一种非参数方法，用于估计对数秩检验统计量之间的相关性，并应用于合取效能计算

Anne Lyngholm Sørensen、Paul Blanche、
HenrikRavn 和 Christian Pipper

2026年1月7日

arXiv:2601.03069v1 [stat.ME] 6 Jan 2026

摘要

我们提出了一种方法，用于估计在两个生存时间终点上分别检验零假设时，对数秩检验统计量之间的相关性。该相关性采用一种非参数方法，基于任意备择假设下对数秩检验统计量的独立同分布分解，利用个体层面数据进行估计。借助独立同分布分解，我们能够在较少依赖假设的情况下估计相关性。文中提供了一个使用所开发方法的示例，并说明如何利用该方法对预期合取效能进行切合实际的量化，从而为使用历史数据设计新的随机临床试验提供指导。最后，我们通过模拟研究考察了该方法的有限样本性质，结果证实所提出的方法具有无偏性和一致性。此外，模拟研究还揭示了删失对对数秩检验统计量之间相关性的影响。

1 引言

大多数临床试验都包含不止一个关注终点。对于多终点试验，通常会在确证性终点之间设定正式的层级，以体现其相对重要性并控制 I 类错误。共同主要终点是一个常见示例。在这种情况下，只有所有主要终点均显示治疗有效，试验才算成功。另一个常见示例是采用一个主要终点和多个确证性次要终点。在这种情况下，各终点按严格的层级排序：试验只有在主要终点上显示疗效后，才能继续评估确证性次要终点上的疗效。本文的示例聚焦于后一种情形。具体而言，我们以多个生存时间终点为背景，展示所开发的方法；这些终点按一个主要终点和多个确证性次要终点的层级排列，且通过门控程序或“固定序贯法´´ 来评估治疗效果并控制总体 I 类错误（Bauer 等，1998；Westfall 和 Krishen，2001；美国食品药品监督管理局，2022）。

在包含确证性次要终点的试验设计阶段，一个重要目标是优化预期效能，既要能够证明主要终点上的疗效，也要能够证明主要终点及所有确证性次要终点上的疗效。我们将同时拒绝所有关注假设的效能称为合取效能（Senn 和 Bretz，2007）。众所周知，检验统计量之间的相关性会影响合取效能；检验之间的相关性越高，合取效能越高（Senn 和 Bretz，2007）。因此，要恰当地评估合取效能，我们需要对所涉及检验统计量之间的相关性作出可靠估计。

本文关注生存数据，以及使用对数秩检验比较不同治疗组之间的风险率。生存时间终点之间的相关性难以概括，因此检验统计量之间的相关性也同样难以概括（Hougaard，2000，第4章）。在生存研究中，一个挑战是随访时间有限所导致的删失，这会使估计过程变得复杂。尽管存在这些挑战，估计生存时间终点及其检验统计量之间相关性的研究仍取得了若干进展。在肿瘤学试验中，多篇研究论文着重估计无进展生存期与总生存期之间的相关性。综述可参见 Meller 等（2019）。这些方法中有许多依赖于对事件时间联合分布进行参数建模，例如通过 Copula。Meller 等（2019）提出、并随后在 Danzer 等（2025）中再次采用的方法，基于疾病-死亡模型对联合终点分布进行建模。直接使用 Copula 对生存时间终点之间的相关性进行参数建模的一个具体示例，是 Schemper 等（2013）提出的迭代多重插补法。该方法假设采用高斯 Copula，并通过蒙特卡罗模拟估计相关性。另一个利用 Copula 估计对数秩检验统计量之间相关性的示例见 Sugimoto 等（2013）。与我们的工作类似，所提出的方法旨在解决基于历史数据进行

效能和样本量计算的问题。Copula模型以及其他用于刻画受试者层面事件时间联合行为的参数方法存在一个局限，即假设了特定的依赖结构。这意味着，要使相关性等指标的估计正确，所选的 Copula模型必须设定正确。

本文提出一种非参数方法，用于估计对数秩检验统计量之间的相关性。具体而言，该方法无需对个体层面数据的联合行为作出假设，即可获得无偏的相关性估计值。我们的方法完全不必对生存时间终点作出建模假设，而是直接评估对数秩检验统计量的同时行为。为此，我们将对数秩检验统计量分解为独立同分布随机变量之和，从而识别其在任意备择假设下的影响函数。随后，可利用这些影响函数，将检验统计量的渐近联合分布刻画为一种特定的多元正态分布（Pipper等人，2012）。我们利用这一刻画，基于个体层面数据（例如来自历史试验的数据）估计检验统计量间的相关性。该方法在计算上也很高效，因为估计相关性时无需对个体层面数据进行模拟或自助法。我们通过一个示例说明该方法的应用：基于从类似历史试验中估计的相关性，使用预期合取效能评估一项即将开展的试验可能采用的检验层级。结果提供了有价值的见解，可为层级决策提供参考。最后，我们通过一项模拟研究表明，所提出的方法能够提供无偏且一致的估计值。

本文结构如下。第2节将介绍一个示例，说明如何利用四个终点的边际效能及检验统计量之间的相关性来计算合取效能。第3节将介绍我们提出的用于估计对数秩检验统计量相关性的方法。本节将先在计数过程框架下介绍对数秩检验统计量，然后对其进行分解。此外，本节还将给出相关性估计量的表达式。随后，第4节将把所提出的方法应用于该示例。第5节介绍模拟研究及其结果，表明该估计量无偏且一致；我们还指出，其计算时间很短。第6节对所提出的方法进行讨论，包括其局限性和未来计划开展的研究，并作为本文的结论。

2 动机：合取效能

设有一项即将开展的随机对照试验（RCT），试验 X，用于比较新治疗与对照组。该试验计划设置一个确证性主要终点和三个确证性次要终点。主要终点和次要终点均为时间-事件终点，试验将采用层级检验程序检验这四个假设。这意味着，试验中的假设将按照预先规定的顺序，在 I 类错误水平 α 下进行检验，并控制家族错误率。检验方案从层级中的第一个（主要）零假设开始。如果在水平 α 下拒绝该零假设，则按顺序在水平 α 下检验第二个假设。若第二个零假设被拒绝，则可继续检验第三个假设；若第三个零假设被拒绝，则可继续检验第四个假设。因此，当某个零假设未被拒绝，或所有零假设均已完成检验时，检验即停止（Bauer 等，1998；美国食品药品监督管理局，2022）。试验 X 将与已完成的 SELECT 试验（Lincoff等人，2023）十分相似，并采用相同类型的主要终点和次要终点。SELECT 是诺和诺德开展的一项心血管结局试验，旨在研究司美格鲁肽对2型糖尿病患者心血管结局的影响，并与安慰剂进行比较。与 SELECT 一样，主要终点为首次发生三点主要不良心血管事件（3-point MACE）的时间，包括心血管死亡（CVD）、非致死性卒中和非致死性心肌梗死（MI）。确证性次要终点为心血管死亡（CVD）、全因死亡（ACD），以及心力衰竭住院和心血管死亡的复合终点（简称 HFC）。我们将利用 SELECT 的数据，为试验 X设计中的一些决策提供参考。

在试验 X 中，层级检验程序将从主要终点——主要不良心血管事件开始。次要终点的顺序尚未确定，我们希望通过研究哪种排序能够使被拒绝的零假设的期望数量最大，来提出一种排序方案。为此，我们将估计四个终点的所有检验子集的合取效能。在此，较高的合取效能意味着拒绝零假设的概率较高。现在，我们将以一般形式定义同时拒绝多个零假设的概率，针对的是我们的四个终点；该公式并不特定于对数秩检验，也可以适用于任何标准化检验统计量。

我们用 H_0^{MACE}、H_0^{CVD}、H_0^{ACD}、和 H_0^{HFC} 表示四个零假设。令 $R^j \in \{0,1\}$ 表示 H_0^j 是否被拒绝，以支持备择假设 H_A^j，其中 $j \in \{MACE, CVD, ACD, HFC\}$。使用此记号，拒绝全部四个零假设的合取效能表示为以下概率：

$$P\left(\{R^{MACE}=1\} \cap \{R^{CVD}=1\} \cap \{R^{ACD}=1\} \cap \{R^{HFC}=1\}\right).$$

请注意，这一表述不依赖于特定的检验顺序。它表示无论采用何种具体的检验顺序，拒绝所有零假设的概率。因此，这等同于在终点为共同主要终点时拒绝全部

四个零假设的概率。我们稍后会加入特定顺序这一要素。采用正式的符号表示后，现在可以说明如何计算合取效能。对于拒绝全部四个零假设的情形，令 $\mathbf{Z} = (Z^{MACE}, Z^{CVD}, Z^{ACD}, Z^{HFC})^T$ 为标准化检验统计量构成的向量。请注意，$\mathbf{Z}$ 可以是任意一组标准化检验统计量。我们预期，渐近地，$\mathbf{Z}$ 服从多元正态分布，即:

$$\mathbf{Z} = \begin{pmatrix} Z^{MACE} \\ Z^{CVD} \\ Z^{ACD} \\ Z^{HFC} \end{pmatrix} \sim N\left[\begin{pmatrix} \delta^{MACE} \\ \delta^{CVD} \\ \delta^{ACD} \\ \delta^{HFC} \end{pmatrix}, \begin{pmatrix} 1 & \rho^{MACE,CVD} & \rho^{MACE,ACD} & \rho^{MACE,HFC} \\ & 1 & \rho^{CVD,ACD} & \rho^{CVD,HFC} \\ & & 1 & \rho^{ACD,HFC} \\ & & & 1 \end{pmatrix} \right].$$

这里，$\rho^{j,k}$ 表示终点 j 和 k 的检验统计量之间的相关性，δ^j 表示第 j个检验统计量的非中心参数均值。对于每个终点 j，我们有 $\delta^j = \mathbf{E}(Z^j) = z_\alpha + z_{\beta_j}$，其中$z_x$ 是标准正态分布的第 x个分位数。因此，δ^j 是达到效能 $1-\beta_j$、在显著性水平为 α 时进行单侧检验所需的预期 z-分数（Proschan，2021）。效能 $1-\beta_j$ 取决于结局的事件数。对于一项主要终点的检验显著性水平为 2.5%、效能为 90% 的试验，备择假设下的预期 z 分数为 $\delta^j = z_{0.025} + z_{0.1} = 3.24$。

根据**Z**的分布，我们可以近似计算拒绝全部四个假设的合取效能$P_{\mathbf{Z}}\left(\cap_j\{R^j=1\}\right) = P_{\mathbf{Z}}\left(\cap_j\{Z^j \in B^j\}\right)$，其中 B^j 表示 Z^j的拒绝域。给定 δ^j 和 $\rho^{j,k}$，可以优化三步程序中确证性次要终点检验的顺序，例如:

$$\begin{aligned}
&\text{step 1: } \operatorname*{argmax}_x P_{\mathbf{Z}}\left(\cap_{j\in\{MACE,x\}}\{R^j=1\} \,\middle|\, x \in \{CVD, ACD, HFC\}\right) \\
&\text{step 2: } \operatorname*{argmax}_y P_{\mathbf{Z}}\left(\cap_{j\in\{MACE,x,y\}}\{R^j=1\} \,\middle|\, y \in \{CVD, ACD, HFC\} \setminus \{x\}\right) \\
&\text{step 3: add the remaining endpoint } w \in \{CVD, ACD, HFC\} \setminus \{x, y\}
\end{aligned}$$

上述方案优先考虑最大化被拒绝的零假设数量；其他方案或许更具相关性，但本文不予考虑，因为这类方案通常基于临床相关性或类似因素，而这些与相关性无关。

2.1 试验 X的合取效能估计

我们希望使用上述方案来确定试验X中确证性次要终点检验的顺序。我们希望采用能使预期被拒绝的零假设数量最大化的顺序。要估计合取效能，我们需要估计预期检验分数 δ^j 以及 $\rho^{j,\,k}$之间的预期相关性。

试验设计的目标是在显著性水平为2.5%的情况下，对备择假设下主要终点（主要不良心血管事件）的零假设进行检验，并达到90%的效能。因此，主要终点的预期检验分数为 $\delta^P = z_{0.025} + z_{0.1} = 3.24$。由于我们关注风险率，可以写出 δ^P 与风险比（HR）之间的关系。对于 $HR = 0.83$，由于$\log(HR)\sqrt{d/4} \approx 3.24$，因此 $d = 1211$，这就是达到90%效能所需的事件数（Proschan，2021）。假设我们预期风险比和事件数将与SELECT试验中观察到的情况非常相似（Lincoff等人，2023年）。那么我们将获得表1所示的信息。此处我们还列出了各终点的边际效能，计算方法为 $\Phi(\delta - 1.96)$，其中 Φ 是标准正态分布的累积分布函数。表中还列出了事件中属于心血管死亡的数量，因为所有主要和次要终点均包含心血管死亡。事件的共享很可能会导致各终点及其检验统计量之间存在较高相关性。

终点	HR	事件	心血管疾病事件	δ	边际效能
MACE	0.83	1211	410	3.24	90%
心血管死亡（CVD）	0.85	485	485	1.79	43%
全因死亡（ACD）	0.80	830	485	3.21	90 %
心力衰竭复合终点（HFC）	0.80	660	445	2.87	82 %

表1：风险比、事件、估计的检验分数 δ 及相应的边际效能。风险比和事件数根据 SELECT 的观测数据估算。

如果合取效能仅取决于边际效能，那么层级中的第二个终点最优选择应为 ACD，因为在三个终点中，该终点的边际效能最高。然而，我们知道合取效能取决于检验统计量之间的相关性，因此，为了根据合取效能确定层级检验顺序，我们需要表2中所述的输入数据。

我们希望无需依赖终点或检验统计量的多变量生存分布来估计检验统计量之间的相关性。第3节介绍了一种估计相关性的新方法，用于估计

次要终点	心血管死亡（CVD）	全因死亡（ACD）	心力衰竭复合终点（HFC）
$\delta = \log(HR)\sqrt{d/4}$	1.79	3.21	2.87
与 MACE-3 的相关性	?	?	?
与心血管死亡的相关性	-	?	?
与 ACD 的相关性	-	-	?

表2：三个确证性终点的检验分数。要计算合取效能，我们需要估计终点之间的相关性。缺失的估计值以问号（?）表示。

两个对数秩检验统计量。该方法利用备择假设下对数秩检验统计量的分解，非参数地估计相关性。

3 利用对数秩检验统计量的分解估计相关性

3.1 设定与符号

考虑一项比较新治疗（$A = 1$）与对照治疗（$A = 0$）的随机对照试验（RCT），其中 n_1 名参与者被随机分配至新治疗组，n_0 名参与者被分配至对照治疗组，因此参与者总数为 $n = n_0 + n_1$。令 T_{ij} 表示参与者 i（$i = 1, \ldots\ldots, n$）自随机化以来至终点 $j \in \{1, 2\}$ 发生事件的时间。为便于说明，本节仅考虑两个终点。将结果推广至两个以上终点并不困难。

我们考虑一项事件驱动试验，当主要事件的比例达到某个百分比 q 时，试验结束。令停止时间 ξ 为达到 q 的日期。对于受试者 i 的入组日期 U_i，ξ 的极限值 ξ 可定义为满足 $P(U_i + T_i \le \xi) = q$ 的分析日期。右删失可能发生在某个日历时间 D_i，并在试验结束时强制实施，使得受试者 i 从随机化到删失的时间定义为 $C_i(\xi) = \xi \wedge D_i - U_i$，因此我们观察到事件时间 $T_{ij}(\xi) = T_{ij} \wedge C_i(\xi)$ 的右删失版本。暂且用未知但固定的 ξ 代替 ξ。这样做是为了从形式上避免对未来进行条件化。本节最后将说明这只是一个技术细节：即使以 ξ 代替 ξ，所推导出的对数秩检验统计量分解仍然成立。

令 $t \in [0, \tau(\xi)]$ 为研究开始与研究时长 $\tau(\xi)$之间的一个时间点，其中 $\tau(\xi)$ 是试验结束时间 ξ 与入组时间 U_i之间的最大差值，因此 $\tau(\xi) = \max(\xi - U_i)$。随后，将第 j 个事件类型在第 i 个受试者中的计数过程 $N_{ij}(t, \xi)$、风险集过程 $Y_{ij}(t, \xi)$ 和滤过 $\mathcal{F}_{ij}(t, \xi)$ 定义如下：

$$
\begin{aligned}
N_{ij}(t, \xi) &= I(T_{ij} \le t,\, T_{ij} \le C_i(\xi)),\\
Y_{ij}(t, \xi) &= I(T_{ij} \ge t,\, C_i(\xi) \ge t),\\
\mathcal{F}_{ij}(t, \xi) &= \sigma\{N_{ij}(s, \xi), Y_{ij}(s, \xi), A_i, U_i :\, s \le t\}.
\end{aligned}
$$

其中，$\mathcal{F}_{ij}(t, \xi)$ 是由 $\{N_{ij}(s, \xi)\ Y_{ij}(s, \xi)\ A_i\}_{0 \le s \le t}$ 张成的最小 σ-代数。

假设入组时间以及删失与潜在事件时间和随机分配的治疗 A 相互独立，则相对于滤过 $\mathcal{F}_{ij}$ 的计数过程强度为：

$$
\lambda_{ij}(t, \xi) = Y_{ij}(t, \xi)\{\gamma_{0j}(t)(1 - A_i) + \gamma_{1j}(t)A_i\},
$$

其中，$\gamma_{Aj}(t)$ 表示治疗组 A、事件类型 j 在时间 t 的风险率。采用上述记号，第 j 个终点的对数秩检验统计量定义为：

$$
G_j(\xi) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau(\xi)}\{A_i - E_j(s, \xi)\}N_{ij}(ds, \xi), \quad \text{with,} \quad E_j(t, \xi) = \frac{\sum_{i=1}^{n} A_i Y_{ij}(t, \xi)}{\sum_{i=1}^{n} Y_{ij}(t, \xi)}. \tag{1}
$$

这里，$G_j(\xi)$ 是事件类型 j 的对数秩检验统计量的分子，$E_j(t, \xi)$ 是在时间 t 处被分配至 $A = 1$ 且面临事件 j 风险的参与者比例，而 $N_{ij}(ds, \xi) = N_{ij}(s, \xi) - \lim_{h \searrow 0} N_{ij}(s - h, \xi)$。为便于后续使用，我们还指出，$E_j(t, \xi)$ 依概率一致收敛至：

$$
e_j(t, \xi) = \frac{P(T_{ij} \ge t, C_i(\xi) \ge t,\, A_i = 1)}{P(T_{ij} \ge t, C_i(\xi) \ge t)}.
$$

对数秩检验用于检验两个累积风险之间是否存在差异。我们将感兴趣的零假设定义为 H_{0j}：$\gamma_{0j}(t) = \gamma_{1j}(t)$ 对所有 t 均成立，而备择假设为 H_{Aj}：$\gamma_{0j}(t) \ne \gamma_{1j}(t)$ 对某个 t 成立。

3.2 分解

我们的相关性估计基于一种分解；经过适当缩放和中心化后，该分解可将$G_j(\xi)$近似表示为所谓的影响函数 Φ_{ij} 之和。这些影响函数是均值为零的独立同分布随机变量（ Tsiatis，2006 ）。具体而言，我们证明：

$$\sqrt{n}\{G_j(\xi) - \mathbf{E}(G_j(\xi))\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{ij}(\xi) + o_P(1).$$

我们在附录6中推导出 $\Phi_{ij}(\xi)$：

$$\begin{aligned}\Phi_{ij}(\xi) =& [A_i - e_j(\overline{T}_{ij},\xi)]I(T_{ij} \le C_i(\xi)) - \int_0^{\tau(\xi)}[A_i - e_j(t,\xi)]Y_{ij}(t,\xi)\gamma_{0j}(t)dt \\ &- \int_0^{\tau(\xi)} Y_{ij}(t,\xi)A_i[1-e_j(t,\xi)][\gamma_{1j}(t)-\gamma_{0j}(t)]dt \\ &+ \int_0^{\tau(\xi)}[A_iY_{ij}(t,\xi) - P(T_{ij}\ge t,\, C_i(\xi)\ge t,\, A_i=1)][1-e_j(t,\xi)]^2[\gamma_{1j}(t)-\gamma_{0j}(t)]dt \\ &+ \int_0^{\tau(\xi)}[(1-A_i)Y_{ij}(t,\xi) - P(T_{ij}\ge t,\, C_i(\xi)\ge t,\, A_i=0)][e_j(t,\xi)]^2[\gamma_{1j}(t)-\gamma_{0j}(t)]dt.\end{aligned} \tag{2}$$

为了估计检验统计量间的相关性，我们使用影响函数的标准化版本。给定 $P(A=1)=\pi(A)$，影响函数的标准化版本为：

$$\Phi^*_{ij}(\xi) = \frac{1}{\sqrt{\pi_A(1-\pi_A)P(T_{ij}\le\tau(\xi))}}\times\Phi_{ij}(\xi). \tag{3}$$

通过堆叠独立同分布分解（Pipper等人，2012），我们可以估计对数秩检验统计量之间的相关性：

$$corr(G_1(\xi),G_2(\xi)) = \rho_{1,2}(\xi) = \frac{\mathbf{E}(\Phi^*_{i1}(\xi)\Phi^*_{i2}(\xi))}{\sqrt{\mathbf{E}\big[\{\Phi^*_{i1}(\xi)\}^2\big]\mathbf{E}\big[\{\Phi^*_{i2}(\xi)\}^2\big]}}. \tag{4}$$

最后，我们放宽停止日期已知这一假设；$\hat{}\,\xi$。相反，我们假设$\hat{}_\xi \xrightarrow{P} \xi$，并将对数秩检验视为 ξ 中的一个过程，即：

$$\mathcal{W}_n(\xi) = \sqrt{n}\{G_j(\xi) - \mathbf{E}(G_j(\xi))\}.$$

根据（van der Vaart，1998，第19章），我们认为上述过程作为 ξ 的过程是紧的。因此，

$$\mathcal{W}_n(\hat{\xi}) - \mathcal{W}_n(\xi) = o_p(1).$$

$\hat{}$ 我们得出结论：将 ξ 替换为 ξ后，分解仍然成立。

3.3 代入估计

为估计　　式 (4)，在实践中我们将使用代入估计值。估计相关性的公式为　　：

$$\hat{\rho}^{j,k}(\hat{\xi}) = \frac{1}{n-1}\frac{\sum_{i=1}^n(\hat{\Phi}^*_{ij}(\hat{\xi}) - \overline{\Phi}_{ij}(\hat{\xi}))(\hat{\Phi}^*_{ik}(\hat{\xi}) - \overline{\Phi}_{ik}(\hat{\xi}))}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n(\hat{\Phi}^*_{ij}(\hat{\xi}) - \overline{\Phi}_{ij}(\hat{\xi}))^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^n(\hat{\Phi}^*_{ik}(\hat{\xi}) - \overline{\Phi}_{ik}(\hat{\xi}))^2}}.$$

其中 $\Phi_{ij}(\hat{}\,\xi) = \frac{1}{n}\sum_{i=1}^n\hat{}\,\Phi^*_{ij}(\hat{}\,\xi) \approx 0$。

令$\hat{e}(t,\hat{}\,\xi) = \frac{\sum_i Y_{ij}(\quad t\,\xi\, A_i,\hat{}\,)}{\sum_i Y_{ij}\, t\,\xi}(\,,\hat{}\,)$。对于风险率，我们使用 $\hat{}\,\gamma_{Aj}(t)dt = d\hat{}\,\Gamma_{Aj}(t)$，其中 $\hat{}\,\Gamma_{Aj}(t)$ 是治疗 A 终点 j在时间 t 处的累积风险函数的Nelson-Aalen估计量。因此，对于所有 t ，$\hat{}\,\Gamma_{Aj}(t)$ 是 $\sum_t \frac{d_{Aj}(t)}{Y_{\cdot Aj}}(t)$，而 $\gamma_{Aj}(t)dt = \frac{d_{Aj}(t)}{Y_{\cdot Aj}(t)}$，其中 $d_{Aj}(t)$ 是治疗组 A 在终点 j的 t 时发生的失败事件数，$\hat{}$， $Y_{\cdot Aj(t)}$ 是治疗 A 在终点 j的 t 时的风险集。最后，两个概率 $P(T_{ij}\ge t,\, C_i(\xi)\ge\hat{}\,t,\, A_i=1)$ 和 $P(T_{ij}\ge t,\, C_i(\xi)\ge t,\, A_i=0)$ 表示各组处于风险状态的概率，其估计方法为

使用 $\frac{\sum_i Y_{ij}(t\ \xi\ A_i,\hat{})}{}$和 $\frac{\sum_i Y_{ij}(t,\hat{}\xi)(1-A_i)}{n}$。其中，$\hat{}\pi_A = \frac{\sum_{i=1}^n A_i}{n}$是分配至$\hat{}$治疗 $A=1$的观测概率， Φ_{ij}^* 的完整表达式为:

$$
\begin{aligned}
\hat{\Phi}_{ij}^* = & \frac{1}{\sqrt{\hat{\pi}_A\left(1-\hat{\pi}_A\right)\frac{\sum_{t=0}^{\tau(\hat{\xi})} d_{0j}(t)+d_{1j}(t)}{n}}} \times \\
& \left(\left[A_i - \frac{\sum_{i=1}^n A_i Y_{ij}(t,\hat{\xi})}{\sum_{i=1}^n Y_{ij}(t,\hat{\xi})}\right] I(T_{ij} \le C_i(\hat{\xi})) - A_i \sum_{t=0}^{\overline{T}_{ij}(\hat{\xi})} \frac{d_{0j}(t)}{Y_{\cdot 0j}(t,\hat{\xi})} + \sum_{t=0}^{\overline{T}_{ij}(\hat{\xi})} \hat{e}(t,\hat{\xi}) \frac{d_{0j}(t)}{Y_{\cdot 0j}(t,\hat{\xi})}\right. \\
& - A_i \sum_{t=0}^{\overline{T}_{ij}(\hat{\xi})} [1-\hat{e}(t,\hat{\xi})] \left[\frac{d_{1j}(t)}{Y_{\cdot 1j}(t,\hat{\xi})} - \frac{d_{0j}(t)}{Y_{\cdot 0j}(t,\hat{\xi})}\right] \\
& + A_i \sum_{t=0}^{\overline{T}_{ij}(\hat{\xi})} [1-\hat{e}(t,\hat{\xi})]^2 \left[\frac{d_{1j}(t)}{Y_{\cdot 1j}(t,\hat{\xi})} - \frac{d_{0j}(t)}{Y_{\cdot 0j}(t,\hat{\xi})}\right] \\
& - \sum_{t=0}^{\overline{T}_{ij}(\hat{\xi})} \frac{Y_{\cdot 1j}(t,\hat{\xi})}{n} [1-\hat{e}(t,\hat{\xi})]^2 \left[\frac{d_{1j}(t)}{Y_{\cdot 1j}(t,\hat{\xi})} - \frac{d_{0j}(t)}{Y_{\cdot 0j}(t,\hat{\xi})}\right] \\
& + (1-A_i) \sum_{t=0}^{\overline{T}_{ij}(\hat{\xi})} \hat{e}(t,\hat{\xi})^2 \left[\frac{d_{1j}(t)}{Y_{\cdot 1j}(t,\hat{\xi})} - \frac{d_{0j}(t)}{Y_{\cdot 0j}(t,\hat{\xi})}\right] \\
& \left. - \sum_{t=0}^{\overline{T}_{ij}(\hat{\xi})} \frac{Y_{\cdot 0j}(t,\hat{\xi})}{n} \hat{e}(t,\hat{\xi})^2 \left[\frac{d_{1j}(t)}{Y_{\cdot 1j}(t,\hat{\xi})} - \frac{d_{0j}(t)}{Y_{\cdot 0j}(t,\hat{\xi})}\right]\right).
\end{aligned}
$$

4 优化逐步合取效能

现在，我们可以回到使用合取效能设计新试验的示例。我们希望利用 SELECT 试验的数据估计四个检验统计量之间的相关性。我们可以通过第3.3小节中介绍的代入估计量，使用我们的建议的估计量（4）所得到的相关性估计值来更新表2。更新后的表2及相关性估计值见表3。所有终点之间都具有较高的相关性，这是很自然的，因为这些结局共享事件；心血管死亡包含在主要终点（主要不良心血管事件）、HFC 结局和 ACD 结局中。

次要终点	CVD	ACD	HFC
δ	1.79	3.21	2.87
与 MACE-3 的相关性	0.60	0.48	0.56
与心血管死亡的相关性	-	0.76	0.85
与 ACD 的相关性	-	-	0.67

表3：使用建议的估计量计算相关性估计值后的更新表。

根据表3中的估计值，并假设以 2.5% 的α 水平拒绝主要零假设的效能为 90%（即 $\delta^{MACE}=3.24$），使用第2节所述的多元正态分布，拒绝全部四个零假设的合取效能为 42%。为了解相关性对效能的影响，我们可以在假设独立性的情况下计算合取效能，即令所有 $j \neq k$ 的相关性 $\rho^{j,\ k}=0$ 。由此得到的合取效能为 28%。同样，也可以将四个零假设的边际效能相乘，得到 28% 的效能（即 $0.9\times0.43\times0.90\times0.82\approx0.28$）。相反，假设它们完全相关，即所有 $j \neq k$ 的相关性 $\rho^{j,\ k}=1$ ，则合取效能为 43%，等于所考虑的四个终点中最低的边际效能。

也可以针对零假设的子集计算合取效能。通过计算不同集合的效能，我们可以使用第2节中描述的三步程序，找出在层级的每一层使合取效能最大化的检验顺序。图1展示了最优检验顺序：层级顶部为主要检验，而次要终点的检验顺序则根据合取效能最高的原则确定。根据表3所示的估计值，就合取效能而言，最优层级为 $MACE \to ACD \to HFC \to CVD$试验 X对应的估计层级合取效能分别为90% $\to$ 83% $\to$ 74% $\to$ 42%。这意味着，在备择假设为真的情况下，拒绝主要终点和第一个次要终点的合取效能为83%。

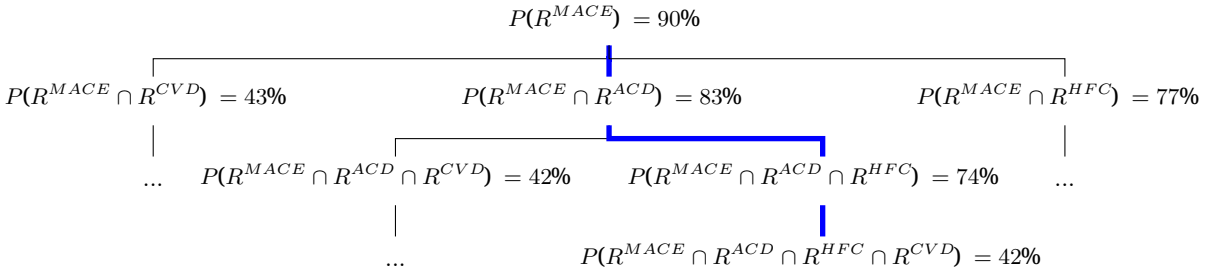

图1：利用合取效能确定确证性次要终点的排序。蓝线表示以最大化合取效能为目标时，确证性次要终点检验的最优排序。

在实践中，计算得出的最优层级可作为安排各终点的一种统计视角。因此，它更多地是设计时的参考信息，而非严格规则，因为其他考量也很重要。此外，由于计算基于历史数据，用户需要评估他们认为新的试验数据与旧数据有多相似。因此，使用略高或略低的相关性取值进行多次计算通常很有价值。我们在第6节中对这一主题进行了更详细的讨论。

5 模拟研究

5.1 模拟目的、设置与评估

我们将通过模拟研究考察该估计量的经验性质，重点关注在样本量、数据生成分布、删失水平和相关性各不相同的情形下，该方法的准确性。部分模拟研究的设计受到 SELECT试验的启发，同时也将涵盖更一般的情形。本文模拟研究的主要目的是验证式（4）中定义的估计量是否无偏且一致。因此，我们将重点关注使用本方法对数秩检验统计量相关性的估计。

本节中的模拟数据通过参数模拟生成。我们模拟一项试验，以1:1的比例将参与者随机分配至对照组或治疗组；所关注的两个终点均为生存时间终点，其中一个归类为主要终点，另一个归类为次要终点。我们假设参与者在1.5年内均匀入组。该试验将以事件为驱动，并在主要终点的事件数达到与目标删失水平相对应的特定数量时停止。

两个终点之间存在相关性，我们将使用不同的 Copula模型生成相关的生存数据。所用的 Copula模型包括：高斯、Clayton、Frank 和 Gumbel（Hougaard，2000）。对于高斯 Copula，我们将考察 Copula参数为 0、0.5 和 0.8 时的相关性；这些参数对应于未发生删失时终点之间的相关性。对于 Frank、Clayton 和 Gumbel 模型，我们分别将 Copula参数设为 4、1 和 1。在本文所述的特定模拟设置下，这些取值会产生大小不一的相关性，具体结果见下文。我们选择考察不同的 Copula模型，以探究不同的相关结构。将从 Copula生成的均匀分布数据转换为指数分布。治疗组的发生率取决于给定的风险比。数据将采用两种指数分布形式：恒定发生率和分段恒定发生率。我们考虑两种删失水平：80% 和 93%。部分模拟参数参考了 SELECT试验，包括 93% 的删失水平；其他参数包括风险率、模拟数据量 n_{obs} （例如 $n_{obs} = 17600$）以及风险比。不过，这些模拟参数中的大多数也采用了其他取值，以探索更一般的情景。

每种感兴趣的情景均进行 10,000 次模拟。根据模拟研究的重点，将采用表 4 中所列模拟参数的不同组合。此外，受 SELECT 试验启发，我们将通过使各终点在分布上相关（借助 Copula 模型）来引入相关性，同时也会将一个终点设为包含另一终点事件的复合终点。复合终点定义如下。再次令 T_{ij} 表示受试者 i终点 j的潜在右删失事件时间。复合终点的时间定义为 $V_{i1} = \min\left(T_{i1}, T_{i2}\right)$。在复合终点的模拟中，相关性将基于 V_{i1} 和 T_{i2}。复合终点的一个例子是第 2.1 节所述的 MACE-3。MACE-3 包括非致死性心肌梗死（MI）、非致死性卒中和心血管死亡（CVD）；因此，在该试验中，CVD 是次要终点，而 MACE-3 则是由 CVD、非致死性 MI 和卒中构成的复合终点。表 4 列出了所有模拟参数。模拟研究中使用这些终点作为记号，相关内容也见表 4。

模拟参数	取值
n_{sim}	10,000
n_{obs}	400、4000、15000、17600、40000
- 每组	200、2000、7500、8800、20000
Copula模型	高斯、Clayton、Frank 和 Gumbel
- θ (Copula 参数)	高斯：0、0.5、0.8 Clayton：1 Frank：4 Gumbel：1
风险率	指数常数和分段常数
- 发生率	常数： $\gamma^{plcb}_{mi,stroke} = 0.017$ 和 $\gamma^{plcb}_{cvd} = 0.009$ 分段常数：风险在 $t = 730.5$ 天时发生变化 $\gamma^{plcb}_{mi,stroke}(0-t) = 0.017$ 和 $\gamma^{plcb}_{mi,stroke}(t-\infty) = 0.017$ $\gamma^{plcb}_{cvd}(0-t) = 0.007$ 和 $\gamma^{plcb}_{cvd}(t-\infty) = 0.012$
风险比	0.8
复合终点	是，否
删失水平	80%、93%

表 4：模拟研究的模拟参数。 n_{sim} 是每种情景下的模拟次数， n_{obs}是每个数据集模拟的参与者人数， γ^x_y 是治疗组 x 在终点y的风险率。模拟分段常数风险率时，记法则为 $\gamma^x_y(z-w)$，其中 z 表示所考虑时间段的开始时间，w 表示结束时间。请注意， $plcb$ 代表安慰剂。

我们希望检验式（4）中给出的两个对数秩检验统计量之间相关性的估计量是否无偏且一致。我们将通过偏倚以及观测到的 2.5% 和 97.5% 百分位数来考察相关性估计的表现。为评估偏倚，我们将真实相关性设为模拟得到的对数秩检验统计量的经验相关性。因此，在每次模拟中，我们收集两个终点 j和 k在 $i = 1, \ldots,$ n_{sim}对应的对数秩检验 z^j_i和 z^k_i。真实相关性即模拟所得对数秩检验分数之间的皮尔逊相关系数，记作 $\tilde{\rho}$。每次模拟中，我们还会收集使用式（4）定义的独立同分布分解得到的相关性估计，记作 $\hat{}\rho^{iid}_i$；报告其均值 $\underline{\rho}^{iid} = 1/n_{sim} \cdot \sum_{i=1}^{n_{sim}} \hat{}\rho^{iid}_i$。偏倚估计计算如下：偏倚 $= \rho^{iid} - \tilde{\rho}$。我们将通过百分位数范围确认其是否会随 n_{obs}增加而缩小，以考察该估计量是否一致。

5.2 模拟结果

在介绍模拟结果之前，我们先强调所提出方法的速度。使用一个示例数据集（该数据集也可在 https://github.com/AnneLyng/cor logrank 获取），相关性估计的平均计算时间为 0.02 秒。计时使用统计软件 R（R 核心团队，2024）中的 microbenchmark 软件包（Mersmann，2024）进行。此平均计算时间包含调用 cor 函数所需的时间。因此，与自助法等方法相比，该方法能够更快地提供结果，使其更适用于模拟研究等场景。由于计算时间取决于运行代码的系统，读者可以在本地运行代码以了解计算时间；不过，鉴于该方法的实现相对简单，我们预计所有用户都能快速运行该方法。

现在我们将证明该方法是无偏且一致的。表5和表6展示了模拟结果。表5考察了恒定风险率、没有复合终点且试验组样本量为8800的情景（灵感来自SELECT）。我们发现，对于所考察的所有情景，无论采用哪种Copula模型或删失水平如何，所提出的方法均无偏。我们还发现，在所考察的情景中，删失会降低检验统计量之间的相关性，但降低程度因Copula模型而异。这是预期之中的，因为不同的Copula模型会将相关性集中在不同的时间段。例如，Clayton Copula的相关性集中得比Gumbel Copula更早；这也体现在删失对Gumbel Copula生成的数据的影响更大（相关性的大小被削弱得更多）这一点上。

表6以SELECT试验为蓝本，其中终点为复合终点，风险率为分段常数，试验组样本量为8800。SELECT试验中主要MACE终点的删失水平为$\sim$ 93%。与之前类似，我们发现该方法是无偏的。我们注意到，复合终点会影响相关性，因为我们估计得到的相关性更高。

相关性估计值的变异程度取决于样本量，此处以每组参与者人数来表示。不出所料，我们发现样本量越大，变异程度越小。图2展示了四种不同Copula模型下，2.5和97.5百分位数如何随样本量变化而降低

Copula	θ	删失	bias	$\tilde{\rho}$	ρ^{iid} （2.5%，97.5%百分位数）
高斯	0	80%	0.006	- 0.006	< 0.001 (-0.015;0.016)
-	-	93%	0.009	-0.009	< 0.001 (-0.015;0.015)
-	0.5	80%	0.013	0.265	0.279 (0.260;0.297)
-	-	93%	0.002	0.203	0.205（0.178；0.232）
-	0.8	80%	0.015	0.515	0.530 (0.531;0.546)
-	-	93%	0.006	0.452	0.458（0.429；0.486）
Clayton	1	80%	-0.024	0.473	0.449 (0.431;0.467)
-	-	93%	-0.024	0.477	0.454（0.424；0.483）
Frank	4	80%	-0.018	0.283	0.265（0.246；0.284）
-	-	93%	-0.022	0.133	0.111 (0.086;0.137)
Gumbel	1	80%	-0.019	0.365	0.346（0.328；0.364）
	-	93%	-0.018	0.250	0.232（0.203；0.261）

表5：在不同Copula模型下，指数风险率保持不变、风险比为0.8且采用非复合终点时的 $n_{obs} = 8800$ 模拟结果。其中， θ 为Copula参数， $\tilde{\rho}$ 为对数秩检验之间模拟得到的真实相关性， ρ^{iid} 为使用独立同分布分解估计的平均相关性，括号内列出了观测到的第2.5%和第97.5%百分位数。模拟次数为 $n_{sim} = 10,000$。

Copula	θ	删失	bias	$\tilde{\rho}$	ρ^{iid} （第2.5%和第97.5%百分位数）
高斯	0	93%	0.014	0.611	0.624 (0.602;0.647)
-	0.8	93%	0.002	0.568	0.570 (0.547;0.592)
Clayton	1	93%	-0.005	0.528	0.523 (0.499;0.546)
Frank	4	93%	-0.008	0.594	0.585 (0.863;0.882)
Gumbel	1	93%	-0.010	0.586	0.576（0.826；0.845）

表6：在不同Copula模型下，采用分段常数指数风险率、风险比为0.8及复合终点时， $n_{obs} = 8800$ 的模拟结果。此处， θ 为Copula参数， $\tilde{\rho}$ 为对数秩检验之间模拟得到的真实相关性， ρ^{iid} 为使用独立同分布分解得到的平均相关性估计值，括号中同时列出了观测到的2.5%和97.5%百分位数。模拟次数为 $n_{sim} = 10,000$

样本量。因此，我们的估计量表现出一致性。

6 讨论

通过对数秩检验统计量的独立同分布分解和堆叠，我们开发了一种估计对数秩检验统计量之间相关性的新方法。我们将该方法应用于一个示例，展示了如何利用它为新试验提供信息，从而在层级检验程序中确定统计学上最优的检验顺序。通过模拟，我们证明了该方法具有无偏性和一致性。接下来，我们将讨论本文所述方法的一些局限性，并进而探讨拓展该方法应用范围的计划研究。

检验统计量之间相关性的用途既包括设计新试验，也包括在试验结束时分析观测数据。本文阐述了如何将其用于设计，并将说明这种做法的一些局限性。在我们的示例中，我们在新试验的设计阶段应用该方法，为层级提出一种具体的排列顺序。该顺序根据合取效能进行优化，而合取效能取决于预期的检验统计量大小以及检验统计量之间的相关性。相关性基于一项历史试验。我们用于设计新试验的合取效能估计，只有在假设新试验的特征（如发生率和治疗效应）与历史试验相似时，才最具相关性。对某些研究而言，这一假设或许成立，但对另一些研究而言则不切实际。如果预期存在差异，了解该方法的适应性就很有价值。虽然新试验中预期的检验分数可以依据医学专业知识设定，也很容易调整，但在保持其他特征不变的情况下，根据新治疗的预期发生率和治疗效应调整用于计算相关性的历史数据，则更为复杂。我们计划研究两方面内容：一是利用历史数据调整操作特征有多容易（或困难），二是哪些操作特征会影响相关性估计以及影响程度。因此，我们希望研究这些特征，以便：a) 了解它们是否能够适应预期治疗效应等发生变化的新试验；b) 弄清哪些因素对相关性的影响最大。如果难以将历史数据调整到新的情景，那么评估在其他条件不变的情况下相关性如何随操作特征变化，或许可以作为

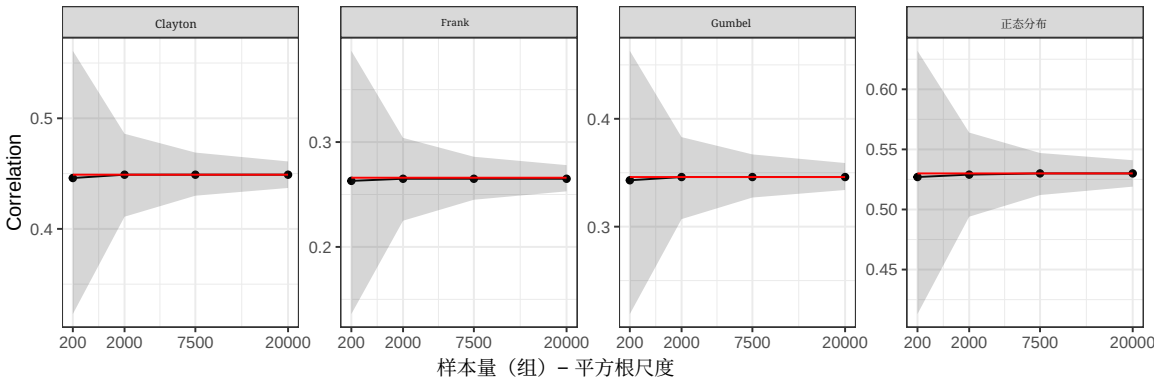

图2：阴影区域表示相关性估计值的第2.5和第97.5百分位数。红线表示真实值，黑线表示所提出的方法的平均相关性估计值 $\bar{\rho}^{iid}$。

另一种量化未来试验中不确定性的方法，例如联合效能计算中的不确定性。 al.

另一项计划中的未来研究是利用相关性估计量，在组序贯试验中优化次要终点的检验。已有多篇论文探讨如何通过 α消耗优化具有多个终点的组序贯试验，例如（Glimm 等人，2010；Tamhane 等人，2012；Li 等人，2018；Danzer 等，2025）。其中，（Glimm 等人，2010；Li 等人，2018）均假设相关性为常数。我们希望研究：对于生存时间检验统计量，当相关性不应被视为常数时，最优 α消耗程序是否会发生变化。Tamhane 等人（2012）的论文没有这一假设，但其方法不适用于生存时间检验统计量。因此，该领域仍需进一步研究，而我们提出的方法可为此提供支持。Danzer 等（2025）近期发表的论文不仅给出了两个生存时间终点的协方差矩阵表达式，还提出了在肿瘤学试验情境下，如何优化两个主要终点的 I 类错误消耗的检验策略。他们在疾病-死亡模型下推导协方差矩阵，并以无进展生存期和总生存期作为终点进行举例。我们认为，Danzer 等（2025）的研究结果与本文关于相关性估计量的结果相辅相成，因为本文提出的方法可用于其他可能不符合疾病-死亡模型的终点，并能提供切合实际的效能计算。对于计划开展的、旨在优化组序贯试验中多个终点检验的未来研究，我们希望探讨的最优设计并非针对多个主要终点，而是针对只有一个主要终点且至少有一个次要终点的情形。

总之，我们所提出的方法无需自助法、模拟或分布假设，即可对相关性进行无偏且一致的估计。与自助法相比，该方法使用起来更快。此外，如果想了解对删失等因素的敏感性，可以通过该方法的实现轻松（且快速）地进行考察。代码及一份展示如何使用该代码处理模拟试验数据的 R 示例脚本，请参阅补充信息或 GitHub 上的代码仓库 https://github.com/AnneLyng/cor logrank。

利益冲突

AnneL. Sørensen、*Henrik Ravn* 和 *Christian B. Pipper* 是诺和诺德公司的员工。*Henrik Ravn* 和 *Christian B. Pipper* 持有诺和诺德公司的股票。

附录

本附录部分将概述影响函数 Φ_{ij} 的推导，该影响函数是未经标准化的对数秩检验统计量的独立同分布分解的一部分。为便于阅读，本节不强调停止日期 ξ 的依赖性，因为它在分解的推导中不会发挥明显作用。

$$\sqrt{n}\{G_j - \mathbf{E}(G_j)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{ij} + o_P(1).$$

The 的推导 Φ_{ij} 可分为四个部分，下面将逐一介绍。这四个步骤是 e:

1. 运用鞅理论重新表述 G_j2. 处理 G_j 的补
偿子部分3. 推导 $\mathbf{E}(G_j)$ 的表达式 4. 综合以
上内容

A.1. 运用鞅理论重新表述 G_j

请记住， G_j 的表述如下:

$$G_j = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{A_i - E_j(s)\}dN_{ij}(s),$$

Wit$\mathrm{h}$ $M_{ij}(t) = N_{ij}(t) - \int_0^t \lambda_{ij}(s)ds$ 表示计数过程鞅，我们将 G_j 改写为:

$$\sqrt{n}G_j = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \{A_i - e_j(t)\}dM_{ij}(t) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \{A_i - E_j(t)\}\lambda_{ij}(t)dt + o_P(1).$$

这里的 $o_P(1)$ 项来自于将方程鞅部分中的 $E_j(t)$ 改为 $e_j(t)$。

A.2. G_j 的补偿子部分

我们首先利用 $\sum_{i=1}^n A_i Y_{ij} = E_j Y_{\cdot jl}$，并写出 $\lambda_{ij}(t)$ 的表达式，然后两次加上 0：($e_j(t) - e_j(t)$ 和 $-nP($ $T_{ij} \geq t, \quad C_i \geq t, \quad A_i = 1) + nP(T_{ij} \geq t, \quad C_i \geq t, \quad A_i = 1)$)，以此改写第二项 $\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \{A_i - E_j(t)\}\lambda_{ij}(t)$ dt。我们进一步引入 $Y_{\cdot jl}(t) = \sum_{i=1}^n Y_{ij}(t)$。

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \{A_i - E_j(t)\}\lambda_{ij}(t)dt = \\
&\frac{1}{\sqrt{n}}\int_0^\tau [e_j(t) - E_j(t)]\times \\
&[E_j(t)Y_{\cdot jl}(t) - nP(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)]dt \qquad (5)\\
&+ \frac{1}{\sqrt{n}}\int_0^\tau [e_j(t) - E_j(t)]nP(T_{ij} \geq t,\, C_i \geq t, A_i = 1)[\gamma_{1j}(t) - \gamma_{0j}(t)]dt \qquad (6)\\
&+ \frac{1}{\sqrt{n}}\int_0^\tau [1 - e_j(t)][E_j(t)Y_{\cdot jl}(t) - nP(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)]dt \qquad (7)\\
&+ \sqrt{n}\int_0^\tau [1 - e_j(t)]P(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)]dt \qquad (8)
\end{aligned}$$

在计数过程的补偿子部分中，我们最终得到四项。现在注意，最后一个等式中的第一项 (5) 是 $o_P(1)$，因为 $[e_j(t) - E_j(t)] = o_P(1)$，而 $[E_j(t)Y_{\cdot jl}(t) - nP(T_{ij} \geq t, \quad C_i \geq t, \quad A_i = 1)] = O_P(1)$。对于第二项 (6)，我们展开 $e_j(t)$和 $E_j(t)$的表达式，并使用“猴子吃自己的尾巴”技巧:

$$\frac{a}{b} - \frac{a_n}{b_n} = \frac{ab_n - ba_n}{bb_n} = \frac{ab_n - n\cdot ab + n\cdot ab - ba_n}{bb_n}$$

其中 $a_n = \sum_{i=1}^n A_i Y_{ij}(t)$ $a = P(T_{ij} \geq t,\ C_i \geq t,\ A_i = 1)$ $b_n = Y_{\cdot jl}(t)$，以及 $b = P(T_{ij} \geq t,\ C_{ij} \geq t)$。因此，对于第二项 (6)：

$$\frac{1}{\sqrt{n}} \int_0^\tau [e_j(t) - E_j(t)] n P(T_{ij} \geq t,\, C_i \geq t, A_i = 1)[\gamma_{1j}(t) - \gamma_{0j}(t)]dt =$$
$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau [Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t)] e_j(t)^2 [\gamma_{1j}(t) - \gamma_{0j}(t)]dt$$
$$-\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau [A_i Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t,\, A_i = 1)] e_j(t)[\gamma_{1j}(t) - \gamma_{0j}(t)]dt + o_P(1).$$

对于第三项 (7)，我们有

$$\frac{1}{\sqrt{n}} \int_0^\tau [1 - e_j(t)][E_j(t) Y_{\cdot jl}(t) - n P(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)]dt =$$
$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau [1 - e_j(t)][A_i Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)]dt$$

我们现在需要定义 $\mathbf{E}(G_{ij})$。

A.3. 求出 $\mathbf{E}(G_j)$ 的表达式

这里用到非中心性项近似为：

$$\sqrt{n}\mathbf{E}(G_j) = \sqrt{n} \int_0^\tau [1 - e_j(t)] P(T_{ij} \geq t,\, C_i \geq t,\, A_i = 1)[\gamma_{1j}(t) - \gamma_{0j}(t)]dt + o_P(1).$$

这是通过对 $\sqrt{n}\bar{G}_j$ 取期望推导得出的：

$$\sqrt{n}\bar{G}_j = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \{A_i - e_j(t)\} dM_{ij}(t) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \{A_i - E_j(t)\} \lambda_{ij}(t)dt + o_P(1).$$

第一部分的期望为 0，因为它是一个鞅。第二部分（补偿子部分）在第6节中给出，取期望后可化为 (8) 和一个余项。

A.4. 综合运用

可以看出，第四项（8）完全是确定性的，并且近似于 $\sqrt{n}\mathbf{E}(G_j)$，这意味着在 Φ_{ij} 中，我们加上并减去 $\sqrt{n}\mathbf{E}(G_j)$。因此，合并所有项后可得：

$$\begin{aligned}
\sqrt{n}\{\bar{G}_j - \mathbf{E}(G_j)\} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \{A_i - e_j(t)\} dM_{ij}(t) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau [Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t)] e_j(t)^2 [\gamma_{1j}(t) - \gamma_{0j}(t)]dt \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau [A_i Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t,\, A_i = 1)] e_j(t)[\gamma_{1j}(t) - \gamma_{0j}(t)]dt \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau [1 - e_j(t)][A_i Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)]dt \\
&+ \sqrt{n} \int_0^\tau [1 - e_j(t)] P(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)]dt \\
&- \sqrt{n} \int_0^\tau [1 - e_j(t)] P(T_{ij} \geq t,\, C_i \geq t,\, A_i = 1)[\gamma_{1j}(t) - \gamma_{0j}(t)]dt \\
&+ o_P(1)
\end{aligned}$$

其中：

$$\begin{aligned}\Phi_{ij} = &\int_0^\tau \{A_i - e_j(t)\} dM_{ij}(t) \\ &+ \int_0^\tau [Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t)] e_j(t)^2 [\gamma_{1j}(t) - \gamma_{0j}(t)] dt \\ &- \int_0^\tau [A_i Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t,\, A_i = 1)] e_j(t) [\gamma_{1j}(t) - \gamma_{0j}(t)] dt \\ &+ \int_0^\tau [1 - e_j(t)][A_i Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t, A_i = 1)][\gamma_{1j}(t) - \gamma_{0j}(t)] dt\end{aligned}$$

利用 $P(T_{ij} \geq t,\, C_i \geq t) = P(T_{ij} \geq t,\, C_i \geq t, A_i = 1) + P(T_{ij} \geq t,\, C_i \geq t, A_i = 0)$ 以及 $Y_{ij}(t) = A_i Y_{ij}(t) + (1 - A_i) Y_{ij}(t)$ 在 $[Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t)]$ 中的第二项，并利用 $(1 - e_j(t))^2 = (1 + e_j(t)^2 - 2e_j(t))$，我们可以简化这个表达式。我们还将鞅改写为计数过程和补偿子部分，因此再次使用 $M_{ij}(t) = N_{ij}(t) - \int_0^t \lambda_{ij}(s) ds$。最终得到如下影响函数表达式：

$$\begin{aligned}\Phi_{ij} \quad = \quad & [A_i - e_j(T_{ij})] I(T_{ij} \leq C_i) - \int_0^\tau [A_i - e_j(t)] Y_{ij}(t) \gamma_{0j}(t) dt \\ &- \int_0^\tau Y_{ij}(t) A_i [1 - e_j(t)][\gamma_{1j}(t) - \gamma_{0j}(t)] dt \\ &+ \int_0^\tau [A_i Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t,\, A_i = 1)][1 - e_j(t)]^2 [\gamma_{1j}(t) - \gamma_{0j}(t)] dt \\ &+ \int_0^\tau [(1 - A_i) Y_{ij}(t) - P(T_{ij} \geq t,\, C_i \geq t,\, A_i = 0)][e_j(t)]^2 [\gamma_{1j}(t) - \gamma_{0j}(t)] dt.\end{aligned}$$

参考文献

Bauer, P., R¨ohmel, J., Maurer, W., and Hothorn, L. (1998). 多剂量试验（包括活性对照）的检验策略。医学统计, 17(18):2133–2146。

Danzer, M. F., Rufibach, K., Beyersmann, J. 和 Schmidt, R.（2025）。在事件驱动的组序贯设计中，采用闭合检验程序检验无进展生存期和总生存期，以耗尽 I 类错误率水平。https://arxiv.org/abs/2512.08658。

Glimm, E., Maurer, W. 和 Bretz, F.（2010）。组序贯试验中多个终点的层级检验。医学统计，29(2)：219–228。

Hougaard, P.（2000）。多变量生存数据分析。生物学与健康统计学。施普林格，纽约；。

Li, H., Wang, J., Luo, X., Grechko, J. 和 Jennison, C.（2018）。用于在主要终点达到显著性后检验次要终点的改进型两阶段组序贯程序。生物统计学杂志，60(5)：893–902。

Lincoff, A. M., Brown-Frandsen, K., Colhoun, H. M., Deanfield, J., Emerson, S. S., Esbjerg, S., Hardt-Lindberg, S., Hovingh, G. K., Kahn, S. E., Kushner, R. F., Lingvay, I., Oral, T. K., Michelsen, M. M., Plutzky, J., Tornøe, C. W., 和 Ryan, D. H. (2023年). 肥胖但无糖尿病患者使用司美格鲁肽与心血管结局《新英格兰医学杂志》, 389(24):2221–2232。

Meller, M., Beyersmann, J., 和 Rufibach, K. (2019年). 无进展生存期与总生存期的联合建模及相关性度量的计算. 医学统计, 38(22):4270–4289。

Mersmann, O. (2024). *microbenchmark*：精确计时函数. R 软件包版本 1.5.0 .

Pipper, C. B., Ritz, C. 和 Bisgaard, H.（2012）。一种在多个模型中确认性评估协变量效应的通用方法。皇家统计学会期刊：*C*辑（应用统计），61(2)：315–326。

Proschan, M. A.（2021）。临床试验中的统计思维。Chapman and Hall/CRC，纽约。

R 核心团队（2024）。*R*：用于统计计算的语言和环境。R 统计计算基金会，奥地利维也纳。

Schemper, M.、Kaider, A.、Wakounig, S. 和 Heinze, G.（2013）。删失条件下双变量失效时间相关性的估计。医学统计，32(27):4781–4790。Senn, S. 和 Bretz, F.（2007）。考虑多个终点时的检验效能和样本量。药物统计学，6(3):161–170。Sugimoto, T.、Sozu, T.、Hamasaki, T. 和 Evans, S. R.（2013）。一种基于对数秩检验的临床试验样本量确定方法，适用于两个共同主要时间-事件终点。*Biostatistics*，14(3):409–421。Tamhane, A. C.、Wu, Y. 和 Mehta, C. R.（2012）。用于检验主要和次要终点的两阶段组序贯程序的自适应扩展（ii）：样本量重新估计。医学统计，31(19):2041–2054。Tsiatis, A. A.（2006）。半参数理论与缺失数据。Springer统计学丛书。Springer纽约，纽约州纽约市，第1版，2006年版。美国食品药品监督管理局（2022）。临床试验中的多个终点：行业指南。药品评价与研究中心（CDER）和生物制品评价与研究中心（CBER）。https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials。van der Vaart, A. W.（1998）。渐近统计学。剑桥大学出版社，剑桥。Westfall, P. 和 Krishen, A.（2001）。最优加权、固定序列和守门多重检验程序。统计规划与推断杂志，99:25–40。