良好的 1:1匹配方案下具有因果解释性，我们对两种器械的比较揭示了颇具启发性的结果，有助于改善患者植入术后的生活质量。从技术上讲，我们的方法 PR-MSMaT 允许时间范围 τ 和基线率函数发生变化，这为分析在其他环境中收集的临床数据提供了重要的灵活性。这些特点使 PR-MSMaT 成为一种实用且依据充分的分析工具，可用于评估植入器械之外的治疗，无论是在随机试验还是基于登记系统的观察性研究中。

PR-MSMaT 还有几个值得进一步发展的方向。将该方法扩展至同时处理多个复发性终点（例如出血、感染和器械故障）以及中心层面异质性，将提升其在多中心登记系统和复合安全性监测中的实用性。针对可忽略组件条件可能遭到违反的情况，制定正式的敏感性分析方法，将增强该方法在观察性数据中的应用，因为在此类数据中，路径隔离可能并不完善。最后，纳入协变量自适应加权，并采用针对路径特异性备择方案校准的样本量公式，或可提高试验效率和检验效能；对于实践模式不断演变且与事件动态相互作用的登记系统嵌入式研究或平台研究，这一点尤为重要。

致谢

作者感谢胸外科医师协会提供数据。本研究部分由 NIH R01ES033656资助（Song）。

参考文献

Andersen, P. K., Borgan, O., Gill, R. D. 和 Keiding, N.（2012）。基于计数过程的统计模型。Springer Science & Business Media。Andersen, P. K. 和 Gill, R. D.（1982）。计数过程的Cox回归模型：一项大样本研究。*The Annals of Statistics*第1100–1120页。

Cleveland, J. C., Naftel, D. C., Reece, T. B. 等（2011）。双心室辅助装置植入后的生存：对Intermacs注册登记系统的分析。*Journal of HeartandLungTransplantation* **30,** 862–869。

Fitzpatrick, J. R., Frederick, J. R., Hiesinger, W. 等（2009）。与将左心室辅助装置延迟转换为双心室辅助装置相比，早期计划实施双心室机械循环支持可改善结局。*Journal ofThoracic and Cardiovascular Surgery* **137,** 971–977。

Fleming, T. R. 和 Harrington, D. P.（2013）。计数过程与生存分析。John Wiley & Sons。

Ghosh, D. 和 Lin, D.（2000）。复发事件和死亡的非参数分析。《生物统计学》**56,** 554–562。

Hall, P. 和 Heyde, C. C. (2014). 鞅极限定理及其应用. Academic Press. Hammer, Y., Xie, J., Yang, G., Bitar, A., Haft, J. W., Cascino, T. M., Likosky, D. S., Pagani, F. D., Zhang, M., 和 Aaronson, K. D. (2024). 植入 HeartMate 3 左心室辅助装置后的胃肠道出血：密歇根出血风险模型。*The Journal of HeartandLungTransplantation* **43,** 604–614. Janvin, M., Young, J. G., Ryalen, P. C., 和 Stensrud, M. J. (2023年). 针对复发事件和竞争事件的因果推断。*Lifetime Data Analysis* 第 1–60 页。

Kirklin, J. K., Naftel, D. C., Pagani, F. D., 等（2015）。第七次Intermacs年度报告：患者人数已达15,000，且仍在增加。《心脏与肺移植杂志》**34,** 1495–1504。

Lawless, J. F. 和 Nadeau, C.（1995）。用于分析复发事件的一些简单稳健方法。计量技术学**37,**158–168。Mao, L.（2023）。复发事件一般存活期间估计目标的非参数推断。《生物统计学》**79,** 1749–1760。Pepe, M. S. 和 Cai, J.（1993）。复发失效时间和时变协变量的若干图形展示及边际回归分析。美国统计学会会刊**88,**811–820。Rebolledo, R.（1980）。局部鞅的中心极限定理。概率论与相关领域杂志**51,**269–286。Ryalen, P. C., Stensrud, M. J. 和 Røysland, K.（2018）。变换累积风险估计。生物统计学**105,** 905–916。Schmidli, H., Roger, J. H. 和 Akacha, M.（2021）。存在终末事件时复发事件终点的估量目标。生物制药研究统计学第1–11页。Stensrud, M. J., Hern´an, M. A., Tchetgen Tchetgen, E. J., Robins, J. M., Didelez, V. 和 Young, J. G.（2021）。竞争事件情境下可分离效应的广义理论。*Lifetime Data Analysis***27,**588–631。Stensrud, M. J., Robins, J. M., Sarvet, A., Tchetgen Tchetgen, E. J. 和 Young, J. G.（2023）。条件可分离效应。美国统计学会会刊**118,** 2671–2683。Stensrud, M. J., Young, J. G., Didelez, V., Robins, J. M. 和 Hern´an, M. A.（2022）。存在竞争事件时用于因果推断的可分离效应。美国统计学会会刊**117,** 175–183。Wei, J., Muetze, T., Jahn-Eimermacher, A. 和 Roger, J.（2023）。复发事件的两种存活期间估计目标及其潜在估计量的性质。生物制药研究统计学**15,** 257–267。

补充信息

第6节中引用的汇总表可在《生物统计学》网站的 Wiley OnlineLibrary 上查阅。该表显示了数据中基线因素令人满意的 1:1 匹配结果，从而证明在实证研究中应用所提出的因果推断方法是合理的。

附录I. PR-MSM 中 β^{a_D} 的估计方程推导

将第3节中的方程（1）改写为连续时间形式，并用计数过程记法表示复发事件过程$N(t)$（Fleming和Harrington，2013），可得可分增量估计量：

$$d\hat{\mu}^{a_Y,a_D}(t) = \frac{\sum_{i=1}^{n} Z_i(t)\, I(A_i = a_Y)\, \hat{S}^{a_Y}(t)\, \hat{W}_i^D(a_Y, a_D; t)\, dN_i(t)}{\sum_{j=1}^{n} Z_j(t)\, I(A_j = a_Y)}.$$

令$a_Y = 1$，可得

$$d\hat{\mu}^{a_Y=1,a_D}(t) = \frac{\sum_{i=1}^{n} Z_i(t) A_i\, \hat{S}^{a_Y=1}(t)\, \hat{W}_i^D(1, a_D; t)\, dN_i(t)}{D_1(t)}, \tag{A.1}$$

其中

$$D_1(t) := \sum_{j=1}^{n} Z_j(t) A_j.$$

式（2）中的 PR-MSM 假设

$$\mu^{a_Y=1,a_D}(t) = \mu^{a_Y=0,a_D}(t) \exp(\beta^{a_D}) \quad \Longleftrightarrow \quad d\mu^{a_Y=1,a_D}(t) = \exp(\beta^{a_D})\, d\mu^{a_Y=0,a_D}(t).$$

为将观测数据与反事实均值模型 $\mu^{a_Y,\,a^D}(t)$ 联系起来，我们用其一致估计量 $d\hat{}\mu^{a_Y,\,a^D}(t)$ 替换每个$d\mu^{a_Y,\,a^D}(t)$（Janvin等，2023年）：

$$d\hat{\mu}^{a_Y=1,a_D}(t) = \exp(\beta^{a_D})\, d\hat{\mu}^{a_Y=0,a_D}(t). \tag{A.2}$$

将（A.2）代入（A.1）左侧，同时将右侧保留为估计量，可得矩条件：

$$\exp(\beta^{a_D})\, d\hat{\mu}^{a_Y=0,a_D}(t) = \frac{\sum_{i=1}^{n} Z_i(t) A_i\, \hat{S}^{a_Y=1}(t)\, \hat{W}_i^D(1, a_D; t)\, dN_i(t)}{D_1(t)}. \tag{A.3}$$

将式 (A.3) 两边乘以 $D_1(t)$ 并重新整理，可得

$$\sum_{i=1}^{n} Z_i(t) A_i\, \hat{S}^{a_Y=1}(t)\, \hat{W}_i^D(1, a_D; t)\, dN_i(t) \;-\; \exp(\beta^{a_D})\, D_1(t)\, d\hat{\mu}^{a_Y=0,a_D}(t) \;=\; 0. \tag{A.4}$$

现在，将未知的 $d\mu^{a_Y=0,a^D}(t)$ 替换为其估计量，该估计量由第3节中含 $a_Y=0$ 的同一方程 (1) 得出，可得

$$d\hat{\mu}^{a_Y=0,a_D}(t)=\frac{\sum_{i=1}^{n}Z_i(t)(1-A_i)\hat{S}^{a_Y=0}(t)\hat{W}_i^D(0,a_D;t)\,dN_i(t)}{D_0(t)}, \tag{A.5}$$

其中

$$D_0(t):=\sum_{j=1}^{n}Z_j(t)(1-A_j).$$

将 (A.5) 代入 (A.4)，得到对两个治疗组求和的估计方程:

$$\begin{aligned}&\sum_{i=1}^{n}Z_i(t)A_i\hat{S}^{a_Y=1}(t)\hat{W}_i^D(1,a_D;t)\,dN_i(t)\\&-\exp(\beta^{a_D})\frac{D_1(t)}{D_0(t)}\sum_{i=1}^{n}Z_i(t)(1-A_i)\hat{S}^{a_Y=0}(t)\hat{W}_i^D(0,a_D;t)\,dN_i(t)\;=\;0.\end{aligned}$$

对 $[0,\tau]$ 积分后，可得到 β_{a^D} 的样本估计方程:

$$\begin{aligned}&\int_0^{\tau}\Bigg[\sum_{i=1}^{n}Z_i(t)A_i\hat{S}^{a_Y=1}(t)\hat{W}_i^D(1,a_D;t)\,dN_i(t)\\&-\exp(\beta^{a_D})\frac{D_1(t)}{D_0(t)}\sum_{i=1}^{n}Z_i(t)(1-A_i)\hat{S}^{a_Y=0}(t)\hat{W}_i^D(0,a_D;t)\,dN_i(t)\Bigg]=0.\end{aligned} \tag{A.6}$$

最后，注意到第一个求和项对应于 $d\hat{}\mu^{a_Y=1,a^D}(t)$，而第二个求和项对应于 $d\hat{}\mu^{a_Y=0,a^D}(t)$，

我们可以将 (A.6) 简洁地写为

$$\int_0^{\tau}\sum_{j=1}^{n}Z_j(t)A_j\Big\{d\hat{\mu}^{a_Y=1,a_D}(t)-\exp(\beta^{a_D})d\hat{\mu}^{a_Y=0,a_D}(t)\Big\}=0.$$

附录II. 定理1的证明（PR-MSMaT的渐近正态性）

令$N_i(t)$表示受试者i 在 $[0,\tau]$上的复发事件计数过程，$Z_i(t)$表示风险指示变量，$A_i\in\{0,1\}$ 表示随机分配的治疗。记$M_i(t)=N_i(t)-\Lambda_i(t)$，其中 $\Lambda_i(t)$是$N_i(t)$相对于自然滤过的补偿子。因此，$\{M_i(t)\}$ 是一个平方可积局部鞅，其可预测变差为 $\langle M_i\rangle(t)$（Andersen等人，2012）。

对于$g\in\{0,1\}$，定义$D_g(t)=\sum_{j=1}^{n}Z_j(t)I(A_j=g)$，并假设$D_g(t)/n\to p_g(t)$一致成立，且 $\inf_{t\leqslant\tau}p_g(t)>0$。

设$\hat{S}^{a_Y}(t)$是$S^{a_Y}(t)$的一致相合估计量，且$\hat{W}_{Di}(a_Y,a_D;t)$是相应权重的一致相合估计量。定义可分均值增量估计量

$$d\hat{\mu}^{a_Y,a_D}(t)=\frac{\sum_{i=1}^{n}Z_i(t)\,I(A_i=a_Y)\,\hat{S}^{a_Y}(t)\,\hat{W}_i^{D}(a_Y,a_D;t)\,dN_i(t)}{D_{a_Y}(t)},$$

这对应于第3节中的（1）。

根据第4节中的（4），在H_0下的PR-MSMaT得分为： $\beta^{a_D}=0$是

$$U_n=\int_0^{\tau}\sum_{j=1}^{n}Z_j(t)A_j\Big\{d\hat{\mu}^{1,a_D}(t)-d\hat{\mu}^{0,a_D}(t)\Big\}. \tag{A.7}$$

固定$a_Y\in\{0,1\}$。记

$$d\hat{\mu}^{a_Y,a_D}(t)=\frac{1}{D_{a_Y}(t)}\sum_{i=1}^{n}I(A_i=a_Y)\,Z_i(t)\,\hat{S}^{a_Y}(t)\,\hat{W}_i^{D}(a_Y,a_D;t)\,dN_i(t),$$

将$dN_i(t)$替换为$dM_i(t)+d\Lambda_i(t)$，并在其期望值附近对分子和分母进行展开。对经验过程比值进行标准泰勒展开，结合$\hat{S}^{a_Y}$和$\hat{W}_i^{D}$的一致相合性，以及可预测有界过程的一致大数定律（Andersen等人，2012），可得

$$d\hat{\mu}^{a_Y,a_D}(t)=\mu'^{a_Y,a_D}(t)\,dt+\frac{1}{n}\sum_{i=1}^{n}\phi_i^{a_Y,a_D}(t)\,dM_i(t)+r_n^{a_Y,a_D}(t), \tag{A.8}$$

其中 $\sup_{t\leqslant\tau}|r^{a_Y,a_Dn}(t)|=o_P(n^{-1/2})$，且

$$\phi_i^{a_Y,a_D}(t)=\frac{n\,I(A_i=a_Y)\,Z_i(t)\,S^{a_Y}(t)\,W_i^{D}(a_Y,a_D;t)}{D_{a_Y}(t)}.$$

将 (A.8) 代入 (A.7),

$$U_n=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\psi_i(t)\,dM_i(t)+o_p(n^{1/2}),$$

其中 $\psi_i(t)=\sum_{j=1}^{n}Zj(t)Aj\{\phi^{1,a_D}i(t)-\phi^{0,a_D}i(t)\}$。

定义

$$\xi_{n,i}:=\int_0^{\tau}\frac{\psi_i(t)}{n^{3/2}}\,dM_i(t),\quad\text{so that}\quad\sum_{i=1}^{n}\xi_{n,i}=n^{-1/2}U_n+o_p(1).$$

可预测变差满足

$$V_n=\sum_{i=1}^{n}\langle\xi_{n,i}\rangle=\sum_{i=1}^{n}\int_0^{\tau}\frac{\psi_i^2(t)}{n^3}\,d\langle M_i\rangle(t).$$

假设：（i）受试者独立同分布；（ii）有界可预测 $\phi^{a_Y,a^D i}(t)$；（iii）$\inf_t Dg(t)/n > 0$和 $\sup_t Dg(t)/n < \infty$；（iv）满足鞅林德伯格条件（Andersen等人，2012；Hall 和 Heyde，2014）。

由于增量按 $1/n^{3/2}$ 缩放，且权重有界，因此条件 (iv) 自动成立。因此，$V_{np} \to \sigma^2 \in (0,\infty)$。

根据 Rebolledo 鞅中心极限定理（Rebolledo，1980年），

$$\sum_{i=1}^{n} \xi_{n,i} \xrightarrow{d} N(0,\sigma^2), \quad \text{equivalently} \quad n^{-1/2} U_n \xrightarrow{d} N(0,\sigma^2).$$

一个相合的方差估计量为

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \int_0^{\tau} \frac{\{\hat{\psi}_i(t)\}^2}{n^3} dN_i(t),$$

因此，$\widehat{\mathrm{Var}}(U_n) = n\,\hat{\sigma}^2$ 满足

$$\frac{\widehat{\mathrm{Var}}(U_n)}{n} \xrightarrow{p} \sigma^2.$$

因此，PR-MSMaT 统计量

$$T_n = \frac{U_n}{\{\widehat{\mathrm{Var}}(U_n)\}^{1/2}} = \frac{n^{-1/2} U_n}{\hat{\sigma}} \xrightarrow{d} N(0,1),$$

在H_0 下：$\beta^{a_D} = 0$。

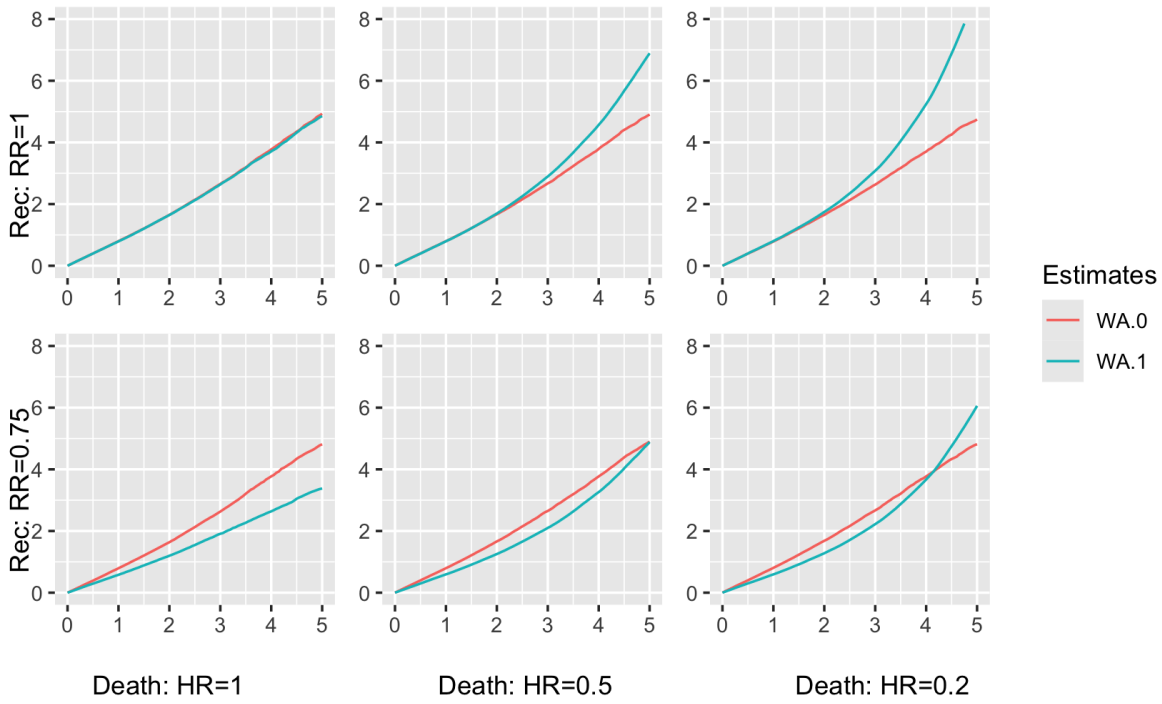


图1：模拟结果说明了存活期间检验的局限性。WA.0：对照组的生存完成累积损失函数；WA.1：治疗组的生存完成累积损失函数；RR：复发事件比率；HR：风险比。

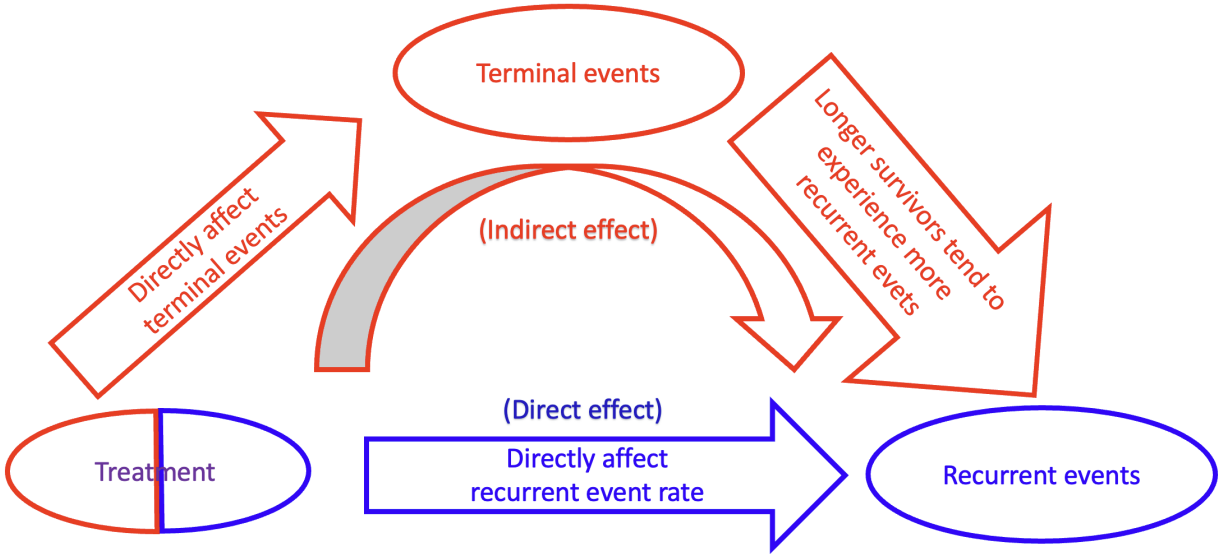
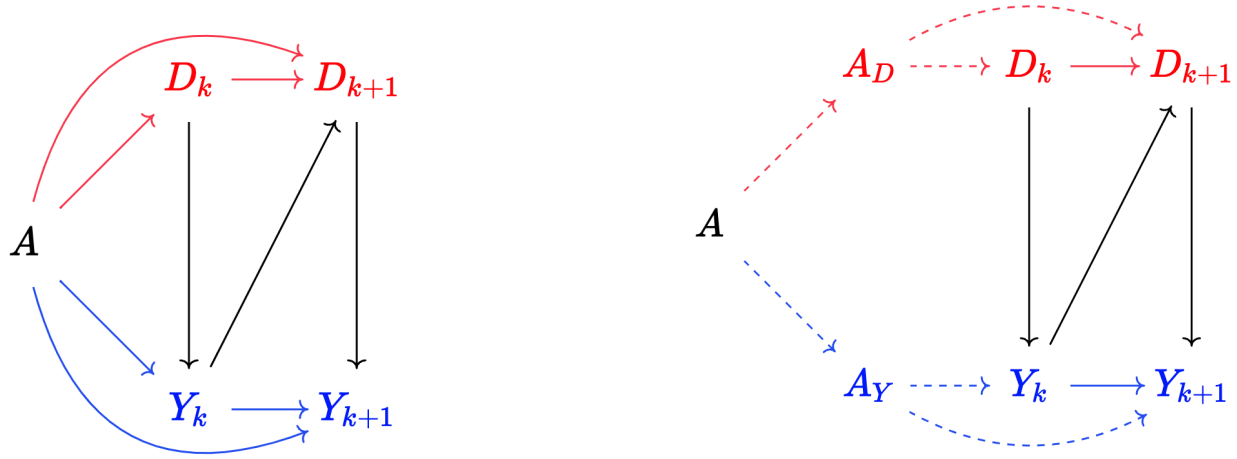


图2：治疗、复发事件与终末事件之间关系的直观解释。



(a) (b)

图3：有向无环图（DAG）展示了：(a) 治疗A、复发事件Y_k和终末事件D_k的因果路径；(b) 将治疗A 假设性地分解为A_Y （直接影响复发事件）和A_D （直接影响终末事件）。

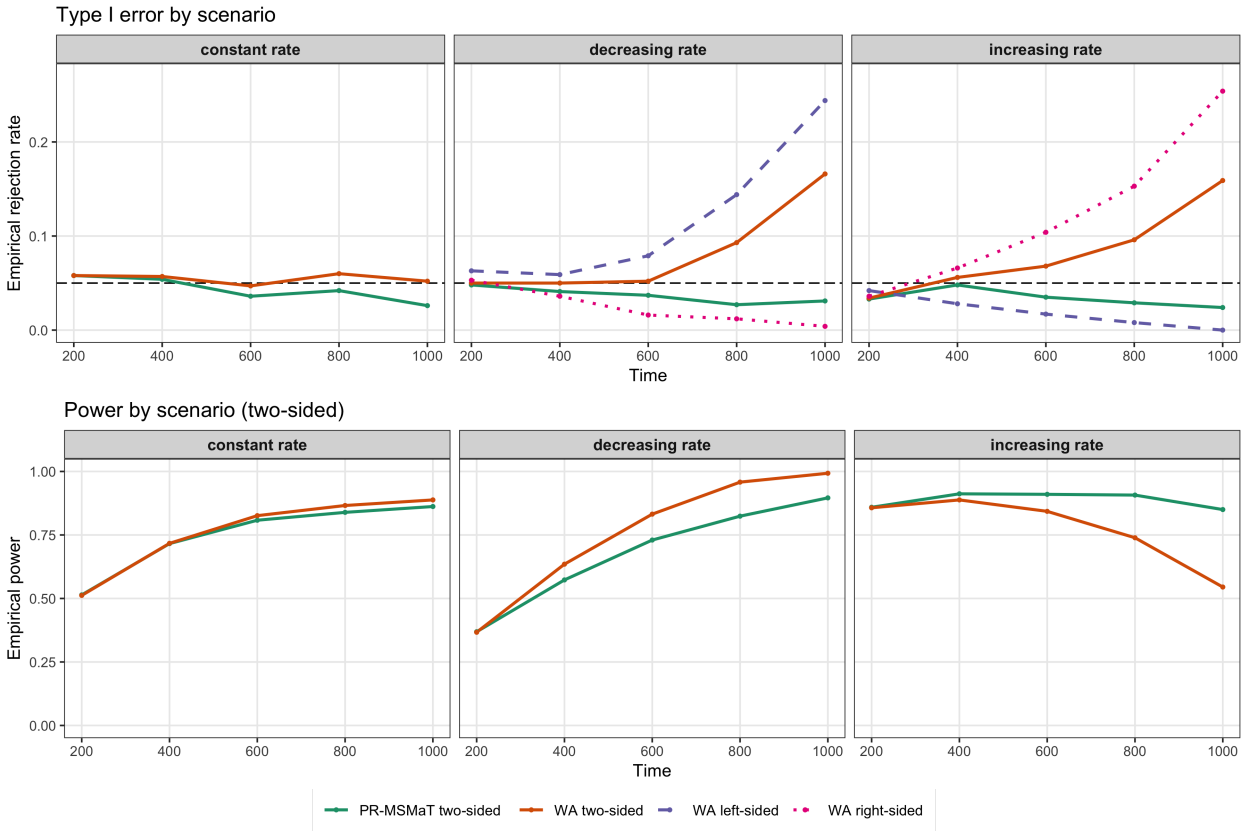


图4：在模拟研究设计的三种情景下，随时间变化的经验I类错误率和检验效能：基线复发事件率恒定、下降和上升。水平虚线表示名义显著性水平0.05。

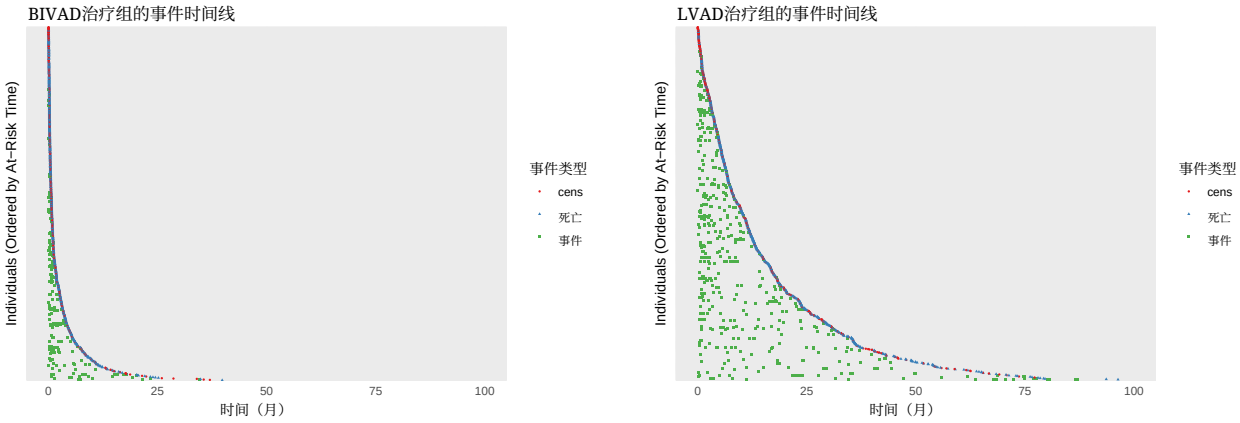


图5：胃肠道出血（GIB）复发事件的发生时间以绿点标记，死亡事件以边缘处的蓝点标记，删失事件以红点标记；图中展示了不同器械类型（LVAD 与双心室辅助装置）下的匹配对。

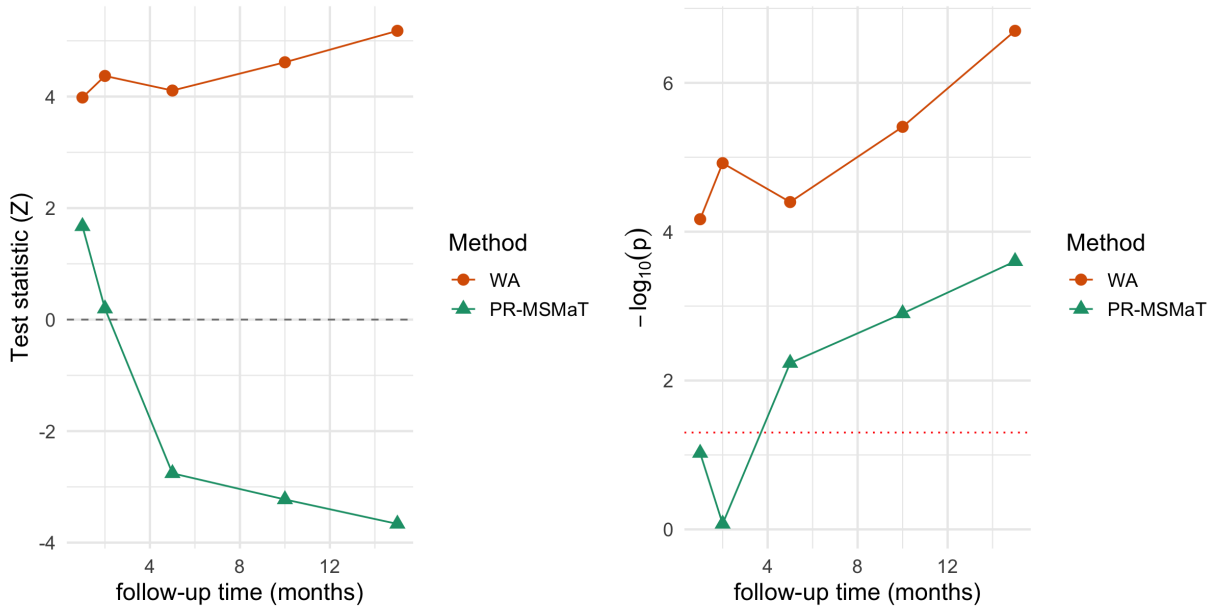


图6：两种检验方法——现有的 WA（基于存活期间的检验）和新提出的 PR-MSMaT（比例率边际结构模型辅助检验）——在五个选定随访时间点（月）的检验统计量（Z）及其 $-\log_{10}(p)$ 值比较。左图中的水平虚线表示 $Z = 0$，右图中的红色点线对应传统的显著性阈值 $p = 0.05$。