基于分层权重校准病例队列数据的相对风险与纯风险 Cox模型推断

短标题（< 70 字符）：基于分层权重校准病例队列数据的 Cox模型推断

Lola Etievant[1]*、Mitchell H. Gail[1]*

lola.etievant@nih.gov

gailm@mail.nih.gov

[1] 美国国家癌症研究所，癌症流行病学与遗传学部，生物统计学处，9609 Medical Center Drive, Rockville, MD 20850-9780。

* 通讯作者

本研究得到美国国立卫生研究院美国国家癌症研究所癌症流行病学与遗传学部内部研究计划的支持。

目录

摘要.3
1. 引言.3
2. 符号说明.5
3. 分层病例队列.6
4. 设计权重的校准.12
5. 缺失数据.15
6. 软件：CRAN 上的 CaseCohortCoxSurvival.18
7. 模拟.19
8 数据分析.23
9 讨论.25
致谢.27
数据可用性声明.28
参考文献.28
附录 A 分层病例队列的校准权重影响函数 31 附录B 协变量信息缺失且估计了设计抽样第三阶段权重的分层病例队列的影响函数.32

补充材料.33

摘要

病例队列抽样设计仅对病例以及整个队列中的一个随机样本（子队列）获取完整协变量数据。后续发表的文献介绍了使用分层和权重校准来提高Cox模型对数相对风险估计的效率，并且已有一些关于估计纯风险的研究。然而，在医学文献中，这些选项的实例很少，而且我们未能找到目前在线提供的、用于分析这些不同选项的程序。因此，我们提出一种统一的方法，并提供R软件，以促进此类分析。我们使用针对不同抽样设计和分析选项调整的影响函数，并结合将两阶段抽样纳入考虑的方差计算方法。这项工作阐明了Barlow广泛使用的“稳健”方差估计何时适用。相应的R软件CaseCohortCoxSurvival支持在有或无分层和/或权重校准的情况下进行分析，也支持有放回或不放回抽样的子队列抽样。对于分层设计，我们还允许第二阶段数据随机缺失。我们不仅针对Cox模型中的对数相对风险提供推断，也针对累积基线风险和协变量特异性纯风险提供推断。我们希望，这些计算方法和软件能够促进在病例队列研究中更广泛地采用更高效且更具原则性的抽样设计与分析选项。

关键词：病例队列、权重校准、分层、调查抽样、方差估计、两阶段抽样、影响函数。

1. 引言

Prentice[1] 描述了针对响应时间结局的病例队列抽样设计，在这种设计中，研究者会获取所有病例（发生事件者）以及随机抽取的子队列（可能是）的协变量信息。

（包括一些病例）来自整个研究队列。这种抽样设计有两大优势：只需为病例和子队列获取难以测量的协变量，而子队列远小于整个研究队列；此外，子队列数据还可用于多种不同类型的响应时间结局。随后，人们又对这种抽样设计进行了改进和扩展。Barlow[2] 提出了一种广泛使用的、基于影响量平方和的对数相对风险（HR）“稳健”方差估计量。Borgan 等人[3] 表明，分层病例队列设计具有更高的效率；Samuelsen、Ånestad和Skrondaand[4]以及Gray[5] 指出，在不放回抽样的情况下，采用分层时，“稳健”方差估计会高估对数相对风险估计值的方差。Breslow 等人[6,7] 提出了调查权重校准，以提高病例队列相对风险估计的效率。尽管这方面的大量文献聚焦于对数相对风险的估计，但一些作者还研究了累积基线风险以及协变量特异性事件“纯”风险的估计[5,8,9,] 第16章和第17章。

Sharp 等人[10] 指出，来自 24 种主要医学和流行病学期刊的 32 项病例队列研究，在分析和报告方面存在差异。这些分析均未使用权重校准；有些分析在分层数据中使用了不恰当的“稳健”方差估计；而针对协变量信息缺失，则采用了各种不同的方法。我们对后续病例队列研究发表成果所做的非正式回顾也表明，分层、权重校准、对缺失子队列数据采取合理的方法，以及纯风险分析，都未得到充分应用。这可能部分是由于相关方法学文献技术性强、内容繁杂，难以理解，也可能是因为缺少便捷的软件。

为促进改进后的病例队列数据抽样设计和分析方法得到更广泛的应用，我们通过在 Cox 比例风险模型下给出对数相对风险和纯风险的经验影响函数，统一上述各种分析方法。这些影响函数针对

上述各种抽样设计和分析方法进行了调整，方差计算同时考虑了从超总体中对队列进行的第一阶段抽样，以及对子队列进行的第二阶段抽样。我们开发了相应软件，使用户能够便捷地分析有或无分层、采用或不采用权重校准的病例队列数据，并能够处理缺失第二阶段数据的分层病例-队列数据。

我们在第2节介绍符号说明，并在第3节介绍分层病例队列设计的推断，其中包括将只有一个层的非分层设计作为特殊情形。我们在第4节描述权重校准，在第5节介绍缺失的第二阶段数据的处理方法，并在第6节介绍软件。第7节和第8节介绍模拟和数据示例，其中我们研究了分层和权重校准对于相对风险和协变量特异性纯风险的比较效率，以及“稳健”方差估计在进行或不进行校准时的表现。结论性意见见第9节。大部分技术推导和细节见网络附录。

2. 符号说明

我们令 J 表示整个队列中的层数，n_j 表示层 j 中的受试者数，$j \in \{1,\ldots, J\}$。于是，$n = \sum_{j=1}^{J} n_j$ 表示整个队列中的受试者数。进一步令 $T_{i,j}$ 表示层 j 中受试者 i 的事件时间（如果分析采用年龄尺度，则为年龄），令 $C_{i,j}$ 表示层 j 中受试者 i 的删失时间，$i \in \{1, \ldots, n_j\}$，$j \in \{1, \ldots, J\}$。采用研究时间尺度时，层 j 中受试者 i 的风险集指标为 $Y_{i,j}(t) = I(\tilde{T}_{i,j} \geq t)$，其中 $\tilde{T}_{i,j} = \min(T_{ij}, C_{ij})$。采用年龄尺度时，$Y_{i,j}(t) = I(\tilde{T}_{i,j} \geq t > E_{i,j})$，其中 $E_{i,j}$ 是层 j 中受试者 i 的入组年龄。令 τ 表示采用年龄尺度进行分析时的最大随访时间或最大年龄。当 $N_{i,j}(t) = I(T_{i,j} \leq t, T_{ij} \leq C_{ij})$ 表示在 t 之前或恰好在该时间发生了观测到的事件时，$\mathrm{d}N_{i,j}(t)$ 表示

如果层 j 中的个体 i 在时间/年龄 t 发生失效（发生事件）。最后，我们令 $\boldsymbol{X}_{i,\ j}$ 表示层 j 中受试者 i 的 p 维基线协变量向量；$\boldsymbol{X}_{i,\ j}$ 包含层指标或层决定因素。

我们假设失效服从 Cox 比例风险模型，其风险函数为 $\lambda(t)\ = \lambda_0(t)\ \exp(\boldsymbol{\beta}'\boldsymbol{X})$，其中协变量为 $\boldsymbol{X}$，而 $\lambda_0(t)$ 是基线风险函数，即具有 $\boldsymbol{X} = \boldsymbol{0}$ 的个体的风险。我们进一步假设 $\lambda_0(t)$ 在各层之间是齐性的，并令 $\Lambda_0(t) = \int\ \lambda_0(_0^t \quad s \quad s\ )\mathrm{d}$ 表示累积基线风险。

完整队列数据的估计见 Web附录B.1，其中相应的影响函数见 Web附录B.2。

3. 分层病例队列

3.1. 相对风险、累积基线风险和纯风险的估计

我们假设从队列中第 j 层（大小为 $n_{(j)}$）不放回抽取固定数量的个体 $m_{(j)}$，且抽样独立于病例状态 $j \in \{1, \ldots, J\}$。各层之间独立抽样。子队列包括从 J 个层中抽取的所有受试者。此外，我们还对队列中的所有病例进行抽样，其中一些病例可能已纳入子队列。所有这些个体构成分层病例队列，我们也称之为第二阶段样本，因为它是队列的一个子集，而队列被视为来自超级总体的第一阶段样本。我们令 $\xi_{i,\ j}$ 表示第 j 层中个体$j\ i$ 的抽样指标，并令 $w_{i,\ j} =$ 为其已知设计权重：若 i 是第 1 $\quad i \qquad\qquad\qquad\qquad j \qquad$ 层中的非病

例，则为 $\{n_{(j)}\ m_{(j)}$；若为该层中的病例，则为 $\quad i \in \{1, \ldots, n_{(j)}\}$，$j \in \{1, \ldots, J\}$。我们假设 $\boldsymbol{X}$ 中的一些协变量仅在第二阶段样本中测量；

我们称这些为“第二阶段协变量”。队列中所有成员的层指标均已知，且不属于第二阶段协变量。由于我们对所有病例进行抽样，$\xi_{i,\ j} w_{i,\ j} = 1$ 对病例成立。非分层病例队列数据对应于特殊情形 $J = 1$。

对数相对数相对风险 $\boldsymbol{\beta}$ 的估计值通过求解以下估计方程得到

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{j=1}^{J} \sum_{i=1}^{n^{(j)}} \int_t \left\{ \boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1(t;\boldsymbol{\beta})}{S_0(t;\boldsymbol{\beta})} \right\} \mathrm{d}N_{i,j}(t) = 0, \tag{1}$$

with

$$S_0(t;\boldsymbol{\beta}) = \sum_{j=1}^{J} \sum_{k=1}^{n^{(j)}} w_{k,j}\ \xi_{k,j}\ Y_{k,j}(t) \exp\left(\boldsymbol{\beta}' \boldsymbol{X}_{k,j}\right), \tag{2}$$

$$\boldsymbol{S}_1(t;\boldsymbol{\beta}) = \sum_{j=1}^{J} \sum_{k=1}^{n^{(j)}} w_{k,j}\ \xi_{k,j}\ Y_{k,j}(t) \exp\left(\boldsymbol{\beta}' \boldsymbol{X}_{k,j}\right) \boldsymbol{X}_{k,j}, \tag{3}$$

并且我们还定义

$$\boldsymbol{S}_2(t;\boldsymbol{\beta}) = \sum_{j=1}^{J} \sum_{k=1}^{n^{(j)}} w_{k,j} \xi_{k,j} Y_{k,j}(t) \exp\left(\boldsymbol{\beta}' \boldsymbol{X}_{k,j}\right) \boldsymbol{X}_{k,j} \boldsymbol{X}_{k,j}'. \tag{4}$$

令 $\widehat{\boldsymbol{\beta}}$ 表示该解。随后，我们以非参数方式[11] 估计时间 t 处的基线风险点质量：

$$\mathrm{d}\widehat{\Lambda}_0\left(t;\widehat{\boldsymbol{\beta}}\right) \equiv \mathrm{d}\widehat{\Lambda}_0(t) = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n^{(j)}} \mathrm{d}N_{i,j}(t)}{S_0\left(t,\widehat{\boldsymbol{\beta}}\right)}, \tag{5}$$

截至时间 t 的累积基线风险为

$$\widehat{\Lambda}_0\left(t;\widehat{\boldsymbol{\beta}},\hat{\lambda}_0\right) \equiv \widehat{\Lambda}_0(t) = \int_0^t \mathrm{d}\widehat{\Lambda}_0(s), \tag{6}$$

以及特征组合 $\boldsymbol{x}$ 在区间（τ_1， τ_2] by 内的纯协变量特异性风险

$$\hat{\pi}\left(\tau_1,\tau_2;\boldsymbol{x},\widehat{\boldsymbol{\beta}},\mathrm{d}\widehat{\Lambda}_0\right)\equiv\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})=1-\exp\left\{-\int_{\tau_1}^{\tau_2}\exp\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{x}\right)\mathrm{d}\widehat{\Lambda}_0(s)\right\}. \tag{7}$$

对于分层病例队列抽样，Lin[12] 的结果表明，$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ 收敛于均值为 0 的正态分布，其方差对应于第一阶段抽样所得正态随机变量与第二阶段抽样所得正态随机变量之和（另见 Borgan 等人[3]）。同样，对于固定的 t，$n^{\frac{1}{2}}\{\widehat{\Lambda}_0(t)-\Lambda_0(t)\}$ 收敛于一个高斯过程，该过程表示两个相互独立的高斯过程之和。在第 3.3 节和第 4.3 节中，我们将说明如何利用影响函数估计 $(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ 以及固定t时 $\{\widehat{\Lambda}_0(t)\ -\Lambda_0(t)\}$ 的方差和协方差。

3.2. 影响函数

我们将 $\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})$ 定义为受试者 i 在层 j 中对参数集合 $\{\widehat{\boldsymbol{\beta}},\mathrm{d}\widehat{\Lambda}_0(t),\ \widehat{\Lambda}_0(t),\ \hat{\pi}(\tau_1,\tau_2;\ \boldsymbol{x})\}$ 中某个参数 $\widehat{\boldsymbol{\theta}}$ 的影响，其中 $i\in\{1,\dots,n_{(j)}\}$, $j\in\{1,\dots,J\}$。根据这些影响，我们通过线性化[13] 估计参数的协方差矩阵，得到

$$\mathrm{var}\left(\widehat{\boldsymbol{\theta}}\right)\approx\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\theta}}\right)\right\}. \tag{8}$$

遵循 Graubard 和 Fears[14] 以及 Pfeiffer 和 Gail[15] 第 4.6 节，我们在网络附录 C.1 中证明，$\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})$ 的一个估计为 $\xi_{i,j}\ w_{i,j}\ \boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})$，其中

$$\begin{aligned}\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\beta}}\right)=&\left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t\left\{\frac{\boldsymbol{S}_2(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})}-\frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'}{S_0(t;\widehat{\boldsymbol{\beta}})^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1}\left[\int_t\left\{\boldsymbol{X}_{i,j}-\right.\right.\\&\left.\left.\frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\left\{\mathrm{d}N_{i,j}(t)-\frac{Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}\right)\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\mathrm{d}N_{k,l}(t)}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\right],\end{aligned} \tag{9}$$

$$IF_{i,j}^{(2)}\{\mathrm{d}\hat{\Lambda}_0(t)\} = \{S_0(t;\hat{\boldsymbol{\beta}})\}^{-1}\Big\{\mathrm{d}N_{i,j}(t) - \mathrm{d}\hat{\Lambda}_0(t)\boldsymbol{S}_1(t;\hat{\boldsymbol{\beta}})'\boldsymbol{IF}_{i,j}^{(2)}(\hat{\boldsymbol{\beta}}) - \mathrm{d}\hat{\Lambda}_0(t)Y_{i,j}(t)\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\Big\}, \tag{10}$$

$$IF_{i,j}^{(2)}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\hat{\Lambda}_0(t)\right\} = \int_{\tau_1}^{\tau_2} IF_{i,j}^{(2)}\{\mathrm{d}\hat{\Lambda}_0(t)\}, \tag{11}$$

and

$$IF_{i,j}^{(2)}\{\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})\} = \left\{\frac{\partial\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\right\}'\boldsymbol{IF}_{i,j}^{(2)}(\hat{\boldsymbol{\beta}}) + \left[\frac{\partial\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\mathrm{d}\Lambda_0(t)=\mathrm{d}\hat{\Lambda}_0(t)}\right]' IF_{i,j}^{(2)}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\hat{\Lambda}_0(t)\right\}. \tag{12}$$

方程 (9)-(12) 依赖于“第二阶段协变量”。因此，我们在 $\boldsymbol{IF}_{i,j}^{(2)}(\hat{\boldsymbol{\theta}})$ 中使用上标 2。方程 (9)-(12) 是通过将 $\hat{\boldsymbol{\beta}}$ 和 $\hat{\Lambda}_0$ 替换为 $\boldsymbol{\beta}$ 和 Λ_0 而得到的理论影响函数的估计。方程 (8) 涉及理论影响函数，其阶为 $O_P(n^{-\frac{1}{2}})$；但由于估计影响函数与理论影响函数之间的差异阶为 $O_P(n^{-1})$，因此可在方程 (8) 中使用估计影响函数[13,16,17]。下文中，我们用 $\boldsymbol{\Delta}_{i,j}(\hat{\boldsymbol{\theta}})$ 表示估计影响函数。

3.3. 基于影响函数的方差分解与估计

方差 $var(\hat{\boldsymbol{\theta}}) \approx var\{\sum \quad \sum \quad \boldsymbol{\Delta}_{i,j}(\hat{\boldsymbol{\theta}})_{j=1}^{J} \quad {}_{i=1}^{n^{(j)}}\}$ 可分解为

$$\mathrm{var}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\hat{\boldsymbol{\theta}})\,|\boldsymbol{C}_1\right\}\right] + \mathrm{E}\left[\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\hat{\boldsymbol{\theta}})\,|\boldsymbol{C}_1\right\}\right], \tag{13}$$

其中 $\boldsymbol{C}_1$ 表示来自整个队列的信息。从队列中抽样得到总体的部分计入了“超总体”（第一阶段方差成分），而

第二个分量用于说明从队列中抽取子队列的抽样（第二阶段分量

方差）。

我们令 $w_{i,k,j}$ 和 $\sigma_{i,k,j}$ 分别表示 $\mathrm{E}(\xi_{i,j}\xi_{k,j}|\boldsymbol{C}_1)^{-1}$ 和 $\mathrm{cov}(\xi_{i,j},\xi_{k,j}|\boldsymbol{C}_1)$，$i,k \in \{1,\ldots,n(j)\}$，$j \in \{1,\ldots,J\}$；它们将在下文中给出。我们知道，给定 $\boldsymbol{C}_1$ 时，$w_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})$是固定的，且 $\mathrm{E}(\xi_{i,j}w_{i,j}|\boldsymbol{C}_1) = 1$。因此，方差 $\{\sum_{j=1}^{J}\sum_{i=1}^{n(j)}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})\}$ = 方差 $\{\sum_{j=1}^{J}\sum_{i=1}^{n(j)}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\}$ +E $\{\sum_{j=1}^{J}\sum_{i=1}^{n(j)}\sum_{k=1}^{n(j)}\sigma_{i,k,j}w_{i,j,k,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(2)}(\widehat{\boldsymbol{\theta}})\}$.

由于只有当层 j 中的个体 i 和 k 均纳入第二阶段样本时，才能计算 $\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})'$ 和 $\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(2)}(\widehat{\boldsymbol{\theta}})'$，因此我们使用“边际”和“联合”抽样设计权重 $w_{i,j}$ 和 $w_{i,k,j}$ 对第二阶段样本中个体的贡献加权，以估计方差$(\widehat{\boldsymbol{\theta}})$，方法是

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n(j)}\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})' + \sum_{j=1}^{J}\sum_{i=1}^{n(j)}\sum_{k=1}^{n(j)}w_{i,k,j}\,\sigma_{i,k,j}\,w_{i,j}\,w_{k,j}\,\xi_{i,j}\,\xi_{k,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(2)}(\widehat{\boldsymbol{\theta}})'. \tag{14}$$

参见网络附录 C.2，其中与 Samuelsen 等人对 var$(\widehat{\boldsymbol{\beta}})$ 的极为相似估计进行了比较[4]。

按照 Barlow[2]，“稳健”方差估计将会是

$$\sum_{j=1}^{J}\sum_{i=1}^{n(j)}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})' = \sum_{j=1}^{J}\sum_{i=1}^{n(j)}\xi_{i,j}\,w_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})'. \tag{15}$$

对于分层数据，方程 (15) 通常过于庞大（另见第 7 节及网络附录 E 中的网络表格 3–20）。方程 (15) 减去方程 (14) 是

$$\frac{1}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\xi_{i,j}w_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})' + \tag{16}$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\k\neq i}}^{n^{(j)}}w_{i,k,j}\sigma_{i,k,j}\,w_{i,j}w_{k,j}\xi_{i,j}\,\xi_{k,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(2)}(\widehat{\boldsymbol{\theta}})'.$$

由于我们在每一层中都进行不放回抽样，因此如果层 j 中的个体 i 和 k 均为非病例，则有 $w_{i,k,j} = {}^{n}(_j)(n(_j)-1)\,{}^{m}(_j)(m(_j)-1)$；否则有 $w_{i,k,j} = w_{i,j} \times w_{k,j}$，其中 $i,k \in \{1,\dots,n(_j)\}$，$k \neq i, j \in \{1,\dots,J\}$。回顾一下，如果层 j 中的个体 i 为非病例，则 $w_{i,i,j} = w_{i,j} = {}^{n}(_j)\,{}^{m}(_j)$；如果层 j 中的个体 i 为病例，则 $w_{i,j} = 1$。于是，如果层 j 中的个体 i 和 k 均为非病例，则 $\sigma_{i,k,j} = {}^{m}(_j)\,{}^{n}(_j)\,\frac{{}^{m}{}_{j}^{-1}}{n}(_j)\,(\,)_{-1} - ({}^{m}(_j)\,{}^{n}(_j)\,)^2$；否则 $\sigma_{i,k,j} = 0$，其中 $i,k \in \{1,\dots,n(_j)\}$，$k \neq i, j \in \{1,\dots,J\}$。类似地，如果层 j 中的个体 i 为非病例，则 $\sigma_{i,i,j} \equiv \sigma_{i,j} = {}^{m}(_j)\,{}^{n}(_j)\,(1 - {}^{m}(_j)\,{}^{n}(_j)\,)$；否则 $\sigma_{i,j} = 0$。因此，在方程 (14) 中，只有抽样得到的非病例会对方差的第二阶段分量作出贡献。对于有放回抽样（即伯努利抽样），个体之间的抽样彼此独立。此时 $w_{i,k,j} = w_{i,j} \times w_{k,j}$，且对于层 j 中任意一对不同个体 (i,k)，有 $\sigma_{i,k,j} = 0$，其中 $i,k \in \{1,\dots,n(_j)\}$，$k \neq i$，$j \in \{1,\dots,J\}$。在这种情况下，方程 (15)) 中“稳健”方差估计与方程 (14)) 之间的差异简化为 $\frac{1}{n-1}\sum_{j=1}^{J}\sum_{i=1}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})^{n\quad\quad\quad

4. 设计权重的校准

4.1. 校准与辅助变量的选择

Breslow等人[6,7] 倡导采用“权重校准”来提高病例队列研究的效率。首先，确定与$\widehat{\boldsymbol{\theta}}$ 的影响函数高度相关且在整个队列中均已知的辅助变量。然后调整设计权重，以获得接近设计权重的校准权重；使用这些校准权重时，第一阶段样本中观测到的辅助变量总和等于第二阶段样本中的加权总和。为获得辅助变量，我们遵循 Breslow等人[6] 和 Shin等[18]的方法。首先，在第二阶段样本中使用加权回归，根据第一阶段数据估计第二阶段协变量的期望值。利用这些期望值，为队列中的所有成员插补第二阶段协变量，包括那些已测量第二阶段协变量的成员。用于插补的第一阶段数据可能包括$\boldsymbol{X}$ 中的协变量，以及在所有队列成员中测量的第二阶段协变量的第一阶段代理变量。辅助变量包括：(i) 基于插补后的队列数据，通过Cox模型估计的对数相对风险参数的影响函数；以及(ii)总随访时间（在拟估计纯风险的时间区间内）与插补后的队列数据的估计相对风险的乘积，其中，对数相对风险参数是基于病例队列数据和使用 (i) 校准的权重，通过 Cox模型估计的。为使权重标准化，我们还对一个恒等于 1 的变量进行校准。更多细节见网络附录D.1；另见Breslow 等人[6] 和 Shin等[18]。已有研究为$\widehat{\Lambda}_0(t)$[8], 提出了其他辅助变量，但在未报告的模拟中，Shin等提出的方法表现更好；另见网络附录 D.3。

我们令$\boldsymbol{A}_{i,j}$ 为个体 i 在层 j中的 q 辅助变量向量，其校准权重为 $w_{i,j}^{*} = w_{i,j}\exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j})$, $i \in \{1, \ldots, n_{(j)}\}$, $j \in \{1, \ldots, J\}$, 这些权重通过求解$\sum_{j=1}^{J}\sum_{i=1}^{n}\{\xi_{ij}w_{ij}\left[\boldsymbol{\eta}'\boldsymbol{A}_{ij}\left(\boldsymbol{A}_{ij}-\boldsymbol{A}_{ij}\right)\right]\exp(\cdot)\}_{(j)} = 0$得到，其中求解对象为 $\widehat{\boldsymbol{\eta}}$。参见网络附录D.1。

4.2. 使用校准权重估计相对风险、累积基线风险和纯风险

$\boldsymbol{\beta}$ 的估计值满足 $\boldsymbol{U}^{*}(\boldsymbol{\beta}) = \sum_{j=1}^{J}\sum_{i=1}^{n_j} w_{ij}^{*}\left[\boldsymbol{X}_{ij} - \frac{\boldsymbol{S}_1^{*}(t,\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})}{S_0^{*}(t,\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})}\right]N_{ij}(t) = 0$, 其中$S_0^{*}(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})$、$\boldsymbol{S}_1^{*}(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})$和$\boldsymbol{S}_2^{*}(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})$由方程 (2)–(4) 得出，其中以$w_{k,j}^{*}$ 替换$w_{k,j}$。令 $\widehat{\boldsymbol{\beta}}^{*}(\widehat{\boldsymbol{\eta}}) \equiv \widehat{\boldsymbol{\beta}}^{*}$，我们估计时间$t$处的基线风险点质量 $\mathrm{d}\widehat{\Lambda}_0^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*}) \equiv \mathrm{d}\widehat{\Lambda}_0^{*}(t)$、截至时间$t$的累积基线风险 $\widehat{\Lambda}_0^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*}) \equiv \widehat{\Lambda}_0^{*}(t)$，以及特征组合$\boldsymbol{x}$ 在区间 $(\tau_1,\tau_2]$内的纯风险 $\hat{\pi}^{*}(\tau_1,\tau_2;\boldsymbol{x},\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*},\mathrm{d}\widehat{\Lambda}_0^{*}) \equiv \hat{\pi}^{*}(\tau_1,\tau_2;\boldsymbol{x})$；这些量由方程 (5)–(7) 得出，其中以 $S_0^{*}(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})$ 和 $\widehat{\boldsymbol{\beta}}^{*}$ 替换 $S_0(t,\widehat{\boldsymbol{\beta}})$ 和 $\widehat{\boldsymbol{\beta}}$。由于所有队列成员的事件时间均已知[18–20]，我们不对 Breslow估计量分子中的病例权重进行校准。

4.3. 基于影响函数的方差估计

我们令$\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}}^{*})$表示层$j$ 中个体i 对参数集合$\{\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*},\mathrm{d}\widehat{\Lambda}_0^{*}(t),\widehat{\Lambda}_0^{*}(t),\hat{\pi}^{*}(\tau_1,\tau_2;\boldsymbol{x})\}$中某个参数 $\widehat{\boldsymbol{\theta}}^{*}$ 的影响，其中 $i \in \{1, \ldots, n_{(j)}\}$，$j \in \{1, \ldots, J\}$，并使用$\mathrm{var}(\widehat{\boldsymbol{\theta}}^{*}) \approx \mathrm{var}\{\sum_{j=1}^{J}\sum_{i=1}^{n_{(j)}}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}}^{*})\}$。遵循Shin等，我们可以证明$\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}}^{*}) = \boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*}) + \xi_{i,j}w_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})$。上标1表示$\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})$仅取决于在所有队列成员中测量的变量。如果层 j 中的个体 i 不在第二阶段样本中，则$\xi_{i,j}w_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})$为零，但此类个体会影响$\widehat{\boldsymbol{\eta}}$，进而通过

$\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})$。$\boldsymbol{IF}_{i}$的显式形式$_{,j}^{(s)}(\widehat{\boldsymbol{\theta}}^{*})$、$s \in\{1,2\}$，见附录 A、网络附录 A.2，并在网络附录 D.2 中推导。

因为 $\boldsymbol{IF}_{i,\ j}^{(1)}$ $(\widehat{\boldsymbol{\theta}})$ 在给定 $\boldsymbol{C}_1$ 的条件下是固定的，类似于方程 (13) 的分解可得 ds

$$\begin{aligned}\operatorname{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}}^{*})\right\} &= \operatorname{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})+\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})\right\}+\\ &\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}\,w_{i,j}\,w_{k,j}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{k,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'\right\},\end{aligned} \tag{17}$$

可通过以下方式估计

$$\begin{aligned}&\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left\{\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})'+2\,\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'+\right.\\ &\left.\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'\right\}+\\ &\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}w_{i,k,j}\,\sigma_{i,k,j}\,\xi_{i,j}\,\xi_{k,j}\,w_{i,j}\,w_{k,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{k,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'.\end{aligned} \tag{18}$$

最后，稳健方差估计[2] 为

$$\begin{aligned}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}}^{*})' &= \sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left\{\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})'+\right.\\ &\left.2\,\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'+\xi_{i,j}\,w_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'\right\}.\end{aligned} \tag{19}$$

以及方程（18）和（19）之间的差异是

$$\begin{aligned}&\frac{1}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left\{\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})'+2\,\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'+\right.\\ &\left.\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'\right\}+\\ &\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\ k\neq i}}^{n^{(j)}}w_{i,k,j}\,\sigma_{i,k,j}\,w_{i,j}\,w_{k,j}\,\xi_{i,j}\,\xi_{k,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})\boldsymbol{IF}_{k,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})'.\end{aligned} \tag{20}$$

对于层 j 中的个体 i 和 k，满足 $\sigma_{i,k\ j}$,且 $\sigma_{i,j}$ 非零（即非病例），$\xi_{i,j}\ w_{i,j}\ \boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})$ 和 $\xi_{k,j}\ w_{k,j}\ \boldsymbol{IF}_{k,\ j}^{(2)}(\widehat{\boldsymbol{\theta}}^{*})$是基于辅助变量进行加权线性回归得到的加权残差；参见网络附录 D.3。对于用于校准的良好辅助变量，可以预期：（i）方差的第二阶段分量，从而总方差会更小；以及（ii）公式（20）中的差异会小于公式（16）中的差异；另请参见 Borgan 等人的第17章。

5. 缺失数据

5.1. 符号说明

第二阶段的个体可能缺少协变量信息。例如，第二阶段个体储存的血液样本可能已被使用或遗失。我们假设此类协变量为随机缺失，并将协变量数据完整的个体集合视为第三阶段样本。更具体地说，令 $V_{i,j}$ 为层 j中受试者i 的第三阶段抽样指标，$i \in \{1, \ldots, n_{(j)}\}, j \in \{1, \ldots, J\}$；我们假设伯努利指标 $V_{i,j}$ 相互独立，且与第二阶段指标 $\xi_{i,j}$独立。令 $w_{i,j}^{(3)} \equiv \frac{1}{\pi_{i}^{\ 3},j}$ $_{0}$为第三阶段设计权重，其中 $\pi_{i,j}^{(3)}$ 为第三阶段设计抽样概率。层 j 中受试者 i 的总体抽样设计权重为 $w_{i,j} = w_{i,j}^{(2)} \times w_{i,j}^{(3)}$。

第三阶段抽样概率可能在总体的$J^{(3)}$ 互斥且完备的子集（第三阶段层）中有所不同，这些子集不一定与J 第二阶段层一致。例如，病例的缺失概率可能不同于非病例。然而，我们按照第3节中的方式对队列成员进行编号。网络附录 H 描述了第三阶段

抽样概率已知。然而，$\pi_{i,j}^{(3)}$ 通常未知，需要进行估计（第 5.2 节）。

5.2 权重估计

当$\pi_{i,j}^{(3)}$ 未知时，可以如下估计$w_{i,j}^{(3)}$，$i \in \{1, \ldots, n(j)\}$，$j \in \{1, \ldots, J\}$。对于第二阶段样本的所有成员，$V_{i,j}$ 均已知，并令 $\boldsymbol{B}_{i,j}$ 为一个 $J^{(3)} \times 1$ 指示变量向量：如果（第二阶段）层j中的受试者i属于相应的第三阶段层，则取值为1，否则取值为 0。令$\exp(\tilde{\boldsymbol{\gamma}})$为通过求解估计方程$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\exp(\boldsymbol{\gamma}'\boldsymbol{B}_{i,j})\xi_{i,j}V_{i,j}\boldsymbol{B}_{i,j} = 0$而获得的 $J^{(3)}$ 第三阶段抽样权重向量。例如，如果第三阶段抽样按病例状态分层，则对于受试者 i 在层 j 中为非病例的情况，我们使用权重 $\tilde{w}_{i,j}^{(3)} = \frac{\sum_{l=1}^{J}\sum_{k=1}^{n_l}\xi_{k,l}V_{k,l}}{\sum_{l=1}^{J}\sum_{k=1}^{n_l}\xi_{k,l}}$，非病例，病例，$\tilde{w}_{i,j}^{(3)} = \frac{\sum_{l=1}^{J}\sum_{k=1}^{n_l}\xi_{k,l}V_{k,l}}{\sum_{l=1}^{J}\sum_{k=1}^{n_l}\xi_{k,l}}$；对于受试者$i$ 在层 j 中为病例的情况，使用 $\tilde{w}_{i,j}^{(3)} =$病例$\tilde{w}_{i,j}^{(3)}\left(1 - \frac{1}{\tilde{w}_{i,j}^{(3)}}\right)$ 估计方差$(V_{i,j}) \equiv \sigma_{i,j}^{(3)}$。

5.3 相对风险、累积基线风险和纯风险的估计

我们通过求解估计方程$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\int_t V_{ij}\tilde{w}_{ij}^{(3)}\left(\boldsymbol{X}_{ij} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})}\right)N_{ij}(t) = 0$，得到对数相对风险估计值$\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\gamma}}) \equiv \tilde{\boldsymbol{\beta}}$，其中$\boldsymbol{\beta}$。我们令$\tilde{w}_{k,j} = w_{i,j}^{(2)} \times \tilde{w}_{i,j}^{(3)}$，并通过在方程(2)-(4) 中分别用$\xi_{k,j}V_{k,j}$替代$\xi_{k,j}$，以及用$\tilde{w}_{k,j}$替代$w_{k,j}$，计算$\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})$、$\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})$和$\tilde{\boldsymbol{S}}_2(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})$。我们估计时间 t处的基线风险点质量，$d\tilde{\Lambda}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}}) \equiv d\tilde{\Lambda}_0(t)$；截至时间 t的累积基线风险，$\tilde{\Lambda}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}}) \equiv \tilde{\Lambda}_0(t)$；以及特征组合 $\boldsymbol{x}$ 的纯风险，在

区间（τ_1，τ_2]，$\tilde{\pi}$（τ_1，τ_2；$\boldsymbol{x}$，$\widetilde{\boldsymbol{\gamma}}$，$\widetilde{\boldsymbol{\beta}}$，$\mathrm{d}\widetilde{\Lambda}_0$）$\equiv\tilde{\pi}$（$\tau_1$，$\tau_2$；$\boldsymbol{x}$），由方程（5）-（7）得出，其中$\tilde{S}_0$（$t$；$\widetilde{\boldsymbol{\gamma}}$，$\boldsymbol{\beta}$）和$\widetilde{\boldsymbol{\beta}}$替换$S_0$（$t$，$\widehat{\boldsymbol{\beta}}$）以及$\widehat{\boldsymbol{\beta}}$。

如果在时间 t 有一例缺失协变量数据的病例，而此时第三阶段样本中没有其他成员处于风险中，那么对累积基线风险的 Breslow估计的贡献未定义。一种做法是通过增大 τ_1 或减小 τ_2 来限制风险投影区间，以避开这些时间点。如果这类时间点数量很少，我们建议在所有计算中忽略它们。

5.4 影响函数

令 $\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})$ 表示层 j 中受试者 i 对参数集合 $\{\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}},\mathrm{d}\widetilde{\Lambda}_0(t),\ \widetilde{\Lambda}_0(t),\ \tilde{\pi}(\tau_1,\tau_2;\ \boldsymbol{x})\}$ 中某个参数 $\widetilde{\boldsymbol{\theta}}$ 的影响，其中 $i\in\{1,\ldots,n(j)\}$，$j\in\{1,\ldots,J\}$。我们可以证明，$\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})=\xi_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\xi_{i,j}V_{i,j}\exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})$。$\boldsymbol{IF}_{i,j}^{(s)}(\widetilde{\boldsymbol{\theta}})$（$s\in\{2,3\}$）的显式形式见附录 B、Web附录A.3，并在 Web附录G.1 中推导。上标 3 强调 $\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})$ 涉及的变量仅在第三阶段样本中的个体上测量。因此，如果层 j 中的个体 i 不属于第三阶段样本，则 $\xi_{i,j}V_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})$ 为零。不过，由于该个体用于估计第三阶段抽样权重，因此他/她会通过其对 $\widetilde{\boldsymbol{\gamma}}$ 的影响（经由 $\xi_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})$）影响 $\widetilde{\boldsymbol{\theta}}$ 。

5.5 方差分解和基于影响函数的估计

由 $\mathrm{var}(\widetilde{\boldsymbol{\theta}})\approx\mathrm{var}\{\sum_{j=1}^{J}\sum_{i=1}^{n(j)}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\}$，$\widetilde{\boldsymbol{\theta}}\in\{\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}},\mathrm{d}\widetilde{\Lambda}_0(t),\ \widetilde{\Lambda}_0(t),\ \tilde{\pi}(\tau_1,\tau_2;\ \boldsymbol{x})\}$，方差 $\mathrm{var}(\widetilde{\boldsymbol{\theta}})$ 可分解为

$$\begin{aligned}&\mathrm{var}\left(\mathrm{E}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right)+\\&\mathrm{E}\left(\mathrm{var}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right)+\\&\mathrm{E}\left(\mathrm{E}\left[\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right),\end{aligned}\tag{21}$$

其中，C_1 表示来自整个队列的信息，C_2 表示来自第二阶段样本的信息。这三项分别对应于从“超总体”中抽样、从队列中抽取子队列，以及从第二阶段样本中抽取第三阶段样本。

我们通过以下方式估计 var($\tilde{\boldsymbol{\theta}}$)

$$\begin{aligned}&\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\Big\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\theta}})'+2\,\xi_{i,j}\,V_{i,j}\,\tilde{w}_{i,j}^{(3)}\,\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})'+\\&\xi_{i,j}\,V_{i,j}\,\tilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})'\Big\}+\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(2)}w_{i,j}^{(2)}\Big\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\theta}})'+\\&2\,\xi_{i,j}\,V_{i,j}\,\tilde{w}_{i,j}^{(3)}\,\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})'+\xi_{i,j}\,V_{i,j}\,\tilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})'\Big\}+\\&\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\k\neq i}}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,k,j}^{(2)}\Big\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\theta}})+\xi_{i,j}\,V_{i,j}\,\tilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})\Big\}\Big\{\xi_{k,j}\,\boldsymbol{IF}_{k,j}^{(2)}(\tilde{\boldsymbol{\theta}})+\\&\xi_{k,j}\,V_{k,j}\,\tilde{w}_{k,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})\Big\}'+\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\tilde{\sigma}_{i,j}^{(3)}\tilde{w}_{i,j}^{(3)}\,\xi_{i,j}\,V_{i,j}\,\tilde{w}_{i,j}^{(3)}\,\tilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})\,\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\theta}})'.\end{aligned}\tag{22}$$

详情请参见网络附录 G.2。第三阶段权重的估计通过 $\xi_{i,\ j}\boldsymbol{IF}_{i,\ j}^{(2)}(\tilde{\boldsymbol{\theta}})$ 的一部分计入方差。

6. 软件：CRAN 上的 CaseCohortCoxSurvival

可从 survival 软件获取的“Dfbetas”近似表示影响函数，可用于估计基于未分层病例-队列数据[21], 的 $\widehat{\boldsymbol{\beta}}$ 方差；此外，还有一段类似的代码，用于

估计分层病例-队列抽样设计[4]中 $\widehat{\boldsymbol{\beta}}$ 的方差（对应于 Borgan 等人[3]提出的估计量 II）。CRAN survival 软件包[22] 中的 cch 函数可处理估计量 I，以及 Borgan 等人[3]提出的估计量 II。CRAN cchs 软件包[23] 是为 Borgan 等人[3], 提出的估计量 III 编写的，但我们不考虑效率低于估计量 II 的估计量 III。CRAN survey 软件包[24] 中的 twophase 函数可根据第二阶段样本估计 $\boldsymbol{\beta}$ 及其方差，因此也可用于未分层或分层病例-队列数据。以往的论文没有讨论纯风险。已有文献提供了用于分层病例-队列研究和纯风险[25], 的估计量 III 的 SAS 代码，但除 cchs[23], 、survival[22] 和 survey[24]，外，我们未能找到仍在维护的在线程序，即使在原始文章提及的网站上也没有。更通用的调查软件除了分层[24], 外，还可进行权重校准，但介绍如何使用这些通用程序估计 $\boldsymbol{\beta}$ 的 R 代码已不再在线[6,7], ， CRAN NestedCohort 软件包[26]也已无法获取。因此，我们需要便捷的软件，以便处理分层、权重校准和缺失的第二阶段数据。

我们创建了一个名为 CaseCohortCoxSurvival 的 CRAN 软件包，以便于开展此类分析。相关详情将在其他地方介绍；第 8 节的表 4 给出了一个示例脚本，用于分析伊朗戈勒斯坦的死亡率数据。

7. 模拟

7.1 模拟设计

我们比较了第 3.3 节和第 4.3 节中的方法在模拟队列中估计对数相对风险方差及纯风险估计方差的表现。我们还评估了使用校准权重而非设计权重所带来的精确度提升。我们考虑了多种

情景，这些情景由下文所述模型以及网络附录 E.1 的网络表 1 和 2 中的参数值定义。

我们模拟了样本量为 $n \in \{5 \times 10^3, 104\}$ 的队列，并以研究时间作为时间尺度。我们模拟了三个协变量 $\boldsymbol{X} = (X_1, X_2, X_3)'$：$X_1 \sim \mathcal{N}(0,1)$，$X_2$ 取值为 $\Omega_{X_2} = \{0,1,2\}$，相应概率为 $\{p_{0|X_1}, p_{1|X_1}, p_{2|X_1}\}$，详见网络附录 E.1 中的网络表格 1；以及 $X_3 \sim \mathcal{N}(\alpha_1 \times X_1 + \alpha_2 \times X_2, 1)$，其中 $\mathcal{N}(a, b)$ 表示均值为 a 、方差为 b 的正态分布。我们根据 Cox 比例模型模拟失效时间 T，其风险为 $\lambda(t; \boldsymbol{X}) = \lambda_0 \times \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$，其中基线风险 $\lambda_0 = \frac{p_Y}{\mathrm{E}\{\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)\} \times 10}$，$p_Y \in \{0.02, 0.05, 0.1\}$，为常数，其取值使得 10 年纯生存概率约为 98%、95% 或 90%。参数 α_1、α_2、β_1、β_2 和 β_3 列于网络附录 E.1 中的网络表格 2。队列进入时间 E 在前 5 年内服从均匀分布；我们假设因失访而被删失的时间 C 服从风险为 $-\frac{1}{10}\log(0.98)$ 的指数分布，对应于 10 年内失访概率为 2%。我们假设 T、E 和 C 相互独立。令 $\tilde{T} = \min(T, 10 - E, C)$ 为观测时间。因此，时间区间 $(\tau_1, \tau_2]$ 内的总随访时间为 $\max\{0, \min(\tilde{T}, \tau_2) - \tau_1\}$。

我们从由 $W = 0 \times I(X_1 \geq 0, X_2 = 0) + 1 \times I(X_1 < 0, X_2 < 2) + 2 \times I(X_1 \geq 0, X_2 > 0) + 3 \times I(X_1 < 0, X_2 = 2)$ 定义的四个层中进行抽样，其中层“0”为低风险，层“1”和“2”均为中风险，层“3”为高风险，$I(\,)$ 为示性函数。我们在各层中独立地不放回抽取固定数量的个体 $m^{(j)}$，$j \in \{0,1,2,3\}$。$m^{(j)}$ 取决于各层中预期的失效数和个体数，具体通过 $m^{(j)} = \left\lfloor \frac{\lambda_0 \times 10 \times \mathrm{E}\{\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) | W = j\}}{1 - \lambda_0 \times 10 \times \mathrm{E}\{\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) | W = j\}} \times \mathrm{E}(n^{(j)}) \times K + \frac{1}{2} \right\rfloor$ 表示，其中 $K \in \{2, 4\}$ 是非-

每个病例中我们希望抽取的病例数为 $\lfloor \rfloor$，其中 $\lfloor \rfloor$ 是向下取整函数。第二阶段样本由子队列和未抽样病例组成。设计权重按照第3.3节计算。权重校准针对第4.1节提出的辅助变量进行，其中整个队列中的协变量根据三个预测模型进行插补，使用 W 和代理变量 $\tilde{\boldsymbol{X}} =$（$\tilde{X}_1$、$\tilde{X}_2$、$\tilde{X}_3$）′。更准确地说，我们假设 $\tilde{\boldsymbol{X}}$ 在所有队列成员中均已测量，并使用 $\tilde{X}_1 = X_1 + \varepsilon_1$、$\tilde{X}_3 = X_3 + \varepsilon_3$，其中 ε_1 和 ε_3 独立地服从正态分布 $\mathcal{N}(0, 0.75^2)$，因此 $\text{corr}(\tilde{X}_1, X_1) = \text{corr}(\tilde{X}_3, X_3) = 0$。8。我们使用 $\tilde{X}_2 = \{2 \times I(X_2 = 0) + 1 \times I(X_2 = 2) + 0 \times I(X_2 = 1)\} \times I(\varepsilon_2 \in [0,0.2[) + \{2 \times I(X_2 = 1) + 1 \times I(X_2 = 0) + 0 \times I(X_2 = 2)\} \times I(\varepsilon_2 \in]0.9,1]) + X_2 \times I(\varepsilon_2 \in [0.1,0.9])$，其中 ε_2 在 $[0,1]$，上服从均匀分布，使得对于约80%的观测，$\tilde{X}_2$ 等于 X_2。为建立第二阶段协变量的插补方程，我们分别对 X_1 关于 $\tilde{X}_1$ 和 W、X_3 关于 $\tilde{X}_1$ 和 $\tilde{X}_3$ 进行加权线性回归，并对 X_2 关于 $\tilde{X}_1$、$\tilde{X}_2$ 和 W 进行加权多项逻辑回归。对于每个情景，我们模拟了5,000个队列。我们估计了时间区间 $(\tau_1, \tau_2] =$（0,8]内的对数相对风险 $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ 和纯风险 $\pi(\tau_1, \tau_2; \boldsymbol{x})$，并针对协变量特征组合 $\boldsymbol{x} \in \{(-1, 1, -0.6)', (1, -1, 0.6)', (1, 1, 0.6)'\}$，使用以下抽样设计和分析方法：采用设计权重的分层病例队列（SCC）；采用校准权重的分层病例队列（SCC.Calib）；采用设计权重的非分层病例队列（USCC）；以及采用校准权重的非分层病例队列（USCC.Calib）。随后我们估计了它们的方差。对于每个模拟实现，我们根据方程（14）获得了SCC的方差估计 $\hat{V}$，并根据方程（15）获得了稳健方差估计（$\hat{V}_{\text{Robust}}$）。对于SCC.Calib，我们在方程（18）中使用 $\hat{V}$，并在方程（19）中使用 $\hat{V}_{\text{Robust}}$。对于USCC和USCC.Calib，我们使用方程（14）、（15）、（18）和（19）中的方差估计，其中 $J = 1$。作为参考，我们还使用整个队列（队列）的数据估计了这些参数。

7.2 模拟结果

$n = 10{,}000$, $p_Y = 0.02$且$K = 2$ 的情景模拟结果显示于图1、表1和表2中；其他情景请参见网络附录 E.2 中的网页表3至表20。对于采用分层设计的大多数对数相对风险和纯风险，以及采用未分层设计的纯风险，稳健方差公式高估了方差（表1），并产生了超名义置信区间覆盖率（表2）。权重校准导致更小的方差，并且在使用校准权重时，稳健方差估计大致有效（表1和表2）。由于正确考虑了抽样特征，方程（14））和方程（18）中的方差估计（分别适用于设计权重或校准权重）在所有抽样设计中均产生了适当的覆盖率（表2）；但当$\boldsymbol{x} = (-1, 1, -0.6)'$时的$\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$除外，在该情形下，全队列分析也具有超名义覆盖率。如图1所示，分层和/或权重校准提高了效率。此外，采用权重校准的未分层病例-队列设计的效率几乎与采用权重校准的分层病例队列设计相当，并且两者的效率均显著高于采用设计权重的分析。使用设计权重时，与未分层病例队列设计相比，分层提高了效率。

下面是几点说明。首先，$\boldsymbol{X}$ 中的三个协变量仅在第二阶段样本中进行了测量。对于整个队列中均可获得的协变量，效率增益会更大（见第8 节）。对于较弱的代理变量，效率增益会更有限，并且稳健方差估计可能过大（见网络附录 E.5）。其次，我们对纯风险进行了对数转换，以基于渐近正态理论提高覆盖率。第三，一些作者使用后分层权重而非设计权重，将病例单独设为一层，并将病例从包含非病例的层中排除[3,4]。这种方法提高了估计的精确度，并且

与使用抽样设计权重相比，SCC 和 USCC 的影响可忽略不计（方差比不超过 1.01）（网络附录 E.4）。

关于缺失的第二阶段数据的模拟显示，第5.5节中的公式（22）以及一个忽略估计权重变异性的较简单公式（网络附录 H.3）对数相对风险和纯风险的置信区间覆盖率均达到名义水平（网络附录 I）；但在未报告的、缺失比例更高的模拟中，后者高估了方差。因此，我们建议使用公式（22），CaseCohortCoxSurvival 中即采用该公式。

8 数据分析

Golestan队列纳入了49,819名年龄为36-81岁的个体，这些个体于2003-2009[27]年招募。为减少计算量，我们随机抽取了$n = 30{,}000$名个体，并对这一子集进行了分析。我们采用年龄尺度，并假设使用Cox比例风险模型，根据基线变量预测死亡率：$x_1 =$ 男性指标、$x_2 =$ 财富评分、$x_3 =$ 既往吸烟者（香烟、鼻烟或鸦片）指标、$x_4 =$ 当前吸烟者指标、$x_5 =$ 发病率指标、$x_6 = x_1x_3$ 以及$x_7 = x_1x_4$。我们以“从不吸烟者”作为参照类别；发病率为二元指标：如果个体在基线时患有以下至少一种疾病，则取值为1：心血管疾病、脑血管意外、高血压、糖尿病、慢性阻塞性肺疾病、结核病和癌症。财富评分根据住房所有权以及家用电器的数量和类型等信息计算得出；参见Islami等人（2009）。我们还估计了区间$(\tau_1, \tau_2] = (52{,}66]$ 以及协变量特征组合$\boldsymbol{x} \in \{(0, -0.4{,}0{,}1, \mathbf{0}_3)', (0{,}0.4{,}0{,}1, \mathbf{0}_3)', (\mathbf{0}_4, 1, \mathbf{0}_2)', \mathbf{0}_7'\}$内的纯风险，其中$\mathbf{0}_a$ 为由零组成的$a \times 1$ 向量；例如，$(0, -0.4{,}0{,}1, \mathbf{0}_3)'$对应于财富评分较低的当前吸烟女性这一特征组合，而$(\mathbf{0}_4, 1, \mathbf{0}_2)'$对应于患有发病率的从不吸烟女性

（基线时）。我们假设队列中每个人的年龄、性别、吸烟状态、发病率、居住地（城市、农村）、族群（Turkmen、其他）、婚姻状况（未婚、已婚、丧偶、离婚/分居、其他）、教育程度（无、少于5^{th}、 6th-8th、 9th-12th、大学）、社会经济状况（低、低至中等、中等至高、高）、死亡状态和随访时间均为已知，但财富评分仅适用于第二阶段的个体。我们分别在由性别（男性、女性）、居住地和四个基线年龄类别（[36,45）、[45,50）、[50,55）和[55,81]）定义的16个层中抽取了33、42、192、246、57、62、313、382、82、86、391、477、565、770、1934和2949名个体，因此预计每个层中约有一名非病例对应一名病例。我们使用SCC、SCC.Calib、USCC和USCC.Calib估计对数相对风险和纯风险（参见第5节中的符号说明和分析方法）。我们将性别、社会经济状况、基线年龄、婚姻状况、族群、教育程度和居住地作为代理变量，为整个队列插补财富评分，然后校准设计权重。我们还分析了整个队列（$n = 30{,}000$名个体）。

表 3 分别展示了对数相对风险和纯风险参数的估计结果。使用设计权重时，对于协变量 x_1 和 x_2 的对数相对风险、分层设计中的所有纯风险，以及非分层设计中具有特征组合 $\boldsymbol{x} \in \{(0, -0.4, 0, 1, \mathbf{0}_3)', (\mathbf{0}_4, 1, \mathbf{0}_2)', \mathbf{0}_7'\}$ 的纯风险，稳健方差估计均较大。在分层设计中，对于 7 个对数相对风险参数中的 5 个，$\hat{V}_{\text{Robust}}$ 与 $\hat{V}$ 吻合良好，这可能是因为分层仅基于 x_1。

权重校准提高了效率，稳健方差估计与所有参数的 $\hat{V}$ 都非常接近。值得注意的是，校准所得估计的精确度几乎与使用完整队列时相当；这不仅适用于整个队列中均有数据的协变量，也适用于财富评分，因为财富评分有良好的代理变量。

对于少数参数，使用 SCC.Calib 或 USCC.Calib 估计的方差略小于使用整个队列时的方差。为了解这是否反映了数据中的某种特殊情况，我们从 Golestan 数据集中有放回抽样，生成了 2,500 个规模为 $n = 20{,}000$ 的队列，并重复进行了分析。完整队列的经验方差和平均估计方差均小于校准程序所得结果；参见网络附录 F。

为展示使用 CaseCohortCoxSurvival CRAN 软件包进行此类分析有多么容易，我们分别在表 4 和网络附录 J 中给出了 SCC 和 SCC.Calib 的脚本与伪代码。

9 讨论

我们提出了一种统一的病例队列数据分析方法，使研究者能够利用自 Prentice[1] 的里程碑论文发表以来，抽样设计和分析方面的各种选择与改进。我们采用了适用于不同抽样设计和分析选项的影响函数，并结合考虑两阶段抽样的方差计算。我们开发了相应的软件 CaseCohortCoxSurvival，便于对分层和/或权重校准进行分析，也支持不放回抽样和有放回抽样的子队列抽样。对于分层设计，我们允许第二阶段数据随机缺失。我们不仅提供 Cox 模型中对数相对风险的推断，还提供协变量特异性累积风险率和纯风险的推断。我们希望这些计算方法和软件能够促进更广泛、更有原则的病例队列数据抽样设计与分析；这方面仍有需求[10]。目前似乎还没有在线提供我们所述这类便捷软件（第 6 节）。

我们发现，无论对子队列进行分层还是非分层抽样，权重校准都能提高效率，这与此前关于非分层设计的研究结果一致。我们从理论和

经验上发现，如果第二阶段抽样指标的协方差[2] $\sigma_{i,\ k,\ j}, i \neq k$为零，稳健方差估计就几乎没有偏倚；例如，对子队列成员进行有放回抽样时便是如此（表5）。不放回抽样时，这些协方差为负，往往会使稳健方差估计偏高。在分层设计中，对数相对风险也曾发现这一点[4,5], ；但我们还发现，在非分层设计中，纯风险也存在这种偏倚。若基于第二阶段协变量的强预测因子进行权重校准，稳健方差的偏倚就很小（表5）。尽管如此，出于理论和经验方面的原因，我们仍建议采用包含完整方差分解的基于影响函数的方法。此外，正如先前的建议[10], ，我们强调在发表成果时完整描述抽样设计的实际重要性，包括分层细节，以及子队列是否采用有放回抽样。我们介绍的方法使用了设计权重。Borgan 等人[3] 和 Samuelsen 等人[4] 建议使用事后分层权重，将其划分为一个病例层和多个非病例层。在我们的模拟中，事后分层带来的效率提升不到2%。将我们推导出的设计权重对应的影响函数用于事后分层权重（病例处于一个层中，非病例则处于原有各层中），所得置信区间具有名义覆盖率（结果未显示）。因此，我们提供的影响函数可用于此类事后分层。若按随访结束的时间区间[4,29,30] 或其他特征[9, 第6章，第16.4.5节] 进行事后分层，效率或许还能进一步提高。

另一种抽样方法是，在每层选择亚队列样本量，使非病例的预期数量为观察到的病例数量的整数倍。使用我们给出的设计权重影响函数，并代入事后分层权重，在未报告的模拟中得到了有效的方差估计和置信区间覆盖率；但如果某一层中的病例和非病例数量较少（例如少于 10 个病例和 20 个非病例），则不然。

正如Keogh等人[31], 所讨论的，处理缺失数据和插补的基于似然的方法可以提高病例-队列分析的效率；但与分层和权重校准不同，如果插补模型设定错误，这些方法会导致风险模型估计产生偏倚。事实上，权重校准的一项关键优势在于，较差的插补模型会降低效率增益，但不会使风险模型参数估计产生偏倚[32]。权重校准估计量属于增广逆概率加权估计量这一类别，同样具有稳健性[32,33]。

在第5.3节中，我们针对病例缺失协变量数据，且该病例发生事件时第三阶段样本中没有其他处于风险中的成员的情况，提出了相应的修改。另一种做法是对Breslow估计量的分子加权，并且仅使用具有完整协变量数据的病例的事件时间t，即 $\mathrm{d}\widetilde{\Lambda}_0(t) = \sum_{j=1}^{J} \sum_{i=1}^{n} V_{i,j} \tilde{w}_{i,j(3)} \mathrm{d}N_{i,j(t)} (j) (t;\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\beta}}) / \tilde{S}_0$。然而，未报告的模拟结果表明，这会导致纯风险估计产生偏倚。

本文讨论的是在基线测量的协变量。尽管对数相对风险率的影响函数同样适用于时变协变量，但纯风险需要进行相应修改，而且大型队列会带来计算上的挑战。此外，除非时变协变量是“外部的”[34]，否则纯风险估计无法解释。我们假设各层具有共同的基线风险。如果层是根据风险模型中的协变量和/或时间尺度定义的，则Cox模型中无需包含层效应；否则，可以纳入层指标，并像其他所有队列成员均已知的协变量一样进行分析。若采用各层基线风险不同的分层Cox模型，则需要对本文给出的影响函数进行修改。

ACKNOWLEDGMENTS

感谢 Barry Graubard博士 提出的宝贵意见，也感谢 Arash Etemadi博士 和 GEMShare 记录并分享 Golestan队列 的数据。本研究得到了

美国国家癌症研究所、美国国立卫生研究院癌症流行病学与遗传学部内部研究计划的支持。本研究使用了美国国立卫生研究院高性能计算 Biowulf 集群的计算资源。（http://hpc.nih.gov）

数据可用性声明

第 7 节模拟所用的 R 代码和函数可在本文的补充材料以及 GitHub 上获取：https://github.com/Etievant/CaseCohort。R软件包 CaseCohortCoxSurvival 整合了 GitHub 上的各种函数，可在 CRAN 上获取[链接]。我们无权公开第 8 节中使用的临床数据。

参考文献

1. Prentice RL. 流行病学队列研究和疾病预防试验的病例队列设计。Biometrika. 1986;73(1):1-11. doi:10.2307/2336266 2. Barlow WE. 病例队列设计的稳健方差估计。Biometrics. 1994;50(4):1064-1072. doi:10.2307/2533444 3. Borgan O, Langholz B, Samuelsen SO, Goldstein L, Pogoda J. 暴露分层病例队列设计。Lifetime Data Anal. 2000;6(1):39-58. doi:10.1023/a:1009661900674 4. Samuelsen SO, Ånestad H, Skrondal A. 一般队列抽样设计的分层病例队列分析。Scandinavian Journal of Statistics. 2007;34(1):103-119. doi:10.1111/j.1467-9469.2006.00552.x 5. Gray RJ. 队列抽样设计的加权分析。Lifetime Data Anal. 2009;15(1):24-40. doi:10.1007/s10985-008-9095-z 6. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. 两阶段分层抽样中模型参数的改进 Horvitz–Thompson 估计：流行病学应用。Stat Biosci. 2009;1(1):32-49. doi:10.1007/s12561-009-9001-6 7. Breslow NE, Lumley T. 在病例队列数据分析中使用整个队列。Am J Epidemiol. 2009;169(11):1398-1405. doi:10.1093/aje/kwp055 8. Breslow NE, Lumley T. 半参数模型和两阶段抽样：Cox 回归的应用。从概率到统计再回归：高维模型与

Processes——献给 Jon A Wellner 的纪念文集。2013;9:65-78。doi:10.1214/12- IMSCOLL906 9.Borgan Ø、Breslow N、ChatterjeeN、Gail MH、Scott A、Wild CJ，编。《病例对照研究统计方法手册》。Chapman and Hall/CRC；2017年。doi:10.1201/9781315154084 10. Sharp SJ、Poulaliou M、ThompsonSG、WhiteIR、Wood AM。病例队列研究已发表分析综述及未来报告建议。PLOS ONE。2014;9(6):e101176。doi:10.1371/journal.pone.010117611. Breslow N。删失survival 数据的协方差分析。Biometrics。1974;30(1):89-99。doi:10.2307/2529620 12.Lin D。将 Cox 比例风险模型拟合于 survey数据。Biometrika。2000年;87(1):37-47。doi:10.1093/biomet/87.1.3713.Deville JC。复杂统计量和估计量的方差估计：线性化和残差技术。SurveyMethodology。1999年;25(2):193-203。14.Graubard BI、Fears TR。简单和复杂抽样设计中归因风险的标准误。Bio metrics。2005;61(3):847-855。doi:10.1111/j.1541-0420.2005.00355.x15. Pfeiffer RM、Gail MH。《绝对风险：临床管理和公共卫生中的方法与应用》。Chapman and Hall/CRC；2017年。doi:10.1201/9781315117539 16. Tsiatis A。《半参数理论与缺失数据》。Springer Science & Business Media；2006。17. van der Vaart AW。《渐近统计学》。Cambridge University Press；2000年。18.Shin YE、PfeifferRM、GraubardBI、Gail MH。通过权重校准提高队列内病例对照样本纯风险估计的效率。Biometrics。2020;76(4):1087-1097。doi:10.1111/biom.1320919.BreslowNE、Wellner JA。半参数模型和两阶段分层抽样的加权似然及其在 Cox 回归中的应用。ScandinavianJournal of Statistics。2007年;34(1):86-102。doi:10.1111/j.1467-9469.2006.00523.x 20. Pugh M、JM R、SR L、Harrington D。Cox 比例风险模型中的推断 Wit

h

缺失协变量数据。1994年1月1日在线发表。21. Therneau TM, Li H. 病例队列设计的Cox模型计算。Lifetime Data Anal

.

1999年;5(2):99-112. doi:10.1023/a:1009691327335 22. Therneau TM（截至2009年）TL（原始S >R 移植和R维护者，Elizabeth A，Cynthia C

.

survival：生存分析。在线发布于2022年3月3日。访问于2022年6月8日。https://CRAN.R-project.org/package=survival

23. Jones E. cchs：采用分层子队列抽样的病例队列数据Cox模型。在线发表于2020年9月10日。https://CRAN.R-project.org/package=cchs24. LumleyT. survey：复杂抽样调查样本分析。在线发表于2021年7月19日。https://CRAN.R-project.org/package=survey 25. Langholz B，Jiao J。病例队列研究的计算方法。计算统计

s

& 数据分析。2007年；51(8)：3737-3748。doi:10.1016/j.csda.2006.12.028 26.Mark SD，Katki HA。缺失病例数据条件下两阶段（嵌套）队列研究中的非参数和半参数生存估计量的指定与实施。期刊

l

《美国统计学会杂志》。2006；101(474)：460–471。doi:10.1198/016214505000000952 27. Pourshams A、Khademi H、Malekshah AF 等。队列概况：Golestan队列研究——一项关于伊朗北部食管癌的前瞻性研究。国际流行病学杂志。2010;39(1):52-59。doi:10.1093/ije/dyp161

28. Islami F, Kamangar F, Nasrollahzadeh D 等。社会经济地位与食管。癌症：来自高风险地区基于人群的病例对照研究的结果。Int J Epidemiol. 2009;38(4):978-988. doi:10.1093/ije/dyp195

29. Chen K. 广义病例队列抽样。《英国皇家统计学会杂志》系列。B（统计方法学）。2001；63(4)：791-809。

30. Ding J、Lu TS、Cai J、Zhou H。结果依赖抽样的最新进展（续）失效时间数据. Lifetime Data Anal. 2017;23(1):57-82. doi:10.1007/s10985-015-9355-7

31. Keogh RH, Seaman SR, Bartlett JW, Wood AM. 病例对照研究和病例队列研究中缺失数据的多重插补。Biometrics. 2018;74(4):1438-1449. doi:10.1111/biom.12910 32. Lumley T, Shaw PA, Dai JY, 等. 不完整数据的调查校准估计量与半参数模型之间的联系 [及讨论]。国际统计评论. 2011;79(2):200-232. 33. Robins JM, Rotnitzky A, Zhao LP. 部分回归变量并非始终可观测时回归系数的估计。Journal of the American Statistical Association. 1994;89(427):846-866. doi:10.2307/2290910 34. Kalbfleisch JD, Prentice RL. 失效时间数据的统计分析。约翰·威立父子出版公司；2011。

附录 A 分层病例队列的影响函数，含

校准权重

$$\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\eta}}) = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\boldsymbol{A}_{k,l}\boldsymbol{A}_{k,l}'\right\}^{-1} \boldsymbol{A}_{i,j},$$

$$\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\eta}}) = -\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\boldsymbol{A}_{k,l}\boldsymbol{A}_{k,l}'\right\}^{-1} \boldsymbol{A}_{i,j},$$

$$\boldsymbol{IF}_{i,j}^{(1)}\left(\widehat{\boldsymbol{\beta}}^{*}\right) = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\boldsymbol{Z}_{k,l}\boldsymbol{A}_{k,l}'\right\}\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\eta}}),$$

and $\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\beta}}^{*}\right) = \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)\boldsymbol{Z}_{i,j} + \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\boldsymbol{Z}_{k,l}\boldsymbol{A}_{k,l}'\right\}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\eta}})$, with

$$\boldsymbol{Z}_{i,j} = \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_{t} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\left\{\frac{\boldsymbol{S}_2^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})}{S_0^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})} - \frac{\boldsymbol{S}_1^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})\boldsymbol{S}_1^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})'}{S_0^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})^{2}}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1}\left[\int_{t}\left\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})}{S_0^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})}\right\}\left\{\mathrm{d}N_{i,j}(t) - \frac{Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{i,j}\right)\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\, w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\mathrm{d}N_{k,l}(t)}{S_0^{*}(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*})}\right\}\right].$$

$$IF_{i,j}^{(1)}\left\{\mathrm{d}\widehat{\Lambda}_0^{*}(t)\right\} = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)H_{k,l}(t)\boldsymbol{A}_{k,l}'\right\}\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\eta}}),$$

$$IF_{i,j}^{(2)}\left\{\mathrm{d}\widehat{\Lambda}_0^{*}(t)\right\} = \left\{S_0^{*}\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*}\right)\right\}^{-1}\Big[\mathrm{d}N_{i,j}(t) +$$

$$\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)H_{i,j}(t) + \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)H_{k,l}(t)\boldsymbol{A}_{k,l}'\right\}\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\eta}})\Big],$$

with $H_{i,j}(t) = -\left\{S_0^{*}\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*}\right)\right\}^{-1}\mathrm{d}\widehat{\Lambda}_0^{*}(t)\left\{\boldsymbol{S}_1^{*}\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^{*}\right)'\boldsymbol{Z}_{i,j} + K_{i,j}(t)\right\}$,

and $K_{i,j}(t) = Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{i,j}\right)$.

Finally, for any $s \in \{1,2\}$, $IF_{i,j}^{(s)}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_0^{*}(t)\right\} = \int_{\tau_1}^{\tau_2} IF_{i,j}^{(s)}\left\{\mathrm{d}\widehat{\Lambda}_0^{*}(t)\right\}$, and $IF_{i,j}^{(s)}\{\hat{\pi}^{*}(\tau_1,\tau_2;\boldsymbol{x})\} =$

$$\left\{\frac{\partial\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{*}}\right\}\boldsymbol{IF}_{i,j}^{(s)}\left(\widehat{\boldsymbol{\beta}}^{*}\right) + \left[\frac{\partial\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\mathrm{d}\Lambda_0(t)=\mathrm{d}\widehat{\Lambda}_0^{*}(t)}\right]IF_{i,j}^{(s)}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_0^{*}(t)\right\}.$$

附录B 协变量信息缺失且估计了设计抽样第三阶段权重的分层病例队列的影响函数

$\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\gamma}}) = \{\sum_{l=1}^{J}\sum_{k=1^{(l)}}^{n} \xi_{k,l} V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\boldsymbol{B}_{k,l}\boldsymbol{B}_{k,l'}\}^{-1} \boldsymbol{B}_{i,j},$

和 $\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\gamma}}) = -\{\sum_{l=1}^{J}\sum_{k=1^{(l)}}^{n} \xi_{k,l} V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\boldsymbol{B}_{k,l}\boldsymbol{B}_{k,l'}\} \boldsymbol{B}_{i,j}$。

然后 $\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\beta}}) = \{\sum_{l=1}^{J}\sum_{k=1^{(l)}}^{n} \xi_{k,l} V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\tilde{\boldsymbol{Z}}_{k,l}\boldsymbol{B}_{k,l'}\} \boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\gamma}}),$

以及 $\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\beta}}) = \tilde{\boldsymbol{Z}}_{i,j} + \{\sum_{l=1}^{J}\sum_{k=1^{(l)}}^{n} \xi_{k,l} V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\tilde{\boldsymbol{Z}}_{k,l}\boldsymbol{B}_{k,l'}\} \boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\gamma}})$，其中

$$\tilde{\boldsymbol{Z}}_{i,j} = w_{i,j}^{(2)} \times \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \int_t \xi_{k,l} V_{k,l} w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}) \left\{\frac{\tilde{\boldsymbol{S}}_2(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})'}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})^2}\right\} \mathrm{d}N_{k,l}(t)\right]^{-1} \times$$

$$\left[\int_t \left\{\boldsymbol{X}_{i,j} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}\right\} \left\{\mathrm{d}N_{i,j}(t) - \frac{Y_{i,j}(t)\exp(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} V_{k,l} w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\mathrm{d}N_{k,l}(t)}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}\right\}\right].$$

$$IF_{i,j}^{(2)}\{\mathrm{d}\tilde{\Lambda}_0(t)\} = \{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})\}^{-1}\mathrm{d}N_{i,j}(t) +$$

$$\{\sum_{l=1}^{J}\sum_{k=1^{(J)}}^{n} \xi_{k,l} V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\tilde{H}_{k,l}(t)\boldsymbol{B}_{k,l'}\} \boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\gamma}}),$$

and $IF_{i,j}^{(3)}\{\mathrm{d}\tilde{\Lambda}_0(t)\} = \tilde{H}_{i,j}(t) + \{\sum_{l=1}^{J}\sum_{k=1^{(J)}}^{n} \xi_{k,l} V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\tilde{H}_{k,l}(t)\boldsymbol{B}_{k,l'}\} \boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\gamma}}),$ with

$\tilde{H}_{i,j}(t) = -\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})^{-1}\mathrm{d}\tilde{\Lambda}_0(t)\{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})'\tilde{\boldsymbol{Z}}_{i,j} + \tilde{K}_{i,j}(t)\}$,

和 $\tilde{K}_{i,j}(t) = w_{i,j}^{(2)} Y_{i,j}(t) \exp(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})$。

最后，对于任意 $s \in \{2,3\}$、$IF_{i,j}^{(s)}\{\int \mathrm{d}\tilde{\Lambda}_0(t)_{\tau_1}^{\tau_2}\} = \int IF_{i、\tau_1}^{\tau_2}{}_{j}^{(2)} \tilde{\Lambda}_0\ t \{\mathrm{d}\ (\ )\}$ 和 $IF_{i,j}^{(s)}\{\hat{\pi}^*(\tau_1,\tau_2;\ \boldsymbol{x})\} = \{\underline{\partial\tilde{\pi}(\tau_1,\tau_2;\ x)\ \partial\boldsymbol{\beta}}\ |\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}\}\boldsymbol{IF}_{i,j}^{(s)}(\hat{\boldsymbol{\beta}}^*) + \left[\partial\underline{\partial\tilde{\pi}(\tau_1,\tau_2;\ x)\{\int \mathrm{d}\Lambda_0(t)_{\tau_1}^{\tau_2}\}}|\mathrm{d}\Lambda_0(t)=\mathrm{d}\tilde{\Lambda}_0(t)\right] IF_{i,j}^{(s)}\{\int \mathrm{d}\tilde{\Lambda}_0(t)_{\tau_1}^{\tau_2}\}$.

补充材料

第 2、3、4、5、7 和 8 节中提及的网络附录可在本文中查阅。网络附录 E 和 I 中模拟所用的 R 代码和函数可在本文的补充材料以及 GitHub 上获取，网址为 https://github.com/Etievant/CaseCohort。

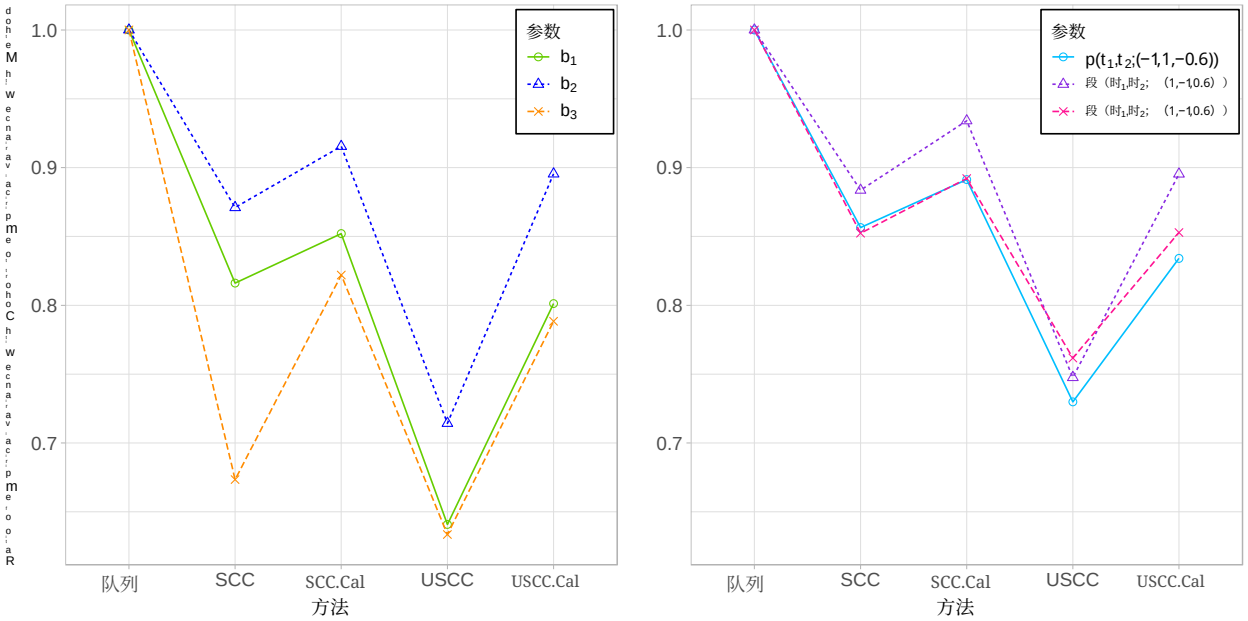

图1- 整个队列中对数相对风险和纯风险估计的经验方差与采用不同抽样设计和分析方法时的经验方差之比。结果来自 5,000个模拟队列，其中 $n = 10{,}000$, $p_Y = 0.02$, $K = 2$.。方差比是衡量相对效率的指标。

Parameter	队列	SCC		SCC.Calib		USCC		USCC.Calib	
		$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$
β^1		0.0069	0.0102	0.0087	0.0085	0.0083	0.0106	0.0106	0.0087	0.0087
		(0.007)	(0.0085)		(0.0082)		(0.0108)		(0.0087)	
β^2		0.0097	0.0139	0.0114	0.0109	0.0109	0.014	0.014	0.0109	0.0109
		(0.01)	(0.0115)		(0.0109)		(0.014)		(0.0112)	
β^3		0.0068	0.0102	0.0102	0.0085	0.0085	0.0107	0.0107	0.0087	0.0087
		(0.0069)	(0.0103)		(0.0084)		(0.0109)		(0.0088)	
$\tau_1(\pi_{gol}(x_{2,60.1.1}=x))$		0.0122	0.0172	0.014	0.0142	0.0137	0.0181	0.0159	0.0145	0.0145
		(0.0119)	(0.0136)		(0.0133)		(0.0158)		(0.014)	
$\tau_2(\pi_{gol}(x_{2,60.1.1}=x))$		0.062	0.0861	0.0697	0.0684	0.0681	0.086	0.0837	0.0696	0.0696
		(0.0618)	(0.0688)		(0.0662)		(0.0823)		(0.0696)	
$\tau_{1,1}(\pi_{gol}(x_{2,60.1.1}=x))$		0.0277	0.0379	0.0333	0.0319	0.0316	0.0386	0.0363	0.0328	0.0328
		(0.0274)	(0.0326)		(0.0309)		(0.0361)		(0.0325)	

表1- 在5,000个模拟队列中，使用不同抽样设计、分析方法和方差估计方法所得的对数相对风险和纯风险估计的估计方差均值，其中 $n = 10{,}000$, $p_Y = 0.02$, $K = 2$。相应的经验方差列于括号内。

Parameter	队列	SCC		SCC.Calib		USCC		USCC.Calib	
		$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$
β^1	0.9476	0.9668*	0.9524	0.9538	0.95	0.947	0.947	0.9452	0.945
β^2	0.9486	0.9716*	0.9522	0.9514	0.9512	0.9554	0.9554	0.9506	0.9506
β^3	0.9422*	0.9492	0.9494	0.9474	0.9474	0.9454	0.9454	0.9476	0.9478
$\{\pi_{go}^1(\tau^2, \tau^1, \boldsymbol{x})\}$	0.956*	0.973*	0.9568*	0.96*	0.9546	0.9656*	0.9526	0.955	0.955
$\{\pi_{go}^1(\tau^2, \tau^1, \boldsymbol{x})\}$	0.952	0.9722*	0.9518	0.9522	0.9516	0.954	0.9522	0.9472	0.947
$\{\pi_{go}^1(\tau^2, \tau^1, \boldsymbol{x})\}$	0.952	0.9666*	0.9542	0.9532	0.9516	0.9574*	0.95	0.9502	0.9502

表2- 使用不同抽样设计、分析方法和方差估计方法，对5,000个模拟队列进行分析所得的对数相对风险和纯风险估计的95%置信区间覆盖率，其中$n = 10{,}000, p_Y = 0.02, K = 2.$。*表示覆盖率超出预期区间[0.9440；0.9560]。

Parameter	队列	SCC		SCC.Calib		USCC		USCC.Calib	
		$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$	$\hat{V}_{稳健}$	$\hat{V}$
β^1	0.0013	0.0018	0.0016	0.0013	0.0013	0.0025	0.0025	0.0013	0.0013
β^2	0.005	0.0075	0.0072	0.0055	0.0055	0.0104	0.0104	0.0061	0.0061
β^3	0.0173	0.0291	0.0291	0.0178	0.0178	0.0394	0.0394	0.017	0.017
β^4	0.0029	0.0057	0.0057	0.0029	0.0029	0.007	0.007	0.003	0.003
β^5	0.0009	0.0013	0.0013	0.0008	0.0008	0.002	0.002	0.0009	0.0009
β^6	0.0217	0.0344	0.0344	0.0219	0.0219	0.0497	0.0497	0.0213	0.0213
β^7	0.0044	0.0077	0.0076	0.0043	0.0043	0.0103	0.0103	0.0045	0.0045
$(\boldsymbol{x}, \boldsymbol{\tau}^2, \boldsymbol{\tau}^1, \boldsymbol{\pi})$ $(\boldsymbol{0}^3, 1, 0, 4, 0, -, 0) = \boldsymbol{x}$	0.00012	0.00023	0.00022	0.00012	0.00012	0.00031	0.00030	0.00013	0.00012
$(\boldsymbol{0}^3, 1, 0, 4, 0, 0) = \boldsymbol{x}$	5.10E-05	9.10E-05	8.80E-05	5.10E-05	5.10E-05	0.00010	0.00010	5.70E-05	5.70E-05
$(\boldsymbol{0}^2, 1, \boldsymbol{0}^4) = \boldsymbol{x}$	2.50E-05	4.00E-05	3.50E-05	2.50E-05	2.50E-05	4.50E-05	4.30E-05	2.70E-05	2.70E-05
$\boldsymbol{0}^7 = \boldsymbol{x}$	6.70E-06	1.10E-05	7.90E-06	6.90E-06	6.80E-06	1.10E-05	9.80E-06	6.90E-06	6.90E-06

表 3- 使用不同抽样设计

、分析方法和方差估计得到的对数相对风险参数及纯风险参数的估计方差，见 Golestan队列（ $n = 30{,}000$）。β_p 表示协变量x_p的对数相对风险参数， $p \in \{1, \dots, 7\}$。

```
library(survival) library(CaseCohortCoxSurvival) library(dplyr) # 加载数据集 ----------
```

表 4- 使用 CaseCohortCoxSurvival R 包获取表 3 中 SCC 和 SCC.Calib 的方差估计的 R 脚本 $\hat{V}$

稳健方差估计的近似有效性					
		不放回抽样			
remarap	抽样 with 替换	未分层的抽样 采用抽样设计权重	非分层抽样 with 校准权重	分层抽样 采用抽样设计权重	分层的抽样 with 校准权重*
相对风险	Yes	Yes	Yes	No	Yes*
纯风险	Yes	No	Yes*	No	Yes*

表5-对于分层权重校准病例-队列数据，在相对风险和纯风险估计中稳健方差估计量近似有效的抽样设计和分析方法。如果理论计算表明偏倚可忽略，和/或在模拟中稳健方差估计的均值接近我们所描述的两阶段方差方法的均值，我们将稳健方差估计归类为近似有效。
* 当第一阶段协变量能够良好预测仅在第二阶段测量的协变量时成立。如果第二阶段协变量的代理变量过弱，稳健方差估计可能不适用。

分层权重校准病例队列数据中相对风险和纯风险的Cox模型推断补充材料

Lola Etievant[1]*、Mitchell H. Gail[1]*

lola.etievant@nih.gov

gailm@mail.nih.gov

[1] 美国国家癌症研究所癌症流行病学与遗传学部生物统计学处,

马里兰州罗克维尔市医疗中心大道9609号，邮编20850-9780。

* 通讯作者

目录

网络附录 A。影响函数的显式形式.4
A.1 使用设计权重的分层病例队列影响函数.4
A.2 使用校准权重的分层病例队列影响函数.5
A.3 协变量信息缺失且设计抽样第三阶段权重为估计值时，分层病例队列的影响函数.6

网络附录 B. 整个队列协变量数据完整时的估计 9 B.1 参数估计.9

B.2 影响函数.9
B.3 基于影响函数的方差估计.11
网络附录 C. 基于分层病例队列设计权重的估计.12

C.1 影响函数的推导.12
C.2 与 Samuelsen、Ånestad 和 Skrondal（2007年）方差的比较 ..... 14 网络附录 D. 使用校准权重估计分层病例队列设计.16

D.1 校准的估计方程.16
D.2 影响函数的推导.17
D.3 关于辅助变量选择的说明.22
网络附录 E. 模拟.24
E.1 主文档第 7 节模拟所使用的固定参数值 24 E.2 所有研究情景的模拟结果.24

E.3 关于稳健方差估计的说明.25
E.4 其他模拟：根据非病例数量对设计权重进行后分层 44 E.5 其他模拟：使用较弱的代理变量进行权重校准.47

网络附录 F. GOLESTAN 数据分析——2,500次重复的结果 50 网络附录 G. 第三阶段设计抽样概率未知时的缺失数据.56

G.1 影响函数的推导.56
G.2 方差分解与估计.61
网络附录 H. 已知第三阶段设计抽样概率时的缺失数据.66

H.1 参数估计.66

H.2 影响函数的推导.67
H.3 基于影响函数的方差分解与估计.69
网络附录 I. 缺失数据模拟.73
I.1 模拟设计.73
I.2 模拟结果.74
网络附录 J. 在主文档表 3 中获取 SCC 和 SCC.Calib 的 $\boldsymbol{V}$ 的逐步伪代码.100

数据可用性声明.103
参考文献 103

网络附录 A。 E影响函数的显式形式

在网络附录 A 中，我们回顾了采用设计权重（A.1）、校准权重（A.2）以及允许存在缺失的第二阶段数据（A.3）的分层病例队列设计的影响函数。将层数设为 1，即可涵盖非分层病例队列设计。推导过程和详细内容见后续网络附录。

A.1 分层病例队列设计权重的影响函数

令$\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&$表示受试者 i 在层j 上对$\boldsymbol{\theta}\%$的影响，$i \in \{1, \ldots, {}^{n}(\#)\}$，$j \in \{1, \ldots, J\}$，$\boldsymbol{\theta}\% \in \{\boldsymbol{\beta}\%$，$\mathrm{d}\Lambda\%_0(t)$，$\Lambda\%_0(t)$，$\pi{:}(\tau_1, \tau_2; \boldsymbol{x})\}$。我们可以证明 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&= \xi_{i,j}\, w_{i,j}\, \boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\%\&$，其中 $\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\%\&$ 涉及仅针对第二阶段样本中的个体测量的变量。这里，我们给出$\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\%\&$ 的显式形式；影响函数的推导见网络附录 C.1。我们有

$$\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\beta}}) = \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t \left\{\frac{\boldsymbol{S}_2(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})\,\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'}{S_0(t;\widehat{\boldsymbol{\beta}})^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1}\left[\int_t \left\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\left\{\mathrm{d}N_{i,j}(t) - \frac{Y_{i,j}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\mathrm{d}N_{k,l}(t)}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\right],$$

with $S_0(t;\boldsymbol{\beta}) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} w_{k,l}\,\xi_{k,l}\,Y_{k,l}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}_{k,l})$,

$\boldsymbol{S}_1(t;\boldsymbol{\beta}) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} w_{k,l}\,\xi_{k,l}\,Y_{k,l}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}_{k,l})\boldsymbol{X}_{k,l}$,

and $\boldsymbol{S}_2(t;\boldsymbol{\beta}) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} w_{k,l}\xi_{k,l}Y_{k,l}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}_{k,l})\boldsymbol{X}_{k,l}\boldsymbol{X}_{k,l}'$,

并且我们有

$$IF_{i,j}^{(2)}\{\mathrm{d}\widehat{\Lambda}_0(t)\} = \{S_0(t;\widehat{\boldsymbol{\beta}})\}^{-1}\left\{\mathrm{d}N_{i,j}(t) - \mathrm{d}\widehat{\Lambda}_0(t)\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\beta}}) - \mathrm{d}\widehat{\Lambda}_0(t)Y_{i,j}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\right\}.$$

注意，$IF_{i,j}^{(2)}\{\mathrm{d}\Lambda_0(t)\}$ 是增量 $\mathrm{d}N_{i,j}(t)$ 和 $\mathrm{d}\Lambda_0(t)$ 的线性组合。因此，我们可以写成 IF_i，$_j^{(2)}\int \mathrm{d}\Lambda_0(t)^{\tau_2}\}=\int IF_{i,j}^{(2)}\{\mathrm{d}\Lambda_0(t)\}$ 。最后，$IF_{i,j}^{(2)}\{\pi:(\tau_1,\tau_2;\boldsymbol{x})\}=\mathrm{N}^{\partial\hat{\pi}(\tau_1,}$ $\mathrm{B}\boldsymbol{\beta}_{|\boldsymbol{\beta}=\boldsymbol{\beta}}\mathrm{O}\,\boldsymbol{IF}_{i,j}^{(2)}\boldsymbol{\beta}+\left[\mathrm{BCD}(\mathrm{A}\#,\mathrm{A}!;\boldsymbol{x})\,\mathrm{BG}\int \mathrm{dI}''(,)^{\tau}!\tau\#\,\mathrm{J}|\mathrm{dI}''(,)=\mathrm{dI}/''(,)_{-}IF_{i,j}^{(2)}\int \mathrm{d}\Lambda_0(t)\right.$

$\left.^{\tau_2}\mathrm{A}\#\right\}$。

A.2 使用校准权重的分层病例队列的影响函数令 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%_*\&$ 表示层 j 中受试者 i 对 $\boldsymbol{\theta}\%_*$, $i\in\{1,\ldots,{}^{n}(\#)\}$, $j\in\{1,\ldots,J\}$,$\boldsymbol{\theta}\%^*\in\{\boldsymbol{\eta}\mathrm{a},\ \boldsymbol{\beta}\%_*,\ \mathrm{d}\Lambda_0^*(t),\ \Lambda_0^*(t),\ \pi:^*(\tau_1,\tau_2;\ \boldsymbol{x})\}$ 的影响。我们可以证明，$\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%_*\&=\boldsymbol{IF}_{i,j}^{(1)}\#\boldsymbol{\theta}\%_*\&+\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\%_*\&$，其中 $\boldsymbol{IF}_{i,j}^{(1)}\#\boldsymbol{\theta}\%_*\&$ 仅涉及在所有队列成员中测量的变量，而 $\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\%_*\&$ 涉及仅针对第二阶段样本中的个体测量的变量。这里，我们给出 $\boldsymbol{IF}_{i,j}^{(s)}\#\boldsymbol{\theta}\%_*\&$, $s\in\{1,2\}$ 的明确形式；影响函数的推导见网络附录 D.2。我们有

$$\boldsymbol{IF}_{i,j}^{(1)}(\boldsymbol{\eta}\mathrm{a})=\mathrm{Z}\Sigma\ \ \Sigma\ \ \xi_{k,l}w_{k,}3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\boldsymbol{A}_{k,l}\boldsymbol{A}_{k,}3'^{\,n}(_j)25(6\,35(\Big\}^{-1}\boldsymbol{A}_{i,j},\boldsymbol{IF}_{i,j}^{(2)}(\boldsymbol{\eta}\mathrm{a})=-\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,j}\&\mathrm{Z}\Sigma\ \ \Sigma\ \ \xi_{k,l}w_{k,}3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\boldsymbol{A}_{k,l}\boldsymbol{A}_{k,}3'^{\,n}(_j)\,25(6\,35(\Big\}^{-1}\boldsymbol{A}_{i,j},\boldsymbol{IF}_{i,j}^{(1)}\#\boldsymbol{\beta}\%_*\&=\mathrm{Z}\Sigma\ \ \Sigma\ \ \xi_{k,l}w_{k,}3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\boldsymbol{Z}_{k,l}\boldsymbol{A}_{k,}3'^{\,n}(_j)\,25(635(\Big\}\boldsymbol{IF}_{i,j}^{(1)}(\boldsymbol{\eta}\mathrm{a}),\text{和}\ \boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\beta}\%_*\&=\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,j}\&\boldsymbol{Z}_{i,j}+\mathrm{Z}\Sigma\ \ \Sigma\ \ \xi_{k,l}w_{k,}3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\boldsymbol{Z}_{k,l}\boldsymbol{A}_{k,}3'^{\,n}(_j)\,25(635(\Big\}\boldsymbol{IF}_{i,j}^{(2)}(\boldsymbol{\eta}\mathrm{a}),\text{其中}\boldsymbol{Z}_{i,j}=\mathrm{D}\Sigma_{l=}^{J}\ \Sigma_{k=}^{n}\ \int \xi_{k,l}w_{k,}3\ \exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,l}\&,(')((\mathrm{G}\,\underline{\boldsymbol{s}^*(\ \boldsymbol{\eta}\boldsymbol{\beta}^*)}_{\boldsymbol{s}}!,;\mathrm{D},/{}''(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/_*)^{-}\#(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/^*)s\#{}^*(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/_*)\$\ s''(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/_*)!\mathrm{I}\ \mathrm{d}N_{k,}3(t)\mathrm{K}^{-1}\left[\int \mathrm{G}\boldsymbol{X}_{i,j}{}^{-},\underline{\boldsymbol{s}^*(\ \boldsymbol{\eta}\boldsymbol{\beta}^*)}_{\boldsymbol{s}}\#,;\mathrm{D},/''(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/_*)\mathrm{I}\ \mathrm{fd}N_{i,j}(t)\ {}_{-}{}^{Y}),_j\right.$$

$\left.(,)9:;\mathrm{N}\boldsymbol{\beta}/_*\$\boldsymbol{X}),_j\mathrm{O}\ \Sigma\ \ \Sigma\mathrm{P},,'\ \mathrm{Q},,'\ {}^{\mathrm{e}}:;(\boldsymbol{\eta}\mathrm{D}\$\ _{\boldsymbol{A}},')^{\mathrm{d}}?,,'(,)-\ (')\ ,.\#/\ '.\#s''(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/_*)\,\mathrm{g}_{-}$, 和 $S_0^*(t;\boldsymbol{\eta}\mathrm{a},\ \boldsymbol{\beta})=\Sigma_{l=}^{J}\ \Sigma_{k=}^{n}\ \xi_{k,l}\ w_{k,}3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\ Y_{k,}3(t)\ \exp\#\boldsymbol{\beta}'\boldsymbol{X}_{k,}3\&\ (')\ ((\ ,$

$\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta}) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) Y_{k,l}(t) \exp\left(\boldsymbol{\beta}'\boldsymbol{X}_{k,l}\right)\boldsymbol{X}_{k,l}$,

and $\boldsymbol{S}_2^*(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta}) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) Y_{k,l}(t) \exp\left(\boldsymbol{\beta}'\boldsymbol{X}_{k,l}\right)\boldsymbol{X}_{k,l}\boldsymbol{X}_{k,l}'$.

Then $IF_{i,j}^{(1)}\left\{\mathrm{d}\widehat{\Lambda}_0^*(t)\right\} = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) H_{k,l}(t)\boldsymbol{A}_{k,l}'\right\} \boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\eta}})$,

and

$IF_{i,j}^{(2)}\left\{\mathrm{d}\widehat{\Lambda}_0^*(t)\right\} = \left\{S_0^*\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*\right)\right\}^{-1}\Big[\mathrm{d}N_{i,j}(t) +$

$\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)H_{i,j}(t) + \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) H_{k,l}(t)\boldsymbol{A}_{k,l}'\right\} \boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\eta}})\Big]$,

with $H_{i,j}(t) = -\left\{S_0^*\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*\right)\right\}^{-1}\mathrm{d}\widehat{\Lambda}_0^*(t)\left\{\boldsymbol{S}_1^*\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*\right)'\boldsymbol{Z}_{i,j} + K_{i,j}(t)\right\}$,

and $K_{i,j}(t) = Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{i,j}\right)$.

注意，$H_{i,j}(t)$ 是$\mathrm{d}\Lambda\%_0^*(t)$ 的倍数，因此$IF_{i,j}^{(s)}\left\{\mathrm{d}\Lambda\%_0^*(t)\right\}$ 是 $\mathrm{d}\Lambda\%_0^*(t)$ 和 $\mathrm{d}^{N_{i,j}}(t)$ 的线性组合。因此$IF_{i,j}^{(s)}$ Z∫ $\mathrm{d}\Lambda\%_0^*(t)^{\tau_2}$ A# $\Big\} = \int$ $IF_{i,j}^{(s)}\left\{\mathrm{d}\Lambda\%_0^*(t)\right\}$ A!A#,$s \in \{1,2\}$。最后，对于任意 $s \in \{1,2\}$，$IF_{i,j}^{(s)}\{\pi{:}^*(\tau_1,\tau_2;\boldsymbol{x})\}$ = N$^{\partial\widehat{\pi}(\tau_1,}$A!;$x$) B$\beta$ $|_{\boldsymbol{\beta}=\boldsymbol{\beta}}$/*O$\boldsymbol{IF}_{i,j}^{(s)}$#$\boldsymbol{\beta}$%*& + BCD(A#,A!; x)BG∫ d I"(,)$^{\tau}$!$^{\tau}$# J|dI"(,)=dI/"*(,)_$IF_{i,j}^{(s)}$ Z∫ $\mathrm{d}\Lambda\%_0^*(t)^{\tau_2}$ A# $\Big\}$。

A.3 具有缺失协变量的分层病例队列的影响函数

信息和估计的设计抽样第三阶段权重。我们令 $\boldsymbol{\Delta}_{i,j}$#$\boldsymbol{\theta}$l& 表示层 j 中受试者 i 对 $\boldsymbol{\theta}$l 的影响，其中 $i \in \{1,\dots,{}^{n}(\#)\}$, $j \in \{1,\dots,J\}$,$\boldsymbol{\theta}$l $\in \{\boldsymbol{\gamma}$n, $\boldsymbol{\beta}$l, $\mathrm{d}\Lambda\mathrm{l}_0(t)$, Λl$_0(t)$, πo$(\tau_1,\tau_2;\ \boldsymbol{x})\}$。我们可以证明，$\boldsymbol{\Delta}_{i,j}$#$\boldsymbol{\theta}$l& = $\xi_{i,j}$ $\boldsymbol{IF}_{i,j}^{(2)}$#$\boldsymbol{\theta}$l& + $\xi_{i,j}$ $V_{i,j}$exp#$\boldsymbol{\gamma}$n$'\boldsymbol{B}_{i,j}$&$\boldsymbol{IF}_{i,j}^{(3)}$#$\boldsymbol{\theta}$l&，其中 $\boldsymbol{IF}_{i,j}^{(3)}$#$\boldsymbol{\theta}$l& 涉及仅在

第三阶段样本中的个体上测量的变量。这里，我们给出 $\boldsymbol{IF}_{i,\ j}^{(s)}\#\boldsymbol{\theta}$l& 的显式形式， $s \in \{2,3\}$；参见

Web Appendix G.1 for derivation of the influence functions. We have

$$\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\gamma}}) = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}\right) \boldsymbol{B}_{k,l}\boldsymbol{B}_{k,l}'\right\}^{-1} \boldsymbol{B}_{i,j},$$

and $\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\gamma}}) = -\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} V_{k,l} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}\right) \boldsymbol{B}_{k,l}\boldsymbol{B}_{k,l}'\right\} \boldsymbol{B}_{i,j}.$

Then $\boldsymbol{IF}_{i,j}^{(2)}\left(\tilde{\boldsymbol{\beta}}\right) = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} V_{k,l} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}\right)\tilde{\boldsymbol{Z}}_{k,l}\boldsymbol{B}_{k,l}'\right\} \boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\gamma}})$,

and $\boldsymbol{IF}_{i,j}^{(3)}\left(\tilde{\boldsymbol{\beta}}\right) = \tilde{\boldsymbol{Z}}_{i,j} + \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} V_{k,l} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}\right)\tilde{\boldsymbol{Z}}_{k,l}\boldsymbol{B}_{k,l}'\right\} \boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\gamma}})$, with

$$\tilde{\boldsymbol{Z}}_{i,j} = w_{i,j}^{(2)} \times \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \int_t \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}) \left\{\frac{\tilde{\boldsymbol{S}}_2(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})'}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})^2}\right\} \mathrm{d}N_{k,l}(t)\right]^{-1} \times$$

$$\left[\int_t \left\{\boldsymbol{X}_{i,j} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}\right\}\left\{\mathrm{d}N_{i,j}(t) - \frac{Y_{i,j}(t)\exp(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\mathrm{d}N_{k,l}(t)}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}\right\}\right],$$

$$\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta}) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, Y_{k,l}(t) \exp\left(\boldsymbol{\beta}'\boldsymbol{X}_{k,l}\right),$$

$$\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta}) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l} w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, Y_{k,l}(t) \exp\left(\boldsymbol{\beta}'\boldsymbol{X}_{k,l}\right) \boldsymbol{X}_{k,l},$$

and $\tilde{\boldsymbol{S}}_2\left(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}}\right) = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l} V_{k,l} w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, Y_{k,l}(t) \exp\left(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l}\right) \boldsymbol{X}_{k,l}\, \boldsymbol{X}_{k,l}'$.

In addition,

$IF_{i,\ j}^{(2)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\} = \{S\mathrm{to}\#t;\ \boldsymbol{\gamma}\mathrm{n},\ \boldsymbol{\beta}\mathrm{l}\&\}^{-1}\mathrm{d}N_{i,\ j}(t)\ +$

$\mathrm{Z}\Sigma\ \ \Sigma\ \ \xi_{k,\ l}V_{k,\ 3}\exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{k,\ 3}\&H\mathrm{l}_{k,l}(t)\boldsymbol{B}_{k,\ 3}{}'^{\,n_{(j)}}25(6\ 35($
$\}\boldsymbol{IF}_{i,j}^{(2)}(\boldsymbol{\gamma}\mathrm{n})$,

and $IF_{i,\ j}^{(3)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\} = H\mathrm{l}_{i,\ j}(t) + \mathrm{Z}\Sigma\ \ \Sigma\ \ \xi_{k,\ l}V_{k,\ 3}\exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{k,\ 3}\&H\mathrm{l}_{k,l}(t)\boldsymbol{B}_{k,\ 3}{}'^{\,n_{(j)}\,k=1J\ l=1}\}\boldsymbol{IF}_{i,j}^{(3)}(\boldsymbol{\gamma}\mathrm{n})$, with

$H\mathrm{l}_{i,\ j}(t) = -S\mathrm{to}\#t;\ \boldsymbol{\gamma}\mathrm{n},\ \boldsymbol{\beta}\mathrm{l}\&^{-1}\mathrm{d}\Lambda\mathrm{l}_0(t)\ \mathrm{Z}\ \boldsymbol{S}\mathrm{u}_1\#t;\ \boldsymbol{\gamma}\mathrm{n},\ \boldsymbol{\beta}\mathrm{l}\&'\boldsymbol{Z}\mathrm{l}_{i,\ j} + K\mathrm{l}_{i,\ j}(t)\}$,

和 $K\mathrm{l}_{i,j}(t) = w_{i,j}^{(2)}\, Y_{i,j}(t)\exp\#\boldsymbol{\beta}\mathrm{l}'\boldsymbol{X}_{i,j}\&$.

注意， $Hli,j(t)$ 是 dΛlo(t) 的倍数，因此 $IF_{i,j}^{(s)}\{d\Lambda l_0(t)\}$ 是 dΛlo(t) 和 $dN_{i,j}(t)$ 的线性组合。因此 $IF_{i,j}^{(s)}$ Z∫ dΛlo(t)$^{\tau_2}$ A# $\} = \int$ $IF_{i,j}^{(2)}\{d\Lambda l_0(t)\}$ A! A#， $s \in \{2,3\}$。最后，对于任意 $s \in \{2,3\}$，$IF_{i,j}^{(s)}\{\pi:^{*}(\tau_1, \tau_2;\ \boldsymbol{x})\}$ = N$^{\partial\tilde{\pi}(\tau_{1'}}$A!; $\boldsymbol{x}$) B$\boldsymbol{\beta}$ $_{|\boldsymbol{\beta}=\boldsymbol{\beta}}$WO $\boldsymbol{IF}_{i,j}^{(s)}$#$\boldsymbol{\beta}$%*& + BCV(A#,A!; $\boldsymbol{x}$)

BG∫ dI"(,)$^{\tau}$!$^{\tau}$# J|dI"(,)=dIW"(,)_$IF_{i,j}^{(s)}$ Z∫ dΛlo(t)$^{\tau_2}$ A# $\}$。

网络附录 B. 整个队列完整协变量数据的估计

B.1 参数估计

这里我们回顾分析具有完整协变量的队列数据的方法，这些协变量在所有队列成员中均已$\boldsymbol{X}$测量。对数相对风险的估计值，$\boldsymbol{\beta}\%_c$, 通过最大化部分对数似然获得

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\int_t\left[\log\{Y_{i,j}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}_{i,j})\}-\log\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}Y_{k,l}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}_{k,l})\right\}\right]\mathrm{d}N_{i,j}(t),$$

或者等价地，通过求解$\boldsymbol{\beta}$估计方程

$$\boldsymbol{U}_c(\boldsymbol{\beta})=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\int_t\left\{\boldsymbol{X}_{i,j}-\frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}Y_{k,l}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}_{k,l})\boldsymbol{X}_{k,l}}{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}Y_{k,l}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}_{k,l})}\right\}\mathrm{d}N_{i,j}(t)=0.$$

时间 t 时的基线风险点质量按非参数方式估计（Breslow，1974年），即

$\mathrm{d}\Lambda\%_{0,c}\#t;\boldsymbol{\beta}\%_c\&\equiv\mathrm{d}\Lambda\%_{0,c}(t)=\frac{\sum\sum\mathrm{d}N),j}{\sum_{\#/\,j.\#}\sum^{-(j)).}Y),j(,\ )\,9:;(\boldsymbol{\beta}\hat{}2\$\boldsymbol{X}),j)}$，以及累积基线风险估计

截至时间 t 的估计值为 $\Lambda\%_{0,c}\#t;\boldsymbol{\beta}\%_c,\hat{\lambda}_{0,c}\&\equiv\Lambda\%_{0,c}(t)=\int\mathrm{d}\Lambda\%_{0,c}(s)^{t}{}_0$.

最后，协变量特征 $\boldsymbol{x}$ 在 $(\tau_1,\tau_2]$ 中的纯风险（即具有特征 $\boldsymbol{X}=\boldsymbol{x}$ 时，在区间 $(\tau_1,\tau_2]$ 内发生事件的概率），$0\le\tau_1<\tau_2\le\tau$，估计为

$\pi{:}_c\#\tau_1,\tau_2;\boldsymbol{x},\boldsymbol{\beta}\%_c,\hat{\lambda}_{0,c}\&\equiv\pi{:}_c(\tau_1,\tau_2;\boldsymbol{x})=1-\exp\mathrm{Z}-\int\exp\#\boldsymbol{\beta}\%_{c}{}'\boldsymbol{x}\&\,\mathrm{d}\Lambda\%_{0,c}(t)\,{}_{\tau_2\,\tau_1}\Big\}.$

B.2 影响函数

令 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%_c\&$ 表示层 j 中受试者 i 对 $\boldsymbol{\theta}\%_c,\ i\in\{1,\dots,{}^{n}{}_{(j)}\},\ j\in\{1,\dots,J\},\boldsymbol{\theta}\%_c\in\{\boldsymbol{\beta}\%_c,\mathrm{d}\Lambda\%_{0,c}(t),\ \Lambda\%_{0,c}(t),\ \pi{:}_c(\tau_1,\tau_2;\boldsymbol{x})\}$ 的影响。引自 Reid 和 Crépeau（1985），

$$\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\beta}}_c\right) = \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t\left\{\frac{\boldsymbol{S}_2^c(t;\widehat{\boldsymbol{\beta}}_c)}{S_0^c(t;\widehat{\boldsymbol{\beta}}_c)} - \frac{\boldsymbol{S}_1^c(t;\widehat{\boldsymbol{\beta}}_c)\,\boldsymbol{S}_1^c(t;\widehat{\boldsymbol{\beta}}_c)'}{S_0^c(t;\widehat{\boldsymbol{\beta}}_c)^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1} \times \left[\int_t\left\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1^c(t;\widehat{\boldsymbol{\beta}}_c)}{S_0^c(t;\widehat{\boldsymbol{\beta}}_c)}\right\}\times \left\{\mathrm{d}N_{i,j}(t) - Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{X}_{i,j}\right)\frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\mathrm{d}N_{k,l}(t)}{S_0(t;\widehat{\boldsymbol{\beta}}_c)}\right\}\right],$$

其中 $S_0^c\#t;\ \boldsymbol{\beta}\%_c\& = \Sigma \quad \Sigma \quad Y_{k,k=} \quad {}_l\ t \quad \left(\boldsymbol{\beta}_c{}'\boldsymbol{X}_{k,l}\right)() \exp \%_{(l)\,1J}$
$_{l=1}$,

$\boldsymbol{S}_1^c\#t;\ \boldsymbol{\beta}\%_c\& = \Sigma_{l=}^{J}\ \Sigma_{k=}\ Y_{k,}{}^{n} \quad {}_l\ t \quad \left(\boldsymbol{\beta}_c{}'\boldsymbol{X}_{k,l}\right)\boldsymbol{X}_{k,l}\ () \exp \%_{(l)\,11}$,

和 $\boldsymbol{S}_2^c\#t;\ \ \boldsymbol{\beta}\%_c\&= \Sigma_{l=}^{J}\ \Sigma_{k=}^{n}\ \ Y_{k,l}(t)\exp\left(\boldsymbol{\beta}\%_c{}'\boldsymbol{X}_{k,l}\right)\boldsymbol{X}_{k,l}\,\boldsymbol{X}_{k,l}{}'_{\ (l)\,11}$.

然后，根据 Pfeiffer 和 Gail（2017）第 4 章,

$$\Delta_{i,j}\left\{\mathrm{d}\widehat{\Lambda}_{0,c}(t)\right\} = \left\{S_0^c\left(t;\widehat{\boldsymbol{\beta}}_c\right)\right\}^{-1}\left[\mathrm{d}N_{i,j}(t) - \mathrm{d}\widehat{\Lambda}_{0,c}(t)\,Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{X}_{i,j}\right) - \mathrm{d}\widehat{\Lambda}_{0,c}(t)\,\boldsymbol{S}_1^c\left(t;\widehat{\boldsymbol{\beta}}_c\right)'\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\beta}}_c\right)\right].$$

注意，$\Delta_{i,j}\left\{\mathrm{d}\Lambda\%_{0,c}(t)\right\}$ 是增量 $\mathrm{d}N_{i,j}(t)$ 和$\mathrm{d}\Lambda\%_{0,c}(t)$ 的线性组合；因此$\Delta_{i,j}\mathrm{Z}\int \mathrm{d}\Lambda\%_{0,c}(t)^{\tau_2}\ \mathrm{A}\#\ \Big\} = \int\ \ \Delta_{i,j}\left\{\mathrm{d}\Lambda\%_{0,c}(t)\right\}\mathrm{A!A}\#$。最后，遵循 Graubard 和 Fears (2005),

$$\Delta_{i,j}\{\hat{\pi}_c(\tau_1,\tau_2;\boldsymbol{x})\} = \left\{\frac{\partial\hat{\pi}_c(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c}\right\}\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\beta}}_c\right) + \left[\frac{\partial\hat{\pi}_c(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\Lambda_0(t)=\widehat{\Lambda}_{0,c}(t)}\right]\Delta_{i,j}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_{0,c}(t)\right\},$$

with

$$\frac{\partial\hat{\pi}_c(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c} = \left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_{0,c}(t)\exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{x}\right)\right\}\exp\left\{-\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_{0,c}(t)\exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{x}\right)\right\}\boldsymbol{x}',$$

$$= \left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_{0,c}(t)\exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{x}\right)\right\}\{1-\hat{\pi}_c(\tau_1,\tau_2;\boldsymbol{x})\}\,\boldsymbol{x}',$$

and

$$\frac{\partial\hat{\pi}_c(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\Lambda_0(t)=\widehat{\Lambda}_{0,c}(t)} = \exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{x}\right)\exp\left\{-\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_{0,c}(t)\exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{x}\right)\right\},$$

$$= \exp\left(\widehat{\boldsymbol{\beta}}_c{}'\boldsymbol{x}\right)\{1-\hat{\pi}_c(\tau_1,\tau_2;\boldsymbol{x})\}.$$

B.3 基于影响函数的方差估计

Reid 和 Crépeau (1985) 估计方差#$\boldsymbol{\beta}$%$_c$&为 $\sum_{=}\sum_{=}\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\beta}\%_c\&\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\beta}\%_c\&$ ^{n} $'$ ($_j$) ! (6 # (。注意其与 Barlow (1994年) 为病例队列设计提出的“稳健方差”的关系。利用整个队列，我们同样可以估计方差#$\boldsymbol{\theta}$%$_c$&，$\boldsymbol{\theta}\%_c \in \{\mathrm{d}\Lambda\%_{0,c}(t),\ \Lambda\%_{0,c}(t),\ \pi{:}_c(\tau_1,\tau_2;\boldsymbol{x})\}$, 通过$\sum_{=}\sum_{=}\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%_c\&\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%_c\&^{n}$ $'$ ($_j$) ! (6 # (。

网络附录 C. 基于分层病例队列设计设计权重的估计

C.1 影响函数的推导

令 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&$ 表示受试者 i 在层 j 中对 $\boldsymbol{\theta}\%$、$i \in \{1, \ldots, {}^{n}(\#)\}$、$j \in \{1, \ldots, J\}$、$\boldsymbol{\theta}\% \in \{\boldsymbol{\beta}\%$、$\mathrm{d}\Lambda\%_0(t)$、$\Lambda\%_0(t)$、$\pi{:}(\tau_1, \tau_2; \boldsymbol{x})\}$的影响。我们可以将估计方程$\sum_{=}^{J} \sum_{=}^{n} \int \mathrm{G}\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1(}{\boldsymbol{S}}$ $\boldsymbol{\beta})$ / "(,;$\boldsymbol{\beta}$/0I ,(j) ! (# ($\mathrm{d}N_{i,j}(t) = 0$ 重写为

$$\boldsymbol{G}_1 - \int_t \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})} \left\{ \sum_{l=1}^{J} \sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l}\, \mathrm{d}N_{k,l}(t) \right\} = 0,$$

其中 $\boldsymbol{G}_1 = \sum_{=}^{J} \sum_{=}^{n} \int_t \xi_{i,j}\, w_{i,j}\, \boldsymbol{X}_{i,j}$ (j) $i1j1\ \mathrm{d}N_{i,j}(t)$，因为 $\xi_{i,j}\, w_{i,j} = 1$ 对于层 j 中的任何受试者 i，满足 $\int \mathrm{d}\ N_{i\,j}\ t\ = 1$,( )t（即，所有病例都纳入第二阶段样本，并且具有单位设计抽样权重）。根据 Graubard 和 Fears（2005），我们有

$$\boldsymbol{\Delta}_{i,j}\{\boldsymbol{G}_1\} + \int_t \left[- \frac{\Delta_{i,j}\left\{ \sum_{l=1}^{J} \sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l}\, \mathrm{d}N_{k,l}(t) \right\} \times \boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})} - \right.$$

$$\left. \frac{\left\{ \sum_{l=1}^{J} \sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l}\, \mathrm{d}N_{k,l}(t) \right\} \times \boldsymbol{\Delta}_{i,j}\{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})\}}{S_0(t;\widehat{\boldsymbol{\beta}})} + \frac{\left\{ \sum_{l=1}^{J} \sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l}\, \mathrm{d}N_{k,l}(t) \right\} \times \boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}}) \times \Delta_{i,j}\{S_0(t;\widehat{\boldsymbol{\beta}})\}}{S_0(t;\widehat{\boldsymbol{\beta}})^2} \right] = 0, \tag{1}$$

with

$$\boldsymbol{\Delta}_{i,j}\{\boldsymbol{G}_1\} = \int_t \xi_{i,j}\, w_{i,j}\, \boldsymbol{X}_{i,j} \mathrm{d}N_{i,j}(t)\ ,$$

$$\Delta_{i,j} \left\{ \sum_{l=1}^{J} \sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l}\, \mathrm{d}N_{k,l}(t) \right\} = \xi_{i,j}\, w_{i,j}\, \mathrm{d}N_{i,j}(t),$$

$$\boldsymbol{\Delta}_{i,j}\left\{ \boldsymbol{S}_1\left(t;\widehat{\boldsymbol{\beta}}\right) \right\} = \xi_{i,j}\, w_{i,j}\, Y_{i,j}(t) \exp\left(\widehat{\boldsymbol{\beta}}' \boldsymbol{X}_{i,j}\right) \boldsymbol{X}_{i,j} + \left\{ \frac{\partial \boldsymbol{S}_1(t,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} \right\} \boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\beta}}\right),$$

$$\frac{\partial \boldsymbol{S}_1(t,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \sum_{l=1}^{J} \sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l}\, Y_{k,l}(t) \exp\left(\widehat{\boldsymbol{\beta}}' \boldsymbol{X}_{k,l}\right) \boldsymbol{X}_{k,l}\, \boldsymbol{X}_{k,l}{}' \equiv \boldsymbol{S}_2\left(t;\widehat{\boldsymbol{\beta}}\right),$$

$$\Delta_{i,j}\{S_0(t;\widehat{\boldsymbol{\beta}})\} = \xi_{i,j}\, w_{i,j}\, Y_{i,j}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}) + \left\{\frac{\partial S_0(t;,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}}),$$

and $\frac{\partial S_0(t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\, w_{k,l}\, Y_{k,l}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l})\,\boldsymbol{X}_{k,l}{}' = \boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'.$

注意，方程 (1) 中方括号内的项是 $\mathrm{d}N_{i,j}(t)$ 的线性组合。因此

$$\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}}) = \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t \xi_{k,l}\, w_{k,l}\left\{\frac{\boldsymbol{S}_2(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'}{S_0(t;\widehat{\boldsymbol{\beta}})^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1} \times \left[\int_t\left\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\times \left\{\xi_{i,j}\, w_{i,j}\,\mathrm{d}N_{i,j}(t) - \xi_{i,j}\, w_{i,j}\, Y_{i,j}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\, w_{k,l}\,\mathrm{d}N_{k,l}(t)}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\right],$$

并且我们可以写作 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\beta}\%\& = \xi_{i,j}\, w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\beta}\%\&$, 其中

$$\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\beta}}) = \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t \xi_{k,l}\, w_{k,l}\left\{\frac{\boldsymbol{S}_2(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'}{S_0(t;\widehat{\boldsymbol{\beta}})^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1}\left[\int_t\left\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\times \left\{\mathrm{d}N_{i,j}(t) - Y_{i,j}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\, w_{k,l}\,\mathrm{d}N_{k,l}(t)}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\right].$$

那么我们知道 $\sum_{\bar{\Sigma}_j}^{J}\sum_{\Sigma^{=}}\{\mathrm{d}^{n} N_{i\,j}\ t\}$, ()(ȷ) ! (# (− dΛ‰(t) × S_0#t; $\boldsymbol{\beta}$%& = 0, 因为我们通过 dΛ‰(t) = ^{d}? 非参数地估计基线风险,ȷ(,)- (ȷ)).#/ .# s"(,;ꞵ/0 . 因此，遵循与$\boldsymbol{\Delta}_i$, ȷ#$\boldsymbol{\beta}$%&类似的论证，我们有 $\mathrm{d}N_{i,j}(t) - \mathrm{d}\Lambda‰(t)\ \Delta_{i,j}\{S_0\#t;\boldsymbol{\beta}\%\&\} - S_0\#t;\boldsymbol{\beta}\%\&\ \Delta_{i,j}\{\mathrm{d}\Lambda‰(t)\} = 0$. 换言之,

$$\mathrm{d}N_{i,j}(t) - \mathrm{d}\widehat{\Lambda}_0(t)\,\xi_{i,j}\, w_{i,j}Y_{i,j}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}) - \mathrm{d}\widehat{\Lambda}_0(t)\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}}) - S_0(t;\widehat{\boldsymbol{\beta}})\Delta_{i,j}\{\mathrm{d}\widehat{\Lambda}_0(t)\} = 0,$$

因此、】【

$$\Delta_{i,j}\{\mathrm{d}\widehat{\Lambda}_0(t)\} = \{S_0(t;\widehat{\boldsymbol{\beta}})\}^{-1}\left[\mathrm{d}N_{i,j}(t) - \mathrm{d}\widehat{\Lambda}_0(t)\,\xi_{i,j}\, w_{i,j}\, Y_{i,j}(t)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}) - \mathrm{d}\widehat{\Lambda}_0(t)\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}})\right].$$

因此，我们可以写作 $\Delta_{i,j}\{\mathrm{d}\Lambda\%0(t)\} = \xi_{i,j}\, w_{i,j}\, IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%0(t)\}$，其中 $IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%0(t)\} = \{S_0\#t;\ \boldsymbol{\beta}\%\&\}^{-1}\mathrm{Z}\mathrm{d}N_{i,j}(t) - \mathrm{d}\Lambda\%0(t)\, Y_{i,j}(t)\exp\#\boldsymbol{\beta}\%'\boldsymbol{X}_{i,j}\& - \mathrm{d}\Lambda\%0(t)\, \boldsymbol{S}_1\#t;\ \boldsymbol{\beta}\%\&'\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\beta}\%\&\}$, , 当 $\xi_{i,j}w_{i,j}$ =1 if $\mathrm{d}N_{i,j}(t)$= 1。

$IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%0(t)\}$ 和 $\Delta_{i,j}\{\mathrm{d}\Lambda\%0(t)\}$ 都是增量 $\mathrm{d}N_{i,j}(t)$ 和 $\mathrm{d}\Lambda\%0(t)$ 的线性组合。因此 $\Delta_{i,j}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\ \mathrm{A}\#\} = \int\ \Delta_{i,j}\{\mathrm{d}\Lambda\%0(t)\}\ \mathrm{A}\#\ \mathrm{A}!\ \mathrm{A}\#$，我们可以将其重写为$\xi_{i,j}\, w_{i,j}\, IF_{i,j}^{(2)}\ \mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\ \mathrm{A}\#\}$, , 其中 $IF_{i,j}^{(2)}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\ \mathrm{A}\#\} = \int\ IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%0(t)\}\ \mathrm{A}!\ \mathrm{A}\#$。最后，

$$\Delta_{i,j}\{\hat{\pi}(\tau_1,\tau_2;\ \boldsymbol{x})\} = \left\{\frac{\partial\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}}) + \left[\frac{\partial\hat{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\mathrm{d}\Lambda_0(t)=\mathrm{d}\widehat{\Lambda}_0(t)}\right]\Delta_{i,j}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_0(t)\right\},$$

且我们可以将 $\Delta_{i,\ j}\{\pi{:}(\tau_1,\tau_2;\ \boldsymbol{x})\} = \xi_{i,\ j}\, w_{i,\ j}\, IF_{i,\ j}^{(2)}\{\pi{:}(\tau_1,\tau_2;\ \boldsymbol{x})\}$ 重写为

$$\frac{\mathrm{BCD}(\mathrm{A}_1,\mathrm{A}_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}{}_{\tau_1}\exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\Big\}\exp\mathrm{Z}-\int \mathrm{d}\Lambda\%0(t)^{\tau_2}{}_{\tau_1}\exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\Big\}\boldsymbol{x}',$$

$$= \mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}{}_{\tau_1}\exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\Big\}\{1-\pi{:}(\tau_1,\tau_2;\ \boldsymbol{x})\}\,\boldsymbol{x}',\ \text{以及}$$

$$\frac{\mathrm{BCD}(\mathrm{A}_1,\mathrm{A}_2;\boldsymbol{x})}{\partial\int_{\tau_1}^{\tau_2}\mathrm{dI}0(,)}_{|\mathrm{dI}0(,)=\mathrm{dI}\ \widehat{}_0(t)} = \exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\ \exp\mathrm{Z}-\int \mathrm{d}\Lambda\%0(t)^{\tau_2}{}_{\tau_1}\exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\Big\},$$

$$= \exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\{1-\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\}.$$

注意，对于任意 $\boldsymbol{\theta}\% \in \{\boldsymbol{\beta}\%,\ \mathrm{d}\Lambda\%0(t),\ \Lambda\%_0(t),\ \pi{:}(\tau_1,\tau_2;\ \boldsymbol{x})\}$，如果第 i 个层中的受试者不属于分层病例队列，那么 $\boldsymbol{\Delta}_{i,\ j}\#\boldsymbol{\theta}\%\&$ 为零，$i \in \{1,\ldots,{}^{n}{}_{(j)}\}, j \in \{1,\ldots,J\}$。

C.2 与 Samuelsen、Ånestad 和 Skrondal（2007年）方差的比较

尽管他们没有提及影响函数，但主文档中第3.3节方程（14）给出的 $\boldsymbol{\beta}\%$ 的方差估计与 Samuelsen 等人（2007年）的结果非常接近。两者有两个细微差异。首先，Samuelsen 等人（2007年）沿用了 Borgan 等人（2000年）的做法，通过排除病例重新定义了各层；而我们没有这样做，非病例层特异性抽样设计

因此，权重取决于每层抽取的个体数，而非每层实际抽取的非病例数。接下来，Samuelsen 等人（2007年）提出的第二阶段分量，经略微修改其符号说明以使其更接近我们的表示法后，为

$$\sum_{j=1}^{J} \frac{\tilde{m}^{(j)}\left(1-\frac{\tilde{m}^{(j)}}{\tilde{n}^{(j)}}\right)}{\tilde{m}^{(j)}-1} \sum_{i=1}^{\tilde{n}^{(j)}} \left(\boldsymbol{D}_{i,j}-\bar{\boldsymbol{D}}_{j}\right)\left(\boldsymbol{D}_{i,j}-\bar{\boldsymbol{D}}_{j}\right)',$$

其中，$n\mathrm{o}_{(j)}$表示移除所有病例后层 j 的大小，$m\mathrm{n}_{(j)}$表示在层 j抽取的$m_{(j)}$个体中抽取的非病例数量。$\boldsymbol{D}_{i,j}$ 是我们的$\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\beta}\%\&$、$\bar{\boldsymbol{D}}_{j} = {}_{1}\, m^{\sim}{}_{(j)} \sum_{=} \boldsymbol{D}_{i,}^{\tilde{n}}{}_{j(j)\, i1}$ 的对应量。该量等于

$$\sum_{j=1}^{J} \frac{\tilde{m}^{(j)}\left(1-\frac{\tilde{m}^{(j)}}{\tilde{n}^{(j)}}\right)}{\tilde{m}^{(j)}-1} \left(\sum_{i=1}^{\tilde{n}^{(j)}} \boldsymbol{D}_{i,j}\boldsymbol{D}_{i,j}{}' - \frac{1}{\tilde{m}^{(j)}} \sum_{i=1}^{\tilde{n}^{(j)}} \sum_{k=1}^{\tilde{n}^{(j)}} \boldsymbol{D}_{i,j}\boldsymbol{D}_{k,j}{}' \right)$$

$$= \sum_{j=1}^{J} \frac{\tilde{m}^{(j)}\left(1-\frac{\tilde{m}^{(j)}}{\tilde{n}^{(j)}}\right)}{\tilde{m}^{(j)}-1} \left(\frac{\tilde{m}^{(j)}-1}{\tilde{m}^{(j)}} \sum_{i=1}^{\tilde{n}^{(j)}} \boldsymbol{D}_{i,j}\boldsymbol{D}_{i,j}{}' - \frac{1}{\tilde{m}^{(j)}} \sum_{i=1}^{\tilde{n}^{(j)}} \sum_{\substack{k=1, \\ k\neq i}}^{\tilde{n}^{(j)}} \boldsymbol{D}_{i,j}\boldsymbol{D}_{k,j}{}' \right),$$

$$= \sum_{j=1}^{J} \left\{ \left(1-\frac{\tilde{m}^{(j)}}{\tilde{n}^{(j)}}\right) \sum_{i=1}^{\tilde{n}^{(j)}} \boldsymbol{D}_{i,j}\boldsymbol{D}_{i,j}{}' - \frac{1}{\tilde{m}^{(j)}-1} \left(1-\frac{\tilde{m}^{(j)}}{\tilde{n}^{(j)}}\right) \sum_{i=1}^{\tilde{n}^{(j)}} \sum_{\substack{k=1, \\ k\neq i}}^{\tilde{n}^{(j)}} \boldsymbol{D}_{i,j}\boldsymbol{D}_{k,j}{}' \right\}.$$

请注意，方差的第二阶段分量中的求和可简化为仅对抽取的非病例求和，但在主文档的公式（3）中，$\left(1-\mathrm{cV}_{(j)}\, 4\mathrm{f}_{(j)}\right)$ 和 $(\mathrm{cV}_{(j)}\text{–}(\left(1-\mathrm{cV}_{(j)}\, 4\mathrm{f}_{(j)}\right)$ 分别应为 $\left(1-\mathrm{c}_{(j)}\, 4_{(j)}\right)$ 和 $(\mathrm{c}_{(j)}\text{–}(\left(1-\mathrm{c}_{(j)}\, 4_{(j)}\right)$。

其次，我们使用影响函数来估计 $\boldsymbol{\beta}\%$ 的第一阶段（超总体）方差分量，而 Samuelsen 等人（2007年）则使用观测信息矩阵的逆矩阵。

网络附录 D. 使用校准权重对分层病例队列设计进行估计

D.1 用于校准的估计方程

令 $\boldsymbol{A}_{i,\ j}$ 为层 j 中个体 i 的 q 辅助变量向量，$i \in \{1, \ldots, n_{(j)}\}, j \in \{1, \ldots, J\}$。按照 Breslow 等人（2009a）和 Shin等人（2020），我们从以下内容中获得辅助变量：（*i*）队列中每位个体的取值均为 1 的变量；（*ii*）基于插补后的队列数据拟合的 Cox模型所得对数相对风险参数的影响函数；以及（*iii*）插补后的队列数据中总随访时间（纯风险待估计的时间区间内）与相对风险的乘积，其中对数相对风险参数是通过使用以（*ii*）校准的权重，对分层病例队列数据拟合 Cox模型而估计得到的。插补后的队列数据是利用第一阶段数据为队列所有成员插补协变量值（对于仅在第二阶段测量的协变量）而得到的。有关辅助变量构造的详情，请参见 Breslow 等人（2009a）和 Shin等人（2020）。

我们希望使用校准权重 $\#w^{*}_{i,j}\&_{i\in\{1,\ldots,n}(j)\mathrm{b},\#\epsilon\{(,\ldots,6\}$，使其尽可能接近设计权重 $\#w_{i,j}\&_{i\in\{1,\ldots,n}(j)\mathrm{b},\#\epsilon\{(,\ldots,6\}$，同时能够正确估计辅助变量的总量（即使得 $\sum^{J}_{=}\ \sum^{n}_{=}\ \xi_{i\,j}\ w^{*}_{i\,j}\ \boldsymbol{A}_{i\,j},\ ,\ (j)\ !\ (\ ,\ \#\ (=\sum_{=}\ \sum_{=}\ \boldsymbol{A}_{i,}{}^{n}\ \ \ \ j(j)\ !\ (6\ \#\ (\ )$）。换言之，对于给定的距离 δ，我们希望在约束条件 $\sum_{=}\ \sum(^{n}_{=}\ \ \xi_{i\,j}\ w^{*}_{i\,j},\ ,\ (j)\ !\ (\ \boldsymbol{A}_{i,j}-\boldsymbol{A}_{i,j})^{\ J}\ \#\ (=0$ 下，最小化 $\sum_{=}\ \sum_{=}\ \xi_{i,}{}^{n}\ \ \ \ j(j)\ !\ (\ \delta(w^{*}_{i,j},w_{i,j})^{\ J}\ \#\ ($。为求解这个约束优化问题，我们考虑 $\boldsymbol{\eta}$—— 一个 q 拉格朗日乘子向量，并寻找拉格朗日函数的驻点

$L\colon \left(w_{1,1}^{*}, \ldots, w_{n^{(J)},}{}^{*}{}_{J}, \eta\right) \mapsto \sum_{=}^{J} \sum\{_{=}^{n} \xi_{i\, j,(J)\, i1}\ \delta(w_{i,j}^{*}, w_{i,j}) + \boldsymbol{\eta}'(\xi_{i,j}\ w_{i,j}^{*}\boldsymbol{A}_{i,j} - \boldsymbol{A}_{i,j})\}_{\ j1}$。为此，我们求解以下方程，它们由 L 的梯度得到：

$$\begin{cases} \xi_{1,1}\left\{\dfrac{\partial\delta(w_{1,1}^{*}, w_{1,1})}{\partial w_{1,1}^{*}} + \boldsymbol{\eta}'\boldsymbol{A}_{1,1}\right\} = 0, \\ \vdots \\ \xi_{n^{(J)},J}\left\{\dfrac{\partial\delta(w_{n^{(J)},J}^{*}, w_{n^{(J)},J})}{\partial w_{n^{(J)},J}^{*}} + \boldsymbol{\eta}'\boldsymbol{A}_{n^{(J)},J}\right\} = 0, \\ \sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}(\xi_{i,j}\ w_{i,j}{}^{*}\,\boldsymbol{A}_{i,j} - \boldsymbol{A}_{i,j}) = 0. \end{cases}$$

我们使用 $\delta\colon (a,b) \mapsto a\log\left(\frac{a}{b}\right) + b - a$ 作为距离度量，这将导出迭代比例拟合程序（Deville 和 Sarndal，1992；Breslow 等人，2009），并求解估计方程

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\{\xi_{i,j}\ w_{i,j}\exp(\boldsymbol{\eta}'\boldsymbol{A}_{i,j})\boldsymbol{A}_{i,j} - \boldsymbol{A}_{i,j}\} = 0.$$

由估计值 $\boldsymbol{\eta}$a，我们进而得到校准权重 $w_{i,}{}^{*}{}_{j} = w_{i,\ j}\exp(\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,\ j}),\quad i \in \{1, \ldots, n^{(j)}\}, j \in \{1, \ldots, J\}.$

D.2 影响函数的推导

令 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%_{*}\&$ 表示第 i 个受试者在第 j 层中对 $\boldsymbol{\theta}\%_{*},\ i \in \{1, \ldots, {}^{n}{}_{(j)}\}, j \in \{1, \ldots, J\}, \boldsymbol{\theta}\%_{}^{*} \in \{\boldsymbol{\eta}\mathrm{a},\ \boldsymbol{\beta}\%_{*},\ \mathrm{d}\Lambda\%_{0}^{*}(t),\ \Lambda\%_{0}^{*}(t),\ \pi\colon{}^{*}(\tau_1, \tau_2;\ \boldsymbol{x})\}$ 的影响。由于 $\sum_{=}^{J}\ \sum_{=}\{\xi_{i,}{}^{n}{}_{\ j}\ w_{i\,j}\quad \boldsymbol{\eta}'\boldsymbol{A}_{i\,j}\ \ \boldsymbol{A}_{i\,j} - \boldsymbol{A}_{i\,j}\}$,exp#a,&$_{,(j)\ i1j1} = 0$，因此遵循 Graubard 和 Fears (2005)

$$\boldsymbol{A}_{i,j} - \xi_{i,j}\ w_{i,j}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)\boldsymbol{A}_{i,j} - \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\ w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\boldsymbol{A}_{k,l}\boldsymbol{A}_{k,l}{}'\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}}) = 0,$$

因此

$$\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}}) = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\ w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\boldsymbol{A}_{k,l}\boldsymbol{A}_{k,l}{}'\right\}^{-1}\left\{\boldsymbol{A}_{i,j} - \xi_{i,j}\ w_{i,j}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)\boldsymbol{A}_{i,j}\right\}.$$

于是我们可以写成

$$\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}}) = \boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\eta}}) + \xi_{i,j}\ w_{i,j}\ \boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\eta}}),$$

with $\boldsymbol{IF}_{i,j}^{(1)}(\widehat{\boldsymbol{\eta}}) = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) \boldsymbol{A}_{k,l}\boldsymbol{A}_{k,l}'\right\}^{-1} \boldsymbol{A}_{i,j}$,

and $\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\eta}}) = -\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) \boldsymbol{A}_{k,l}\boldsymbol{A}_{k,l}'\right\}^{-1} \boldsymbol{A}_{i,j}$.

于是我们可以写出估计方程

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}} \int_t \xi_{i,j}\, w_{i,j} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)\left\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}\right\} \mathrm{d}N_{i,j}(t) = 0,$$

as

$$\boldsymbol{G}_1^*(\widehat{\boldsymbol{\eta}}) - \int_t \frac{\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\, \mathrm{d}N_{k,l}(t)\right\} = 0,$$

with $\boldsymbol{G}_1^*(\widehat{\boldsymbol{\eta}}) = \sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}} \int_t \xi_{i,j}\, w_{i,j} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right) \mathrm{d}N_{i,j}(t)\, \boldsymbol{X}_{i,j}$.

按照与 $\boldsymbol{\Delta}_{i,j}(\boldsymbol{\eta}\mathrm{a})$ 类似的论证，我们有

$$\boldsymbol{\Delta}_{i,j}\{\boldsymbol{G}_1^*(\widehat{\boldsymbol{\eta}})\} + \int_t \left[-\frac{\Delta_{i,j}\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) \mathrm{d}N_{k,l}(t)\right\}\times \boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)} - \right.$$

$$\frac{\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) \mathrm{d}N_{k,l}(t)\right\}\times \boldsymbol{\Delta}_{i,j}\{\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\}}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)} + \tag{2}$$

$$\left.\frac{\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) \mathrm{d}N_{k,l}(t)\right\}\times \boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\times \Delta_{i,j}\{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\}}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)^2}\right] = 0,$$

with

$$\boldsymbol{\Delta}_{i,j}\{\boldsymbol{G}_1^*(\widehat{\boldsymbol{\eta}})\} = \int_t \xi_{i,j}\, w_{i,j} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right) \mathrm{d}N_{i,j}(t)\, \boldsymbol{X}_{i,j} + \left\{\frac{\partial \boldsymbol{G}_1^*(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}}),$$

$$\Delta_{i,j}\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\, \mathrm{d}N_{k,l}(t)\right\} = \xi_{i,j}\, w_{i,j} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right) \mathrm{d}N_{i,j}(t) +$$

$$\left[\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, w_{k,l} \exp\left(\boldsymbol{\eta}'\boldsymbol{A}_{k,l}\right) \mathrm{d}N_{k,l}(t)\right\}}{\partial \boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right]\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}}),$$

$$\frac{\partial \boldsymbol{G}_1^*(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \int_t \xi_{k,l}\, w_{k,l} \exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right) \boldsymbol{X}_{k,l}\boldsymbol{A}_{k,l}'\, \mathrm{d}N_{k,l}(t),$$

$$\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp(\boldsymbol{\eta}'\boldsymbol{A}_{k,l})\,\mathrm{d}N_{k,l}(t)\right\}}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}=\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l})\,\boldsymbol{A}_{k,l}{}'\,\mathrm{d}N_{k,l}(t),$$

$$\boldsymbol{\Delta}_{i,j}\left\{\boldsymbol{S}_1^*\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*\right)\right\}=\xi_{i,j}\,w_{i,j}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{i,j}\right)\boldsymbol{X}_{i,j}+\left\{\frac{\partial\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^*}\right\}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\beta}})+$$

$$\left\{\frac{\partial\boldsymbol{S}_1^*(t;\boldsymbol{\eta},\widehat{\boldsymbol{\beta}}^*)}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}}),$$

$$\frac{\partial\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^*}=\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l})\,Y_{k,l}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{k,l}\right)\boldsymbol{X}_{k,l}\,\boldsymbol{X}_{k,l}{}'\equiv\boldsymbol{S}_2^*\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*\right),$$

$$\frac{\partial\boldsymbol{S}_1^*(t;\boldsymbol{\eta},\widehat{\boldsymbol{\beta}}^*)}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}=\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l})\,Y_{k,l}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{k,l}\right)\boldsymbol{X}_{k,l}\,\boldsymbol{A}_{k,l}{}',$$

$$\Delta_{i,j}\left\{S_0^*\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*\right)\right\}=\xi_{i,j}\,w_{i,j}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{i,j}\right)+\left\{\frac{\partial S_0^*(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^*}\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}}^*)+$$

$$\left\{\frac{\partial S_0^*(t;\boldsymbol{\eta},\widehat{\boldsymbol{\beta}}^*)}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}}),$$

$$\frac{\partial S_0^*(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^*}=\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)Y_{k,l}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{k,l}\right)\boldsymbol{X}_{k,l}{}'=\boldsymbol{S}_1^*\left(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*\right)',$$

and $\frac{\partial S_0^*(t;\boldsymbol{\eta},\widehat{\boldsymbol{\beta}}^*)}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}=\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l})\,Y_{k,l}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{k,l}\right)\boldsymbol{A}_{k,l}{}'.$

Note, $\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\mathrm{d}N_{k,l}(t)$, $\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp(\boldsymbol{\eta}'\boldsymbol{A}_{k,l})\,\mathrm{d}N_{k,l}(t)\right\}}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}$ and thus

$\Delta_{i,j}\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\mathrm{d}N_{k,l}(t)\right\}$ and the component between square brackets in Equation (2) above are linear combinations of the increments $\mathrm{d}N_{i,j}$. Thus

$$\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\beta}})=\left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t\xi_{k,l}\,w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\left\{\frac{\boldsymbol{S}_2^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}-\frac{\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)'}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1}\times$$

$$\left[\int_t\left\{\boldsymbol{X}_{i,j}-\frac{\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}\right\}\times\left\{\xi_{i,j}\,w_{i,j}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)\mathrm{d}N_{i,j}(t)-\right.\right.$$

$$\left.\left.\xi_{i,j}\,w_{i,j}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{i,j}\right)\frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l})\,\mathrm{d}N_{k,l}(t)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}\right\}\right]+$$

$$\left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t \xi_{k,l}\, w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\left\{\frac{\boldsymbol{S}_2^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}-\frac{\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)'}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1}\times\left(\left\{\frac{\partial \boldsymbol{G}_1^*(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right\}+\right.$$

$$-\int_t\frac{\boldsymbol{S}_1^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}\times\left[\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp\left(\boldsymbol{\eta}'\boldsymbol{A}_{k,l}\right)\mathrm{d}N_{k,l}(t)\right\}}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right]-\int_t\frac{\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\mathrm{d}N_{k,l}(t)\right\}}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)}\times$$

$$\left.\left\{\frac{\partial\boldsymbol{S}_1^*(t;\boldsymbol{\eta},\widehat{\boldsymbol{\beta}}^*)}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right\}+\int_t\frac{\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,w_{k,l}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{k,l}\right)\mathrm{d}N_{k,l}(t)\right\}\tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})}{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)^2}\times\left\{\frac{\partial S_0^*(t;\boldsymbol{\eta},\widehat{\boldsymbol{\beta}}^*)}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right\}\right)\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\gamma}}),$$

我们可以写成 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\beta}\%_*\&=\boldsymbol{IF}_{i,j}^{(1)}\#\boldsymbol{\beta}\%_*\&+\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\beta}\%_*\&$，其中

$\boldsymbol{IF}_{i,j}^{(1)}\#\boldsymbol{\beta}\%_*\& = \mathrm{Z}\Sigma\quad\Sigma\quad\xi_{k,l}\,w_{k,}3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\boldsymbol{Z}_{k,l}\boldsymbol{A}_{k,}3'^{\,n}{}_{(j)\;k=1J\;l=1}\Big\}\times\boldsymbol{IF}_{i,j}^{(1)}(\boldsymbol{\eta}\mathrm{a})$, 和

$\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\beta}\%_*\& = \exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,j}\&\boldsymbol{Z}_{i,j}+\mathrm{Z}\Sigma\quad\Sigma\quad\xi_{k,l}\,w_{k,l}\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,l}\&\,\boldsymbol{Z}_{k,l}\boldsymbol{A}_{k,l'}{}^{n}{}_{(j)\;k=1J\;l=1}\Big\}\times\boldsymbol{IF}_{i,j}^{(2)}(\boldsymbol{\eta}\mathrm{a})$,

其中 $\boldsymbol{Z}_{i,j}=\Big[\sum_{l=}^{J}\;\sum_{k=}^{n}\;\int\xi_{k,l}\,w_{k,3}\exp\left(\boldsymbol{\eta}^{\frown\prime}\boldsymbol{A}_{k,}{}^{l}\right)_{,(l)}((\Big\{\frac{S^*(\;\eta\beta^*)}{S}{}_{2,;\mathrm{D},/0}^{*}(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/{}_{*)}-$

$\frac{\boldsymbol{S}_1^*(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/^*)\boldsymbol{S}_1^*(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/^*)'}{S_0^*(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/^*)2}\mathrm{Id}N_{k,l}(t)\mathrm{K}^{\;-1}\;\mathrm{L}\int\mathrm{G}\;\;-\frac{S^*(\;\eta\beta^*)_{1,;}}{S_{\mathrm{D},/0}^{*}(,;\boldsymbol{\eta}\mathrm{D},\boldsymbol{\beta}/^*)}\Big\}\;\times\mathrm{N}\,\xi_{i,\;j}\,w_{i,\;j}\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,\;j}\&\,\mathrm{d}$

$\xi_{i,j}\,w_{i,j}\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,j}\&\,Y_{i,j}(t)\exp\left(\boldsymbol{\beta}\%^{*\prime}\boldsymbol{X}_{i,j}\right)\frac{\Sigma\quad\Sigma\quad\xi_{k,l}\,wk,l\,\mathrm{e}\mathrm{x}\mathrm{p}\left(\boldsymbol{\eta}^{\frown\prime}\;\boldsymbol{A}kl\right)\mathrm{d}_{Nk,l(t)}}{n\,(l)\quad k=1J\;\;l=1\quad\;\;S_0^*(t;\boldsymbol{\eta}^{\frown},\boldsymbol{\beta}^{\frown}{}_*)}\;\mathrm{OP}.$

估计方程 $\sum_{=}^{J}\;\sum_{=}^{n}\;\{\mathrm{d}N_{i,j}(t)\}-\mathrm{d}\Lambda\%_0^*(t)\;{}_{(j)\,i1}\times S_0^*\#t;\;\boldsymbol{\eta}\mathrm{a},\;\boldsymbol{\beta}\%_*\&\;{}_{j1}=0$ 得到 Breslow

估计 $\mathrm{d}\Lambda\%_0^*(t)\;=\frac{\Sigma_j\quad\Sigma\quad\mathrm{d}_{Ni,j(t)n\,(j)\;i=1J}}{=1\,S_0^*(t;\boldsymbol{\eta}^{\frown},\boldsymbol{\beta}^{\frown}{}_*)}$。因此

$$\mathrm{d}N_{i,j}(t)-\mathrm{d}\widehat{\Lambda}_0^*(t)\,\Delta_{i,j}\{S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\}-S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\,\Delta_{i,j}\{\mathrm{d}\widehat{\Lambda}_0^*(t)\}=0,$$

换句话说,

$$\mathrm{d}N_{i,j}(t)-\mathrm{d}\widehat{\Lambda}_0^*(t)\,\xi_{i,j}\,w_{i,j}\exp\left(\widehat{\boldsymbol{\eta}}'\boldsymbol{A}_{i,j}\right)Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}_{i,j}\right)-\mathrm{d}\widehat{\Lambda}_0^*(t)\left\{\frac{\partial S_0^*(t;\widehat{\boldsymbol{\eta}},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^*}\right\}'\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}}^*)$$

$$-\mathrm{d}\widehat{\Lambda}_0^*(t)\left\{\frac{\partial S_0^*(t;\boldsymbol{\eta},\widehat{\boldsymbol{\beta}}^*)}{\partial\boldsymbol{\eta}}_{|\boldsymbol{\eta}=\widehat{\boldsymbol{\eta}}}\right\}'\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\eta}})-S_0^*(t;\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\beta}}^*)\,\Delta_{i,j}\{\mathrm{d}\widehat{\Lambda}_0^*(t)\}=0.$$

因此

$$\Delta_{i,j}\{\mathrm{d}\Lambda\%_0^*(t)\} = \{S_0^*\#t;\ \boldsymbol{\eta}\mathrm{a}, \boldsymbol{\beta}\%_*\&\}^{-1}\mathrm{D}\mathrm{d}N_{i,j}(t)\ -\mathrm{d}\Lambda\%_0^*(t)\,\xi_{i,j}\,w_{i,j}\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,j}\&Y_{i,j}(t)\exp\left(\boldsymbol{\beta}\%^{*'}\boldsymbol{X}_{i,j}\right) -\mathrm{d}\Lambda\%_0^*(t)\,\boldsymbol{S}_1^*\#t;\boldsymbol{\eta}\mathrm{a},\boldsymbol{\beta}\%_*\&'\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\beta}\%_*\& \ -\mathrm{d}\Lambda\%_0^*(t)\mathrm{Z}\sum\ \ \sum\ \ \xi_k,$$

$$_{k=}^{n}\ _l\,w_{k\,l}\ \ \ \boldsymbol{\eta}'\boldsymbol{A}_{k\,l}\ \ Y_{k\,l}\ \ t\ \ \ \left(\boldsymbol{\beta}^{*'}\boldsymbol{X}_{k\,l}\right)\boldsymbol{A}_{k\,l},\exp(\mathrm{a}\ ,\ ),\ (\ )\exp\%\ ,\ ,\ (')\ (6\ 35(\Big\}'\boldsymbol{\Delta}_{i,j}(\boldsymbol{\eta}$$

a)K,因此我们可以改写$\Delta_{i,j}\{\mathrm{d}\Lambda\%_0^*(t)\} = IF_{i,j}^{(1)}\left(\mathrm{d}\Lambda\%_0^*(t)\right) + \xi_{i,j}\,w_{i,j}\,IF_{i,j}^{(2)}\left(\mathrm{d}\Lambda\%_0^*(t)\right)$, 其中$IF_i$,

$$_j^{(1)}\{\mathrm{d}\Lambda\%_0^*(t)\} = \mathrm{Z}\sum\ \ \sum\ \ \xi_{k,l}\,w_{k,}3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\,H_{k,}3(t)\boldsymbol{A}_{k,}3'^{\,n}(')\ 25(635(\Big\}\times\boldsymbol{IF}_{i,j}^{(1)}(\boldsymbol{\eta}\mathrm{a}),IF_i,$$

$$_j^{(2)}\{\mathrm{d}\Lambda\%_0^*(t)\} = \{S_0^*\#t;\boldsymbol{\eta}\mathrm{a},\boldsymbol{\beta}\%_*\&\}^{-1}\mathrm{d}N_{i,j}(t)\ +\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{i,j}\&\,H_{i,j}(t)+\mathrm{Z}\sum\ \ \sum\ \ \xi_{k,l}\,w_{k,}$$

$$3\exp\#\boldsymbol{\eta}\mathrm{a}'\boldsymbol{A}_{k,}3\&\,H_{k,}3(t)\ \boldsymbol{A}_{k,}3'^{n}(')\ 25(6\ 35(\Big\}\times\boldsymbol{IF}_{i,j}^{(2)}(\boldsymbol{\eta}\mathrm{a}),\ \text{其中}\xi_{i,j}w_{i,j}=1\ \text{若}\ \mathrm{d}N_{i,j}(t)=1,\ \text{以}$$

及 $H_{i,j}(t)= -\{S_0^*\#t;\boldsymbol{\eta}\mathrm{a},\ \boldsymbol{\beta}\%_*\&\}^{-1}\mathrm{d}\Lambda\%_0^*(t)\ \mathrm{Z}\boldsymbol{S}_1^*\#t;\boldsymbol{\eta}\mathrm{a},\ \boldsymbol{\beta}\%_*\&'\boldsymbol{Z}_{i,j}+K_{i,j}(t)\Big\}$ 以及 $K_{i,j}(t)=Y_{i,j}(t)$ $\exp\#\boldsymbol{\beta}\%^{*'}\boldsymbol{X}_{i,j}\&$.

注意，$H_{i,j}(t)$，因此 $IF_{i,j}^{(1)}\{\mathrm{d}\Lambda\%_0^*(t)\}$ 和$IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%_0^*(t)\}$，是增量 $\mathrm{d}N_{i,j}(t)$ 和 $\mathrm{d}\Lambda\%_0^*(t)$ 的线性组合。因此$\Delta_{i,j}\,\mathrm{Z}\int\mathrm{d}\Lambda\%_0^*(t)^{\tau_2}\,\mathrm{A}\#\Big\} = \int\ \ \Delta_{i,j}\{\mathrm{d}\Lambda\%_0^*(t)\}\,\mathrm{A}!\mathrm{A}\#$，我们可以将其重写为$IF_{i,j}^{(1)}$ $\mathrm{Z}\int\mathrm{d}\Lambda\%_0^*(t)^{\tau_2}\,\mathrm{A}\#\Big\} + \xi_{i,j}\,w_{i,j}\,IF_{i,j}^{(2)}\,\mathrm{Z}\int\mathrm{d}\Lambda\%_0^*(t)^{\tau_2}\,\mathrm{A}\#\Big\}$，，其中$IF_{i,j}^{(1)}\{\Lambda\%_0^*(\tau_2)\} = \int\ \ IF_{i,j}^{(1)}\{\mathrm{d}\Lambda$ $\%_0^*(t)\}\,\mathrm{A}!\mathrm{A}\#$，以及$IF_{i,j}^{(2)}\{\Lambda\%_0^*(\tau_2)\} = \int\ \ IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%_0^*(t)\}\,\mathrm{A}!\ \mathrm{A}\#$。

最后

$$\Delta_{i,j}\{\hat{\pi}^*(\tau_1,\tau_2;\boldsymbol{x})\} = \left\{\frac{\partial\hat{\pi}^*(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^*}\right\}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\beta}}^*) + \left[\frac{\partial\hat{\pi}^*(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\mathrm{d}\Lambda_0(t)=\mathrm{d}\widehat{\Lambda}_0^*(t)}\right]\Delta_{i,j}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_0^*(t)\right\},$$

$$\text{with}\ \frac{\partial\hat{\pi}^*(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^*} = \left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widehat{\Lambda}_0^*(t)\exp\left(\widehat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{x}\right)\right\}\{1-\hat{\pi}^*(\tau_1,\tau_2;\boldsymbol{x})\}\boldsymbol{x}',$$

$$\text{and}\ \frac{\partial\hat{\pi}^*(\tau_1,\tau_2;\boldsymbol{x})}{\partial\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)}_{|\mathrm{d}\Lambda_0(t)=\mathrm{d}\widehat{\Lambda}_0^*(t)} = \exp\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{x}\right)\{1-\hat{\pi}^*(\tau_1,\tau_2;\boldsymbol{x})\},$$

并且我们可以将 $\Delta_{i,j}\{\pi: {}^{*}(\tau_1,\tau_2;\boldsymbol{x})\}$ 改写为 $IF_{i,j}^{(1)}\{\pi: {}^{*}(\tau_1,\tau_2;\boldsymbol{x})\}+\xi_{i,j}\,w_{i,j}\,IF_{i,j}^{(2)}\{\pi: {}^{*}(\tau_1,\tau_2;\boldsymbol{x})\}$。

D.3 关于辅助变量选择的评论

如主文档第4.3节所述，对于任意 $\boldsymbol{\theta}\%^{*}\in\{\boldsymbol{\eta}\mathrm{a}、\ \boldsymbol{\beta}\%^{*}、\ \mathrm{d}\Lambda\%_0^{*}(t)、\ \Lambda\%_0^{*}(t)、\ \pi: {}^{*}(\tau_1,\tau_2;\boldsymbol{x})\}$，我们通过以下方式估计 $\mathrm{var}\#\boldsymbol{\theta}\%^{*}\&$:

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\Big\{\boldsymbol{IF}_{i,j}^{(1)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)\boldsymbol{IF}_{i,j}^{(1)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)' + 2\,\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)\boldsymbol{IF}_{i,j}^{(2)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)' +$$

$$\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)\boldsymbol{IF}_{i,j}^{(2)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)'\Big\} +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}} w_{i,k,j}\,\sigma_{i,k,j}\,w_{i,j}\,w_{k,j}\xi_{i,j}\,\xi_{k,j}\boldsymbol{IF}_{i,j}^{(2)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)\boldsymbol{IF}_{k,j}^{(2)}\big(\widehat{\boldsymbol{\theta}}^{*}\big)',$$

其中，当层 j 中的个体 i 和 k 均为非病例时，$\sigma_{i,2,j}=\mathrm{c}\,(_j)\,4\,(_j)\,\mathrm{c}\,(_j)$–$(\,4\,(_j)$–$(\,-\binom{m}{2}\,(_j)\,4\,(_j)\big)$；否则为 $\sigma_{i,2,j}=0$；并且，当层 j 中的个体 i 为非病例时，$\sigma_{i,i,j}=\mathrm{c}\,(_j)\,4\,(_j)\,\big(1-\mathrm{c}\,(_j)\,4\,(_j)\big)$，否则为 $\sigma_{i,i,j}=0$；$i,k\in\{1,\dots,n(\#)\}$，$j\in\{1,\dots,J\}$。

注意，网络附录 D.2 中给出的 $\boldsymbol{IF}_{i,j}^{(1)}\#\boldsymbol{\beta}\%^{*}\&$ 是使用校准权重 $w_{i,j}{}^{*}$，$i\in\{1,\dots,n(\#)\}, j\in\{1,\dots,J\}$，将 $\xi_{i,j}\boldsymbol{Z}_{i,j}$ 对 $\xi_{i,j}\boldsymbol{A}_{i,j}$ 进行加权线性回归模型拟合后，得到的第 i 个个体在第 j 层中的预测值。于是，$\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\beta}\%^{*}\&$ 是该加权线性回归的第 $\sum_{l=}^{n(_{j-1}\ _{l}\ +)}$ 个加权残差；如果第 j 层中的个体 i 不属于分层病例队列（即不是病例或抽样选取的非病例），则该值为零，$i\in 21,\dots,n(\#)\}, j\in\{1,\dots,J\}$。类似地，$IF_{i,j}^{(1)}\{\mathrm{d}\Lambda\%_0^{*}(t)\}$ 是使用权重 $w_{i,j}{}^{*}$，将 $\xi_{i,j}H_{i,j}(t)$ 对 $\xi_{i,j}\boldsymbol{A}_{i,j}$ 进行加权线性回归后，得到的第 i 个个体在第 j 层中的预测值；如果第 j 层中的个体 i 不属于分层病例队列，则 $IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%_0^{*}(t)\}$ 为零，$i\in 21,\dots,n(\#)\}, j\in\{1,\dots,J\}$。总之，对于任意 $\boldsymbol{\theta}\%^{*}\in\{\boldsymbol{\eta}\mathrm{a}、\ \boldsymbol{\beta}\%^{*}、\ \mathrm{d}\Lambda\%_0^{*}(t)、\ \Lambda\%_0^{*}(t)、\ \pi: {}^{*}(\tau_1,\tau_2;\boldsymbol{x})\}$，$\xi_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\%^{*}\&$，如果第 j 层中的个体 i 不属于分层病例队列，则该值为零；此类

个体会通过影响 $\boldsymbol{\eta}$a 来影响 $\boldsymbol{\theta}\%^{*}$ ，因为其用于校准设计权重。此外，对于第 j 层中满足 $\sigma_{i,\ k\,j}$，且 $\sigma_{i,\ i\,j}$，均非零（即非病例）的任意个体 i 和 k，$\xi_{i,\ j}\,w_{i,\ j}\,\boldsymbol{IF}_{i,\ j}^{(2)}\#\boldsymbol{\theta}\%^{*}\&$ 和 $\xi_{k,\ j}\,w_{k,\ j}\,\boldsymbol{IF}_{k,\ j}^{(2)}\#\boldsymbol{\theta}\%^{*}\&$ 是加权线性回归的加权残差。

那么，当这些加权残差接近于零时，可以预期 $\boldsymbol{\theta}\%^{*}$ 的方差中的第二阶段分量会更小，进而总体方差也会更小。对于 $\boldsymbol{\theta}\%^{*} \in \{\boldsymbol{\beta}\%^{*},\ \mathrm{d}\Lambda\%_0^{*}(t),\ \Lambda\%_0^{*}(t),\ \pi:^{*}(\tau_1,\tau_2;\ \boldsymbol{x})\}$ ，当 $\boldsymbol{A}$ 包含对数相对风险的影响函数，以及纯风险时间区间内总随访时间乘以相对风险率的影响函数时（详见主文档第4.1节和网络附录D.1），情况应当如此。事实上，注意到网络附录D.2中给出的 $\boldsymbol{Z}_{i,j}$ 是个体 i 在层 j 中对 $\boldsymbol{\beta}\%^{*}$ 的“直接”影响（即，不通过其对校准权重的影响），$i \in \{1, \ldots, {}^{n}(\#)\}, j \in \{1, \ldots, J\}$。其形式与使用设计权重时的影响函数相似；参见网络附录 C.1。另一方面，$\int K_{i,j}(^{\tau_2}\ \ t\ \ t)\mathrm{dA}\#$对应区间 $(\tau_1, \tau_2]$ 内的总随访时间，乘以个体 i 在层 $j, i \in 21, \ldots, {}^{n}(\#)\}, j \in \{1, \ldots, J\}$的估计相对风险率。尽管 Shin等人（2020）没有提及这一点，但这支持他们选择辅助变量而非 Breslow和Lumley（2013）所选变量的做法。

网络附录 E. 模拟

E.1 固定参数值 用于主文档第 7 节模拟的参数值

	$X_1 < -2$	$-2 \leq X_1 < 1$	$1 \leq X_1$
$p_{0\|X_1}$	0.7	0.45	0.4
$p_{1\|X_1}$	0.05	0.2	0.3
$p_{2\|X_1}$	0.25	0.35	0.3

网络表格 1-参数值 $p_{0|X_1} = P(X_2 = 0|X_1)$， $p_{1|X_1} = P(X_2 = 1|X_1)$ 和 $p_{2|X_1} = P(X_2 = 2|X_1)$，用于模拟 X_2。

α_1	α_2	β_1	β_2	β_3
0.05	-0.35	-0.2	0.25	-0.3

网络表格 2-用于模拟 X_3 和 T. 的参数值。

E.2 所研究的情景的模拟结果

请注意，估计分别采用了采用设计权重的分层病例队列（SCC）、采用校准权重的分层病例队列（SCC.Calib）、采用设计权重的非分层病例队列（USCC）以及采用校准权重的非分层病例队列（USCC.Calib）。对于 SCC，我们使用主文档第3.3节方程（14）中给出的、包含超总体和第二阶段方差分量的方差估计（Vfi ）；以及主文档第3.3节方程（15）中给出的稳健方差估计（Vfi$_{\text{Robust}}$）。对于 SCC.Calib，我们使用主文档第4.3节方程（18）中给出的、包含超总体和第二阶段方差分量的方差估计（Vfi ）；以及主文档第4.3节方程（19）中给出的稳健方差估计（Vfi$_{\text{Robust}}$）。对于 USCC 和 USCC.Calib，我们使用了从以下简化版本中得到的方差估计

这些方程，当$J = 1$时。我们还使用整个队列的数据估计了这些参数，作为参考点（队列）。所有12 个情景的模拟结果（其中$n \in \{5 \times 10^3, 10^4\}$、$p_Y \in \{0.02, 0.05, 0.1\}$ 和$N \in \{2, 4\}$）显示在 **网络表 3** 至**网络表 20**中。更具体地说，网络表格 **网络表 3**至 **网络表8** 显示95%置信区间的覆盖率，**网络表 9** 至**网络表 14** 显示估计方差的均值，而 **网络表15** 至 **网络表20**显示经验方差及其与使用整个队列所得结果的经验方差比，针对β_1， β_2、 β_3 和$\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$，分别为$(\tau_1, \tau_2] = (0,8]$ 和$\boldsymbol{x} \in \{(-1,1, -0.6)', (1, -1,0.6)', (1, 1,0.6)'\}$。结果解释和结论与主文档第 7.2 节中的内容相同。

E.3 关于稳健方差估计的评论

Barlow（1994年）也提到了子队列的分层抽样，并指出他的稳健公式可用于方差估计，但未给出具体公式。同样，其他提及分层病例队列数据稳健方差的作者也未提供公式，即使他们比较了方差估计（Samuelsen 等人，2007年；Gray，2009）。另一方面，Jiao（2002年）指出，Barlow（1994年）提出的相对风险“朴素”稳健方差估计（即影响量平方和）在对子队列进行分层抽样时并不成立。她提出了针对分层病例队列设计的稳健方差扩展。但由于我们的模拟中未观察到两种稳健方差估计之间存在差异（结果未显示），因此我们采用了 Barlow 的初始稳健方差公式。Jiao（2002年）推测，Barlow（1994年）可能只是指影响量平方和；我们认为，许多从业者都像 Barlow（1994年）那样使用影响量平方和。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	β_1
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.944	0.9682*	0.9516	0.9528	0.95	0.9558	0.956*	0.9524	0.9524	5000	2	0.02	-0.2
0.9512	0.961*	0.9524	0.953	0.9522	0.9512	0.9514	0.9462	0.9466	5000	4	0.02	-0.2
0.9476	0.9668*	0.9524	0.9538	0.95	0.947	0.947	0.9452	0.945	10000	2	0.02	-0.2
0.948	0.9596*	0.9514	0.9514	0.95	0.9502	0.9502	0.951	0.951	10000	4	0.02	-0.2
0.9476	0.968*	0.9548	0.9558	0.953	0.9528	0.953	0.948	0.948	5000	2	0.05	-0.2
0.954	0.9622*	0.9546	0.9524	0.951	0.9528	0.953	0.9538	0.9538	5000	4	0.05	-0.2
0.9552	0.9654*	0.953	0.9558	0.954	0.9536	0.9536	0.9556	0.9558	10000	2	0.05	-0.2
0.9482	0.9566*	0.9516	0.9526	0.9512	0.9468	0.9468	0.949	0.949	10000	4	0.05	-0.2
0.947	0.9598*	0.9448	0.948	0.9472	0.9522	0.9522	0.948	0.948	5000	2	0.1	-0.2
0.9544	0.9584*	0.9538	0.957*	0.957*	0.955	0.955	0.9568*	0.9568*	5000	4	0.1	-0.2
0.9488	0.9618*	0.9506	0.9504	0.9486	0.9474	0.9476	0.9496	0.9496	10000	2	0.1	-0.2
0.9526	0.9552	0.9496	0.9532	0.9514	0.9478	0.9478	0.9486	0.9486	10000	4	0.1	-0.2

WEB 表格 3- 覆盖率 of 95% CIs for 对数相对 风险 参数 β_1 在 5,000个模拟队列中，不同抽样设计、分析方法和方差估计方法下的覆盖率。* 表示覆盖率超出预期区间 [0.9440；0.9560]。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	β_2
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.9532	0.9714*	0.9568*	0.9552	0.9548	0.9536	0.9536	0.9498	0.9494	5000	2	0.02	0.25
0.9514	0.9666*	0.955	0.9542	0.9538	0.9564*	0.9564*	0.95	0.95	5000	4	0.02	0.25
0.9486	0.9716*	0.9522	0.9514	0.9512	0.9554	0.9554	0.9506	0.9506	10000	2	0.02	0.25
0.9488	0.9606*	0.9504	0.9468	0.9468	0.9488	0.9492	0.948	0.948	10000	4	0.02	0.25
0.9488	0.967*	0.9488	0.9498	0.9492	0.953	0.953	0.9508	0.9508	5000	2	0.05	0.25
0.9524	0.9608*	0.9502	0.9512	0.9508	0.9542	0.9542	0.9528	0.9528	5000	4	0.05	0.25
0.9506	0.9676*	0.95	0.952	0.952	0.9506	0.9506	0.9526	0.9526	10000	2	0.05	0.25
0.951	0.9562*	0.9488	0.9488	0.9484	0.9504	0.9504	0.9514	0.9514	10000	4	0.05	0.25
0.9512	0.9656*	0.95	0.952	0.9518	0.947	0.947	0.9516	0.9516	5000	2	0.1	0.25
0.9496	0.9592*	0.952	0.951	0.951	0.9504	0.9506	0.9532	0.9532	5000	4	0.1	0.25
0.9504	0.9636*	0.9504	0.952	0.9518	0.9524	0.9524	0.9534	0.9534	10000	2	0.1	0.25
0.952	0.9588*	0.953	0.9514	0.9512	0.9496	0.9496	0.9512	0.9512	10000	4	0.1	0.25

WEB 表格 4- 覆盖率 of 95% CIs for 对数相对 风险 参数 β_2 在 5,000个模拟队列中，不同抽样设计、分析方法和方差估计方法的比较。* 表示覆盖率超出预期区间 [0.9440；0.9560]。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	β_3
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.9442	0.9546	0.9548	0.9508	0.9506	0.9482	0.9486	0.9468	0.9468	5000	2	0.02	-0.3
0.9464	0.9478	0.9478	0.9478	0.9474	0.9566*	0.9566*	0.9544	0.9544	5000	4	0.02	-0.3
0.9422*	0.9492	0.9494	0.9474	0.9474	0.9454	0.9454	0.9476	0.9478	10000	2	0.02	-0.3
0.9506	0.9516	0.9516	0.9522	0.9522	0.9524	0.9524	0.9502	0.9502	10000	4	0.02	-0.3
0.9484	0.9462	0.9466	0.9458	0.946	0.9522	0.9522	0.9506	0.9506	5000	2	0.05	-0.3
0.9464	0.95	0.95	0.9482	0.948	0.948	0.948	0.9444	0.9446	5000	4	0.05	-0.3
0.9504	0.9466	0.9468	0.9498	0.9498	0.9456	0.9458	0.95	0.95	10000	2	0.05	-0.3
0.9498	0.9522	0.9524	0.9516	0.9516	0.9512	0.9512	0.9524	0.9524	10000	4	0.05	-0.3
0.9494	0.951	0.951	0.951	0.951	0.949	0.949	0.9514	0.9514	5000	2	0.1	-0.3
0.9508	0.9498	0.9498	0.951	0.951	0.952	0.952	0.9492	0.9492	5000	4	0.1	-0.3
0.9532	0.9518	0.9518	0.9516	0.9516	0.9516	0.9516	0.9532	0.9532	10000	2	0.1	-0.3
0.9482	0.9536	0.9536	0.9512	0.9512	0.949	0.949	0.95	0.9502	10000	4	0.1	-0.3

WEB 表格 5- 覆盖率 of 95% CIs for 对数相对 风险 参数 β_3 在5,000个模拟队列中，不同抽样设计、分析方法和方差估计方法的结果。*表示覆盖率超出预期区间[0.9440；0.9560]。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}$ #$%&'(	$\hat{V}$	$\hat{V}$ #$%&'(	$\hat{V}$	$\hat{V}$ #$%&'(	$\hat{V}$	$\hat{V}$ #$%&'(	$\hat{V}$				
0.9466	0.9722*	0.95	0.9558	0.9516	0.9634*	0.9494	0.9526	0.9526	5000	2	0.02	-3.948
0.9484	0.9602*	0.9492	0.9508	0.9506	0.9558	0.9456	0.95	0.95	5000	4	0.02	-3.948
0.956*	0.973*	0.9568*	0.96	0.9546	0.9656*	0.9526	0.955	0.955	10000	2	0.02	-3.948
0.951	0.9656*	0.9538	0.9538	0.952	0.9602*	0.9526	0.9512	0.9512	10000	4	0.02	-3.948
0.9488	0.9694*	0.9494	0.9528	0.9494	0.958*	0.948	0.9506	0.9506	5000	2	0.05	-3.046
0.9546	0.9634*	0.955	0.956*	0.954	0.9596*	0.9538	0.9524	0.9524	5000	4	0.05	-3.046
0.9556	0.971*	0.9556	0.9602*	0.9562*	0.9656*	0.9554	0.9542	0.9542	10000	2	0.05	-3.046
0.9534	0.9604*	0.9518	0.9554	0.953	0.962*	0.9538	0.9534	0.9534	10000	4	0.05	-3.046
0.9506	0.9658*	0.951	0.954	0.951	0.9622*	0.954	0.9522	0.9522	5000	2	0.1	-2.377
0.9496	0.9566*	0.9512	0.951	0.9504	0.9544	0.9514	0.9506	0.9508	5000	4	0.1	-2.377
0.9484	0.9634*	0.9476	0.9534	0.9504	0.957*	0.9484	0.9444	0.9444	10000	2	0.1	-2.377
0.9538	0.9586*	0.9524	0.9548	0.954	0.9594*	0.9562*	0.9558	0.9558	10000	4	0.1	-2.377

网页表6- 在不同抽样设计、分析方法和方差估计方法下，5,000个模拟队列中对数纯风险 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的95%置信区间覆盖率，其中 $(\tau_1,\tau_2]=(0,8]$ 和 $\boldsymbol{x}=(-1,1,-0.6)'$。*表示覆盖率超出预期区间 [0.9440; 0.9560].

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	$\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.9476	0.9754*	0.9504	0.95	0.9492	0.9552	0.9526	0.9516	0.9516	5000	2	0.02	-5.201
0.946	0.9642*	0.948	0.9472	0.9472	0.953	0.9514	0.9454	0.9454	5000	4	0.02	-5.201
0.952	0.9722*	0.9518	0.9522	0.9516	0.954	0.9522	0.9472	0.947	10000	2	0.02	-5.201
0.946	0.9582*	0.9482	0.9464	0.946	0.9528	0.9508	0.9456	0.9456	10000	4	0.02	-5.201
0.9456	0.968*	0.9462	0.9474	0.9472	0.9512	0.9486	0.946	0.946	5000	2	0.05	-4.288
0.9488	0.962*	0.9508	0.9498	0.9494	0.9516	0.95	0.9478	0.9478	5000	4	0.05	-4.288
0.9508	0.966*	0.949	0.9464	0.9462	0.9558	0.9522	0.9494	0.9494	10000	2	0.05	-4.288
0.9526	0.9576*	0.9508	0.9524	0.9524	0.9526	0.9512	0.9544	0.9544	10000	4	0.05	-4.288
0.9442	0.9666*	0.9492	0.9484	0.948	0.9518	0.9494	0.9514	0.9514	5000	2	0.1	-3.602
0.9548	0.9616*	0.9546	0.9544	0.9544	0.954	0.9534	0.9528	0.9528	5000	4	0.1	-3.602
0.9526	0.9698*	0.9568*	0.9536	0.9532	0.9544	0.9522	0.9508	0.9508	10000	2	0.1	-3.602
0.9514	0.959*	0.9518	0.9512	0.951	0.9524	0.9514	0.9522	0.9522	10000	4	0.1	-3.602

网络 表格 7- 对数纯风险 $\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$ 的 95%置信区间覆盖率，其中 $(\tau_1, \tau_2] = (0,8]$ 和 $\boldsymbol{x} = (1, -1, 0.6)'$，针对不同抽样设计以及 5,000个模拟队列中的分析方法和方差估计。* 表示覆盖率超出预期区间 [0.9440; 0.9560]。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.955	0.9712*	0.9582*	0.9562*	0.9546	0.9658*	0.9596	0.9584*	0.9586	5000	2	0.02	-4.702
0.9468	0.9578*	0.9494	0.9494	0.949	0.9534	0.95	0.9482	0.9482	5000	4	0.02	-4.702
0.952	0.9666*	0.9542	0.9532	0.9516	0.9574*	0.95	0.9502	0.9502	10000	2	0.02	-4.702
0.9492	0.9584*	0.9504	0.9516	0.9512	0.956*	0.9522	0.9482	0.9482	10000	4	0.02	-4.702
0.9488	0.962*	0.9488	0.9498	0.949	0.9572*	0.949	0.9506	0.9506	5000	2	0.05	-3.793
0.95	0.9602*	0.953	0.9526	0.9516	0.9526	0.9502	0.951	0.951	5000	4	0.05	-3.793
0.9462	0.9606*	0.9468	0.9494	0.948	0.953	0.9444	0.946	0.946	10000	2	0.05	-3.793
0.9496	0.9562*	0.9494	0.9538	0.9536	0.9518	0.95	0.9516	0.9516	10000	4	0.05	-3.793
0.9452	0.9608*	0.9486	0.9484	0.9484	0.9514	0.9466	0.9458	0.9458	5000	2	0.1	-3.111
0.9516	0.957*	0.9522	0.9528	0.9524	0.9544	0.9506	0.9508	0.951	5000	4	0.1	-3.111
0.9502	0.9606*	0.9512	0.951	0.9504	0.9522	0.9466	0.9474	0.9474	10000	2	0.1	-3.111
0.9528	0.956*	0.9512	0.9518	0.9516	0.9526	0.9508	0.9524	0.9524	10000	4	0.1	-3.111

网络表8- 对于不同抽样设计、分析方法和方差估计方法，在5,000个模拟队列中，对数纯风险 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的95%置信区间覆盖率，其中（τ_1,τ_2] $=(0,8]$ 和 $\boldsymbol{x}=(1,1,0.6)'$。*表示覆盖率超出预期区间 [0.9440；0.9560]。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	β_1
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.0138	0.0212	0.0182	0.0177	0.0172	0.0222	0.0223	0.0184	0.0184	5000	2	0.02	-0.2
0.0139	0.0174	0.016	0.0157	0.0155	0.0177	0.0178	0.0159	0.0159	5000	4	0.02	-0.2
0.0069	0.0102	0.0087	0.0085	0.0083	0.0106	0.0106	0.0087	0.0087	10000	2	0.02	-0.2
0.0069	0.0085	0.0078	0.0076	0.0075	0.0086	0.0087	0.0077	0.0077	10000	4	0.02	-0.2
0.0056	0.0079	0.0069	0.0067	0.0065	0.0082	0.0082	0.0068	0.0068	5000	2	0.05	-0.2
0.0056	0.0066	0.0061	0.0061	0.006	0.0067	0.0067	0.0061	0.0061	5000	4	0.05	-0.2
0.0028	0.0039	0.0034	0.0033	0.0032	0.004	0.004	0.0033	0.0033	10000	2	0.05	-0.2
0.0028	0.0033	0.0031	0.003	0.003	0.0033	0.0033	0.003	0.003	10000	4	0.05	-0.2
0.0029	0.0037	0.0033	0.0032	0.0032	0.0039	0.0039	0.0033	0.0033	5000	2	0.1	-0.2
0.0029	0.0032	0.003	0.003	0.003	0.0032	0.0032	0.003	0.003	5000	4	0.1	-0.2
0.0014	0.0019	0.0017	0.0016	0.0016	0.0019	0.0019	0.0016	0.0016	10000	2	0.1	-0.2
0.0014	0.0016	0.0015	0.0015	0.0015	0.0016	0.0016	0.0015	0.0015	10000	4	0.1	-0.2

WEB 表格 9- Mean of 估计 方差 of 对数相对 风险 参数 β_1 针对不同抽样设计、分析方法和方差估计，在5,000个模拟队列中。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	β_2
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.0197	0.0286	0.0237	0.0226	0.0225	0.029	0.0291	0.0227	0.0227	5000	2	0.02	0.25
0.0198	0.024	0.0217	0.0212	0.0211	0.0243	0.0243	0.0213	0.0213	5000	4	0.02	0.25
0.0097	0.0139	0.0114	0.0109	0.0109	0.014	0.014	0.0109	0.0109	10000	2	0.02	0.25
0.0097	0.0117	0.0106	0.0103	0.0103	0.0118	0.0118	0.0103	0.0103	10000	4	0.02	0.25
0.0079	0.0108	0.0091	0.0087	0.0087	0.0109	0.0109	0.0087	0.0087	5000	2	0.05	0.25
0.0079	0.0092	0.0084	0.0082	0.0082	0.0092	0.0092	0.0082	0.0082	5000	4	0.05	0.25
0.0039	0.0054	0.0045	0.0043	0.0043	0.0054	0.0054	0.0043	0.0043	10000	2	0.05	0.25
0.0039	0.0046	0.0042	0.0041	0.0041	0.0046	0.0046	0.0041	0.0041	10000	4	0.05	0.25
0.004	0.0052	0.0045	0.0043	0.0043	0.0052	0.0052	0.0043	0.0043	5000	2	0.1	0.25
0.004	0.0044	0.0042	0.0041	0.0041	0.0044	0.0044	0.0041	0.0041	5000	4	0.1	0.25
0.002	0.0026	0.0022	0.0021	0.0021	0.0026	0.0026	0.0021	0.0021	10000	2	0.1	0.25
0.002	0.0022	0.0021	0.0021	0.0021	0.0022	0.0022	0.002	0.002	10000	4	0.1	0.25

WEB 表格 10- Mean of 估计的 方差 of 对数相对 风险 参数 β_2 针对不同抽样设计、分析方法和方差估计，在5,000个模拟队列中。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	β_3
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.0136	0.0214	0.0215	0.0179	0.0179	0.0223	0.0223	0.0183	0.0183	5000	2	0.02	-0.3
0.0136	0.0173	0.0173	0.0157	0.0156	0.0178	0.0178	0.0159	0.0159	5000	4	0.02	-0.3
0.0068	0.0102	0.0102	0.0085	0.0085	0.0107	0.0107	0.0087	0.0087	10000	2	0.02	-0.3
0.0068	0.0084	0.0084	0.0076	0.0076	0.0086	0.0086	0.0077	0.0077	10000	4	0.02	-0.3
0.0055	0.0079	0.0079	0.0067	0.0067	0.0082	0.0082	0.0068	0.0068	5000	2	0.05	-0.3
0.0055	0.0066	0.0066	0.006	0.006	0.0067	0.0067	0.0061	0.0061	5000	4	0.05	-0.3
0.0028	0.0039	0.0039	0.0033	0.0033	0.004	0.004	0.0033	0.0033	10000	2	0.05	-0.3
0.0028	0.0032	0.0032	0.003	0.003	0.0033	0.0033	0.003	0.003	10000	4	0.05	-0.3
0.0028	0.0037	0.0037	0.0032	0.0032	0.0038	0.0038	0.0033	0.0033	5000	2	0.1	-0.3
0.0028	0.0031	0.0031	0.003	0.003	0.0032	0.0032	0.003	0.003	5000	4	0.1	-0.3
0.0014	0.0018	0.0018	0.0016	0.0016	0.0019	0.0019	0.0016	0.0016	10000	2	0.1	-0.3
0.0014	0.0016	0.0016	0.0015	0.0015	0.0016	0.0016	0.0015	0.0015	10000	4	0.1	-0.3

WEB 表格 11- Mean of 估计的 方差 of 对数相对 风险 参数 β_3 采用不同的抽样设计、分析方法和方差估计，对5,000个模拟队列进行分析。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	$\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.0248	0.0354	0.029	0.0296	0.0285	0.0374	0.0329	0.0306	0.0306	5000	2	0.02	-3.948
0.0248	0.0298	0.0268	0.0271	0.0266	0.0307	0.0286	0.0275	0.0275	5000	4	0.02	-3.948
0.0122	0.0172	0.014	0.0142	0.0137	0.0181	0.0159	0.0145	0.0145	10000	2	0.02	-3.948
0.0122	0.0145	0.013	0.0131	0.0129	0.015	0.0139	0.0133	0.0133	10000	4	0.02	-3.948
0.0096	0.013	0.0109	0.011	0.0106	0.0137	0.0122	0.0112	0.0112	5000	2	0.05	-3.046
0.0096	0.0111	0.0102	0.0102	0.0101	0.0114	0.0108	0.0103	0.0103	5000	4	0.05	-3.046
0.0048	0.0064	0.0054	0.0054	0.0052	0.0068	0.006	0.0055	0.0055	10000	2	0.05	-3.046
0.0048	0.0055	0.005	0.0051	0.005	0.0057	0.0053	0.0051	0.0051	10000	4	0.05	-3.046
0.0047	0.0059	0.0052	0.0051	0.005	0.0062	0.0057	0.0052	0.0052	5000	2	0.1	-2.377
0.0047	0.0051	0.0048	0.0048	0.0048	0.0052	0.005	0.0049	0.0049	5000	4	0.1	-2.377
0.0023	0.0029	0.0026	0.0026	0.0025	0.0031	0.0028	0.0026	0.0026	10000	2	0.1	-2.377
0.0023	0.0025	0.0024	0.0024	0.0024	0.0026	0.0025	0.0024	0.0024	10000	4	0.1	-2.377

WEB 表 12- 使用不同抽样设计、分析方法和方差估计方法，在 5,000个模拟队列中得到的对数纯风险 $\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$ 的估计方差均值，其中 $(\tau_1, \tau_2] = (0,8]$ and $\boldsymbol{x} = (-1, 1, -0.6)'$。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.1249	0.1773	0.1441	0.141	0.1401	0.1788	0.1744	0.1453	0.1452	5000	2	0.02	-5.201
0.1258	0.1507	0.1349	0.1335	0.1331	0.1511	0.1489	0.1348	0.1348	5000	4	0.02	-5.201
0.062	0.0861	0.0697	0.0684	0.0681	0.086	0.0837	0.0696	0.0696	10000	2	0.02	-5.201
0.0618	0.0733	0.0654	0.0648	0.0647	0.0731	0.072	0.0653	0.0653	10000	4	0.02	-5.201
0.0496	0.0666	0.0549	0.0538	0.0536	0.0661	0.0645	0.0546	0.0546	5000	2	0.05	-4.288
0.0496	0.0571	0.0518	0.0513	0.0513	0.0569	0.0562	0.0518	0.0518	5000	4	0.05	-4.288
0.0247	0.0329	0.0271	0.0265	0.0265	0.0325	0.0317	0.0268	0.0268	10000	2	0.05	-4.288
0.0247	0.0284	0.0257	0.0255	0.0255	0.0282	0.0278	0.0257	0.0257	10000	4	0.05	-4.288
0.0248	0.0316	0.0268	0.0263	0.0262	0.0311	0.0304	0.0266	0.0266	5000	2	0.1	-3.602
0.0248	0.0273	0.0255	0.0253	0.0253	0.0271	0.0268	0.0254	0.0254	5000	4	0.1	-3.602
0.0124	0.0157	0.0133	0.013	0.013	0.0154	0.0151	0.0132	0.0132	10000	2	0.1	-3.602
0.0124	0.0136	0.0127	0.0126	0.0126	0.0134	0.0133	0.0126	0.0126	10000	4	0.1	-3.602

网页表13- 对数纯风险 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的估计方差的均值，其中 $(\tau_1,\tau_2]=(0,8]$ 和 $\boldsymbol{x}=(1,-1,0.6)'$，针对5,000个模拟队列中的不同抽样设计、分析方法和方差估计方法。

队列	SCC		SCC.Calib		USCC		USCC.Calib		n	K	p_Y	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$				
0.0557	0.0789	0.0698	0.0672	0.0665	0.0806	0.0761	0.0689	0.0689	5000	2	0.02	-4.702
0.0559	0.067	0.0626	0.0613	0.061	0.0679	0.0657	0.0621	0.0621	5000	4	0.02	-4.702
0.0277	0.0379	0.0333	0.0319	0.0316	0.0386	0.0363	0.0328	0.0328	10000	2	0.02	-4.702
0.0277	0.0325	0.0303	0.0297	0.0295	0.0328	0.0317	0.03	0.03	10000	4	0.02	-4.702
0.0221	0.0291	0.0259	0.0249	0.0247	0.0294	0.0278	0.0253	0.0253	5000	2	0.05	-3.793
0.0221	0.0251	0.0237	0.0232	0.0231	0.0253	0.0246	0.0235	0.0235	5000	4	0.05	-3.793
0.011	0.0143	0.0127	0.0122	0.0121	0.0145	0.0136	0.0124	0.0124	10000	2	0.05	-3.793
0.011	0.0125	0.0117	0.0115	0.0115	0.0125	0.0122	0.0116	0.0116	10000	4	0.05	-3.793
0.011	0.0136	0.0123	0.0119	0.0118	0.0137	0.013	0.0121	0.0121	5000	2	0.1	-3.111
0.011	0.0119	0.0114	0.0113	0.0113	0.0119	0.0117	0.0114	0.0114	5000	4	0.1	-3.111
0.0055	0.0068	0.0061	0.0059	0.0059	0.0068	0.0065	0.006	0.006	10000	2	0.1	-3.111
0.0055	0.0059	0.0057	0.0056	0.0056	0.0059	0.0058	0.0056	0.0056	10000	4	0.1	-3.111

网络表14- 对数纯风险 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的估计方差的均值，其中 $(\tau_1,\tau_2]=(0,8]$ 和 $\boldsymbol{x}=(1,1,0.6)'$，针对 5,000个模拟队列中的不同抽样设计、分析方法和方差估计方法。

经验方差					与整体经验方差之比 将队列转换为经验方差，使用的是							
队列	SCC	SCC.Calib	USCC	USCC.Calib	SCC	SCC.Calib	USCC	USCC.Calib	n	K	p_Y	β_1
0.0138	0.0181	0.017	0.0223	0.0177	0.763	0.8141	0.62	0.7794	5000	2	0.02	-0.2
0.0138	0.0159	0.0152	0.0177	0.0158	0.8678	0.9054	0.7769	0.8711	5000	4	0.02	-0.2
0.007	0.0085	0.0082	0.0108	0.0087	0.8161	0.8521	0.6408	0.8013	10000	2	0.02	-0.2
0.0069	0.0076	0.0074	0.0086	0.0076	0.9055	0.9322	0.8017	0.9039	10000	4	0.02	-0.2
0.0056	0.0068	0.0064	0.0082	0.0067	0.8327	0.8769	0.6897	0.8401	5000	2	0.05	-0.2
0.0055	0.006	0.0059	0.0066	0.006	0.917	0.9363	0.8331	0.9207	5000	4	0.05	-0.2
0.0027	0.0033	0.0031	0.004	0.0033	0.8367	0.8832	0.6895	0.839	10000	2	0.05	-0.2
0.0028	0.003	0.0029	0.0033	0.003	0.9275	0.9547	0.8424	0.9402	10000	4	0.05	-0.2
0.003	0.0034	0.0033	0.0039	0.0033	0.8631	0.9076	0.7548	0.8862	5000	2	0.1	-0.2
0.0028	0.003	0.0029	0.0032	0.003	0.9535	0.9759	0.8931	0.9574	5000	4	0.1	-0.2
0.0014	0.0016	0.0016	0.0019	0.0016	0.8596	0.9023	0.7398	0.8725	10000	2	0.1	-0.2
0.0014	0.0015	0.0015	0.0016	0.0015	0.9603	0.973	0.8908	0.9485	10000	4	0.1	-0.2

网络表 15- 在 5,000个模拟队列中，采用不同抽样设计、分析方法和方差估计方法时，对数相对风险参数的经验方差及经验方差比（相对于整个队列的方差）。β_1，

经验方差					与整体相比的经验方差比率 队列的经验方差之比与							
队列	SCC	SCC.Calib	USCC	USCC.Calib	SCC	SCC.Calib	USCC	USCC.Calib	n	K	p_Y	β_2
0.0197	0.0235	0.0223	0.0287	0.0224	0.8415	0.8867	0.6885	0.883	5000	2	0.02	0.25
0.0195	0.0213	0.0205	0.0239	0.0211	0.9165	0.9497	0.816	0.9248	5000	4	0.02	0.25
0.01	0.0115	0.0109	0.014	0.0112	0.871	0.9155	0.7143	0.8955	10000	2	0.02	0.25
0.0099	0.0108	0.0105	0.012	0.0105	0.9173	0.9499	0.8279	0.9447	10000	4	0.02	0.25
0.008	0.0092	0.0088	0.0108	0.0088	0.868	0.91	0.7415	0.909	5000	2	0.05	0.25
0.008	0.0085	0.0083	0.0093	0.0083	0.9404	0.9647	0.8621	0.9594	5000	4	0.05	0.25
0.0039	0.0045	0.0042	0.0054	0.0042	0.863	0.9241	0.725	0.9221	10000	2	0.05	0.25
0.0038	0.0042	0.0041	0.0044	0.004	0.9263	0.9501	0.8763	0.9694	10000	4	0.05	0.25
0.004	0.0045	0.0043	0.0052	0.0043	0.8895	0.9347	0.7727	0.947	5000	2	0.1	0.25
0.0041	0.0042	0.0042	0.0045	0.0042	0.9647	0.9732	0.9184	0.9858	5000	4	0.1	0.25
0.002	0.0022	0.0021	0.0025	0.0021	0.8994	0.9384	0.7814	0.9437	10000	2	0.1	0.25
0.002	0.0021	0.002	0.0022	0.002	0.9564	0.9701	0.8997	0.9801	10000	4	0.1	0.25

网络表格 16- 在 5,000个模拟队列中，针对不同抽样设计、分析方法和方差估计方法，对数相对风险参数的经验方差及其与整个队列情形下经验方差的比值β_2。

经验方差					整个队列的经验方差与队列经验方差之比							
队列	SCC	SCC.Calib	USCC	USCC.Calib	SCC	SCC.Calib	USCC	USCC.Calib	n	K	p_Y	β_3
0.014	0.0217	0.0179	0.0223	0.0182	0.6443	0.7801	0.6254	0.7671	5000	2	0.02	-0.3
0.0135	0.0176	0.0155	0.0175	0.0154	0.7652	0.8709	0.7707	0.8774	5000	4	0.02	-0.3
0.0069	0.0103	0.0084	0.0109	0.0088	0.6734	0.822	0.6335	0.7884	10000	2	0.02	-0.3
0.0067	0.0085	0.0075	0.0085	0.0076	0.7921	0.8962	0.7916	0.8881	10000	4	0.02	-0.3
0.0056	0.0081	0.0067	0.0082	0.0068	0.6875	0.8256	0.6781	0.818	5000	2	0.05	-0.3
0.0057	0.0066	0.0061	0.0069	0.0062	0.8593	0.9308	0.8276	0.9152	5000	4	0.05	-0.3
0.0028	0.004	0.0033	0.0041	0.0034	0.696	0.8445	0.6862	0.8338	10000	2	0.05	-0.3
0.0027	0.0032	0.0029	0.0032	0.0029	0.8468	0.9291	0.8443	0.9269	10000	4	0.05	-0.3
0.0029	0.0037	0.0032	0.0039	0.0032	0.7675	0.8895	0.7349	0.881	5000	2	0.1	-0.3
0.0028	0.0031	0.0029	0.0032	0.003	0.9157	0.9606	0.8901	0.9391	5000	4	0.1	-0.3
0.0014	0.0018	0.0016	0.0019	0.0016	0.7728	0.8803	0.7393	0.8669	10000	2	0.1	-0.3
0.0014	0.0016	0.0015	0.0016	0.0015	0.9	0.9536	0.8925	0.9475	10000	4	0.1	-0.3

网络 表格 17- 在5,000个模拟队列中，不同抽样设计、分析方法和方差估计方法下，对数相对风险参数β_3的经验方差及经验方差比（相对于整个队列的情况）。

经验方差					与整体经验方差的比值 队列与经验方差之比				n	K	p_Y	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
队列	SCC	SCC.Calib	USCC	USCC.Calib	SCC	SCC.Calib	USCC	USCC.Calib				
0.0254	0.0295	0.0286	0.0332	0.03	0.8606	0.8887	0.764	0.8475	5000	2	0.02	-3.948
0.0255	0.0275	0.0269	0.0293	0.0278	0.9255	0.9458	0.8689	0.9146	5000	4	0.02	-3.948
0.0119	0.0136	0.0133	0.0158	0.014	0.87	0.895	0.7522	0.8482	10000	2	0.02	-3.948
0.0119	0.0127	0.0124	0.0137	0.0129	0.9384	0.9558	0.8689	0.9234	10000	4	0.02	-3.948
0.0097	0.0108	0.0105	0.0124	0.0112	0.8983	0.9279	0.7826	0.8691	5000	2	0.05	-3.046
0.0094	0.0099	0.0098	0.0105	0.0101	0.952	0.9621	0.899	0.9363	5000	4	0.05	-3.046
0.0046	0.0052	0.005	0.0058	0.0053	0.8931	0.9251	0.796	0.8749	10000	2	0.05	-3.046
0.0047	0.0049	0.0048	0.0052	0.0049	0.9441	0.9617	0.9008	0.9519	10000	4	0.05	-3.046
0.0046	0.0052	0.005	0.0056	0.0051	0.9003	0.9276	0.8338	0.9061	5000	2	0.1	-2.377
0.0047	0.0048	0.0048	0.005	0.0049	0.9741	0.9824	0.9289	0.9628	5000	4	0.1	-2.377
0.0023	0.0026	0.0025	0.0028	0.0026	0.9076	0.9394	0.8249	0.8978	10000	2	0.1	-2.377
0.0023	0.0023	0.0023	0.0024	0.0024	0.9735	0.9836	0.9366	0.9703	10000	4	0.1	-2.377

网络表 18- 对数纯风险的经验方差及经验方差比（相对于整个队列） $log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$, 其中 $(\tau_1,\tau_2] = (0,8]$ 和 $\boldsymbol{x} = (-1, 1, -0.6)'$, 针对5,000个模拟队列中不同抽样设计以及分析和方差估计方法。

经验方差					整个队列的经验方差与队列到经验方差与				n	K	p_Y	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
队列	SCC	SCC.Calib		USCC.Calib	SCC	SCC.Calib		USCC.Calib				
0.1248	0.1423	0.1423	0.1388	0.1725	0.8772	0.8991	0.7233	0.8799	5000	2	0.02	-5.201
0.1294	0.1382	0.1382	0.1343	0.1523	0.9363	0.9635	0.85	0.9352	5000	4	0.02	-5.201
0.0618	0.0688	0.0688	0.0662	0.0823	0.8983	0.934	0.7518	0.8891	10000	2	0.02	-5.201
0.062	0.0657	0.0657	0.0646	0.0721	0.9438	0.96	0.8604	0.9481	10000	4	0.02	-5.201
0.0514	0.0568	0.0568	0.0551	0.0647	0.9047	0.9337	0.795	0.919	5000	2	0.05	-4.288
0.0499	0.052	0.052	0.0513	0.0562	0.9599	0.9719	0.8882	0.9577	5000	4	0.05	-4.288
0.0241	0.0266	0.0266	0.026	0.0313	0.9064	0.9295	0.7713	0.9185	10000	2	0.05	-4.288
0.0245	0.0255	0.0255	0.0252	0.0275	0.9636	0.9742	0.8908	0.9717	10000	4	0.05	-4.288
0.025	0.027	0.027	0.0261	0.0305	0.9256	0.9544	0.8188	0.9382	5000	2	0.1	-3.602
0.0243	0.0248	0.0248	0.0247	0.0259	0.9789	0.9813	0.9362	0.9815	5000	4	0.1	-3.602
0.0125	0.0133	0.0133	0.013	0.0151	0.9411	0.9608	0.8257	0.9424	10000	2	0.1	-3.602
0.0122	0.0125	0.0125	0.0124	0.0132	0.9762	0.982	0.92	0.9778	10000	4	0.1	-3.602

网络表格19- 对数纯风险的经验方差及经验方差与采用整个队列时所得方差之比 $log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$，其中（$\tau_1,\tau_2] = (0,8]$ 和 $\boldsymbol{x} = (1,-1,0.6)'$，针对不同抽样设计、分析方法和方差估计方法，在5,000个模拟队列中。

经验方差					整个队列的经验方差与队列至经验方差与				n	K	p_Y	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
队列	SCC	SCC.Calib			SCC	SCC.Calib						
0.0537	0.0668	0.0631	0.0721	0.0645	0.8032	0.85	0.7448	0.8327	5000	2	0.02	-4.702
0.058	0.0648	0.0622	0.0678	0.0634	0.8962	0.9327	0.8557	0.9156	5000	4	0.02	-4.702
0.0274	0.0326	0.0309	0.0361	0.0325	0.8421	0.8865	0.7589	0.8439	10000	2	0.02	-4.702
0.0277	0.0303	0.0292	0.0315	0.03	0.914	0.9481	0.8793	0.9248	10000	4	0.02	-4.702
0.0226	0.0265	0.0252	0.028	0.0253	0.8506	0.8934	0.8069	0.8907	5000	2	0.05	-3.793
0.0218	0.0233	0.0227	0.0246	0.0232	0.935	0.9592	0.8868	0.9397	5000	4	0.05	-3.793
0.0111	0.0128	0.0121	0.014	0.0126	0.8618	0.9145	0.7903	0.8766	10000	2	0.05	-3.793
0.0111	0.0118	0.0114	0.0123	0.0116	0.9405	0.9741	0.9017	0.9548	10000	4	0.05	-3.793
0.0114	0.0126	0.0122	0.0135	0.0123	0.9051	0.938	0.8434	0.928	5000	2	0.1	-3.111
0.0109	0.0113	0.0112	0.0116	0.0114	0.9625	0.9744	0.9405	0.9582	5000	4	0.1	-3.111
0.0055	0.0061	0.0058	0.0066	0.006	0.9004	0.9369	0.8318	0.9183	10000	2	0.1	-3.111
0.0054	0.0056	0.0055	0.0058	0.0057	0.963	0.9815	0.9306	0.9582	10000	4	0.1	-3.111

网络表格 20- 对数纯风险的经验方差及经验方差比（相对于整个队列） $log\{\pi(\tau_1,\tau_2;\ \boldsymbol{x})\}$，其中 $(\tau_1,\tau_2]=(0,8]$ 和 $\boldsymbol{x}=(1,1,0.6)'$， 在5,000个模拟队列中，不同抽样设计、分析方法和方差估计方法下的结果。

E.4 额外模拟，其中设计权重按非病例数量进行后分层

在主文档第 7 节和网络附录 E.2的模拟中，SCC 和USCC 的结果是使用设计权重获得的。在文献中，几位作者通过排除病例后重新定义层来计算设计权重（Borgan等人，2000年；Samuelsen等人，2007年）；这相当于根据非病例数对权重进行事后分层。在本网络附录中，我们进行了与之前相同的模拟，但除了使用 SCC和 USCC 进行估计外，还使用具有事后分层权重的分层病例队列（SCC.Poststrat）以及具有事后分层权重的未分层病例-队列（USCC.Poststrat）进行了估计。

SCC 的经验方差与 SCC.Poststrat 的经验方差之比，以及USCC 的经验方差与 USCC.Poststrat 的经验方差之比，分别显示于 β_1, β_2 和β_3 （**网络表格21)**）以及 π（τ_1, τ_2；$\boldsymbol{x}$），其中（τ_1, τ_2] = （0,8] 和$\boldsymbol{x} \in \{(-1,1,-0.6)', (1,-1,0.6)', (1,1,0.6)'\}$（**网络表格 22**）。这些比值衡量了在分层病例队列和非分层病例队列抽样设计中，使用事后分层权重而非设计权重所带来的效率增益。比值均接近于 1，表明根据非病例数量对权重进行事后分层几乎没有改善。

β_1		β_2		β_3		n	K	p_Y
SCC / SCC.Poststrat	USCC / USCC.Poststrat	SCC / SCC.Poststrat	USCC / USCC.Poststrat	SCC / SCC.Poststrat	USCC / USCC.Poststrat			
1.002	1	1.001	1	1	1	5000	2	0.02
1.003	1	1.003	1	1	1	5000	4	0.02
1.001	1	1.004	1	1	1	10000	2	0.02
1	1	1.002	1	1	1	10000	4	0.02
1.005	1	1.004	1	1	1	5000	2	0.05
1.001	1	1	1	1	1	5000	4	0.05
1.006	1	1.006	1	1	1	10000	2	0.05
1.004	1	1.004	1	1	1	10000	4	0.05
1.015	1	1.011	1	1	1	5000	2	0.1
1	1	1.006	1	1	1	5000	4	0.1
1.012	1	1.01	1	1	1	10000	2	0.1
1.004	1	1.003	1	1	1	10000	4	0.1

网络 表格 21- 使用 SCC 的经验方差与使用 SCC.Poststrat 的经验方差之比，以及使用 USCC 的经验方差与使用 USCC.Poststrat 的经验方差之比，针对对数相对风险参数 β_1, β_2 和 β_3，基于 5,000个模拟队列。该比值衡量使用后分层权重所带来的效率增益。

$\boldsymbol{x} = (-1, 1, -0.6)'$		$\boldsymbol{x} = (1, -1, 0.6)'$		$\boldsymbol{x} = (1, 1, 0.6)'$		n	K	p_Y
SCC / SCC.Poststrat	USCC / USCC.Poststrat	SCC / SCC.Poststrat	USCC / USCC.Poststrat	SCC / SCC.Poststrat	USCC / USCC.Poststrat			
1.005	1.005	1.006	1.001	1.003	1	5000	2	0.02
1.001	1.002	1.005	0.999	1.002	1	5000	4	0.02
1.002	1.003	1.006	1.001	1	1.001	10000	2	0.02
1.002	1.003	1.002	1.001	1.001	1	10000	4	0.02
1.009	1.005	1.007	1.002	1.008	1.002	5000	2	0.05
1.008	1.005	1	1.001	1.002	1.001	5000	4	0.05
1.005	1.005	1.008	1.001	1.007	1.003	10000	2	0.05
1.004	1	1.005	1.001	1.005	1.002	10000	4	0.05
1.018	1.011	1.012	1.003	1.011	1.006	5000	2	0.1
1.005	1.001	1.006	1	1.002	1.002	5000	4	0.1
1.019	1.007	1.01	1.001	1.006	1.001	10000	2	0.1
1.002	1.003	1.004	1	1.004	1.002	10000	4	0.1

网络 表格 22- 使用SCC时的经验方差与使用SCC. 后分层时的经验方差之比，以及使用USCC时的经验方差与使用USCC.Poststrat时的经验方差之比，对于纯风险 $\pi(\tau_1, \tau_2; \boldsymbol{x})$，其中（$\tau_1, \tau_2] = (0,8]$ 和 $\boldsymbol{x} \in \{(-1,1,-0.6)', (1,-1,0.6)', (1,1,0.6)'\}$，来自5,000个模拟队列。该比值是使用后分层权重所带来的效率增益的度量。

E.5 使用较弱代理变量校准权重的额外模拟

在主文档第 7 节和网络附录 E.2 的模拟中，SCC.Calib 的结果是通过使用代理变量 $\boldsymbol{X}_1 = (X_{u1}, X_{u2}, X_{u3})'$ 得到的，其中 $X_{u1} = X_1 + \varepsilon_1, \varepsilon_1 \sim \mathcal{N}(0, 0.75^2)$，因此 $\mathrm{corr}(X_{u1}, X_1) = 0.8$。在本节中，我们进行了相同的模拟，但使用了 $\varepsilon_1 \sim \mathcal{N}(0, 1.2^2)$，因此 $\mathrm{corr}(X_{u1}, X_1) \approx 0.64$。

网络表 23 和 **网络表 24** 分别展示了 SCC.Calib 对数相对风险 β_1 以及特征组合为 $\boldsymbol{x} = (-1, 1, -0.6)'$ 的纯风险的估计结果。由于代理变量 X_1，较弱，稳健方差公式高估了方差，因此在许多情景下置信区间覆盖率高于名义水平。此外，与较强代理变量相比，效率增益较小；参见网络附录 E.2 中的经验方差：**网络表 15** 和 **网络表 18**。

SCC.Calib								
经验方差	估计方差的均值方差		95%置信区间覆盖率		n	K	p_Y	β_1
	VfiRobust	$\hat{V}$	VfiRobust	$\hat{V}$				
0.0184	0.0193	0.0182	0.9582*	0.9524	5000	2	0.02	-0.2
0.0161	0.0165	0.016	0.954	0.9506	5000	4	0.02	-0.2
0.0085	0.0091	0.0086	0.9584*	0.9528	10000	2	0.02	-0.2
0.0076	0.008	0.0077	0.9562*	0.9532	10000	4	0.02	-0.2
0.0069	0.0072	0.0068	0.954	0.9478	5000	2	0.05	-0.2
0.0061	0.0063	0.0061	0.9516	0.9486	5000	4	0.05	-0.2
0.0033	0.0035	0.0033	0.9586*	0.9526	10000	2	0.05	-0.2
0.0031	0.0031	0.003	0.9526	0.9494	10000	4	0.05	-0.2
0.0033	0.0034	0.0033	0.9582*	0.9536	5000	2	0.1	-0.2
0.0031	0.0031	0.003	0.9482	0.9466	5000	4	0.1	-0.2
0.0016	0.0017	0.0016	0.959*	0.9528	10000	2	0.1	-0.2
0.0015	0.0015	0.0015	0.952	0.9496	10000	4	0.1	-0.2

网络表 23- 使用不同方差估计 方法对 5,000 个模拟队列中的对数相对风险 β_1, 进行估计的结果（使用较弱的代理变量），采用 SCC.Calib。 * 表示覆盖率超出预期区间 [0.9440; 0.9560]。

经验方差	SCC.Calib				n	K	p_Y	$\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$
	均值估计的方差		95%覆盖率 CIs					
	VfiRobust	$\hat{V}$	VfiRobust	$\hat{V}$				
0.0292	0.031	0.0292	0.9576*	0.9514	5000	2	0.02	-3.948
0.027	0.0278	0.027	0.9526	0.949	5000	4	0.02	-3.948
0.014	0.0148	0.014	0.959*	0.9524	10000	2	0.02	-3.948
0.0128	0.0135	0.0131	0.9584*	0.9536	10000	4	0.02	-3.948
0.0106	0.0114	0.0108	0.9618*	0.9562*	5000	2	0.05	-3.046
0.0105	0.0104	0.0102	0.9478	0.945	5000	4	0.05	-3.046
0.0053	0.0056	0.0053	0.954	0.9494	10000	2	0.05	-3.046
0.005	0.0051	0.005	0.9546	0.9516	10000	4	0.05	-3.046
0.005	0.0053	0.0051	0.9532	0.9486	5000	2	0.1	-2.377
0.005	0.0049	0.0048	0.9494	0.9472	5000	4	0.1	-2.377
0.0024	0.0026	0.0025	0.959*	0.9548	10000	2	0.1	-2.377
0.0025	0.0024	0.0024	0.948	0.9462	10000	4	0.1	-2.377

网络表 24- 使用不同方差估计 方法，在 5,000个模拟队列中，采用 SCC.Calib（使用较弱的代理变量）估计特征组合 $\boldsymbol{x} = (-1, 1, -0.6)'$对应的纯风险的结果。，* 表示覆盖率超出预期区间 [0.9440；0.9560]。

网络附录 F. GOLESTAN 数据分析——2,500次重复的结果

我们在主文档第 8 节中，使用从 Golestan 数据集中有放回抽取的 2,500 个队列（每个队列的大小为 n =20,000），重复了数据分析。由此，我们能够比较估计方差的均值与经验方差。回顾一下，我们使用年龄时间尺度，并拟合 Cox 比例风险模型，根据基线变量预测死亡率：x_1 =男性指标、x_2 = 财富评分、x_3 = 既往吸烟者指标（香烟、鼻烟或鸦片）、x_4 = 当前吸烟者指标、x_5 = 发病率指标、$x_6 = x_1x_3$ 和 $x_7 = x_1x_4$。对于分层抽样，**网络表 25**和**网络表 26**分别展示了对数相对风险和协变量特异性纯风险的估计结果。非分层抽样的结果展示于**网络表 27**和**网络表 28**。

在分层设计中使用设计权重时，稳健方差公式高估了协变量x_1 和x_2 的对数相对风险以及具有特征组合$\boldsymbol{x} \in \{(0,-0.4,0,1,\mathbf{0}_3)', (\mathbf{0}_4,1,\mathbf{0}_2)', \mathbf{0}_7'\}$的纯风险的方差。对于$x_6$ 的对数相对风险以及具有特征组合$\boldsymbol{x} = (0,0.4,0,1,\mathbf{0}_3)'$的纯风险，稳健方差估计的均值实际上最接近经验方差，但两种方差估计的均值之间的差异可以忽略不计。这可能是由于违反了比例风险假设，或者是因为可用的吸烟变量过于粗略（例如，每天分别吸食2支和30支香烟的两个个体可能具有相同的协变量特征组合）。此外，在此数据示例中，稳健方差公式为分层设计中的许多参数带来了有效推断，这可能是因为分层仅基于7个协变量中的一个，即x_1。另一方面，对于具有特征组合$\boldsymbol{x}$ ∈的纯风险，稳健方差公式略微高估了方差。

$\{(\mathbf{0}_4, 1, \mathbf{0}_2)', \mathbf{0}_7'\}$，在采用设计权重的非分层设计中。对于另外两个纯风险，稳健方差估计的均值与经验方差最为接近。分层提高了效率，但校准设计权重使估计效率进一步提高。具体而言，在非分层病例队列设计中使用校准权重所获得的精确度，与使用整个队列或在分层病例队列中使用校准权重所获得的精确度相当，但协变量 x_2 的对数相对风险除外。

Parameter		队列		SCC			SCC.Calib		
		Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$
Log-relative hazard	β_1	0.002	0.002	0.0024	0.0028	0.0024	0.002	0.002	0.002
	β_2	0.0071	0.0073	0.0103	0.0107	0.0104	0.0076	0.0077	0.0077
	β_3	0.0253	0.0254	0.0434	0.0419	0.0419	0.0253	0.0261	0.0261
	β_4	0.0042	0.0042	0.0081	0.008	0.008	0.0042	0.0042	0.0042
	β_5	0.0013	0.0013	0.002	0.002	0.002	0.0013	0.0013	0.0013
	β_6	0.0321	0.0319	0.0518	0.0501	0.05	0.0321	0.0326	0.0326
	β_7	0.0063	0.0063	0.0109	0.0109	0.0109	0.0064	0.0064	0.0064

网络表 25- 在来自Golestan队列的2,500次重复中，采用不同的分析方法和方差估计方法进行分层抽样所得的对数相对风险参数的经验方差及估计方差的均值。β_p 表示协变量x_p的对数相对风险参数，$p \in \{1, \dots, 7\}$。

Parameter		队列		SCC			SCC.Calib		
		Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$
$\pi(\tau_1, \tau_2; \boldsymbol{x})$	$\boldsymbol{x} = (0, -0.4, 0, 1, \mathbf{0}_3)'$	0.0035	0.0036	0.0063	0.0066	0.0064	0.0036	0.0037	0.0037
	$\boldsymbol{x} = (0, 0.4, 0, 1, \mathbf{0}_3)'$	0.0053	0.0052	0.0091	0.009	0.0086	0.0054	0.0053	0.0053
	$\boldsymbol{x} = (\mathbf{0}_4, 1, \mathbf{0}_2)'$	0.0014	0.0014	0.002	0.0022	0.0019	0.0014	0.0014	0.0014
	$\boldsymbol{x} = \mathbf{0}_7'$	0.0012	0.0012	0.0014	0.0018	0.0014	0.0013	0.0013	0.0013

网络表 26- 使用不同分析方法和方差估计方法对分层样本进行2,500次重复，从Golestan队列获得的纯风险参数的经验方差和估计方差的均值。

Parameter		队列		USCC			USCC.Calib		
		Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$
Log-relative hazard	β_1	0.002	0.002	0.0039	0.0039	0.0039	0.002	0.002	0.002
	β_2	0.0071	0.0073	0.015	0.0146	0.0146	0.0081	0.0083	0.0083
	β_3	0.0253	0.0254	0.0566	0.0541	0.0541	0.0254	0.0277	0.0277
	β_4	0.0042	0.0042	0.0097	0.0093	0.0093	0.0042	0.0042	0.0042
	β_5	0.0013	0.0013	0.0029	0.0028	0.0028	0.0013	0.0013	0.0013
	β_6	0.0321	0.0319	0.0707	0.0686	0.0686	0.0322	0.0343	0.0343
	β_7	0.0063	0.0063	0.0148	0.014	0.0141	0.0064	0.0064	0.0064

网络表 27- 在 Golestan队列的 2,500次重复中，采用非分层抽样时，使用不同分析方法和方差估计方法得到的对数相对风险参数的经验方差和估计方差的均值。β_p 表示协变量 x_p的对数相对风险参数，$p \in \{1, \ldots, 7\}$。

Parameter		队列		USCC			USCC.Calib		
		Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	Emp. 方差	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$
$\pi(\tau_1, \tau_2; \boldsymbol{x})$	$\boldsymbol{x} = (0, -0.4, 0, 1, \mathbf{0}_3)'$	0.0035	0.0036	0.0078	0.0077	0.0076	0.0036	0.0037	0.0037
	$\boldsymbol{x} = (0, 0.4, 0, 1, \mathbf{0}_3)'$	0.0053	0.0052	0.0115	0.0109	0.0108	0.0056	0.0055	0.0055
	$\boldsymbol{x} = (\mathbf{0}_4, 1, \mathbf{0}_2)'$	0.0014	0.0014	0.0024	0.0025	0.0024	0.0014	0.0015	0.0014
	$\boldsymbol{x} = \mathbf{0}_7'$	0.0012	0.0012	0.0018	0.0019	0.0018	0.0013	0.0013	0.0013

网络表28-在来自Golestan队列的2,500次重复中，采用非分层抽样的不同分析方法和方差估计方法所得纯风险参数的经验方差和估计方差的均值。

网络附录 G. 第三阶段设计抽样概率未知时的缺失数据

G.1 影响函数的推导

我们令 $\boldsymbol{\Delta}_{i,j}(\boldsymbol{\theta})$ 表示层 j 中受试者 i 对 $\boldsymbol{\theta}$ 的影响，$i \in \{1, \ldots, n_{(j)}\}, j \in \{1, \ldots, J\}$，$\boldsymbol{\theta} \in \{\boldsymbol{\gamma}, \boldsymbol{\beta}, d\Lambda_0(t), \Lambda_0(t), \pi_0(\tau_1, \tau_2; \boldsymbol{x})\}$.

由于 $\sum_{j=1}^{J} \sum_{i=1}^{n_{(j)}} \boldsymbol{\Delta}_{i,j}(\boldsymbol{\theta}) = 0$，然后，遵循 Graubard 和 Fears (2005)，我们知道

$$\xi_{i,j}\boldsymbol{B}_{i,j} - \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j}\right)\xi_{i,j}\, V_{i,j}\, \boldsymbol{B}_{i,j} - \left\{\sum_{l=1}^{J}\sum_{k=1}^{n_{(l)}} \xi_{k,l}\, V_{k,l} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}\right)\, \boldsymbol{B}_{k,l}\boldsymbol{B}_{k,l}'\right\}\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\gamma}}) = 0,$$

因此

$$\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\gamma}}) = \left\{\sum_{l=1}^{J}\sum_{k=1}^{n_{(l)}} \xi_{k,l}\, V_{k,l} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}\right) \boldsymbol{B}_{k,l}\boldsymbol{B}_{k,l}'\right\}^{-1} \left\{\xi_{i,j}\boldsymbol{B}_{i,j} - \xi_{i,j}\, V_{i,j} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j}\right)\boldsymbol{B}_{i,j}\right\}.$$

于是我们可以写出

$$\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\gamma}}) = \xi_{i,j}\, \boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\gamma}}) + \xi_{i,j}\, V_{i,j} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j}\right) \boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\gamma}}),$$

其中 $\boldsymbol{IF}_{i,j}^{(2)}(\tilde{\boldsymbol{\gamma}}) = \left\{\sum \sum \xi_{k,l}\, V_{k,3} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,3}\right)\boldsymbol{B}_{k,l}\boldsymbol{B}_{k,3}'^{\,n_{(l)}\,25(6\,35(}\right\}^{-1} \boldsymbol{B}_{i,j},$

和 $\boldsymbol{IF}_{i,j}^{(3)}(\tilde{\boldsymbol{\gamma}}) = -\left\{\sum \sum \xi_{k,l}\, V_{k,3} \exp\left(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,3}\right)\boldsymbol{B}_{k,l}\boldsymbol{B}_{k,3}'^{\,n_{(l)}\,25(6\,35(}\right\}^{-1} \boldsymbol{B}_{i,j_0}$

那么，我们知道 $\int \sum_{=} \sum_{=} V_{i,}^{\,n}{}_{j(j)}$! (6 # (, exp#$\boldsymbol{\gamma}$n′$\boldsymbol{B}_{i,j}$& G$X_{i,j} - {}^{S}$T#$\underline{(\ \gamma\beta)}$,;V,W X"(,;$\boldsymbol{\gamma}$V,$\boldsymbol{\beta}$W0I d$N_{i,j}(t) = 0$. 我们可以将此估计方程改写为 $\int \sum_{=} \sum_{=} \xi_{i,}^{\,n}{}_{j} V_{i\,j}\, w_{i\,j}^{(2)}$, , (ȷ) ! (6 # (, exp#$\boldsymbol{\gamma}$n′$\boldsymbol{B}_{i,j}$& G$X_{i,j} - {}^{S}$T#$\underline{(\ \gamma\beta)}$,;V,W X"(,;$\boldsymbol{\gamma}$V,$\boldsymbol{\beta}$W0I d$N_{i,j}(t) = 0$, 因为对于层 j 中满足 $\int$ d $N_{i\,j}\ t\ = 1$, ( ) 的任意受试者 i，都有 $\xi_{i,j}\, w_{i,j}^{(2)} = 1$（即所有病例均纳入第二阶段样本，且第二阶段抽样权重为 1）。因此，按照与 $\boldsymbol{\Delta}_{i,j}(\boldsymbol{\gamma}$n) 类似的论证，我们知道

$$\boldsymbol{\Delta}_{i,j}\{\widetilde{\boldsymbol{G}}_1(\widetilde{\boldsymbol{\gamma}})\} + \int_t \Bigg[- \frac{\Delta_{i,j}\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}) \mathrm{d}N_{k,l}(t)\right\} \times \tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})} -$$

$$\frac{\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}) \mathrm{d}N_{k,l}(t)\right\} \times \boldsymbol{\Delta}_{i,j}\{\tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})\}}{\tilde{S}_0(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})} + \tag{3}$$

$$\frac{\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l}) \mathrm{d}N_{k,l}(t)\right\} \times \tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}}) \times \Delta_{i,j}\{\tilde{S}_0(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})\}}{\{\tilde{S}_0(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})\}^2} \Bigg] = 0,$$

with

$$\widetilde{\boldsymbol{G}}_1(\widetilde{\boldsymbol{\gamma}}) = \int_t \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, \boldsymbol{X}_{k,l}\, \mathrm{d}N_{k,l}(t),$$

$$\boldsymbol{\Delta}_{i,j}\{\widetilde{\boldsymbol{G}}_1(t,\widetilde{\boldsymbol{\gamma}})\} = \xi_{i,j}\, V_{i,j}\, w_{i,j}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\, \mathrm{d}N_{i,j}(t)\, \boldsymbol{X}_{i,j} + \left\{\frac{\partial \widetilde{\boldsymbol{G}}_1(t,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}}\right\} \boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\gamma}}),$$

$$\Delta_{i,j}\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, \mathrm{d}N_{k,l}(t)\right\} = \xi_{i,j}\, V_{i,j}\, w_{i,j}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\, \mathrm{d}N_{i,j}(t) +$$

$$\left[\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\boldsymbol{\gamma}'\boldsymbol{B}_{k,l}) \mathrm{d}N_{k,l}(t)\right\}}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}}\right] \boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\gamma}}),$$

$$\frac{\partial \widetilde{\boldsymbol{G}}_1(t,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, \mathrm{d}N_{k,l}(t)\, \boldsymbol{X}_{k,l}\boldsymbol{B}_{k,l}{}',$$

$$\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\boldsymbol{\gamma}'\boldsymbol{B}_{k,l}) \mathrm{d}N_{k,l}(t)\right\}}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, \mathrm{d}N_{k,l}(t)\, \boldsymbol{B}_{k,l}{}',$$

$$\boldsymbol{\Delta}_{i,j}\{\tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})\} = \xi_{i,j}\, V_{i,j}\, w_{i,j}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\, Y_{i,j}(t) \exp(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j})\, \boldsymbol{X}_{i,j} + \left\{\frac{\partial \tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}\right\} \boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\beta}}) +$$

$$\left\{\frac{\partial \tilde{\boldsymbol{S}}_1(t;\boldsymbol{\gamma},\widetilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}}\right\} \boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\gamma}}),$$

$$\frac{\partial \tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, Y_{k,l}(t) \exp(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l})\, \boldsymbol{X}_{k,l}\, \boldsymbol{X}_{k,l}{}' \equiv \tilde{\boldsymbol{S}}_2(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}}),$$

$$\frac{\partial \tilde{\boldsymbol{S}}_1(t;\boldsymbol{\gamma},\widetilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, Y_{k,l}(t) \exp(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l})\, \boldsymbol{X}_{k,l}\, \boldsymbol{B}_{k,l}{}',$$

$$\Delta_{i,j}\{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})\} = \xi_{i,j}\, V_{i,j} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\, w_{i,j}^{(2)}\, Y_{i,j}(t) \exp(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}) + \left\{\frac{\partial \tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}_{|\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}\right\}\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\beta}}) +$$

$$\left\{\frac{\partial \tilde{S}_0(t;\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}\right\}\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\gamma}}),$$

$$\frac{\partial \tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}_{|\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, w_{k,l}^{(2)}\, Y_{k,l}(t) \exp(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l})\, \boldsymbol{X}_{k,l}{}' = \tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})',$$

and $\frac{\partial \tilde{S}_0(t;\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}} = \sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, w_{k,l}^{(2)}\, Y_{k,l}(t) \exp(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l})\, \boldsymbol{B}_{k,l}{}'$.

Note, $\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, \mathrm{d}N_{k,l}(t)$, $\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\boldsymbol{\gamma}'\boldsymbol{B}_{k,l})\mathrm{d}N_{k,l}(t)\right\}}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}$

and thus $\Delta_{i,j}\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\, \mathrm{d}N_{k,l}(t)\right\}$ and the term between brackets in Equation (3) are linear combinations of $\mathrm{d}N_{i,j}(t)$. Hence

$$\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\beta}}) = \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \int_t \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\left\{\frac{\tilde{\boldsymbol{S}}_2(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})'}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1} \times$$

$$\left[\int_t \left\{\boldsymbol{X}_{i,j} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}\right\} \times \left\{\xi_{i,j}\, V_{i,j}\, w_{i,j}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\, \mathrm{d}N_{i,j}(t) -\right.\right.$$

$$\left.\left.\xi_{i,j}\, V_{i,j}\, w_{i,j}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\, Y_{i,j}(t) \exp(\tilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}) \times \frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\mathrm{d}N_{k,l}(t)}{\tilde{S}_0(t;\tilde{\gamma},\tilde{\beta})}\right\}\right] +$$

$$\left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}} \int_t \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\tilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\left\{\frac{\tilde{\boldsymbol{S}}_2(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})} - \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})'}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1} \times$$

$$\left(\frac{\partial \tilde{\boldsymbol{G}}_1(t,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}} - \int_t \frac{\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})} \times \left[\frac{\partial\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\boldsymbol{\gamma}'\boldsymbol{B}_{k,l})\mathrm{d}N_{k,l}(t)\right\}}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}\right] -\right.$$

$$\left.\int_t \frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\boldsymbol{\gamma}'\boldsymbol{B}_{k,l})\mathrm{d}N_{k,l}(t)}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})} \times \left\{\frac{\partial \tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\boldsymbol{\beta})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}\right\} +\right.$$

$$\left.\int_t \frac{\left\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)} \exp(\boldsymbol{\gamma}'\boldsymbol{B}_{k,l})\mathrm{d}N_{k,l}(t)\right\}\tilde{\boldsymbol{S}}_1(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})}{\tilde{S}_0(t;\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\beta}})^2} \times \left\{\frac{\partial \tilde{S}_0(t;\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\gamma}}_{|\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}\right\}\right)\boldsymbol{\Delta}_{i,j}(\tilde{\boldsymbol{\gamma}}),$$

and we can write

$$\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\beta}}) = \xi_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\beta}}) + \xi_{i,j}\, V_{i,j}\exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\,\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\beta}}),$$

其中 $\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\beta}\mathrm{l}\& = \mathrm{Z}\Sigma \quad \Sigma \quad \xi_{k,l}\, V_{k,l}\exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{k,l}\&\boldsymbol{Z}\mathrm{l}_{k,l}\boldsymbol{B}_{k,l'}{}^{n_{(j)}\ k=1J\ l=1}\Big\}\times\boldsymbol{IF}_{i,j}^{(2)}(\boldsymbol{\gamma}\mathrm{n})$,

和 $\boldsymbol{IF}_{i,j}^{(3)}\#\boldsymbol{\beta}\mathrm{l}\& = \boldsymbol{Z}\mathrm{l}_{i,j} + \mathrm{Z}\Sigma \quad \Sigma \quad \xi_{k,l}\, V_{k,l}\exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{k,l}\&\boldsymbol{Z}\mathrm{l}_{k,l}\boldsymbol{B}_{k,l'}{}^{n_{(j)}\ k=1J\ l=1}\Big\}\times\boldsymbol{IF}_{i,j}^{(3)}(\boldsymbol{\gamma}\mathrm{n})$,

and 其中 $\widetilde{\boldsymbol{Z}}_{!,\#} = w_{i,j}^{(2)}{}_{\mathrm{W0}}{}^{\times}\Big[\sum_{l=}^{J}\ \sum_{k=}^{n}\ \int \xi_{k,l}\, V_{k,l}\, w_{k,l}^{(2)}\exp(\boldsymbol{\gamma}^{\sim}\boldsymbol{B}_{k,l})\ {}_{,(l)}\ ((\Big\{\frac{\boldsymbol{s}\ (\ \boldsymbol{\gamma}\boldsymbol{\beta})}{S}\ \mathrm{T}_{2},;\mathrm{V},\mathrm{W}\ \mathrm{X}_{0}(,;\boldsymbol{\gamma}\mathrm{V},\boldsymbol{\beta} -$

$\frac{\boldsymbol{s}^{\mathrm{T}_1}(,;\boldsymbol{\gamma}\mathrm{V},\boldsymbol{\beta}\mathrm{W})\boldsymbol{s}^{\mathrm{T}_1}(,;\boldsymbol{\gamma}\mathrm{V},\boldsymbol{\beta}\mathrm{W}0'}{1\mathrm{X}_0(,;\boldsymbol{\gamma}\mathrm{V},\boldsymbol{\beta}\ \mathrm{W}0_2}\mathrm{Id}N_{k,l}(t)\mathrm{K}^{\ -1}\times\mathrm{L}\int_{j}^{t}\mathrm{G}\boldsymbol{X}_{i,} - \frac{\boldsymbol{s}\ (\ \boldsymbol{\gamma}\boldsymbol{\beta})^{\mathrm{T}_{1},;}}{\tilde{S}_{\mathrm{V},\mathrm{W}}\mathrm{X}_0(,;\boldsymbol{\gamma}\mathrm{V},\boldsymbol{\beta}\ \mathrm{W}0\Big\}}\times\mathrm{Nd}N_{i,j}(t) - Y_{i,j}(t)\exp\#\boldsymbol{\beta}\mathrm{l}'\boldsymbol{X}_{i,j}\&$

$\underline{\sum_{n\,(j)}\ \sum_{k=1J\ l=1}\ \xi_{k,l}\, V_{k,l}\, w_{k,l(2)}\exp(\boldsymbol{\gamma}^{\sim\prime}\ \boldsymbol{B}_{kl})\mathrm{d}N_{k,l}(t)}{S^{\sim}{}_0(t;\boldsymbol{\gamma}^{\sim},\boldsymbol{\beta}^{\sim})}\ \mathrm{OP.}$

那么我们知道 $\sum_{=}\Sigma\mathrm{Zd}^{n\ j}_{=}\ N_{i\ j}\ t\ -\ \Lambda_0\ t\ \xi_{i\ j}V_{i\ j}\ w_{i\ j}^{(2)}\ \boldsymbol{\gamma}'\boldsymbol{B}_{i\ j}\ Y_{i\ j}\ t\ \boldsymbol{\beta}'\boldsymbol{X}_{i\ j}\Big\},(\,)\,\mathrm{dl}\,(\,),\ ,\ ,\ \exp(\mathrm{n},\ ),\ (\,)\,\exp\#\mathrm{l},\ \&_{()\ i1J\ j1} = 0,$

由于我们通过$\mathrm{d}\Lambda\mathrm{l}_0(t) = {}^{\Sigma_j\ \ \Sigma\ \ \mathrm{d}}{}_{Ni,j(t)n\,(j)\ i=1J\ =1\tilde{S}^{\sim}0(t;\boldsymbol{\gamma}^{\sim},\boldsymbol{\beta}^{\sim})}$ 非参数地估计基线风险，因此

$$\mathrm{d}N_{i,j}(t) - \mathrm{d}\widetilde{\Lambda}_0(t)\,\xi_{i,j}V_{i,j}\,w_{i,j}^{(2)}\exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\,Y_{i,j}(t)\exp(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}) - \mathrm{d}\widetilde{\Lambda}_0(t)\,\tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})'\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\beta}})$$

$$-\mathrm{d}\widetilde{\Lambda}_0(t)\Big\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}}\xi_{k,l}\,V_{k,l}\,w_{k,l}^{(2)}\exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\,Y_{k,l}(t)\exp(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l})\,\boldsymbol{B}_{k,l}\Big\}'\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\gamma}})$$

$$-\tilde{S}_0(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})\,\Delta_{i,j}\{\mathrm{d}\widetilde{\Lambda}_0(t)\} = 0.$$

因此,

$$\Delta_{i,j}\{\mathrm{d}\widetilde{\Lambda}_0(t)\} = \{\tilde{S}_0(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})\}^{-1}\Big[\mathrm{d}N_{i,j}(t) - \mathrm{d}\widetilde{\Lambda}_0(t)\,\xi_{i,j}V_{i,j}\,w_{i,j}^{(2)}\exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\,Y_{i,j}(t)\exp(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}) -$$

$$\mathrm{d}\widetilde{\Lambda}_0(t)\,\tilde{\boldsymbol{S}}_1(t;\widetilde{\boldsymbol{\gamma}},\widetilde{\boldsymbol{\beta}})'\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\beta}}) -$$

$$\mathrm{d}\widetilde{\Lambda}_0(t)\Big\{\sum_{l=1}^{J}\sum_{k=1}^{n^{(j)}}\xi_{k,l}\,V_{k,l}\,w_{k,l}^{(2)}\exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{k,l})\,Y_{k,l}(t)\exp(\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}_{k,l})\,\boldsymbol{B}_{k,l}\Big\}'\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\gamma}})\Big],$$

并且我们可以写出

$$\Delta_{i,j}\{\mathrm{d}\widetilde{\Lambda}_0(t)\} = \xi_{i,j}\,IF_{i,j}^{(2)}\{\mathrm{d}\widetilde{\Lambda}_0(t)\} + \xi_{i,j}\,V_{i,j}\exp(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j})\,IF_{i,j}^{(3)}\{\mathrm{d}\widetilde{\Lambda}_0(t)\},$$

with

$IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\} = \{S\mathrm{t0\#}t;\boldsymbol{\gamma}\mathrm{n},\ \boldsymbol{\beta}\mathrm{l\&}\}^{-1}\mathrm{d}N_{i,j(t)} + \mathrm{z}\Sigma \quad \Sigma \quad \xi_{k,l}\ V_{k,3}\ \exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{k,3\&}H\mathrm{l}^{k,l}(t)\boldsymbol{B}_{k,3'}{}^{n}{}_{(j)\ 25(6\ 35(}\Big\}\times$

$\boldsymbol{IF}_{i,j}^{(2)}(\boldsymbol{\gamma}\mathrm{n}),$

和$IF_{i,j}^{(3)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\} = H\mathrm{l}_{i,j}(t) + \mathrm{z}\Sigma \quad \Sigma \quad \xi_{k,l}\ V_{k,l}\ \exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{k,l\&}H\mathrm{l}^{k,l}(t)\boldsymbol{B}_{k,l'}{}^{n}{}_{(j)\ k=1J\ l=1}\Big\}\times \boldsymbol{IF}_{i,j}^{(3)}(\boldsymbol{\gamma}\mathrm{n}),$

以及其中 $H\mathrm{l}_{i,j}(t) = -S\mathrm{t0\#}t;\boldsymbol{\gamma}\mathrm{n},\ \boldsymbol{\beta}\mathrm{l\&}^{-1}\mathrm{d}\Lambda\mathrm{l}_0(t)\ \mathrm{z}\boldsymbol{S}\mathrm{u}_1\#t;\ \boldsymbol{\gamma}\mathrm{n},\ \boldsymbol{\beta}\mathrm{l\&}'\boldsymbol{Z}\mathrm{l}_{i,j} + K\mathrm{l}_{i,j}(t)\Big\},$

以及 $K\mathrm{l}_{i,\ j}(t) = w_{i,\ j}^{(2)}\ Y_{i,\ j}(t)\ \exp\#\boldsymbol{\beta}\mathrm{l}'\boldsymbol{X}_{i,\ j\&}$。

注意，$H\mathrm{l}_{i、\ j}(t)$，因此 $IF_{i、\ j}^{(2)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\}$ 和 $IF_{i、\ j}^{(3)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\}$，都是增量 $\mathrm{d}N_{i、\ j}(t)$ 和 $\mathrm{d}\Lambda\mathrm{l}_0(t)$ 的线性组合。因此 Δ_i 、 $_j\mathrm{z}\int \mathrm{d}\Lambda\mathrm{l}_0(t)^{\tau_2}\ \mathrm{A\#}\Big\} = \int \ \Delta_{i,j}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\}\ \mathrm{A!}\ \mathrm{A\#}$，我们可以写成

$$\xi_{i,j}\ IF_{i,j}^{(2)}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widetilde{\Lambda}_0(t)\right\} + \xi_{i,j}\ V_{i,j}\ \exp\left(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j}\right)\ IF_{i,j}^{(3)}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widetilde{\Lambda}_0(t)\right\},$$

其中 $IF_{i,\ j}^{(2)}\mathrm{z}\int \mathrm{d}\Lambda\mathrm{l}_0(t)^{\tau_2\ \tau_1}\Big\} = \sum_{=}\ IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\}_{\ \tau!\ ,\ '}$，

以及 $IF_{i,\ j,\ }^{(2\ 3)}\mathrm{z}\int \mathrm{d}\Lambda\mathrm{l}_0(t)^{\tau_2\ \tau_1}\Big\} = \sum_{=}\ IF_{i,j}^{(3)}\{\mathrm{d}\Lambda\mathrm{l}_0(t)\}_{\ \tau!\ ,\ '}$，最终得到

$$\Delta_{i,j}\{\tilde{\pi}(\tau_1,\tau_2;\boldsymbol{x})\} = \left\{\frac{\partial\tilde{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}\right\}\boldsymbol{\Delta}_{i,j}\left(\widetilde{\boldsymbol{\beta}}\right) + \left[\frac{\partial\tilde{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\mathrm{d}\Lambda_0(t)=\mathrm{d}\widetilde{\Lambda}_0(t)}\right]\Delta_{i,j}\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widetilde{\Lambda}_0(t)\right\},$$

我们可以写成

$$\xi_{i,j}\ IF_{i,j}^{(2)}\{\tilde{\pi}(\tau_1,\tau_2;\boldsymbol{x})\} + \xi_{i,j}\ V_{i,j}\ \exp\left(\widetilde{\boldsymbol{\gamma}}'\boldsymbol{B}_{i,j}\right)IF_{i,j}^{(3)}\{\tilde{\pi}(\tau_1,\tau_2;\boldsymbol{x})\},$$

with $\frac{\partial\tilde{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\boldsymbol{\beta}}_{|\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} = \left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\widetilde{\Lambda}_0(t)\ \exp\left(\widetilde{\boldsymbol{\beta}}'\boldsymbol{x}\right)\right\}\{1-\tilde{\pi}(\tau_1,\tau_2;\ \boldsymbol{x})\}\boldsymbol{x}',$

and $\frac{\partial\tilde{\pi}(\tau_1,\tau_2;\boldsymbol{x})}{\partial\left\{\int_{\tau_1}^{\tau_2}\mathrm{d}\Lambda_0(t)\right\}}_{|\mathrm{d}\Lambda_0(t)=\mathrm{d}\widetilde{\Lambda}_0(t)} = \exp\left(\widetilde{\boldsymbol{\beta}}'\boldsymbol{x}\right)\{1-\tilde{\pi}(\tau_1,\tau_2;\ \boldsymbol{x})\}.$

注意，对于任意 $\boldsymbol{\theta}\mathrm{l} \in \{\boldsymbol{\gamma}\mathrm{n},\ \boldsymbol{\beta}\mathrm{l},\ \mathrm{d}\Lambda\mathrm{l}_0(t),\ \Lambda\mathrm{l}_0(t),\ \pi\mathrm{o}\ (\tau_1,\tau_2;\ \boldsymbol{x})\}$，如果层 j 中的受试者 i 属于第二阶段样本，但不属于第三阶段样本，那么 $\xi_{i,\ j}\ V_{i,\ j}\ \exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{i,\ j\&}\ \boldsymbol{IF}_{i,\ j}^{(3)}\#\boldsymbol{\theta}\mathrm{l\&}$ 为零。然而，

$\xi_{i,\ j}\boldsymbol{IF}_{i,\ j}^{(2)}\#\boldsymbol{\theta}$l& 通常不为零。具体而言，由于该个体被用于估计第三阶段抽样权重，因此他/她会通过对 $\boldsymbol{\gamma}$n 的影响而影响 $\boldsymbol{\theta}$l。

G.2 方差分解与估计

如主文档第5.5节所述，对于任意 $\boldsymbol{\theta}$l $\in\{\boldsymbol{\gamma}$n、$\boldsymbol{\beta}$l、dΛl$_0(t)$、Λl$_0(t)$、πo$(\tau_1,\tau_2;\boldsymbol{x})\}$，利用全协方差公式和全期望公式，我们可以将方差Z$\Sigma_{\ =\ \Sigma_{\ =\ }\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}}$l&$^{n_{(j)}\quad i1J\ j1}\quad\}$分解为

$$\mathrm{var}\left(\mathrm{E}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right)+\ \mathrm{E}\left(\mathrm{var}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right)+$$

$$\mathrm{E}\left(\mathrm{E}\left[\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right),$$

其中，C_1 表示来自整个队列的信息，C_2 表示来自第二阶段样本的信息。

对于任意 $j,l\in\{1,\ldots,J\}$，我们知道，如果个体 i 和 k, $i\neq k$在层 j中均为非病例，则 $E\#\xi_{i,j}|C_1\&\ =$(Q),$_j$(!), $E\#\xi_{i,j}\xi_{k,j}|C_1\&\ =$(Q),,,$_j$(!) $=$c ($_j$)(c ($_j$)–() 4 ($_j$)(4 ($_j$)–()；如果 i和/或 k, $i\neq k$在层 j中为病例，则 $E\#\xi_{i,j}\xi_{k,j}|C_1\&\ =$(Q),$_j$(!) ×(Q,,$_j$(!)。此外，如果个体 i 和k 分别属于层 j 和层 l，且 $j\neq l$，则 $E\#\xi_{i,j}\xi_{k,l}|C_1\&\ =$(Q),$_j$(!) × (Q,,'(!)。因此，如果 i 和 k, $i\neq k$在层 j中均为非病例，则 cov$\#\xi_{i,j},\xi_{k,j}|C_1\&=\sigma_{i,k\ j}^{(2)}=$, (Q),,,$_j$(!) −(Q),$_j$(!) × (Q,,$_j$(!)；如果 i 和/或 k 在层 j中为病例，则 cov$\#\xi_{i,j},\xi_{k,j}|C_1\&\ =0$；如果 i 和 k 分别属于层 j 和层 l，且 $j\neq l$，则 cov$\#\xi_{i,j},\xi_{k,l}|C_1\&\ =0$；如果 i 在层 j中为非病例，则 var$\#\xi_{i,j}|C_1\&=\sigma_{i,}$ $_{i\ j}^{(2)}\equiv\sigma_{i,,j}^{(2)}=$(Q),$_j$(!)$\left(1-\right.$(Q),$_j$(!)$\left.\right)$；如果 i 在层 j中为病例，则 var$\#\xi_{i,j}|C_1\&\ =0$。此外，由于第三阶段抽样是

伯努利, $E\#V_{i,j}|C_1,C_2\& = (\mathrm{Q}),_j(6)$ 和 $E\#V_{i,j}V_{k,j}|C_1,C_2\& = (\mathrm{Q}),_j(6) \times (\mathrm{Q},,_j(6)$ 若 $i \neq k$, 则 $\mathrm{cov}\#V_{i,j},V_{k,j}|C_1,C_2\& = \sigma_{i,k\,j}^{(3)} = 0$, 若 $i \neq k$ 且方差$\#V_{i,j}|C_1,C_2\& = \sigma_{i,i\,j}^{(3)} \equiv \sigma_{i,,j}^{(3)} = (\mathrm{Q}),_j(6)\left(1-(\mathrm{Q}),_j(6)\right)$; 并且显然 $E\#V_{i,j}V_{k,l}|C_1,C_2\& = (\mathrm{Q}),_j(6) \times (\mathrm{Q},,'(6)$ 且 $\mathrm{cov}\#V_{i,j},V_{k,l}|C_1,C_2\& = 0$ 在层 j 和层 $l, j \neq l$. 对于我们的符号说明, 我们还将使用 $w_{i,k\,j} = w_{i,,k\,j}^{(2)} \times w_{i,,j}^{(3)} \times w_{k,j}^{(3)}$, 表示 $i \neq k$, 因为$E\#\xi_{i,j}\,\xi_{k,j}\,V_{i,j}\,V_{k,j}|C_1\& = E\{E\#\xi_{i,j}\,\xi_{k,j}\,V_{i,j}\,V_{k,j}|C_1,C_2\&|C_1\},= E\{\xi_{i,j}\,\xi_{k,j}E\#V_{i,j}|C_1,C_2\&E\#V_{k,j}|C_1,C_2\&|C_1\},= (\mathrm{Q}),_j(6) \times (\mathrm{Q},,_j(6) \times E\#\xi_{i,j}\,\xi_{k,j}\,|C_1\& = (\mathrm{Q}),_j(6) \times (\mathrm{Q},,_j(6) \times (\mathrm{Q}),,,_j(!)$, 并且我们还将使用 $w_{i,k\,j}^{(3)} = w_{i,,j}^{(3)} \times w_{k,j}^{(3)}$, 表示 $i \neq k$, $w_{i,j} = w_{i,,j}^{(2)} \times w_{i,j}^{(3)}$, $\sigma_{i,j} = (\mathrm{Q}),_j(!) \times (\mathrm{Q}),_j(6)\left(1-\right.$我们已在网络附录 G.1 中证明, 对于任意 $\boldsymbol{\theta}\mathrm{l} \in \{\boldsymbol{\gamma}\,\mathrm{n},\ \boldsymbol{\beta}\mathrm{l},\ \mathrm{d}\Lambda\mathrm{l}_0(t),\ \Lambda\mathrm{l}_0(t),\ \pi\mathrm{o}(\tau_1,\tau_2;\ \boldsymbol{x})\},\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\mathrm{l}\& = \xi_{i,j}\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\mathrm{l}\& + \xi_{i,j}\,V_{i,j}\,w\mathrm{n}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}\#\boldsymbol{\theta}\mathrm{l}\&$, 其中 $w\mathrm{n}_{i,j}^{(3)} = \exp\#\boldsymbol{\gamma}\mathrm{n}'\boldsymbol{B}_{i,j}\&$. 我们有 $\mathrm{E}\,\mathrm{Z}\sum_{=}\sum_{=}\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\mathrm{l}\&^{n}(_j)\,!\,(6\,\#\,(|C_1,C_2\} = \sum_{=}^{J}\sum\mathrm{Z}$ $\xi_{i\,j}\boldsymbol{IF}_{i\,j}^{(2)}\,\boldsymbol{\theta} + V_{i\,j}\,|C_1,C_2\;\xi_{i\,j}\,w_{i\,j}^{(3)}\boldsymbol{IF}_{i\,j}^{(3)}\,\boldsymbol{\theta}\,\},,\#\mathrm{l}\&\ \mathrm{E}\#\,,\&\,,\mathrm{n}\,,,\#\mathrm{l}\&\ (\,)\,!\,(\#\,(\,,$ $= \overline{\sum_{=}\sum\mathrm{N}\xi_{i,j}\boldsymbol{IF}_{i,j}^{(2)}\#\boldsymbol{\theta}\mathrm{l}\& + \xi_{i,j}\,\mathrm{QV}),_j(6)\,\mathrm{Q}),_j(6)\,\boldsymbol{IF}_{i,j}^{(3)}\#\boldsymbol{\theta}\mathrm{l}\&\mathrm{O}^{n}(_j)\,!=(6\,\#\,(\,,$ 并且我们有

$\frac{1}{w_{i,j}^{(2)}} \times \frac{1}{w_{i,j}^{(3)}}\Big)$ and $\sigma_{i,k,j} = \frac{1}{w_{i,k,j}} - \frac{1}{w_{i,j}} \times \frac{1}{w_{k,j}} = \frac{1}{w_{i,j}^{(3)}} \times \frac{1}{w_{k,j}^{(3)}} \times \sigma_{i,k,j}^{(2)}$, for $i \neq k$.

$$\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}\left(\widetilde{\boldsymbol{\theta}}\right)|C_1,C_2\right\} =$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\mathrm{var}(V_{i,j}|\,C_1,C_2)\left\{\xi_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}\left(\widetilde{\boldsymbol{\theta}}\right)\right\}\left\{\xi_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}\left(\widetilde{\boldsymbol{\theta}}\right)\right\}',$$

$$= \sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(3)}\;\xi_{i,j}\,\widetilde{w}_{i,j}^{(3)}\;\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}\left(\widetilde{\boldsymbol{\theta}}\right)\boldsymbol{IF}_{i,j}^{(3)}\left(\widetilde{\boldsymbol{\theta}}\right)',$$

事实上，$\xi_{i,\ j}\boldsymbol{IF}_{i,}^{(2)}{}_{j}\#\boldsymbol{\theta}$l& 在给定 C_2 的条件下是固定的。

于是我们有

$$\mathrm{E}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\mathrm{E}(\xi_{i,j}\,|C_1)\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+$$

$$\mathrm{E}(\xi_{i,j}|C_1)\frac{\widetilde{w}_{i,j}^{(3)}}{w_{i,j}^{(3)}}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}}),$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\left\{\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\frac{\widetilde{w}_{i,j}^{(3)}}{w_{i,j}^{(3)}}\,\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\},$$

and

$$\mathrm{var}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]=\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\xi_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\xi_{i,j}\frac{\widetilde{w}_{i,j}^{(3)}}{w_{i,j}^{(3)}}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\},$$

$$=\sum_{j=1}^{J}\mathrm{var}\left[\sum_{i=1}^{n^{(j)}}\xi_{i,j}\left\{\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\frac{\widetilde{w}_{i,j}^{(3)}}{w_{i,j}^{(3)}}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}\right],$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\left\{\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\frac{\widetilde{w}_{i,j}^{(3)}}{w_{i,j}^{(3)}}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}\left\{\boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\frac{\widetilde{w}_{k,j}^{(3)}}{w_{k,j}^{(3)}}\boldsymbol{IF}_{k,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}',$$

and

$$\mathrm{E}\left[\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\mathrm{E}(\xi_{i,j}|C_1)\,\sigma_{i,j}^{(3)}\,\widetilde{w}_{i,j}^{(3)}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})',$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\sigma_{i,j}^{(3)}\,\widetilde{w}_{i,j}^{(3)}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'.$$

最后，

$$\mathrm{var}\left(\mathrm{E}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right)=\mathrm{var}\left[\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\left\{\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\frac{\widetilde{w}_{i,j}^{(3)}}{w_{i,j}^{(3)}}\,\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}\right],$$

可通过以下方式估计

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left(\frac{1}{w_{i,j}^{(2)}}\right)^2\left\{\xi_{i,j}\,w_{i,j}^{(2)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})'+2\,\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'+\right.$$

$$\left.\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\},$$

即通过

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\Big\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})' + 2\,\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})' +$$

$$\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\Big\},$$

或者再次通过

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\Bigg\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})' + 2\,\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})' +$$

$$\frac{1}{\widetilde{w}_{i,j}^{(3)}}\xi_{i,j}\,V_{i,j}\widetilde{w}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\Bigg\}.$$

Then

$$\mathrm{E}\left(\mathrm{var}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right) = \mathrm{E}\left[\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\left\{\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) + \frac{\widetilde{w}_{i,j}^{(3)}}{w_{i,j}^{(3)}}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}\left\{\boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) + \frac{\widetilde{w}_{k,j}^{(3)}}{w_{k,j}^{(3)}}\boldsymbol{IF}_{k,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}'\right],$$

可以通过……估计的

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(2)}\left\{\xi_{i,j}\,w_{i,j}^{(2)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})' + 2\,\xi_{i,j}\,V_{i,j}\,w_{i,j}^{(2)}\,\widetilde{w}_{i,j}^{(3)}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})' +\right.$$

$$\left.\xi_{i,j}\,V_{i,j}\,w_{i,j}^{(2)}\,\widetilde{w}_{i,j}^{(3)}\,\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\} +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\ k\neq i}}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\left\{\xi_{i,j}\,\xi_{k,j}\,w_{i,k,j}^{(2)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}})' +\right.$$

$$\xi_{i,j}\,\xi_{k,j}\,V_{k,j}\,w_{i,k,j}^{(2)}\,\widetilde{w}_{k,j}^{(3)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(3)}(\widetilde{\boldsymbol{\theta}})' + \xi_{i,j}\,\xi_{k,j}\,V_{i,j}\,w_{i,k,j}^{(2)}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}})' +$$

$$\left.\xi_{i,j}\,\xi_{k,j}\,V_{i,j}\,V_{k,j}\,\widetilde{w}_{i,k,j}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\},$$

也就是说，
通过

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(2)}w_{i,j}^{(2)}\left\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})'+2\,\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'+\right.$$

$$\left.\frac{1}{\widetilde{w}_{i,j}^{(3)}}\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\}+\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\k\neq i}}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,k,j}^{(2)}\left\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\right.$$

$$\left.\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}\left\{\xi_{k,j}\,\boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\xi_{k,j}\,V_{k,j}\,\widetilde{w}_{k,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}',$$

事实上 $w\mathrm{n}_{i,k\,j}=w_{i,k\,j}^{(2)}\times w_{,\mathrm{n}i,j}^{(3)}\times w_{\mathrm{n}k,j}^{(3)}$, 以及

$$\mathrm{E}\left(\mathrm{E}\left[\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widetilde{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right)=\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\sigma_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\},$$

可由……估计

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\tilde{\sigma}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'.$$

因此，方差#$\boldsymbol{\theta}$l&可估计为

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\left\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})'+2\,\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'+\right.$$

$$\left.\frac{1}{\widetilde{w}_{i,j}^{(3)}}\xi_{i,j}\,V_{i,j}\widetilde{w}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\}+\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(2)}w_{i,j}^{(2)}\left\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})'+\right.$$

$$\left.2\,\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'+\frac{1}{\widetilde{w}_{i,j}^{(3)}}\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\}+$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\k\neq i}}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,k,j}^{(2)}\left\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}\left\{\xi_{k,j}\,\boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\right.$$

$$\left.\xi_{k,j}\,V_{k,j}\,\widetilde{w}_{k,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\right\}'+\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\tilde{\sigma}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\,\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'.$$

注意，我们可以将这个量改写为

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\left\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})'+2\,\xi_{i,j}\,V_{i,j}\,\widetilde{w}_{i,j}^{(3)}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'+\right.$$

$$\left.\frac{1}{\widetilde{w}_{i,j}^{(3)}}\xi_{i,j}\,V_{i,j}\widetilde{w}_{i,j}^{(3)}\widetilde{w}_{i,j}^{(3)}\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\right\}+\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,k,j}^{(2)}\left\{\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})+\right.$$

$$\xi_{i,j}\, V_{i,j}\, \widetilde{w}_{i,j}^{(3)} \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\Big\}\Big\{\xi_{k,j}\, \boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) + \xi_{k,j}\, V_{k,j}\, \widetilde{w}_{k,j}^{(3)} \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\Big\}' +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}} \frac{1}{w_{i,j}^{(2)}} \tilde{\sigma}_{i,j}^{(3)} \widetilde{w}_{i,j}^{(3)} \Big\{\xi_{i,j}\, V_{i,j}\, \widetilde{w}_{i,j}^{(3)} \widetilde{w}_{i,j}^{(3)} \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\Big\},$$

也可以写成

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}} \frac{1}{w_{i,j}^{(2)}} \Bigg\{\xi_{i,j}\, \boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})' + 2\, \xi_{i,j}\, V_{i,j}\, \widetilde{w}_{i,j}^{(3)}\, \boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})' +$$

$$\frac{1}{\widetilde{w}_{i,j}^{(3)}} \xi_{i,j}\, V_{i,j} \widetilde{w}_{i,j}^{(3)} \widetilde{w}_{i,j}^{(3)} \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})' \Bigg\} + \sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}} \tilde{\sigma}_{i,k,j}\, \widetilde{w}_{i,k,j} \Big\{\xi_{i,j}\, \boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) +$$

$$\xi_{i,j}\, V_{i,j}\, \widetilde{w}_{i,j}^{(3)} \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\Big\}\Big\{\xi_{k,j}\, \boldsymbol{IF}_{k,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) + \xi_{k,j}\, V_{k,j}\, \widetilde{w}_{k,j}^{(3)} \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})\Big\}' -$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}} \frac{1}{w_{i,j}^{(2)}} \tilde{\sigma}_{i,j}^{(3)} \widetilde{w}_{i,j}^{(3)} \Big\{\xi_{i,j}\, \boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}})' + 2\, \xi_{i,j}\, V_{i,j}\, \widetilde{w}_{i,j}^{(3)}\, \boldsymbol{IF}_{i,j}^{(2)}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{IF}_{i,j}^{(3)}(\widetilde{\boldsymbol{\theta}})'\Big\}.$$

网络附录 H. 已知第三阶段设计抽样概率时的缺失数据

H.1 参数估计

这里，我们假设第三阶段设计抽样概率已知；在这种情况下，第三阶段设计权重 $w_{i,j}^{(3)}$ 和方差 $\sigma_{i,j}^{(3)}$ 也已知，无需估计,$i \in \{1, \dots, n_{(j)}\}$, $j \in \{1, \dots, J\}$。根据可用数据，我们通过求解以下估计方程来估计对数相对风险：$\boldsymbol{\beta}$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}} \int_t V_{i,j}\, w_{i,j}^{(3)} \Big\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S_1}(t;\boldsymbol{\beta})}{S_0(t;\boldsymbol{\beta})}\Big\} \mathrm{d}N_{i,j}(t) = 0,$$

with $S_0(t;\boldsymbol{\beta}) = \sum_{j=1}^{J}\sum_{k=1}^{n^{(j)}} \xi_{k,j}\, V_{k,j}\, w_{k,j}\, Y_{k,j}(t) \exp\big(\boldsymbol{\beta}'\boldsymbol{X}_{k,j}\big)$ and $\boldsymbol{S}_1(t;\boldsymbol{\beta}) =$ $\sum^{J}_{=}\ \sum_{k=}^{n}\ \xi_{k,j}\, V_{k,j}\, w_{k,j}\, Y_{k,j}(t)\exp\#\boldsymbol{\beta}' X_{k,j}\&_{(j)}\quad {}_{1j1}\quad \boldsymbol{X}_{k,j}$，并且我们估计基线风险点质量

在时间 t 时，$\Lambda\%_0\#t$ 的 $\mathrm{d}\Lambda\%_0\#t$；$\boldsymbol{\beta}\%\& \equiv \mathrm{d}\Lambda\%_0(t) = {}^{\Sigma_j \quad \Sigma \quad \mathrm{d}}_{N_{i,j}(t) n\,(j)\, i=1 J=1 S_0(t;\boldsymbol{\beta}\hat{})}$. 随后，我们估计截至时间 t 的累积基线风险，方法是 $\Lambda\%_0\#t;\boldsymbol{\beta}\%\& \equiv \Lambda\%_0(t) = \int \mathrm{d}\Lambda\%_0(s)^t{}_0$，最后估计纯风险为

$$\hat{\pi}\left(\tau_1, \tau_2;\, \boldsymbol{x}, \widehat{\boldsymbol{\beta}}, \mathrm{d}\widehat{\Lambda}_0\right) \equiv \hat{\pi}(\tau_1, \tau_2; \boldsymbol{x}) = 1 - \exp\left\{-\int_{\tau_1}^{\tau_2} \exp\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{x}\right)\mathrm{d}\widehat{\Lambda}_0(t)\right\}.$$

我们假设队列中所有n 个体的失效时间和病例状态均已知；然后，与主文档中一样，我们保持修改后的非参数Breslow估计量的分子不变。此外，注意到上面给出的估计方程可以改写为$\sum_{=}\ \sum_{=}\ \int \xi_{i,}{}^{n\ j}{}_{\ j}\ V_{i\ j}\ w_{i\ j}\ \ X_{i\ j} - \frac{\boldsymbol{S}_1\ \ \boldsymbol{\beta}}{S\quad\beta}\Big\}, , ,\mathrm{Z}, (,;)\ "(,;\ )()!\ (6\#\ (\mathrm{d}N_{i,j}(t) = 0$ ，并且 $\mathrm{d}\Lambda\%_0(t)$可以改写为 $\Sigma\quad\Sigma\mathrm{P}),_j\,\mathrm{Q}),_j(!)\mathrm{d}?),_j(,)\ -\ (_j)\ ).\#/\ _j.\#s"(,;\boldsymbol{\beta}/0$，因为事实上所有病例最初都已纳入分层病例队列（即纳入第二阶段样本），并且具有单位第二阶段抽样权重，因此对于层j 中满足$\int dN_{i,j}(t)\ = 1$的任何受试者i ，均有$\xi_{i,j}\ w_{i,j}^{(2)} = 1$ 。

H.2 影响函数的推导

我们令 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&$ 表示层 j 中受试者 i 对 $\boldsymbol{\theta}\%, i \in \{1, \dots, {}^{n}{}_{(j)}\}, j \in \{1, \dots, J\}, \boldsymbol{\theta}\% \in \{\boldsymbol{\beta}\%, \mathrm{d}\Lambda\%_0(t),\ \Lambda\%_0(t), \pi{:}(\tau_1, \tau_2; \boldsymbol{x})\}$ 的影响。由于推导与网络附录 C.1 和 D.2 中的推导类似，我们不再详述。我们有

$$\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\beta}}\right) = \ \xi_{i,j}\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\beta}}\right) + \ \xi_{i,j}\ V_{i,j}\ w_{i,j}\ \boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\beta}}\right),$$

使用 $\boldsymbol{IF}_{i,\ j}{}^{(2)}\#\boldsymbol{\beta}\%\& \ = 0,$

$$\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\beta}}\right) = \left[\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\int_t \xi_{k,l}\ V_{k,l}\ w_{k,l}\left\{\frac{\boldsymbol{S}_2(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})\,\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})'}{S_0(t;\widehat{\boldsymbol{\beta}})^2}\right\}\mathrm{d}N_{k,l}(t)\right]^{-1} \times \left[\int_t \left\{\boldsymbol{X}_{i,j} - \frac{\boldsymbol{S}_1(t;\widehat{\boldsymbol{\beta}})}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\left\{\mathrm{d}N_{i,j}(t) - Y_{i,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,j}\right)\frac{\sum_{l=1}^{J}\sum_{k=1}^{n^{(l)}}\xi_{k,l}\,V_{k,l}\,w_{k,l}\,\mathrm{d}N_{k,l}(t)}{S_0(t;\widehat{\boldsymbol{\beta}})}\right\}\right],$$

and $\boldsymbol{S}_2\left(t;\widehat{\boldsymbol{\beta}}\right) = \sum_{j=1}^{J}\sum_{k=1}^{n^{(j)}}\ \xi_{k,j}\ V_{k,j}\ w_{k,j}\ Y_{k,j}(t)\exp\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}_{k,j}\right)\boldsymbol{X}_{k,j}\ \boldsymbol{X}_{k,j}{}'.$

Then $\Delta_{i,j}\{\mathrm{d}\Lambda\%0(t)\} = \xi_{i,j} IF_{i,j}^{(0)}\{\mathrm{d}\Lambda\hat{}_0(t)\} + \xi_{i,j}\, V_{i,j}\, w_{i,j}\, IF_{i}$ $_{,j}^{(3)}2\mathrm{d}\Lambda\%0(t)\}$，其中 $IF_{ij}^{(2)}2\mathrm{d}\Lambda\%0(t)\} = \{S_0\#t;\hat{\beta}\&\}^{-1}\mathrm{d}N_{i,j}$ 和 $IF_{i,j}^{(3)}\{\mathrm{d}\Lambda\%0(t)\} = -\{S_0\#t;\ \hat{\beta}\&\}^{-1}\mathrm{Z}\mathrm{d}\Lambda\%0(t)\ \boldsymbol{S}_1\#t;\boldsymbol{\beta}\%\&' \boldsymbol{IF}_{i,j}^{(2,3)}\#\boldsymbol{\beta}\%\& + \mathrm{d}\Lambda\%0(t)\ Y_{i,j}(t)\exp\#\boldsymbol{\beta}\%'\boldsymbol{X}_{i,j}\&\}$。注意，$IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%0(t)\}$ 和 $IF_{i,j}^{(3)}\{\mathrm{d}\Lambda\%0(t)\}$ 是增量 $\mathrm{d}N_{i,j}(t)$ 和 $\mathrm{d}\Lambda\mathrm{l}_0(t)$ 的线性组合。因此 $\Delta_{i,j}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\mathrm{A}\#\} = \xi_{i,j} IF_{i,j}^{(2)}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\mathrm{A}\#\} + \xi_{i,j}\, V_{i,j}\, w_{i,j}\, IF_{i,j}^{(3)}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\mathrm{A}\#\},$，其中 $IF_{i,j}^{(s)}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\mathrm{A}\#\} = \int IF_{i,j}^{(s)}\{\mathrm{d}\Lambda\%0(s)\}\mathrm{A}!\mathrm{A}\#$，$s \in \{2,3\}$。最后，$\Delta_{i,j}\{\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\} = \xi_{i,j}\, IF_{i,j}^{(2)}\{\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\} + \xi_{i,j}\, V_{i,j}\, w_{i,j}\, IF_{i,j}^{(3)}\{\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\}$，其中 $IF_{i,j}^{(s)}\{\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\} = \mathrm{N}^{\partial\hat{\pi}(\tau_1,}\mathrm{A}!;x)\mathrm{B}\beta_{|\boldsymbol{\beta}=\boldsymbol{\beta}/}\ \mathrm{O}\ \boldsymbol{IF}_{i,j}^{(s)} + \Big[\mathrm{BCD}(\mathrm{A}\#,\mathrm{A}!;\ x)\mathrm{BG}\int \mathrm{d}\mathrm{I}''(,)^{\tau}!\ ^{\tau}\#\ \mathrm{J}_{|}\mathrm{d}\mathrm{I}''(,)=\mathrm{d}\mathrm{I}/''(,)_{\_}\ \Delta_{i,j}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\mathrm{A}\#\Big\}$，$s \in \{2,3\}$，并且对于 $\mathrm{BCD}(\mathrm{A}\#,\mathrm{A}!;x)\ \mathrm{B}\boldsymbol{\beta}_{|\boldsymbol{\beta}=\boldsymbol{\beta}/} = \{\sum_{=}^{\tau_2}\ \hat{\lambda}_0(t),\ \mathrm{A}\#\ \exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\}\{1-\pi{:}(\tau_1,\tau_2;\ \boldsymbol{x})\}\boldsymbol{x}'$，$\mathrm{BCV}(\mathrm{A}\#,\mathrm{A}!;\ x)\mathrm{B}\}''(\mathrm{A}\#)_{|}\}''(\mathrm{A}\#)_{=}\}\mathrm{W}''(\mathrm{A}\#) = \exp\#\boldsymbol{\beta}\mathrm{l}'\boldsymbol{x}\&\{1-\pi\mathrm{o}(\tau_1,\tau_2;\boldsymbol{x})\}$，以及 $\mathrm{BCD}(\mathrm{A}\#,\mathrm{A}!;\ x)\ \mathrm{BG}\int \mathrm{d}\mathrm{I}''(,)^{\tau}!^{\tau}\#\ \mathrm{J}_{|}\mathrm{d}\mathrm{I}''(,)=\mathrm{d}\mathrm{I}/''(,) = -\exp\#\boldsymbol{\beta}\%'\boldsymbol{x}\&\{1-\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\}$。

这里，如果受试者 i 在层 j 中未同时在第二阶段抽样和第三阶段抽样中被抽中，则 $\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\beta}\%\&$ 为零，$i \in \{1,\dots,{}^{n}(\#)\}, j \in \{1,\dots,J\}$；因此，病例可能对 $\boldsymbol{\beta}\%$ 没有影响。另一方面，如果受试者 i 在层 j 中未同时在第二阶段抽样和第三阶段抽样中被抽中，则 $\xi_{i,j}\, V_{i,j}\, w_{i,j}\, IF_{i,}$ $_{j}^{(3)}\{\mathrm{d}\Lambda\%0(t)\}$，$\xi_{i,j}\, V_{i,j}\, w_{i,j}\, IF_{i,j}^{(3)}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\ \mathrm{A}\#\}$ 和 $\xi_{i,j}\, V_{i,j}\, w_{i,j}\, IF_{i,j}^{(3)}\{\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\}$ 也为零，但 $\xi_{i,j}\, IF_{i,j}^{(2)}\{\mathrm{d}\Lambda\%0(t)\}$，$\xi_{i,j}\, IF_{i,j}^{(2)}\mathrm{Z}\int \mathrm{d}\Lambda\%0(t)^{\tau_2}\ \mathrm{A}\#\}$ 和 $\xi_{i,j}\, IF_{i,j}^{(2)}\{\pi{:}(\tau_1,\tau_2;\boldsymbol{x})\}$ 可能非零。事实上，由于我们选择不对 Breslow估计量的分子重新加权（即，我们选择使用实际观察到的病例的失效时间），

即使病例i 在层 j 中缺失协变量数据（即未在第三阶段抽样中被抽样），其对 dΛ‰(t)仍有非零影响，等于 $\{S_0\#t;\ \boldsymbol{\beta}\%\&\}^{-1}\mathrm{d}N_{i,j}(t)$，在其失败时间 t，随后可能对 $\Lambda\%_0(t)$ 和 π:($\tau_1, \tau_2; x)$ 产生非零影响。

H.3 基于影响函数的方差分解和估计 对于任意 $\boldsymbol{\theta}\% \in \{\boldsymbol{\beta}\%,\ \mathrm{d}\Lambda\%_0(t),\ \Lambda\%_0(t),\ \pi:(\tau_1, \tau_2;\ \boldsymbol{x})\}$，基于影响函数的方差为 var Z$\Sigma_{=}\ \Sigma_{=}\ \boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&^{n}(_j)$! (6 # (}。利用全协方差公式和全期望公式，可以将其分解为 var $\left(\mathrm{E}\ \mathrm{iE}\ \mathrm{Z}\Sigma_{=}\ \Sigma_{=}\ \boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&^{n}(_j)\ !\ (6\ \#\ (\ |C_1, C_2\}\ |C_1\mathrm{j}\right) +\mathrm{E}\ \left(\mathrm{var}\ \mathrm{iE}\ \mathrm{Z}\Sigma_{=}\ \Sigma_{=}\ \boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&^{n}(_j)\ !\ (6\ \#\ (|C_1, C_2\}\ |C_1\mathrm{j}\right) +\ \mathrm{E}\ \left(\mathrm{E}\ \mathrm{ivar}\ \mathrm{Z}\ \Sigma_{=}\ \Sigma_{=}\ \boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\&^{n}(_j)\ !\ (6\ \#\ (|C_1, C_2\}\ |C_1\mathrm{j}\right)$，其中 C_1 表示来自整个队列的信息，而 C_2 表示来自第二阶段样本的信息。

我们已在网络附录 H.2 中证明，$\boldsymbol{\Delta}_{i,j}\#\boldsymbol{\theta}\%\& = \xi_{i,j}\boldsymbol{IF}^{(2)}_{i,j}\#\boldsymbol{\theta}\%\&+\ \xi_{i,j}\ V_{i,j}\ w_{i,j}\ \boldsymbol{IF}^{(3)}_{i,j}\#\boldsymbol{\theta}\%\&$，其中 $\boldsymbol{IF}^{(2)}_{i,j}\#\boldsymbol{\beta}\%\& = 0$。利用与网络附录 G.2 中类似的论证，我们有

E$\{\Sigma_{=}\ \Sigma_{=}\ \boldsymbol{\Delta}_{i,j}(\boldsymbol{\theta}\frown)^{n}{}_{(j)}\ !\ (6\ \#\ (|C_1, C_2\} = \Sigma^{J}_{=}\ \Sigma\{^{n}_{=}\ \ \xi_{i\,j}\boldsymbol{IF}^{(2)}_{i\,j}\ \boldsymbol{\theta}\ \ +,,(\frown)\ {}_{(J)}\ !$
(# (
$\mathrm{E}\left(V_{i,j}\ w^{(3)}_{i,j}|C_1, C_2\right)\xi_{i,j}\ w^{(2)}_{i,j}\boldsymbol{IF}^{(3)}_{i,j}\#\boldsymbol{\theta}\%\&\},$

$=\Sigma_{=}\ \Sigma_{=}\ \xi_{i,}{}^{n\ j}\ {}_{j}\ \boldsymbol{IF}^{(2)}_{i\,j}\ \boldsymbol{\theta}\ \ + w^{(2)}_{i\,j}\boldsymbol{IF}^{(3)}_{i\,j}\ \boldsymbol{\theta}\ \}\{,(\frown),,($
$\frown)_{()}\ !\ (6\ \#\ (,$

并且我们有

$$\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})\,|C_1, C_2\right\} =$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\mathrm{var}(V_{i,j}|\,C_1, C_2)\left\{\xi_{i,j}\ w_{i,j}\boldsymbol{IF}^{(3)}_{i,j}(\widehat{\boldsymbol{\theta}})\right\}\left\{\xi_{i,j}\ w_{i,j}\boldsymbol{IF}^{(3)}_{i,j}(\widehat{\boldsymbol{\theta}})\right\}',$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma^{(3)}_{i,j}\ \xi_{i,j}\ w_{i,j}\ w_{i,j}\boldsymbol{IF}^{(3)}_{i,j}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}^{(3)}_{i,j}(\widehat{\boldsymbol{\theta}})',$$

正如 $\xi_{i,\ j}\boldsymbol{IF}_{i,\ j}^{(2)}\#\boldsymbol{\theta}\%\&$ 在给定 C_2 的条件下是固定的。

于是我们有

$$\mathrm{E}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\theta}}\right)|C_1,C_2\right\}|C_1\right]=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\mathrm{E}\left(\xi_{i,j}\,|C_1\right)\left\{\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)+w_{i,j}^{(2)}\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\right\},$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left\{\frac{1}{w_{i,j}^{(2)}}\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)+\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\right\},$$

and

$$\mathrm{var}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\theta}}\right)|C_1,C_2\right\}|C_1\right]=\sum_{j=1}^{J}\mathrm{var}\left[\sum_{i=1}^{n^{(j)}}\xi_{i,j}\left\{\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)+w_{i,j}^{(2)}\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\right\}\right],$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\left\{\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)+w_{i,j}^{(2)}\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\right\}\left\{\boldsymbol{IF}_{k,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)+w_{k,j}^{(2)}\boldsymbol{IF}_{k,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\right\}',$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,j}^{(2)}w_{k,j}^{(2)}\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{IF}_{k,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)',$$

事实上，层 j（ $j\in\{1,\ldots,J\}$）中的任意一对个体 i 和 k ，只要满足 $\sigma_{i,\ k\,j}^{(2)}\neq 0$ ，，就必然满足 $\boldsymbol{IF}_{i,\ j}^{(2)}\#\boldsymbol{\theta}\%\&=\boldsymbol{IF}_{k,\ j}^{(3)}\#\boldsymbol{\theta}\%\&=0$，并且

$$\mathrm{E}\left[\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\theta}}\right)|C_1,C_2\right\}|C_1\right]=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\mathrm{E}\left(\xi_{i,j}|C_1\right)\sigma_{i,j}^{(3)}\,w_{i,j}\,w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)',$$

$$=\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(3)}\,w_{i,j}^{(3)}w_{i,j}\,\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)'.$$

最后，

$$\mathrm{var}\left(\mathrm{E}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}\left(\widehat{\boldsymbol{\theta}}\right)|C_1,C_2\right\}|C_1\right]\right)=\mathrm{var}\left[\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\frac{1}{w_{i,j}^{(2)}}\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)+\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\right],$$

并且可以估计为

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left\{\xi_{i,j}\,\frac{1}{w_{i,j}^{(2)}}\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)'+2\,\xi_{i,j}\,V_{i,j}\,\frac{w_{i,j}}{w_{i,j}^{(2)}}\,\boldsymbol{IF}_{i,j}^{(2)}\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)'+\right.$$

$$\left.\xi_{i,j}\,V_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{IF}_{i,j}^{(3)}\left(\widehat{\boldsymbol{\theta}}\right)'\right\},$$

即通过

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left\{\frac{1}{w_{i,j}^{(2)}}\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})' + 2\frac{1}{w_{i,j}^{(2)}}\,\xi_{i,j}\,V_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})' + \right.$$

$$\left.\frac{1}{w_{i,j}}\xi_{i,j}\,V_{i,j}\,w_{i,j}\,w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})'\right\},$$

and

$$\mathrm{E}\left(\mathrm{var}\left[\mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right) =$$

$$\mathrm{E}\left[\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,j}^{(2)}w_{k,j}^{(2)}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(3)}(\widehat{\boldsymbol{\theta}})'\right],$$

可通过

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{k=1}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,k,j}\,\xi_{i,j}\,\xi_{k,j}\,V_{i,j}\,V_{k,j}\,w_{i,j}^{(2)}w_{k,j}^{(2)}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{k,j}^{(3)}(\widehat{\boldsymbol{\theta}})',$$

即通过

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(2)}\frac{w_{i,j}^{(2)}}{w_{i,j}^{(3)}}\xi_{i,j}\,V_{i,j}\,w_{i,j}w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})' +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\ k\neq i}}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,k,j}^{(2)}\,\xi_{i,j}\,V_{i,j}\,\xi_{k,j}\,V_{k,j}\,w_{i,j}w_{k,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\,\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})',$$

and

$$\mathrm{E}\left(\mathrm{E}\left[\mathrm{var}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\boldsymbol{\Delta}_{i,j}(\widehat{\boldsymbol{\theta}})\,|C_1,C_2\right\}|C_1\right]\right) = \mathrm{E}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(3)}\,w_{i,j}^{(3)}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})'\right\},$$

可通过以下方式估计的

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\xi_{i,j}\,V_{i,j}\,w_{i,j}\,\sigma_{i,j}^{(3)}\,w_{i,j}^{(3)}\,w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})',$$

即通过

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(3)}w_{i,j}^{(3)}\,\xi_{i,j}\,V_{i,j}\,w_{i,j}\,w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\,\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})'.$$

因此，方差#$\boldsymbol{\theta}$%& 可以通过以下方式估计

$$\frac{n}{n-1}\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\left\{\frac{1}{w_{i,j}^{(2)}}\xi_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})' + 2\frac{1}{w_{i,j}^{(2)}}\,\xi_{i,j}\,V_{i,j}\,w_{i,j}\,\boldsymbol{IF}_{i,j}^{(2)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})' + \right.$$

$$\left.\frac{1}{w_{i,j}}\xi_{i,j}\,V_{i,j}\,w_{i,j}\,w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})'\right\} +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(2)}\frac{w_{i,j}^{(2)}}{w_{i,j}^{(3)}}\xi_{i,j}\,V_{i,j}\,w_{i,j}w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})' +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sum_{\substack{k=1,\\ k\neq i}}^{n^{(j)}}\sigma_{i,k,j}^{(2)}\,w_{i,k,j}^{(2)}\,\xi_{i,j}\,V_{i,j}\,\xi_{k,j}\,V_{k,j}\,w_{i,j}w_{k,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\,\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})' +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n^{(j)}}\sigma_{i,j}^{(3)}w_{i,j}^{(3)}\,\xi_{i,j}\,V_{i,j}\,w_{i,j}\,w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})\,\boldsymbol{IF}_{i,j}^{(3)}(\widehat{\boldsymbol{\theta}})',$$

我们可以将其改写为

$$\frac{}{n^{J}_{=}}\frac{4\,47(\sum_{=}\sum \mathrm{N}\,(\,\mathrm{Q})_{,j}(!)}{}\xi_{i\,j}\,\boldsymbol{IF}_{i\,j}^{(2)}\;\boldsymbol{\theta}\;\boldsymbol{IF}_{i\,j}^{(2)}\;\boldsymbol{\theta}\;' + 2\frac{}{}\,\xi_{i\,j}\,V_{i\,j}\,w_{i\,j}\,\boldsymbol{IF}_{i\,j}^{(2)}\;\boldsymbol{\theta}\;\boldsymbol{IF}_{i\,j}^{(3)}\;\boldsymbol{\theta}\;' +, , \#\%\&\,, \#\%\&\,(\,\mathrm{Q})_{,j}$$

$$(!)\,,\,,\,,,\,\#\%\&,\,\#\%\&\,(\,)!\;(6\#\;((\,\mathrm{Q})_{,j}\,\xi_{i}{}_{,j}\,V_{i,j}\,w_{i,j}\,w_{i,j}\boldsymbol{IF}_{i,j}^{(3)}\#\boldsymbol{\theta}\%\&\boldsymbol{IF}_{i,j}^{(3)}\#\boldsymbol{\theta}\%\&\;'\mathrm{O} +$$

$$\frac{}{}\sum_{=}\sum_{=}\sum_{k=}\sigma_{i,k\,j}\,w_{i,\,,k\,j}\,\xi_{i,\,,j}\,V_{i,j}\,\xi_{k,j}\,V_{k,j}\,w_{i,j}w_{k,j}\,\boldsymbol{IF}_{i,j}^{(3)}\#\boldsymbol{\theta}\%\&\,{}^{n}(_{j})\;(4\;(_{j})\;!\;(6\#\;(\,\boldsymbol{IF}_{i,j}^{(3)}\#\boldsymbol{\theta}\%\&\;'\text{, 确实 }\sigma_{i,j}^{(2)}\;{}_{,j}\,\mathrm{Q})_{,j}(6)\;+\sigma_{i,j}^{(3)}w_{i,j}^{(3)}=1-(\mathrm{Q})_{,j}=\sigma_{i}{}_{,j}\,w_{i,j},\text{ 和 }\sigma_{i,k\,j}^{(2)}\,w_{i,,k\,j}^{(2)}=\sigma_{i,\,,k\,j}\times,(\,\mathrm{Q})_{,j}(6)\times(\mathrm{Q},_{,j}(6)\;\times w_{i,k\,j}^{(2)}\times w_{i,\,,j}^{(3)}\times w_{k,j}^{(3)}=\sigma_{i,k\,j}w_{i,\,,k\,j,}\,,\;i,k\in\{1,\ldots,{}^{n}(\#)\},\,i\neq k,j\in\{1,\ldots,$$

$$J\}.\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad(4)$$

注意，当我们假设已知 $w_i^{(3)}$，且由于第三阶段抽样采用伯努利抽样，因此可以将第二阶段和第三阶段合并为一个抽样阶段。上文公式（4）给出的方差估计仅由两个部分构成，对应于将第二阶段和第三阶段抽样合并为一个抽样阶段后直接得到的形式。

网络附录 I. 缺失数据模拟

I.1 模拟设计

我们按照主文档第 7 节的方法，模拟了具有 $n \in \{5 \times 10^3, 10^4\}$ 的队列。我们在由 $W = 0 \times I(X_1 \geq 0, X_2 = 0) + 1 \times I(X_1 < 0, X_2 < 2) + 2 \times I(X_1 \geq 0, X_2 > 0) + 3 \times I(X_1 < 0, X_2 = 2)$ 定义的 $J = 4$ 层内独立抽样，抽取固定数量的个体：$m^{(j)} = \lfloor$ λ"×('×E{9:;(β#o#+β!o!+β₃o6)|W=#} (− λ"×('×E{9:;(β#o#+β!o!+β₃o6)|W=#} × E#$n^{(j)}$& × K + ( ) $\rfloor$，其中 $K \in \{2, 4\}$，$j \in \{0,1,2,3\}$，且 $\lfloor \rfloor$ 为向下取整函数。这些个体连同其余病例构成了第二阶段样本。同时，我们从按病例状态定义的 $J^{(3)} = 2$层中抽取个体，第三阶段抽样概率为 $\boldsymbol{\pi}$(S)∈{(0.9, 0.8), (0.98, 0.9)}。换言之，我们假设病例的协变量信息缺失概率始终较高，例如，储存的血液样本可能已被此前使用。最后，我们将同时属于两个样本的个体（即协变量数据完整的个体）视为第三阶段样本。

第二阶段抽样设计权重按照主文档第3.3节中的方法计算，第三阶段抽样设计权重则取第三阶段抽样概率的倒数。随后，通过将二者相乘得到总体抽样设计权重。然而，在实际应用中，第三阶段抽样设计权重 $\frac{\mathbf{1}}{\boldsymbol{\pi}^{3}}$0 通常是未知的。因此，我们按照主文档第7.2节中的方法估计第三阶段抽样设计权重；随后，将估计得到的第三阶段抽样设计权重与已知的第二阶段抽样设计权重相乘，得到总体估计抽样设计权重。

对于每种情景，我们模拟了 5,000 个队列。我们估计了时间区间（τ_1, τ_2] = (0,8] 内的对数相对风险 $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ 和纯风险 $\pi(\tau_1, \tau_2; \boldsymbol{x})$，以及协变量特征组合

$\boldsymbol{x} \in \{(-1, 1, -0.6)', (1, -1, 0.6)', (1, 1, 0.6)'\}$，采用以下方法：使用估计设计权重的分层病例队列（SCC.Est）；使用真实设计权重的分层病例队列（SCC.True）；使用估计设计权重的未分层病例-队列（USCC.Est）；以及使用真实设计权重的未分层病例-队列（USCC.True）。对于每个模拟实现，我们随后估计其方差。对于 SCC.Est，我们使用：公式（22）中的超总体、第二阶段和第三阶段方差分量（$V\text{fi}$）方差估计，该公式见主文档第 5.5 节或网络附录 G.2；以及稳健方差估计（$V\text{fi}_{\text{Robust}}$），其计算方法与主文档第 3.3 节或第 4.3 节相同，但影响函数采用主文档第 5.4 节或网络附录 G.1 中给出的形式。对于 SCC.True，我们使用：网络附录 H.3 中公式（4）的超总体、第二阶段和第三阶段方差分量（$V\text{fi}$）方差估计；以及按主文档第 3.3 节或第 4.3 节的方法计算的 $V\text{fi}_{\text{Robust}}$，但影响函数采用网络附录 H.2 中给出的形式。对于 USCC.Est 和 USCC.True，我们使用类似的方差估计，其中 $J = 1$。另一种做法是将估计得到的设计权重视为已知真实权重，并按照网络附录 H.3 估计方差（SCC.Naive 和 USCC.Naive）。作为参照，我们还使用整个队列（Cohort）的数据估计了这些参数。

I.2 模拟结果

网络表29至**网络表40**展示了95%置信区间的覆盖率，而**网络表41**至**网络表52** 展示了估计方差的均值和经验方差，分别对应β_1, β_2、β_3 和 $\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$、$(\tau_1, \tau_2] = (0,8]$ 和 $\boldsymbol{x} \in \{(-1, 1, -0.6)', (1, -1, 0.6)', (1, 1, 0.6)'\}$。为简洁起见，我们省略了$n = 10{,}000$和$p_Y = 0.1$的结果。同样，在大多数情景下，稳健方差高估了分层设计下对数相对风险和纯风险的方差，并导致置信区间覆盖率超过名义水平；对于纯

在大多数情景下，非分层设计的风险估计亦如此。另一方面，适当考虑抽样特征的方差估计在大多数情景下均获得了合理的覆盖率。使用估计的设计权重与使用真实设计权重的表现相近。具体而言，将估计权重视为已知的真实权重，并使用网络附录 H.3 给出的较简单方差公式，其表现通常与使用主文档中公式（22）给出的、更复杂的基于影响的方差估计相近。因此，或许无需考虑第三阶段权重估计所带来的变异性。在本次模拟中，与 SCC.True（分别为 USCC.True）相比，在 SCC.Est 和 SCC.Naive（分别为 USCC.Est 和 USCC.Naive）中使用估计权重几乎没有带来效率提升。

队列	SCC.True		SCC.Est		SCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_1
	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$					
0.9504	0.969*	0.9574*	0.969*	0.9574*	0.969*	0.9574*	5000	2	0.02	(0.9,0.8)	-0.2
0.9504	0.9694*	0.9554	0.9694*	0.9554	0.9694*	0.9554	5000	2	0.02	(0.98,0.9)	-0.2
0.943*	0.9534	0.9474	0.9536	0.9474	0.9536	0.9474	5000	4	0.02	(0.9,0.8)	-0.2
0.943*	0.9568*	0.948	0.957*	0.948	0.957*	0.948	5000	4	0.02	(0.98,0.9)	-0.2
0.9478	0.9652*	0.953	0.9652*	0.953	0.9652*	0.953	10000	2	0.02	(0.9,0.8)	-0.2
0.9478	0.9664*	0.9534	0.9662*	0.9534	0.9662*	0.9534	10000	2	0.02	(0.98,0.9)	-0.2
0.9482	0.9542	0.9462	0.9542	0.9462	0.9542	0.9462	10000	4	0.02	(0.9,0.8)	-0.2
0.9482	0.9572*	0.9474	0.9572*	0.9474	0.9572*	0.9474	10000	4	0.02	(0.98,0.9)	-0.2
0.9486	0.9638*	0.9548	0.9638*	0.9546	0.9638*	0.9546	5000	2	0.05	(0.9,0.8)	-0.2
0.9486	0.96*	0.9472	0.96*	0.947	0.96*	0.947	5000	2	0.05	(0.98,0.9)	-0.2
0.9462	0.9506	0.944	0.9506	0.9442	0.9506	0.9442	5000	4	0.05	(0.9,0.8)	-0.2
0.9462	0.955	0.9474	0.9548	0.9476	0.9548	0.9476	5000	4	0.05	(0.98,0.9)	-0.2
0.9486	0.9654*	0.9536	0.9654*	0.9536	0.9654*	0.9536	10000	2	0.05	(0.9,0.8)	-0.2
0.9486	0.966*	0.9498	0.966*	0.95	0.966*	0.95	10000	2	0.05	(0.98,0.9)	-0.2
0.948	0.9554	0.9496	0.9554	0.9496	0.9554	0.9496	10000	4	0.05	(0.9,0.8)	-0.2
0.948	0.9594*	0.9514	0.9596*	0.9514	0.9596*	0.9514	10000	4	0.05	(0.98,0.9)	-0.2
0.949	0.9574*	0.947	0.9574*	0.9474	0.9574*	0.9474	5000	2	0.1	(0.9,0.8)	-0.2
0.949	0.9592*	0.9484	0.9594*	0.9484	0.9594*	0.9484	5000	2	0.1	(0.98,0.9)	-0.2
0.9494	0.952	0.9486	0.9518	0.9484	0.9518	0.9484	5000	4	0.1	(0.9,0.8)	-0.2
0.9494	0.954	0.9494	0.9536	0.9494	0.9536	0.9494	5000	4	0.1	(0.98,0.9)	-0.2

网络表29-在不同协变量数据缺失概率下，采用不同分析方法和方差估计方法，对5,000个模拟队列中通过分层抽样获得的对数相对风险参数β_1的95%置信区间覆盖率。*表示覆盖率超出预期区间[0.9440；0.9560]。

队列	USCC.True		USCC.Est		USCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_1
	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$					
0.9504	0.9576*	0.9576*	0.9576*	0.9578*	0.9576*	0.9578*	5000	2	0.02	(0.9,0.8)	-0.2
0.9504	0.9564*	0.9566*	0.956	0.9566*	0.956	0.9566*	5000	2	0.02	(0.98,0.9)	-0.2
0.943*	0.9514	0.9514	0.9514	0.9514	0.9514	0.9514	5000	4	0.02	(0.9,0.8)	-0.2
0.943*	0.9486	0.9486	0.9486	0.9486	0.9486	0.9486	5000	4	0.02	(0.98,0.9)	-0.2
0.9478	0.9542	0.9542	0.9542	0.9542	0.9542	0.9542	10000	2	0.02	(0.9,0.8)	-0.2
0.9478	0.9512	0.9512	0.9512	0.9512	0.9512	0.9512	10000	2	0.02	(0.98,0.9)	-0.2
0.9482	0.9458	0.9458	0.9458	0.946	0.9458	0.946	10000	4	0.02	(0.9,0.8)	-0.2
0.9482	0.948	0.948	0.948	0.948	0.948	0.948	10000	4	0.02	(0.98,0.9)	-0.2
0.9486	0.9504	0.9504	0.9506	0.9508	0.9506	0.9508	5000	2	0.05	(0.9,0.8)	-0.2
0.9486	0.9506	0.9506	0.9506	0.9506	0.9506	0.9506	5000	2	0.05	(0.98,0.9)	-0.2
0.9462	0.9492	0.9492	0.9492	0.9492	0.9492	0.9492	5000	4	0.05	(0.9,0.8)	-0.2
0.9462	0.9472	0.9472	0.9472	0.9472	0.9472	0.9472	5000	4	0.05	(0.98,0.9)	-0.2
0.9486	0.9474	0.9476	0.9476	0.9476	0.9476	0.9476	10000	2	0.05	(0.9,0.8)	-0.2
0.9486	0.9502	0.9504	0.9502	0.9504	0.9502	0.9504	10000	2	0.05	(0.98,0.9)	-0.2
0.948	0.9514	0.9514	0.9512	0.9512	0.9512	0.9512	10000	4	0.05	(0.9,0.8)	-0.2
0.948	0.95	0.95	0.95	0.95	0.95	0.95	10000	4	0.05	(0.98,0.9)	-0.2
0.949	0.9506	0.9506	0.9508	0.9508	0.9508	0.9508	5000	2	0.1	(0.9,0.8)	-0.2
0.949	0.95	0.95	0.9498	0.9498	0.9498	0.9498	5000	2	0.1	(0.98,0.9)	-0.2
0.9494	0.9482	0.9482	0.9484	0.9484	0.9484	0.9484	5000	4	0.1	(0.9,0.8)	-0.2
0.9494	0.9496	0.9496	0.9496	0.9496	0.9496	0.9496	5000	4	0.1	(0.98,0.9)	-0.2

网络表格 30-在 5,000个模拟队列中，针对不同的协变量数据缺失概率，采用不同分析方法和方差估计方法，对非分层抽样所得对数相对风险参数 β_1 的 95%置信区间覆盖率。* 表示覆盖率不在预期区间 [0.9440; 0.9560] 内。

队列	SCC.True		SCC.Est		SCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_2
	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$					
0.9464	0.9632*	0.9486	0.9634*	0.9488	0.9634*	0.9488	5000	2	0.02	(0.9,0.8)	0.25
0.9464	0.9662*	0.947	0.9662*	0.9468	0.9662*	0.947	5000	2	0.02	(0.98,0.9)	0.25
0.9508	0.9614*	0.9524	0.9614*	0.9524	0.9614*	0.9524	5000	4	0.02	(0.9,0.8)	0.25
0.9508	0.9614*	0.95	0.9614*	0.95	0.9614*	0.95	5000	4	0.02	(0.98,0.9)	0.25
0.9474	0.9676*	0.9502	0.9676*	0.9502	0.9676*	0.9502	10000	2	0.02	(0.9,0.8)	0.25
0.9474	0.9682*	0.9512	0.9682*	0.9512	0.9682*	0.9512	10000	2	0.02	(0.98,0.9)	0.25
0.9488	0.9612*	0.949	0.961*	0.949	0.961*	0.949	10000	4	0.02	(0.9,0.8)	0.25
0.9488	0.9634*	0.951	0.9636*	0.9512	0.9636*	0.9512	10000	4	0.02	(0.98,0.9)	0.25
0.9508	0.9658*	0.9542	0.9658*	0.9546	0.9658*	0.9546	5000	2	0.05	(0.9,0.8)	0.25
0.9508	0.9714*	0.9574*	0.9712*	0.9574*	0.9712*	0.9574*	5000	2	0.05	(0.98,0.9)	0.25
0.9438*	0.9546	0.9466	0.9546	0.9464	0.9546	0.9464	5000	4	0.05	(0.9,0.8)	0.25
0.9438*	0.9556	0.9474	0.9556	0.9478	0.9556	0.9478	5000	4	0.05	(0.98,0.9)	0.25
0.95	0.9594*	0.9464	0.9594*	0.9464	0.9594*	0.9464	10000	2	0.05	(0.9,0.8)	0.25
0.95	0.9652*	0.9506	0.965*	0.9506	0.965*	0.9506	10000	2	0.05	(0.98,0.9)	0.25
0.9458	0.959*	0.9506	0.9588*	0.9504	0.9588*	0.9504	10000	4	0.05	(0.9,0.8)	0.25
0.9458	0.9584*	0.9498	0.9582*	0.9498	0.9582*	0.9498	10000	4	0.05	(0.98,0.9)	0.25
0.9492	0.963*	0.9492	0.9628*	0.9488	0.9628*	0.9488	5000	2	0.1	(0.9,0.8)	0.25
0.9492	0.9664*	0.9514	0.9664*	0.9514	0.9664*	0.9514	5000	2	0.1	(0.98,0.9)	0.25
0.954	0.954	0.9494	0.9542	0.9496	0.9542	0.9496	5000	4	0.1	(0.9,0.8)	0.25
0.954	0.9552	0.9496	0.9554	0.9494	0.9556	0.9494	5000	4	0.1	(0.98,0.9)	0.25

网页表31- 在不同协变量数据缺失概率下，对于5,000个模拟队列，采用不同分析方法和方差估计方法进行分层抽样时，对数相对风险参数 β_2 的95%置信区间覆盖率。*表示覆盖率超出预期区间 [0.9440；0.9560]。

队列	USCC.True		USCC.Est		USCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_2
	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$					
0.9464	0.9484	0.9486	0.9484	0.9486	0.9484	0.9486	5000	2	0.02	(0.9,0.8)	0.25
0.9464	0.95	0.9506	0.95	0.9506	0.95	0.9506	5000	2	0.02	(0.98,0.9)	0.25
0.9508	0.9518	0.9518	0.9518	0.9518	0.9518	0.9518	5000	4	0.02	(0.9,0.8)	0.25
0.9508	0.9534	0.9534	0.9534	0.9534	0.9534	0.9534	5000	4	0.02	(0.98,0.9)	0.25
0.9474	0.9512	0.9512	0.9512	0.9512	0.9512	0.9512	10000	2	0.02	(0.9,0.8)	0.25
0.9474	0.9486	0.9486	0.9486	0.9488	0.9486	0.9488	10000	2	0.02	(0.98,0.9)	0.25
0.9488	0.951	0.951	0.9508	0.951	0.9508	0.951	10000	4	0.02	(0.9,0.8)	0.25
0.9488	0.9532	0.9532	0.953	0.9532	0.953	0.9532	10000	4	0.02	(0.98,0.9)	0.25
0.9508	0.9514	0.9516	0.9518	0.952	0.9518	0.952	5000	2	0.05	(0.9,0.8)	0.25
0.9508	0.9538	0.954	0.9538	0.9538	0.9538	0.9538	5000	2	0.05	(0.98,0.9)	0.25
0.9438*	0.9472	0.9472	0.947	0.947	0.947	0.947	5000	4	0.05	(0.9,0.8)	0.25
0.9438*	0.945	0.9452	0.9454	0.9454	0.9454	0.9454	5000	4	0.05	(0.98,0.9)	0.25
0.95	0.95	0.95	0.95	0.95	0.95	0.95	10000	2	0.05	(0.9,0.8)	0.25
0.95	0.9482	0.9482	0.9482	0.9482	0.9482	0.9482	10000	2	0.05	(0.98,0.9)	0.25
0.9458	0.954	0.954	0.9538	0.9538	0.9538	0.9538	10000	4	0.05	(0.9,0.8)	0.25
0.9458	0.9522	0.9522	0.952	0.952	0.952	0.952	10000	4	0.05	(0.98,0.9)	0.25
0.9492	0.9488	0.949	0.9488	0.9488	0.9488	0.9488	5000	2	0.1	(0.9,0.8)	0.25
0.9492	0.9504	0.9504	0.9504	0.9504	0.9504	0.9504	5000	2	0.1	(0.98,0.9)	0.25
0.954	0.9494	0.9494	0.9492	0.9492	0.9492	0.9492	5000	4	0.1	(0.9,0.8)	0.25
0.954	0.95	0.95	0.9498	0.9498	0.9498	0.9498	5000	4	0.1	(0.98,0.9)	0.25

网络表32-95%置信区间覆盖率：对于来自非分层抽样的对数相对风险参数β_2，采用不同的分析方法和方差估计方法，在不同协变量数据缺失概率下，基于5,000个模拟队列。*表示覆盖率超出预期区间 [0。9440；0.9560]。

队列	SCC.True		SCC.Est		SCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_3
	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$					
0.9456	0.9524	0.9526	0.9524	0.9526	0.9524	0.9526	5000	2	0.02	(0.9,0.8)	-0.3
0.9456	0.9538	0.9548	0.9534	0.9546	0.9534	0.9546	5000	2	0.02	(0.98,0.9)	-0.3
0.9412*	0.9492	0.9494	0.9492	0.9496	0.9492	0.9496	5000	4	0.02	(0.9,0.8)	-0.3
0.9412*	0.9462	0.9462	0.9462	0.9462	0.9462	0.9462	5000	4	0.02	(0.98,0.9)	-0.3
0.9478	0.9484	0.949	0.9486	0.9492	0.9486	0.9492	10000	2	0.02	(0.9,0.8)	-0.3
0.9478	0.9522	0.9524	0.9522	0.9522	0.9522	0.9522	10000	2	0.02	(0.98,0.9)	-0.3
0.9448	0.9514	0.9514	0.951	0.951	0.951	0.951	10000	4	0.02	(0.9,0.8)	-0.3
0.9448	0.949	0.949	0.949	0.949	0.949	0.949	10000	4	0.02	(0.98,0.9)	-0.3
0.945	0.949	0.949	0.9488	0.949	0.9488	0.949	5000	2	0.05	(0.9,0.8)	-0.3
0.945	0.944	0.9442	0.944	0.9442	0.944	0.9442	5000	2	0.05	(0.98,0.9)	-0.3
0.952	0.9556	0.9556	0.956	0.956	0.956	0.956	5000	4	0.05	(0.9,0.8)	-0.3
0.952	0.9558	0.9558	0.9558	0.9558	0.9558	0.9558	5000	4	0.05	(0.98,0.9)	-0.3
0.9462	0.952	0.9518	0.9524	0.9524	0.9524	0.9524	10000	2	0.05	(0.9,0.8)	-0.3
0.9462	0.9488	0.949	0.9488	0.9488	0.9488	0.9488	10000	2	0.05	(0.98,0.9)	-0.3
0.9532	0.9514	0.9514	0.9514	0.9514	0.9514	0.9514	10000	4	0.05	(0.9,0.8)	-0.3
0.9532	0.949	0.949	0.9486	0.9486	0.9486	0.9486	10000	4	0.05	(0.98,0.9)	-0.3
0.9512	0.9494	0.9496	0.9494	0.9496	0.9496	0.9496	5000	2	0.1	(0.9,0.8)	-0.3
0.9512	0.9476	0.9476	0.9476	0.9476	0.9476	0.9476	5000	2	0.1	(0.98,0.9)	-0.3
0.9518	0.9506	0.9506	0.9502	0.9502	0.9502	0.9502	5000	4	0.1	(0.9,0.8)	-0.3
0.9518	0.9512	0.9512	0.9514	0.9514	0.9514	0.9514	5000	4	0.1	(0.98,0.9)	-0.3

网络表33-采用不同分析方法和方差估计方法，对各种协变量数据缺失概率下分层抽样所得的对数相对风险参数 β_3 来自5,000个模拟队列的95%置信区间覆盖率。*表示覆盖率超出预期区间 [0.9440；0.9560]。

队列	USCC.True		USCC.Est		USCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_3
	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$					
0.9456	0.951	0.9512	0.951	0.9512	0.951	0.9512	5000	2	0.02	(0.9,0.8)	-0.3
0.9456	0.9522	0.9522	0.9522	0.9522	0.9522	0.9522	5000	2	0.02	(0.98,0.9)	-0.3
0.9412	0.9456	0.9458	0.9456	0.9456	0.9456	0.9456	5000	4	0.02	(0.9,0.8)	-0.3
0.9412	0.9468	0.947	0.9468	0.947	0.9468	0.947	5000	4	0.02	(0.98,0.9)	-0.3
0.9478	0.9516	0.9516	0.9514	0.9518	0.9514	0.9518	10000	2	0.02	(0.9,0.8)	-0.3
0.9478	0.9496	0.9496	0.9496	0.9496	0.9496	0.9496	10000	2	0.02	(0.98,0.9)	-0.3
0.9448	0.9488	0.9488	0.9488	0.9488	0.9488	0.9488	10000	4	0.02	(0.9,0.8)	-0.3
0.9448	0.947	0.947	0.947	0.947	0.947	0.947	10000	4	0.02	(0.98,0.9)	-0.3
0.945	0.947	0.947	0.9468	0.9468	0.9468	0.9468	5000	2	0.05	(0.9,0.8)	-0.3
0.945	0.9508	0.9508	0.9508	0.9508	0.9508	0.9508	5000	2	0.05	(0.98,0.9)	-0.3
0.952	0.949	0.949	0.9488	0.9488	0.9488	0.9488	5000	4	0.05	(0.9,0.8)	-0.3
0.952	0.954	0.954	0.954	0.954	0.954	0.954	5000	4	0.05	(0.98,0.9)	-0.3
0.9462	0.9482	0.9482	0.9482	0.9482	0.9482	0.9482	10000	2	0.05	(0.9,0.8)	-0.3
0.9462	0.9474	0.9476	0.9472	0.9472	0.9472	0.9472	10000	2	0.05	(0.98,0.9)	-0.3
0.9532	0.9546	0.9546	0.9544	0.9544	0.9544	0.9544	10000	4	0.05	(0.9,0.8)	-0.3
0.9532	0.9574*	0.9574*	0.9574*	0.9574*	0.9574*	0.9574*	10000	4	0.05	(0.98,0.9)	-0.3
0.9512	0.9558	0.9558	0.956	0.956	0.956	0.956	5000	2	0.1	(0.9,0.8)	-0.3
0.9512	0.9554	0.9554	0.9556	0.9556	0.9556	0.9556	5000	2	0.1	(0.98,0.9)	-0.3
0.9518	0.95	0.95	0.9496	0.9496	0.9496	0.9496	5000	4	0.1	(0.9,0.8)	-0.3
0.9518	0.9498	0.9498	0.9496	0.9496	0.9496	0.9496	5000	4	0.1	(0.98,0.9)	-0.3

网页表34-在不同协变量数据缺失概率下，使用不同分析方法和方差估计方法对非分层抽样所得对数相对风险参数 β_3 的95%置信区间覆盖率；共涉及5,000个模拟队列。*表示覆盖率超出预期区间 [0.9440；0.9560].

队列	SCC.True		SCC.Est		SCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$
	$\hat{V}$ #$%&'(	$\hat{V}$	$\hat{V}$ #$%&'(	$\hat{V}$	$\hat{V}$ #$%&'(	$\hat{V}$					
0.949	0.9724*	0.9548	0.9716*	0.9554	0.973*	0.9566	5000	2	0.02	(0.9,0.8)	-3.948
0.949	0.9714*	0.9524	0.9706*	0.9522	0.9708*	0.9528	5000	2	0.02	(0.98,0.9)	-3.948
0.9494	0.9594*	0.9486	0.96*	0.9498	0.9608*	0.9512	5000	4	0.02	(0.9,0.8)	-3.948
0.9494	0.9598*	0.9492	0.9596*	0.949	0.9596*	0.949	5000	4	0.02	(0.98,0.9)	-3.948
0.949	0.9688*	0.9508	0.9692*	0.9508	0.9698*	0.953	10000	2	0.02	(0.9,0.8)	-3.948
0.949	0.969*	0.9498	0.9684*	0.9498	0.9688*	0.9504	10000	2	0.02	(0.98,0.9)	-3.948
0.9476	0.9572*	0.9446	0.9564*	0.9444	0.9582*	0.946	10000	4	0.02	(0.9,0.8)	-3.948
0.9476	0.9606*	0.9474	0.962*	0.9476	0.9622*	0.9482	10000	4	0.02	(0.98,0.9)	-3.948
0.9532	0.9688*	0.9528	0.9696*	0.9552	0.9716*	0.9564*	5000	2	0.05	(0.9,0.8)	-3.046
0.9532	0.9712*	0.955	0.9708*	0.9556	0.9712*	0.9558	5000	2	0.05	(0.98,0.9)	-3.046
0.9438*	0.9586*	0.95	0.9578*	0.9498	0.959*	0.9508	5000	4	0.05	(0.9,0.8)	-3.046
0.9438*	0.9566*	0.9472	0.9554	0.947	0.9558	0.947	5000	4	0.05	(0.98,0.9)	-3.046
0.9496	0.9702*	0.9544	0.9712*	0.954	0.9722*	0.955	10000	2	0.05	(0.9,0.8)	-3.046
0.9496	0.9708*	0.954	0.9704*	0.955	0.9706*	0.955	10000	2	0.05	(0.98,0.9)	-3.046
0.9474	0.9578*	0.9486	0.958*	0.9488	0.9592*	0.95	10000	4	0.05	(0.9,0.8)	-3.046
0.9474	0.9562*	0.9464	0.9566*	0.947	0.957*	0.948	10000	4	0.05	(0.98,0.9)	-3.046
0.9482	0.9622*	0.9484	0.9634*	0.949	0.9642*	0.9524	5000	2	0.1	(0.9,0.8)	-2.377
0.9482	0.963*	0.9494	0.964*	0.9484	0.9642*	0.9492	5000	2	0.1	(0.98,0.9)	-2.377
0.9488	0.9522	0.9478	0.9522	0.9472	0.9536	0.9484	5000	4	0.1	(0.9,0.8)	-2.377
0.9488	0.9526	0.9492	0.9542	0.9512	0.9542	0.9514	5000	4	0.1	(0.98,0.9)	-2.377

网页表 35- 纯风险参数 $\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$ 其中 $\boldsymbol{x} = (-1, 1, -0.6)'$，在不同分析方法和方差估计方法下，通过分层抽样获得的数据中，针对不同协变量数据缺失概率，在5,000个模拟队列中的95%置信区间覆盖率。*表示覆盖率超出预期区间 [0.9440；0.9560]。

队列	USCC.True		USCC.Est		USCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.949	0.9638*	0.952	0.966*	0.9532	0.9668*	0.9546	5000	2	0.02	(0.9,0.8)	-3.948
0.949	0.9658*	0.951	0.9652*	0.953	0.9656*	0.9536	5000	2	0.02	(0.98,0.9)	-3.948
0.9494	0.959*	0.951	0.9584*	0.9522	0.9596*	0.9534	5000	4	0.02	(0.9,0.8)	-3.948
0.9494	0.959*	0.9516	0.9592*	0.952	0.9592*	0.9522	5000	4	0.02	(0.98,0.9)	-3.948
0.949	0.9624*	0.9518	0.962*	0.951	0.9634*	0.9532	10000	2	0.02	(0.9,0.8)	-3.948
0.949	0.9628*	0.9506	0.9628*	0.9524	0.9632*	0.9528	10000	2	0.02	(0.98,0.9)	-3.948
0.9476	0.9556	0.9486	0.9548	0.9492	0.9556	0.9504	10000	4	0.02	(0.9,0.8)	-3.948
0.9476	0.956	0.9474	0.9554	0.9476	0.9556	0.9478	10000	4	0.02	(0.98,0.9)	-3.948
0.9532	0.9668*	0.9576*	0.9672*	0.9568*	0.9688*	0.9586*	5000	2	0.05	(0.9,0.8)	-3.046
0.9532	0.9688*	0.9566*	0.9682*	0.957*	0.9682*	0.9574*	5000	2	0.05	(0.98,0.9)	-3.046
0.9438*	0.947	0.9384*	0.946	0.9394*	0.9466	0.9404*	5000	4	0.05	(0.9,0.8)	-3.046
0.9438*	0.947	0.9402*	0.9462	0.9392*	0.9462	0.9394*	5000	4	0.05	(0.98,0.9)	-3.046
0.9496	0.9634*	0.9526	0.9628*	0.9506	0.9638*	0.9522	10000	2	0.05	(0.9,0.8)	-3.046
0.9496	0.9636*	0.9522	0.9644*	0.9524	0.9644*	0.9526	10000	2	0.05	(0.98,0.9)	-3.046
0.9474	0.9568*	0.9508	0.958*	0.9504	0.9584*	0.9514	10000	4	0.05	(0.9,0.8)	-3.046
0.9474	0.956	0.952	0.9558	0.9516	0.9564*	0.952	10000	4	0.05	(0.98,0.9)	-3.046
0.9482	0.9572*	0.9454	0.9584*	0.9494	0.9598*	0.9506	5000	2	0.1	(0.9,0.8)	-2.377
0.9482	0.9582*	0.9446	0.9588*	0.9448	0.9588*	0.9452	5000	2	0.1	(0.98,0.9)	-2.377
0.9488	0.9484	0.9426	0.9498	0.9446	0.9498	0.9452	5000	4	0.1	(0.9,0.8)	-2.377
0.9488	0.9528	0.9492	0.9532	0.9492	0.9532	0.9494	5000	4	0.1	(0.98,0.9)	-2.377

WEB TABLE 36- 纯风险参数的95%置信区间覆盖率 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 其中 $\boldsymbol{x}=(-1,1,-0.6)'$，采用不同分析方法和方差估计方法进行非分层抽样，在不同协变量数据缺失概率下，对5,000个模拟队列的结果。*表示覆盖率超出预期区间 [0.9440; 0.9560].

队列	SCC.True		SCC.Est		SCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}$	$\hat{V}$	$\hat{V}$	$\hat{V}$	$\hat{V}$	$\hat{V}$					
0.9484	0.967*	0.9488	0.9664*	0.9488	0.9664*	0.949	5000	2	0.02	(0.9,0.8)	-5.201
0.9484	0.971*	0.9498	0.9716*	0.9514	0.9716*	0.9514	5000	2	0.02	(0.98,0.9)	-5.201
0.9448	0.9606*	0.9522	0.96*	0.9518	0.96*	0.9522	5000	4	0.02	(0.9,0.8)	-5.201
0.9448	0.9608*	0.95	0.9606*	0.9506	0.9606*	0.9506	5000	4	0.02	(0.98,0.9)	-5.201
0.95	0.9662*	0.9472	0.9664*	0.9474	0.9668*	0.9482	10000	2	0.02	(0.9,0.8)	-5.201
0.95	0.9698*	0.9492	0.9694*	0.9492	0.9694*	0.9494	10000	2	0.02	(0.98,0.9)	-5.201
0.9494	0.9608*	0.9502	0.9618*	0.9508	0.9618*	0.951	10000	4	0.02	(0.9,0.8)	-5.201
0.9494	0.965*	0.9538	0.965*	0.9542	0.965*	0.9542	10000	4	0.02	(0.98,0.9)	-5.201
0.949	0.9694*	0.9498	0.969*	0.9496	0.969*	0.9502	5000	2	0.05	(0.9,0.8)	-4.289
0.949	0.9692*	0.9502	0.969*	0.9502	0.969*	0.9504	5000	2	0.05	(0.98,0.9)	-4.289
0.9432*	0.9614*	0.9516	0.9626*	0.9514	0.9626*	0.9514	5000	4	0.05	(0.9,0.8)	-4.289
0.9432*	0.957*	0.9478	0.957*	0.9476	0.957*	0.9476	5000	4	0.05	(0.98,0.9)	-4.289
0.944	0.9656*	0.949	0.9662*	0.9474	0.9662*	0.9478	10000	2	0.05	(0.9,0.8)	-4.289
0.944	0.97*	0.9492	0.9708*	0.9508	0.971*	0.9508	10000	2	0.05	(0.98,0.9)	-4.289
0.9486	0.9556	0.9498	0.9564*	0.95	0.9568*	0.9502	10000	4	0.05	(0.9,0.8)	-4.289
0.9486	0.9588*	0.9486	0.9596*	0.949	0.9596*	0.949	10000	4	0.05	(0.98,0.9)	-4.289
0.9526	0.9652*	0.9474	0.9646*	0.9494	0.9652*	0.9498	5000	2	0.1	(0.9,0.8)	-3.602
0.9526	0.9664*	0.9494	0.9666*	0.95	0.9668*	0.95	5000	2	0.1	(0.98,0.9)	-3.602
0.9462	0.9562*	0.9498	0.9572*	0.9484	0.9574*	0.9486	5000	4	0.1	(0.9,0.8)	-3.602
0.9462	0.9538	0.9474	0.9538	0.948	0.9538	0.948	5000	4	0.1	(0.98,0.9)	-3.602

网络 表格 37- 使用不同分析方法和方差估计方法进行分层抽样时，在缺失协变量数据的不同概率下，5,000个模拟队列中纯风险参数 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的95%置信区间覆盖率，其中 $\boldsymbol{x}=(1,-1,0.6)'$，。* 表示覆盖率超出预期区间 [0.9440; 0.9560]。
</final

队列	USCC.True		USCC.Est		USCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$	$\widehat{V}_{\#\$\%\&'(}$	$\widehat{V}$					
0.9484	0.9578*	0.9552	0.9584*	0.9552	0.9584*	0.9556	5000	2	0.02	(0.9,0.8)	-5.201
0.9484	0.9554	0.9528	0.9556	0.9536	0.9556	0.9536	5000	2	0.02	(0.98,0.9)	-5.201
0.9448	0.9478	0.9462	0.946	0.9452	0.9466	0.9454	5000	4	0.02	(0.9,0.8)	-5.201
0.9448	0.948	0.9466	0.9482	0.9466	0.9482	0.9466	5000	4	0.02	(0.98,0.9)	-5.201
0.95	0.956	0.951	0.956	0.952	0.9562*	0.9522	10000	2	0.02	(0.9,0.8)	-5.201
0.95	0.9564*	0.953	0.9552	0.953	0.9554	0.953	10000	2	0.02	(0.98,0.9)	-5.201
0.9494	0.9534	0.9518	0.9526	0.951	0.9532	0.9512	10000	4	0.02	(0.9,0.8)	-5.201
0.9494	0.953	0.9516	0.953	0.9516	0.953	0.9516	10000	4	0.02	(0.98,0.9)	-5.201
0.949	0.9524	0.9508	0.9518	0.9508	0.9518	0.951	5000	2	0.05	(0.9,0.8)	-4.289
0.949	0.9544	0.952	0.9546	0.9518	0.9546	0.9518	5000	2	0.05	(0.98,0.9)	-4.289
0.9432*	0.9476	0.9462	0.9486	0.9474	0.9488	0.9476	5000	4	0.05	(0.9,0.8)	-4.289
0.9432*	0.9512	0.9498	0.9502	0.9488	0.9502	0.9488	5000	4	0.05	(0.98,0.9)	-4.289
0.944	0.952	0.9504	0.9526	0.9504	0.9528	0.9508	10000	2	0.05	(0.9,0.8)	-4.289
0.944	0.9534	0.95	0.954	0.951	0.954	0.951	10000	2	0.05	(0.98,0.9)	-4.289
0.9486	0.9508	0.9496	0.9522	0.951	0.9524	0.9514	10000	4	0.05	(0.9,0.8)	-4.289
0.9486	0.952	0.9504	0.9522	0.9504	0.9522	0.9504	10000	4	0.05	(0.98,0.9)	-4.289
0.9526	0.9546	0.952	0.9546	0.9522	0.9548	0.9526	5000	2	0.1	(0.9,0.8)	-3.602
0.9526	0.9586*	0.9566*	0.9596*	0.9562*	0.9596*	0.9564*	5000	2	0.1	(0.98,0.9)	-3.602
0.9462	0.9504	0.9492	0.9514	0.9508	0.9518	0.9512	5000	4	0.1	(0.9,0.8)	-3.602
0.9462	0.9512	0.9504	0.9514	0.95	0.9514	0.95	5000	4	0.1	(0.98,0.9)	-3.602

网络表38-不同分析方法和方差估计方法下，基于非分层抽样，在协变量数据缺失概率不同的情况下，纯风险参数$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$的95%置信区间覆盖率，其中$\boldsymbol{x}=(1,-1,0.6)'$，来自5,000个模拟队列。*表示覆盖率超出预期区间 [0.9440; 0.9560]

队列	SCC.True		SCC.Est		SCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.944	0.962	0.952	0.964	0.952	0.9644	0.9522	5000	2	0.02	(0.9,0.8)	-4.702
0.944	0.9652	0.949	0.9652	0.9484	0.9658	0.949	5000	2	0.02	(0.98,0.9)	-4.702
0.9426	0.955	0.9498	0.9546	0.9496	0.9548	0.9504	5000	4	0.02	(0.9,0.8)	-4.702
0.9426	0.9542	0.9476	0.954	0.9468	0.954	0.947	5000	4	0.02	(0.98,0.9)	-4.702
0.9458	0.9598	0.9498	0.9614	0.9486	0.9618	0.949	10000	2	0.02	(0.9,0.8)	-4.702
0.9458	0.9678	0.9524	0.9666	0.951	0.9666	0.9514	10000	2	0.02	(0.98,0.9)	-4.702
0.9514	0.9602	0.9528	0.96	0.953	0.9604	0.9534	10000	4	0.02	(0.9,0.8)	-4.702
0.9514	0.9594	0.952	0.9602	0.9522	0.9602	0.9526	10000	4	0.02	(0.98,0.9)	-4.702
0.9468	0.96	0.9456	0.9598	0.9462	0.9604	0.9476	5000	2	0.05	(0.9,0.8)	-3.793
0.9468	0.9558	0.9452	0.9564	0.944	0.9564	0.944	5000	2	0.05	(0.98,0.9)	-3.793
0.9526	0.9554	0.9512	0.9572	0.9522	0.9576	0.9528	5000	4	0.05	(0.9,0.8)	-3.793
0.9526	0.9594	0.9536	0.9602	0.9534	0.9602	0.9534	5000	4	0.05	(0.98,0.9)	-3.793
0.9464	0.9658	0.9546	0.9654	0.9538	0.9662	0.9546	10000	2	0.05	(0.9,0.8)	-3.793
0.9464	0.9654	0.952	0.9648	0.9522	0.9648	0.9526	10000	2	0.05	(0.98,0.9)	-3.793
0.9532	0.9602	0.9562	0.9592	0.9548	0.9594	0.955	10000	4	0.05	(0.9,0.8)	-3.793
0.9532	0.9616	0.956	0.9616	0.954	0.9618	0.9542	10000	4	0.05	(0.98,0.9)	-3.793
0.9532	0.9616	0.9542	0.9614	0.9532	0.9632	0.9542	5000	2	0.1	(0.9,0.8)	-3.111
0.9532	0.9638	0.951	0.9638	0.9518	0.964	0.9524	5000	2	0.1	(0.98,0.9)	-3.111
0.9532	0.9556	0.951	0.956	0.952	0.9564	0.9522	5000	4	0.1	(0.9,0.8)	-3.111
0.9532	0.9556	0.9512	0.9546	0.9518	0.9546	0.9518	5000	4	0.1	(0.98,0.9)	-3.111

网络表格 39-纯风险参数 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的95%置信区间覆盖率：在5,000个模拟队列中，针对不同的协变量数据缺失概率，采用不同的分析方法和方差估计方法进行分层抽样，且$\boldsymbol{x}=(1,1,0.6)'$，。*表示覆盖率超出预期区间[0.9440; 0.9560]。

队列	USCC.True		USCC.Est		USCC.Naive		n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.944	0.9604*	0.9524	0.9578*	0.9534	0.9584*	0.9538	5000	2	0.02	(0.9,0.8)	-4.702
0.944	0.959*	0.954	0.9588*	0.9536	0.9588*	0.9536	5000	2	0.02	(0.98,0.9)	-4.702
0.9426*	0.95	0.947	0.9488	0.9466	0.9492	0.9468	5000	4	0.02	(0.9,0.8)	-4.702
0.9426*	0.9518	0.9474	0.9512	0.947	0.9512	0.947	5000	4	0.02	(0.98,0.9)	-4.702
0.9458	0.9608*	0.9554	0.9608*	0.957*	0.9616*	0.958*	10000	2	0.02	(0.9,0.8)	-4.702
0.9458	0.9608*	0.9548	0.9614*	0.9544	0.9614*	0.9546	10000	2	0.02	(0.98,0.9)	-4.702
0.9514	0.956	0.9516	0.9542	0.9504	0.9542	0.951	10000	4	0.02	(0.9,0.8)	-4.702
0.9514	0.9572*	0.9528	0.9558	0.952	0.956	0.952	10000	4	0.02	(0.98,0.9)	-4.702
0.9468	0.9526	0.948	0.9556	0.9494	0.9562*	0.95	5000	2	0.05	(0.9,0.8)	-3.793
0.9468	0.9536	0.9478	0.953	0.948	0.953	0.948	5000	2	0.05	(0.98,0.9)	-3.793
0.9526	0.9538	0.9512	0.9554	0.9532	0.9554	0.954	5000	4	0.05	(0.9,0.8)	-3.793
0.9526	0.9604*	0.9586*	0.9604*	0.958	0.9604*	0.958	5000	4	0.05	(0.98,0.9)	-3.793
0.9464	0.9586*	0.9522	0.9584*	0.9532	0.9594*	0.9536	10000	2	0.05	(0.9,0.8)	-3.793
0.9464	0.9596*	0.9538	0.9602*	0.9542	0.9602*	0.9542	10000	2	0.05	(0.98,0.9)	-3.793
0.9532	0.9564*	0.9546	0.958*	0.9566*	0.9586*	0.9566*	10000	4	0.05	(0.9,0.8)	-3.793
0.9532	0.9564*	0.9536	0.9572*	0.953	0.9574*	0.953	10000	4	0.05	(0.98,0.9)	-3.793
0.9532	0.9566*	0.9522	0.9578*	0.9536	0.9586*	0.9544	5000	2	0.1	(0.9,0.8)	-3.111
0.9532	0.9562*	0.9518	0.9558	0.9518	0.956	0.9518	5000	2	0.1	(0.98,0.9)	-3.111
0.9532	0.9474	0.9454	0.9484	0.947	0.9486	0.9476	5000	4	0.1	(0.9,0.8)	-3.111
0.9532	0.954	0.952	0.9532	0.9522	0.9532	0.9522	5000	4	0.1	(0.98,0.9)	-3.111

网络表40- 纯风险参数 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的95%置信区间覆盖率：基于非分层抽样，在5,000个模拟队列中，针对不同的协变量数据缺失概率，采用不同的分析方法和方差估计方法计算。*表示覆盖率超出了预期区间 [0.9440; 0.9560]。

队列		SCC.True			SCC.Est			SCC.Naive			n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_1
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.0135	0.014	0.0224	0.0261	0.0232	0.0224	0.0261	0.0232	0.0224	0.0261	0.0232	5000	2	0.02	(0.9,0.8)	-0.2
0.0135	0.014	0.0195	0.0232	0.0202	0.0195	0.0232	0.0202	0.0195	0.0232	0.0202	5000	2	0.02	(0.98,0.9)	-0.2
0.0141	0.0139	0.0207	0.0214	0.02	0.0207	0.0214	0.02	0.0207	0.0214	0.02	5000	4	0.02	(0.9,0.8)	-0.2
0.0141	0.0139	0.0181	0.019	0.0176	0.0181	0.019	0.0176	0.0181	0.019	0.0176	5000	4	0.02	(0.98,0.9)	-0.2
0.007	0.007	0.0111	0.0124	0.0109	0.0111	0.0124	0.0109	0.0111	0.0124	0.0109	10000	2	0.02	(0.9,0.8)	-0.2
0.007	0.007	0.0096	0.0111	0.0096	0.0096	0.0111	0.0096	0.0096	0.0111	0.0096	10000	2	0.02	(0.98,0.9)	-0.2
0.0071	0.0069	0.01	0.0104	0.0097	0.01	0.0104	0.0097	0.01	0.0104	0.0097	10000	4	0.02	(0.9,0.8)	-0.2
0.0071	0.0069	0.0086	0.0093	0.0086	0.0086	0.0093	0.0086	0.0086	0.0093	0.0086	10000	4	0.02	(0.98,0.9)	-0.2
0.0056	0.0056	0.0084	0.0095	0.0085	0.0084	0.0095	0.0085	0.0084	0.0095	0.0085	5000	2	0.05	(0.9,0.8)	-0.2
0.0056	0.0056	0.0075	0.0085	0.0075	0.0075	0.0085	0.0075	0.0075	0.0085	0.0075	5000	2	0.05	(0.98,0.9)	-0.2
0.0058	0.0056	0.0078	0.0081	0.0077	0.0078	0.0081	0.0077	0.0078	0.0081	0.0077	5000	4	0.05	(0.9,0.8)	-0.2
0.0058	0.0056	0.0069	0.0072	0.0068	0.0069	0.0072	0.0068	0.0069	0.0072	0.0068	5000	4	0.05	(0.98,0.9)	-0.2
0.0028	0.0028	0.0041	0.0047	0.0042	0.0041	0.0047	0.0042	0.0041	0.0047	0.0042	10000	2	0.05	(0.9,0.8)	-0.2
0.0028	0.0028	0.0036	0.0042	0.0037	0.0036	0.0042	0.0037	0.0036	0.0042	0.0037	10000	2	0.05	(0.98,0.9)	-0.2
0.0028	0.0028	0.0038	0.004	0.0038	0.0038	0.004	0.0038	0.0038	0.004	0.0038	10000	4	0.05	(0.9,0.8)	-0.2
0.0028	0.0028	0.0034	0.0036	0.0034	0.0034	0.0036	0.0034	0.0034	0.0036	0.0034	10000	4	0.05	(0.98,0.9)	-0.2
0.0029	0.0029	0.0042	0.0045	0.0041	0.0042	0.0045	0.0041	0.0042	0.0045	0.0041	5000	2	0.1	(0.9,0.8)	-0.2
0.0029	0.0029	0.0037	0.0041	0.0037	0.0037	0.0041	0.0037	0.0037	0.0041	0.0037	5000	2	0.1	(0.98,0.9)	-0.2
0.0029	0.0029	0.0038	0.0039	0.0038	0.0038	0.0039	0.0038	0.0038	0.0039	0.0038	5000	4	0.1	(0.9,0.8)	-0.2
0.0029	0.0029	0.0033	0.0035	0.0033	0.0033	0.0035	0.0033	0.0033	0.0035	0.0033	5000	4	0.1	(0.98,0.9)	-0.2

网络表41- 分层抽样所得对数相对风险参数 β_1 的经验方差和估计方差的均值

使用不同的分析方法和方差估计方法，在协变量数据缺失概率各异的情况下，对 5,000 次模拟

队列。

队列		USCC.True			USCC.Est			USCC.Naive			n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_1
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.0135	0.014	0.0264	0.0273	0.0273	0.0264	0.0273	0.0273	0.0264	0.0273	0.0273	5000	2	0.02	(0.9,0.8)	-0.2
0.0135	0.014	0.0235	0.0242	0.0243	0.0235	0.0242	0.0243	0.0235	0.0242	0.0243	5000	2	0.02	(0.98,0.9)	-0.2
0.0141	0.0139	0.0223	0.0221	0.0221	0.0223	0.0221	0.0221	0.0223	0.0221	0.0221	5000	4	0.02	(0.9,0.8)	-0.2
0.0141	0.0139	0.0197	0.0196	0.0196	0.0197	0.0196	0.0196	0.0197	0.0196	0.0196	5000	4	0.02	(0.98,0.9)	-0.2
0.007	0.007	0.0129	0.0128	0.0128	0.0129	0.0128	0.0128	0.0129	0.0128	0.0128	10000	2	0.02	(0.9,0.8)	-0.2
0.007	0.007	0.0115	0.0115	0.0115	0.0115	0.0115	0.0115	0.0115	0.0115	0.0115	10000	2	0.02	(0.98,0.9)	-0.2
0.0071	0.0069	0.0111	0.0106	0.0106	0.0111	0.0106	0.0106	0.0111	0.0106	0.0106	10000	4	0.02	(0.9,0.8)	-0.2
0.0071	0.0069	0.0099	0.0095	0.0095	0.0099	0.0095	0.0095	0.0099	0.0095	0.0095	10000	4	0.02	(0.98,0.9)	-0.2
0.0056	0.0056	0.0097	0.0098	0.0098	0.0097	0.0098	0.0098	0.0097	0.0098	0.0098	5000	2	0.05	(0.9,0.8)	-0.2
0.0056	0.0056	0.0087	0.0088	0.0088	0.0087	0.0088	0.0088	0.0087	0.0088	0.0088	5000	2	0.05	(0.98,0.9)	-0.2
0.0058	0.0056	0.0084	0.0083	0.0083	0.0084	0.0083	0.0083	0.0084	0.0083	0.0083	5000	4	0.05	(0.9,0.8)	-0.2
0.0058	0.0056	0.0075	0.0074	0.0074	0.0075	0.0074	0.0074	0.0075	0.0074	0.0074	5000	4	0.05	(0.98,0.9)	-0.2
0.0028	0.0028	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	10000	2	0.05	(0.9,0.8)	-0.2
0.0028	0.0028	0.0044	0.0044	0.0044	0.0044	0.0044	0.0044	0.0044	0.0044	0.0044	10000	2	0.05	(0.98,0.9)	-0.2
0.0028	0.0028	0.004	0.0041	0.0041	0.004	0.0041	0.0041	0.004	0.0041	0.0041	10000	4	0.05	(0.9,0.8)	-0.2
0.0028	0.0028	0.0036	0.0037	0.0037	0.0036	0.0037	0.0037	0.0036	0.0037	0.0037	10000	4	0.05	(0.98,0.9)	-0.2
0.0029	0.0029	0.0047	0.0047	0.0047	0.0047	0.0047	0.0047	0.0047	0.0047	0.0047	5000	2	0.1	(0.9,0.8)	-0.2
0.0029	0.0029	0.0041	0.0042	0.0042	0.0041	0.0042	0.0042	0.0041	0.0042	0.0042	5000	2	0.1	(0.98,0.9)	-0.2
0.0029	0.0029	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	5000	4	0.1	(0.9,0.8)	-0.2
0.0029	0.0029	0.0036	0.0035	0.0035	0.0036	0.0035	0.0035	0.0036	0.0035	0.0035	5000	4	0.1	(0.98,0.9)	-0.2

网络 表格 42- 非分层抽样的对数相对风险参数 β_1 的经验方差和估计方差的均值

在不同协变量数据缺失概率下，采用不同的分析方法和方差估计方法，对 5,000 个模拟的

队列。

队列		SCC.True			SCC.Est			SCC.Naive			n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_2
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.021	0.0197	0.0321	0.0353	0.0303	0.0321	0.0353	0.0303	0.0321	0.0353	0.0303	5000	2	0.02	(0.9,0.8)	0.25
0.021	0.0197	0.0281	0.0314	0.0264	0.0281	0.0314	0.0264	0.0281	0.0314	0.0264	5000	2	0.02	(0.98,0.9)	0.25
0.02	0.0198	0.0274	0.0298	0.0275	0.0274	0.0298	0.0275	0.0274	0.0298	0.0275	5000	4	0.02	(0.9,0.8)	0.25
0.02	0.0198	0.0243	0.0265	0.0241	0.0243	0.0265	0.0241	0.0243	0.0265	0.0241	5000	4	0.02	(0.98,0.9)	0.25
0.0097	0.0098	0.0145	0.017	0.0145	0.0145	0.017	0.0145	0.0145	0.017	0.0145	10000	2	0.02	(0.9,0.8)	0.25
0.0097	0.0098	0.0126	0.0151	0.0127	0.0126	0.0151	0.0127	0.0126	0.0151	0.0127	10000	2	0.02	(0.98,0.9)	0.25
0.0099	0.0097	0.0135	0.0144	0.0133	0.0135	0.0144	0.0133	0.0135	0.0144	0.0133	10000	4	0.02	(0.9,0.8)	0.25
0.0099	0.0097	0.0119	0.0129	0.0117	0.0119	0.0129	0.0117	0.0119	0.0129	0.0117	10000	4	0.02	(0.98,0.9)	0.25
0.0079	0.0079	0.0116	0.0132	0.0114	0.0116	0.0132	0.0114	0.0116	0.0132	0.0114	5000	2	0.05	(0.9,0.8)	0.25
0.0079	0.0079	0.01	0.0118	0.01	0.01	0.0118	0.01	0.01	0.0118	0.01	5000	2	0.05	（0.98,0.9）	0.25
0.0083	0.0079	0.0109	0.0113	0.0106	0.0109	0.0113	0.0106	0.0109	0.0113	0.0106	5000	4	0.05	(0.9,0.8)	0.25
0.0083	0.0079	0.0096	0.0101	0.0093	0.0096	0.0101	0.0093	0.0096	0.0101	0.0093	5000	4	0.05	(0.98,0.9)	0.25
0.004	0.0039	0.0058	0.0065	0.0056	0.0058	0.0065	0.0056	0.0058	0.0065	0.0056	10000	2	0.05	(0.9,0.8)	0.25
0.004	0.0039	0.0051	0.0058	0.005	0.0051	0.0058	0.005	0.0051	0.0058	0.005	10000	2	0.05	(0.98,0.9)	0.25
0.004	0.0039	0.0053	0.0056	0.0052	0.0053	0.0056	0.0052	0.0053	0.0056	0.0052	10000	4	0.05	(0.9,0.8)	0.25
0.004	0.0039	0.0046	0.005	0.0046	0.0046	0.005	0.0046	0.0046	0.005	0.0046	10000	4	0.05	(0.98,0.9)	0.25
0.004	0.004	0.0056	0.0063	0.0056	0.0056	0.0063	0.0056	0.0056	0.0063	0.0056	5000	2	0.1	(0.9,0.8)	0.25
0.004	0.004	0.0049	0.0056	0.005	0.0049	0.0056	0.005	0.0049	0.0056	0.005	5000	2	0.1	(0.98,0.9)	0.25
0.004	0.004	0.0052	0.0054	0.0052	0.0052	0.0054	0.0052	0.0052	0.0054	0.0052	5000	4	0.1	(0.9,0.8)	0.25
0.004	0.004	0.0046	0.0049	0.0046	0.0046	0.0049	0.0046	0.0046	0.0049	0.0046	5000	4	0.1	(0.98,0.9)	0.25

网页表43- 对数相对风险参数 β_2 分层抽样的经验方差和估计方差的均值

在不同协变量数据缺失概率下，采用不同的分析方法和方差估计方法，进行了 5,000 次模拟。

cohorts.

队列		USCC.True			USCC.Est			USCC.Naive			n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_2
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.021	0.0197	0.0368	0.0357	0.0357	0.0368	0.0357	0.0357	0.0368	0.0357	0.0357	5000	2	0.02	(0.9,0.8)	0.25
0.021	0.0197	0.0332	0.0318	0.0318	0.0332	0.0318	0.0318	0.0332	0.0318	0.0318	5000	2	0.02	(0.98,0.9)	0.25
0.02	0.0198	0.0301	0.03	0.03	0.0301	0.03	0.03	0.0301	0.03	0.03	5000	4	0.02	(0.9,0.8)	0.25
0.02	0.0198	0.0268	0.0267	0.0267	0.0268	0.0267	0.0267	0.0268	0.0267	0.0267	5000	4	0.02	(0.98,0.9)	0.25
0.0097	0.0098	0.017	0.0171	0.0171	0.017	0.0171	0.0171	0.017	0.0171	0.0171	10000	2	0.02	(0.9,0.8)	0.25
0.0097	0.0098	0.0152	0.0153	0.0153	0.0152	0.0153	0.0153	0.0152	0.0153	0.0153	10000	2	0.02	(0.98,0.9)	0.25
0.0099	0.0097	0.0144	0.0145	0.0145	0.0144	0.0145	0.0145	0.0144	0.0145	0.0145	10000	4	0.02	(0.9,0.8)	0.25
0.0099	0.0097	0.0129	0.0129	0.0129	0.0129	0.0129	0.0129	0.0129	0.0129	0.0129	10000	4	0.02	(0.98,0.9)	0.25
0.0079	0.0079	0.0131	0.0132	0.0133	0.0131	0.0132	0.0133	0.0131	0.0132	0.0133	5000	2	0.05	(0.9,0.8)	0.25
0.0079	0.0079	0.0119	0.0118	0.0119	0.0119	0.0118	0.0119	0.0119	0.0118	0.0119	5000	2	0.05	（0.98,0.9）	0.25
0.0083	0.0079	0.0117	0.0113	0.0113	0.0117	0.0113	0.0113	0.0117	0.0113	0.0113	5000	4	0.05	(0.9,0.8)	0.25
0.0083	0.0079	0.0104	0.0101	0.0101	0.0104	0.0101	0.0101	0.0104	0.0101	0.0101	5000	4	0.05	(0.98,0.9)	0.25
0.004	0.0039	0.0066	0.0065	0.0065	0.0066	0.0065	0.0065	0.0066	0.0065	0.0065	10000	2	0.05	(0.9,0.8)	0.25
0.004	0.0039	0.0059	0.0058	0.0058	0.0059	0.0058	0.0058	0.0059	0.0058	0.0058	10000	2	0.05	(0.98,0.9)	0.25
0.004	0.0039	0.0055	0.0056	0.0056	0.0055	0.0056	0.0056	0.0055	0.0056	0.0056	10000	4	0.05	(0.9,0.8)	0.25
0.004	0.0039	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	10000	4	0.05	(0.98,0.9)	0.25
0.004	0.004	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063	5000	2	0.1	(0.9,0.8)	0.25
0.004	0.004	0.0055	0.0057	0.0057	0.0055	0.0057	0.0057	0.0055	0.0057	0.0057	5000	2	0.1	(0.98,0.9)	0.25
0.004	0.004	0.0054	0.0055	0.0055	0.0054	0.0055	0.0055	0.0054	0.0055	0.0055	5000	4	0.1	(0.9,0.8)	0.25
0.004	0.004	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	5000	4	0.1	(0.98,0.9)	0.25

网页表格44- 非分层抽样下对数相对风险参数 β_2 的经验方差和估计方差的均值

使用不同的分析方法和方差估计方法，在协变量数据缺失概率各异的情况下，进行 5,000 次模拟。

队列。

队列		SCC.True			SCC.Est			SCC.Naive			n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_3
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.0137	0.0136	0.027	0.0262	0.0263	0.027	0.0262	0.0263	0.027	0.0262	0.0263	5000	2	0.02	(0.9,0.8)	-0.3
0.0137	0.0136	0.0236	0.0233	0.0234	0.0236	0.0233	0.0234	0.0236	0.0233	0.0234	5000	2	0.02	(0.98,0.9)	-0.3
0.0141	0.0136	0.0218	0.0214	0.0214	0.0218	0.0214	0.0214	0.0218	0.0214	0.0214	5000	4	0.02	(0.9,0.8)	-0.3
0.0141	0.0136	0.0194	0.019	0.019	0.0194	0.019	0.019	0.0194	0.019	0.019	5000	4	0.02	(0.98,0.9)	-0.3
0.0069	0.0068	0.0126	0.0124	0.0125	0.0126	0.0124	0.0125	0.0126	0.0124	0.0125	10000	2	0.02	(0.9,0.8)	-0.3
0.0069	0.0068	0.011	0.0111	0.0111	0.011	0.0111	0.0111	0.011	0.0111	0.0111	10000	2	0.02	(0.98,0.9)	-0.3
0.0068	0.0068	0.0104	0.0103	0.0103	0.0104	0.0103	0.0103	0.0104	0.0103	0.0103	10000	4	0.02	(0.9,0.8)	-0.3
0.0068	0.0068	0.0094	0.0092	0.0092	0.0094	0.0092	0.0092	0.0094	0.0092	0.0092	10000	4	0.02	(0.98,0.9)	-0.3
0.0057	0.0055	0.01	0.0096	0.0096	0.01	0.0096	0.0096	0.01	0.0096	0.0096	5000	2	0.05	(0.9,0.8)	-0.3
0.0057	0.0055	0.009	0.0086	0.0086	0.009	0.0086	0.0086	0.009	0.0086	0.0086	5000	2	0.05	(0.98,0.9)	-0.3
0.0054	0.0056	0.0078	0.0081	0.0081	0.0078	0.0081	0.0081	0.0078	0.0081	0.0081	5000	4	0.05	(0.9,0.8)	-0.3
0.0054	0.0056	0.007	0.0072	0.0072	0.007	0.0072	0.0072	0.007	0.0072	0.0072	5000	4	0.05	(0.98,0.9)	-0.3
0.0028	0.0028	0.0046	0.0047	0.0047	0.0046	0.0047	0.0047	0.0046	0.0047	0.0047	10000	2	0.05	(0.9,0.8)	-0.3
0.0028	0.0042	0.0042	0.0042	0.0042	0.0042	0.0042	0.0042	0.0042	0.0042	0.0042	10000	2	0.05	(0.98,0.9)	-0.3
0.0028	0.0039	0.0039	0.004	0.004	0.0039	0.004	0.004	0.0039	0.004	0.004	10000	4	0.05	(0.9,0.8)	-0.3
0.0028	0.0035	0.0035	0.0036	0.0036	0.0035	0.0036	0.0036	0.0035	0.0036	0.0036	10000	4	0.05	(0.98,0.9)	-0.3
0.0028	0.0044	0.0044	0.0045	0.0045	0.0044	0.0045	0.0045	0.0044	0.0045	0.0045	5000	2	0.1	(0.9,0.8)	-0.3
0.0028	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	5000	2	0.1	(0.98,0.9)	-0.3
0.0028	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	0.0038	5000	4	0.1	(0.9,0.8)	-0.3
0.0028	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	5000	4	0.1	(0.98,0.9)	-0.3

网页表45- 分层抽样下对数相对风险参数 β_3 的经验方差和估计方差的均值

针对不同的协变量数据分析方法和方差估计方法，以及各种协变量数据缺失概率，在 5,000 次模拟中

队列。

队列		USCC.True			USCC.Est			USCC.Naive			n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	β_3
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$					
0.0137	0.0136	0.028	0.0274	0.0274	0.028	0.0274	0.0274	0.028	0.0274	0.0274	5000	2	0.02	(0.9,0.8)	-0.3
0.0137	0.0136	0.0249	0.0244	0.0244	0.0249	0.0244	0.0244	0.0249	0.0244	0.0244	5000	2	0.02	(0.98,0.9)	-0.3
0.0141	0.0136	0.0232	0.0219	0.0219	0.0232	0.0219	0.0219	0.0232	0.0219	0.0219	5000	4	0.02	(0.9,0.8)	-0.3
0.0141	0.0136	0.0205	0.0195	0.0195	0.0205	0.0195	0.0195	0.0205	0.0195	0.0195	5000	4	0.02	(0.98,0.9)	-0.3
0.0069	0.0068	0.0131	0.0129	0.0129	0.0131	0.0129	0.0129	0.0131	0.0129	0.0129	10000	2	0.02	(0.9,0.8)	-0.3
0.0069	0.0068	0.0119	0.0115	0.0115	0.0119	0.0115	0.0115	0.0119	0.0115	0.0115	10000	2	0.02	(0.98,0.9)	-0.3
0.0068	0.0068	0.0108	0.0105	0.0105	0.0108	0.0105	0.0105	0.0108	0.0105	0.0105	10000	4	0.02	(0.9,0.8)	-0.3
0.0068	0.0068	0.0096	0.0094	0.0094	0.0096	0.0094	0.0094	0.0096	0.0094	0.0094	10000	4	0.02	(0.98,0.9)	-0.3
0.0057	0.0055	0.0101	0.0099	0.0099	0.0101	0.0099	0.0099	0.0101	0.0099	0.0099	5000	2	0.05	(0.9,0.8)	-0.3
0.0057	0.0055	0.0089	0.0088	0.0089	0.0089	0.0088	0.0089	0.0089	0.0088	0.0089	5000	2	0.05	(0.98,0.9)	-0.3
0.0054	0.0056	0.008	0.0082	0.0082	0.008	0.0082	0.0082	0.008	0.0082	0.0082	5000	4	0.05	(0.9,0.8)	-0.3
0.0054	0.0056	0.0071	0.0073	0.0074	0.0071	0.0073	0.0074	0.0071	0.0073	0.0074	5000	4	0.05	(0.98,0.9)	-0.3
0.0028	0.0028	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	0.0049	10000	2	0.05	(0.9,0.8)	-0.3
0.0028	0.0042	0.0045	0.0044	0.0044	0.0045	0.0044	0.0044	0.0045	0.0044	0.0044	10000	2	0.05	(0.98,0.9)	-0.3
0.0028	0.0039	0.004	0.0041	0.0041	0.004	0.0041	0.0041	0.004	0.0041	0.0041	10000	4	0.05	(0.9,0.8)	-0.3
0.0028	0.0035	0.0035	0.0036	0.0036	0.0035	0.0036	0.0036	0.0035	0.0036	0.0036	10000	4	0.05	(0.98,0.9)	-0.3
0.0028	0.0044	0.0044	0.0046	0.0046	0.0044	0.0046	0.0046	0.0044	0.0046	0.0046	5000	2	0.1	(0.9,0.8)	-0.3
0.0028	0.004	0.0039	0.0042	0.0042	0.0039	0.0042	0.0042	0.0039	0.0042	0.0042	5000	2	0.1	(0.98,0.9)	-0.3
0.0028	0.0038	0.0039	0.0039	0.0039	0.0039	0.0039	0.0039	0.0039	0.0039	0.0039	5000	4	0.1	(0.9,0.8)	-0.3
0.0028	0.0034	0.0034	0.0035	0.0035	0.0034	0.0035	0.0035	0.0034	0.0035	0.0035	5000	4	0.1	(0.98,0.9)	-0.3

网络表46- 非分层抽样下对数相对风险参数 β_3 的经验方差和估计方差的均值

采用不同的分析方法和方差估计方法，在协变量数据缺失概率各不相同的情况下，进行了 5,000 次模拟。

队列。

队列		SCC.True			SCC.Est			SCC.Naive							
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
0.0256	0.025	0.0337	0.0398	0.0334	0.0332	0.0393	0.0329	0.0332	0.0398	0.0334	5000	2	0.02	(0.9,0.8)	-3.948
0.0256	0.025	0.0312	0.0371	0.0307	0.0311	0.037	0.0306	0.0311	0.0371	0.0307	5000	2	0.02	(0.98,0.9)	-3.948
0.0255	0.0248	0.031	0.0332	0.0301	0.0307	0.0329	0.0299	0.0307	0.0332	0.0301	5000	4	0.02	(0.9,0.8)	-3.948
0.0255	0.0248	0.0289	0.0311	0.0281	0.0288	0.0311	0.028	0.0288	0.0311	0.0281	5000	4	0.02	(0.98,0.9)	-3.948
0.0123	0.0123	0.0162	0.0191	0.0159	0.0159	0.0188	0.0156	0.0159	0.0191	0.0159	10000	2	0.02	(0.9,0.8)	-3.948
0.0123	0.0123	0.015	0.0179	0.0147	0.0149	0.0179	0.0147	0.0149	0.0179	0.0147	10000	2	0.02	(0.98,0.9)	-3.948
0.0127	0.0122	0.0154	0.0161	0.0146	0.0152	0.016	0.0145	0.0152	0.0161	0.0146	10000	4	0.02	(0.9,0.8)	-3.948
0.0127	0.0122	0.0142	0.0152	0.0137	0.0141	0.0152	0.0137	0.0141	0.0152	0.0137	10000	4	0.02	(0.98,0.9)	-3.948
0.0095	0.0096	0.012	0.0144	0.0122	0.0118	0.0142	0.012	0.0118	0.0144	0.0122	5000	2	0.05	(0.9,0.8)	-3.046
0.0095	0.0096	0.0111	0.0135	0.0114	0.0111	0.0134	0.0113	0.0111	0.0135	0.0114	5000	2	0.05	(0.98,0.9)	-3.046
0.0103	0.0097	0.0117	0.0122	0.0113	0.0116	0.0121	0.0112	0.0116	0.0122	0.0113	5000	4	0.05	(0.9,0.8)	-3.046
0.0103	0.0097	0.0111	0.0115	0.0106	0.0111	0.0115	0.0106	0.0111	0.0115	0.0106	5000	4	0.05	（0.98,0.9）	-3.046
0.0047	0.0048	0.0059	0.0071	0.006	0.0058	0.007	0.0059	0.0058	0.0071	0.006	10000	2	0.05	(0.9,0.8)	-3.046
0.0047	0.0048	0.0055	0.0067	0.0056	0.0055	0.0067	0.0056	0.0055	0.0067	0.0056	10000	2	0.05	(0.98,0.9)	-3.046
0.0049	0.0048	0.0056	0.006	0.0056	0.0056	0.006	0.0055	0.0056	0.006	0.0056	10000	4	0.05	(0.9,0.8)	-3.046
0.0049	0.0048	0.0053	0.0057	0.0053	0.0053	0.0057	0.0052	0.0053	0.0057	0.0053	10000	4	0.05	(0.98,0.9)	-3.046
0.0048	0.0047	0.0059	0.0065	0.0058	0.0058	0.0065	0.0057	0.0058	0.0065	0.0058	5000	2	0.1	(0.9,0.8)	-2.377
0.0048	0.0047	0.0055	0.0062	0.0054	0.0055	0.0061	0.0054	0.0055	0.0062	0.0054	5000	2	0.1	(0.98,0.9)	-2.377
0.0048	0.0047	0.0055	0.0056	0.0054	0.0055	0.0056	0.0053	0.0055	0.0056	0.0054	5000	4	0.1	(0.9,0.8)	-2.377
0.0048	0.0047	0.0051	0.0053	0.0051	0.0051	0.0053	0.0051	0.0051	0.0053	0.0051	5000	4	0.1	(0.98,0.9)	-2.377

网页表47- 纯风险参数 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 的经验方差和估计方差的均值，其中 $\boldsymbol{x}=(-1,1,-0.6)'$，

在5,000个模拟队列中，针对不同的协变量数据缺失概率，采用分层抽样以及不同的分析方法和方差估计方法。

队列		USCC.True			USCC.Est			USCC.Naive							
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$
0.0256	0.025	0.038	0.042	0.0375	0.0374	0.0415	0.037	0.0374	0.042	0.0375	5000	2	0.02	(0.9,0.8)	-3.948
0.0256	0.025	0.0352	0.0391	0.0347	0.0351	0.039	0.0346	0.0351	0.0391	0.0347	5000	2	0.02	(0.98,0.9)	-3.948
0.0255	0.0248	0.0325	0.0343	0.0322	0.0321	0.0341	0.0319	0.0321	0.0343	0.0322	5000	4	0.02	(0.9,0.8)	-3.948
0.0255	0.0248	0.0304	0.0322	0.03	0.0303	0.0321	0.03	0.0303	0.0322	0.03	5000	4	0.02	(0.98,0.9)	-3.948
0.0123	0.0123	0.018	0.0201	0.0179	0.0178	0.0199	0.0176	0.0178	0.0201	0.0179	10000	2	0.02	(0.9,0.8)	-3.948
0.0123	0.0123	0.0167	0.0188	0.0166	0.0167	0.0188	0.0165	0.0167	0.0188	0.0166	10000	2	0.02	(0.98,0.9)	-3.948
0.0127	0.0122	0.0162	0.0166	0.0155	0.0161	0.0165	0.0154	0.0161	0.0166	0.0155	10000	4	0.02	(0.9,0.8)	-3.948
0.0127	0.0122	0.0153	0.0156	0.0146	0.0153	0.0156	0.0145	0.0153	0.0156	0.0146	10000	4	0.02	(0.98,0.9)	-3.948
0.0095	0.0096	0.0132	0.0151	0.0136	0.0131	0.0149	0.0134	0.0131	0.0151	0.0136	5000	2	0.05	(0.9,0.8)	-3.046
0.0095	0.0096	0.0125	0.0142	0.0126	0.0125	0.0141	0.0126	0.0125	0.0142	0.0126	5000	2	0.05	(0.98,0.9)	-3.046
0.0103	0.0097	0.0127	0.0126	0.0119	0.0126	0.0125	0.0118	0.0126	0.0126	0.0119	5000	4	0.05	(0.9,0.8)	-3.046
0.0103	0.0097	0.0119	0.0119	0.0112	0.0119	0.0119	0.0112	0.0119	0.0119	0.0112	5000	4	0.05	（0.98,0.9）	-3.046
0.0047	0.0048	0.0066	0.0075	0.0067	0.0065	0.0074	0.0066	0.0065	0.0075	0.0067	10000	2	0.05	(0.9,0.8)	-3.046
0.0047	0.0048	0.0061	0.007	0.0062	0.0061	0.007	0.0062	0.0061	0.007	0.0062	10000	2	0.05	(0.98,0.9)	-3.046
0.0049	0.0048	0.0058	0.0062	0.0059	0.0058	0.0062	0.0058	0.0058	0.0062	0.0059	10000	4	0.05	(0.9,0.8)	-3.046
0.0049	0.0048	0.0056	0.0059	0.0055	0.0056	0.0059	0.0055	0.0056	0.0059	0.0055	10000	4	0.05	(0.98,0.9)	-3.046
0.0048	0.0047	0.0066	0.0069	0.0063	0.0065	0.0068	0.0062	0.0065	0.0069	0.0063	5000	2	0.1	(0.9,0.8)	-2.377
0.0048	0.0047	0.0061	0.0065	0.0059	0.0061	0.0065	0.0059	0.0061	0.0065	0.0059	5000	2	0.1	(0.98,0.9)	-2.377
0.0048	0.0047	0.0057	0.0058	0.0056	0.0057	0.0057	0.0055	0.0057	0.0058	0.0056	5000	4	0.1	(0.9,0.8)	-2.377
0.0048	0.0047	0.0054	0.0055	0.0053	0.0054	0.0055	0.0052	0.0054	0.0055	0.0053	5000	4	0.1	(0.98,0.9)	-2.377

网页表48- 纯风险参数 $\log\{\pi(\tau_1, \tau_2; \boldsymbol{x})\}$ 的经验方差和估计方差的均值，其中 $\boldsymbol{x} = (-1, 1, -0.6)'$，

在5,000个模拟队列中，针对不同的协变量数据缺失概率，采用非分层抽样以及不同的分析方法和方差估计方法。

队列		SCC.True			SCC.Est			SCC.Naive						
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
0.1304	0.126	0.1913	0.216	0.1829	0.1904	0.2154	0.1823	0.1904	0.216	0.1829	5000	2	0.02	(0.9,0.8)	-5.201
0.1304	0.126	0.1644	0.1933	0.1602	0.1642	0.1932	0.16	0.1642	0.1933	0.1602	5000	2	0.02	(0.98,0.9)	-5.201
0.1311	0.126	0.1699	0.1831	0.1672	0.1697	0.1828	0.1669	0.1697	0.183	0.1672	5000	4	0.02	(0.9,0.8)	-5.201
0.1311	0.126	0.1528	0.1642	0.1484	0.1527	0.1642	0.1483	0.1527	0.1642	0.1484	5000	4	0.02	(0.98,0.9)	-5.201
0.0623	0.062	0.0891	0.1031	0.0868	0.0888	0.1028	0.0865	0.0888	0.1031	0.0868	10000	2	0.02	(0.9,0.8)	-5.201
0.0623	0.062	0.0779	0.0928	0.0764	0.0778	0.0928	0.0764	0.0778	0.0928	0.0764	10000	2	0.02	(0.98,0.9)	-5.201
0.0617	0.0618	0.08	0.0886	0.0807	0.0797	0.0884	0.0805	0.0797	0.0886	0.0807	10000	4	0.02	(0.9,0.8)	-5.201
0.0617	0.0618	0.0707	0.0797	0.0718	0.0707	0.0797	0.0718	0.0707	0.0797	0.0718	10000	4	0.02	(0.98,0.9)	-5.201
0.0504	0.0496	0.0691	0.0796	0.0679	0.0689	0.0794	0.0677	0.0689	0.0796	0.0679	5000	2	0.05	(0.9,0.8)	-4.289
0.0504	0.0496	0.0621	0.0719	0.0601	0.0621	0.0718	0.0601	0.0621	0.0719	0.0601	5000	2	0.05	(0.98,0.9)	-4.289
0.0519	0.0498	0.0661	0.0691	0.0638	0.0659	0.069	0.0637	0.0659	0.0691	0.0638	5000	4	0.05	(0.9,0.8)	-4.289
0.0519	0.0498	0.0589	0.0623	0.057	0.0588	0.0623	0.057	0.0588	0.0623	0.057	5000	4	0.05	(0.98,0.9)	-4.289
0.0246	0.0247	0.0335	0.0393	0.0335	0.0334	0.0392	0.0333	0.0334	0.0393	0.0335	10000	2	0.05	(0.9,0.8)	-4.289
0.0246	0.0247	0.0294	0.0355	0.0297	0.0294	0.0355	0.0297	0.0294	0.0355	0.0297	10000	2	0.05	(0.98,0.9)	-4.289
0.0248	0.0247	0.0318	0.0341	0.0315	0.0317	0.0341	0.0315	0.0317	0.0341	0.0315	10000	4	0.05	(0.9,0.8)	-4.289
0.0248	0.0247	0.0281	0.0308	0.0282	0.028	0.0308	0.0282	0.028	0.0308	0.0282	10000	4	0.05	(0.98,0.9)	-4.289
0.0246	0.0248	0.0328	0.0376	0.0329	0.0325	0.0375	0.0328	0.0325	0.0376	0.0329	5000	2	0.1	(0.9,0.8)	-3.602
0.0246	0.0248	0.029	0.034	0.0293	0.0289	0.034	0.0293	0.0289	0.034	0.0293	5000	2	0.1	(0.98,0.9)	-3.602
0.0253	0.0248	0.0313	0.0328	0.031	0.0312	0.0328	0.031	0.0312	0.0328	0.031	5000	4	0.1	(0.9,0.8)	-3.602
0.0253	0.0248	0.0283	0.0296	0.0278	0.0283	0.0296	0.0278	0.0283	0.0296	0.0278	5000	4	0.1	(0.98,0.9)	-3.602

网络表49- 纯风险参数 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})$ 的经验方差和估计方差的均值，其中 $\boldsymbol{x}=(1,-1,0.6)'$,

在5,000个模拟队列中，针对不同的协变量数据缺失概率，采用不同的分析方法和方差估计方法进行分层抽样。

队列		USCC.True			USCC.Est			USCC.Naive							
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
0.1304	0.126	0.2136	0.2179	0.2135	0.2133	0.2174	0.2129	0.2133	0.2179	0.2135	5000	2	0.02	(0.9,0.8)	-5.201
0.1304	0.126	0.1925	0.1951	0.1906	0.1924	0.195	0.1905	0.1924	0.1951	0.1906	5000	2	0.02	（0.98,0.9）	-5.201
0.1311	0.126	0.1909	0.1855	0.1833	0.1904	0.1852	0.183	0.1904	0.1854	0.1833	5000	4	0.02	(0.9,0.8)	-5.201
0.1311	0.126	0.1718	0.166	0.1639	0.1716	0.166	0.1638	0.1716	0.166	0.1638	5000	4	0.02	(0.98,0.9)	-5.201
0.0623	0.062	0.0988	0.1027	0.1004	0.0986	0.1025	0.1002	0.0986	0.1027	0.1004	10000	2	0.02	(0.9,0.8)	-5.201
0.0623	0.062	0.0893	0.0925	0.0902	0.0893	0.0924	0.0902	0.0893	0.0925	0.0902	10000	2	0.02	(0.98,0.9)	-5.201
0.0617	0.0618	0.0874	0.0886	0.0875	0.0873	0.0884	0.0873	0.0873	0.0886	0.0875	10000	4	0.02	(0.9,0.8)	-5.201
0.0617	0.0618	0.0784	0.0796	0.0784	0.0784	0.0795	0.0784	0.0784	0.0796	0.0784	10000	4	0.02	(0.98,0.9)	-5.201
0.0504	0.0496	0.0773	0.0788	0.0772	0.0769	0.0786	0.077	0.0769	0.0788	0.0772	5000	2	0.05	(0.9,0.8)	-4.289
0.0504	0.0496	0.0698	0.0712	0.0695	0.0697	0.0711	0.0695	0.0697	0.0712	0.0695	5000	2	0.05	(0.98,0.9)	-4.289
0.0519	0.0498	0.0703	0.0689	0.0681	0.0702	0.0687	0.068	0.0702	0.0688	0.0681	5000	4	0.05	(0.9,0.8)	-4.289
0.0519	0.0498	0.0624	0.062	0.0613	0.0623	0.062	0.0613	0.0623	0.062	0.0613	5000	4	0.05	(0.98,0.9)	-4.289
0.0246	0.0247	0.0379	0.039	0.0381	0.0377	0.0389	0.038	0.0377	0.039	0.0381	10000	2	0.05	(0.9,0.8)	-4.289
0.0246	0.0247	0.0341	0.0352	0.0344	0.034	0.0352	0.0344	0.034	0.0352	0.0344	10000	2	0.05	(0.98,0.9)	-4.289
0.0248	0.0247	0.0333	0.034	0.0336	0.0333	0.0339	0.0335	0.0333	0.034	0.0336	10000	4	0.05	(0.9,0.8)	-4.289
0.0248	0.0247	0.0302	0.0307	0.0303	0.0302	0.0306	0.0303	0.0302	0.0307	0.0303	10000	4	0.05	(0.98,0.9)	-4.289
0.0246	0.0248	0.0365	0.0371	0.0365	0.0364	0.037	0.0363	0.0364	0.0371	0.0365	5000	2	0.1	(0.9,0.8)	-3.602
0.0246	0.0248	0.0324	0.0336	0.0329	0.0324	0.0335	0.0329	0.0324	0.0336	0.0329	5000	2	0.1	(0.98,0.9)	-3.602
0.0253	0.0248	0.0326	0.0326	0.0323	0.0325	0.0325	0.0323	0.0325	0.0326	0.0323	5000	4	0.1	(0.9,0.8)	-3.602
0.0253	0.0248	0.0296	0.0294	0.0292	0.0296	0.0294	0.0292	0.0296	0.0294	0.0292	5000	4	0.1	(0.98,0.9)	-3.602

网络表格 50- 纯风险参数 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ with $\boldsymbol{x}=(1,-1,0.6)'$,

在5,000个模拟队列中，针对不同的协变量数据缺失概率，采用不同的分析方法和方差估计方法进行非分层抽样。

队列		SCC.True			SCC.Est			SCC.Naive							
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
0.0569	0.0559	0.0876	0.0938	0.0847	0.0867	0.0932	0.0841	0.0867	0.0938	0.0847	5000	2	0.02	(0.9,0.8)	-4.702
0.0569	0.0559	0.0771	0.0849	0.0758	0.0768	0.0848	0.0757	0.0768	0.0849	0.0758	5000	2	0.02	(0.98,0.9)	-4.702
0.0577	0.0559	0.0762	0.0793	0.075	0.0763	0.0791	0.0747	0.0763	0.0793	0.075	5000	4	0.02	（0.9，0.8）	-4.702
0.0577	0.0559	0.07	0.0721	0.0678	0.0699	0.0721	0.0677	0.0699	0.0721	0.0678	5000	4	0.02	(0.98,0.9)	-4.702
0.0274	0.0277	0.0401	0.0444	0.0398	0.0399	0.0441	0.0396	0.0399	0.0444	0.0398	10000	2	0.02	(0.9,0.8)	-4.702
0.0274	0.0277	0.0357	0.0404	0.0358	0.0355	0.0403	0.0358	0.0355	0.0404	0.0358	10000	2	0.02	(0.98,0.9)	-4.702
0.0267	0.0276	0.0354	0.0382	0.036	0.0353	0.0381	0.0359	0.0353	0.0382	0.036	10000	4	0.02	(0.9,0.8)	-4.702
0.0267	0.0276	0.0319	0.0349	0.0327	0.0319	0.0348	0.0326	0.0319	0.0349	0.0327	10000	4	0.02	(0.98,0.9)	-4.702
0.0228	0.0221	0.0317	0.034	0.0308	0.0314	0.0338	0.0306	0.0314	0.034	0.0308	5000	2	0.05	(0.9,0.8)	-3.793
0.0228	0.0221	0.0291	0.0311	0.0278	0.0291	0.031	0.0278	0.0291	0.0311	0.0278	5000	2	0.05	(0.98,0.9)	-3.793
0.0221	0.0222	0.0285	0.0296	0.0282	0.0283	0.0295	0.0281	0.0283	0.0296	0.0282	5000	4	0.05	(0.9,0.8)	-3.793
0.0221	0.0222	0.0257	0.0271	0.0256	0.0256	0.0271	0.0256	0.0256	0.0271	0.0256	5000	4	0.05	(0.98,0.9)	-3.793
0.011	0.011	0.0147	0.0167	0.0151	0.0147	0.0166	0.015	0.0147	0.0167	0.0151	10000	2	0.05	(0.9,0.8)	-3.793
0.011	0.011	0.0133	0.0153	0.0137	0.0133	0.0153	0.0136	0.0133	0.0153	0.0137	10000	2	0.05	(0.98,0.9)	-3.793
0.0108	0.011	0.0136	0.0146	0.0139	0.0136	0.0146	0.0138	0.0136	0.0146	0.0139	10000	4	0.05	(0.9,0.8)	-3.793
0.0108	0.011	0.0123	0.0134	0.0126	0.0123	0.0134	0.0126	0.0123	0.0134	0.0126	10000	4	0.05	(0.98,0.9)	-3.793
0.0108	0.011	0.0142	0.0158	0.0145	0.0141	0.0157	0.0144	0.0141	0.0158	0.0145	5000	2	0.1	(0.9,0.8)	-3.111
0.0108	0.011	0.013	0.0145	0.0132	0.013	0.0145	0.0132	0.013	0.0145	0.0132	5000	2	0.1	(0.98,0.9)	-3.111
0.0108	0.011	0.0133	0.0139	0.0134	0.0133	0.0139	0.0134	0.0133	0.0139	0.0134	5000	4	0.1	(0.9,0.8)	-3.111
0.0108	0.011	0.0122	0.0127	0.0122	0.0121	0.0127	0.0122	0.0121	0.0127	0.0122	5000	4	0.1	(0.98,0.9)	-3.111

网络表格 51- 经验方差和纯风险参数估计方差的均值 $\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$ 使用$\boldsymbol{x}=(1,1,0.6)'$, 来自

在不同协变量数据缺失概率下，采用不同的分析方法和方差估计方法进行分层抽样,

5,000个模拟队列。

队列		USCC.True			USCC.Est			USCC.Naive							
经验 var	$\hat{V}_{\#\$\%\&'(}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	经验 var	$\hat{V}_{\#\$\%\&'(}$	$\hat{V}$	n	K	p_Y	$\boldsymbol{\pi}^{(3)}$	$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$
0.0569	0.0559	0.0922	0.0962	0.0917	0.0916	0.0957	0.0911	0.0916	0.0962	0.0917	5000	2	0.02	(0.9,0.8)	-4.702
0.0569	0.0559	0.083	0.087	0.0825	0.0828	0.0869	0.0824	0.0828	0.087	0.0825	5000	2	0.02	(0.98,0.9)	-4.702
0.0577	0.0559	0.0822	0.0808	0.0786	0.082	0.0805	0.0783	0.082	0.0808	0.0786	5000	4	0.02	(0.9,0.8)	-4.702
0.0577	0.0559	0.074	0.0732	0.071	0.0738	0.0731	0.0709	0.0738	0.0732	0.071	5000	4	0.02	（0.98,0.9）	-4.702
0.0274	0.0277	0.0418	0.0449	0.0426	0.0415	0.0446	0.0423	0.0415	0.0449	0.0426	10000	2	0.02	(0.9,0.8)	-4.702
0.0274	0.0277	0.038	0.041	0.0387	0.038	0.0409	0.0386	0.038	0.041	0.0387	10000	2	0.02	(0.98,0.9)	-4.702
0.0267	0.0276	0.0371	0.0385	0.0374	0.037	0.0384	0.0373	0.037	0.0385	0.0374	10000	4	0.02	(0.9,0.8)	-4.702
0.0267	0.0276	0.0334	0.0351	0.034	0.0333	0.0351	0.034	0.0333	0.0351	0.034	10000	4	0.02	(0.98,0.9)	-4.702
0.0228	0.0221	0.0332	0.0343	0.0327	0.033	0.0341	0.0325	0.033	0.0343	0.0327	5000	2	0.05	(0.9,0.8)	-3.793
0.0228	0.0221	0.0301	0.0313	0.0297	0.0301	0.0313	0.0297	0.0301	0.0313	0.0297	5000	2	0.05	(0.98,0.9)	-3.793
0.0221	0.0222	0.0285	0.0299	0.0292	0.0284	0.0298	0.0291	0.0284	0.0299	0.0292	5000	4	0.05	(0.9,0.8)	-3.793
0.0221	0.0222	0.0261	0.0273	0.0266	0.0261	0.0272	0.0265	0.0261	0.0273	0.0266	5000	4	0.05	(0.98,0.9)	-3.793
0.011	0.011	0.016	0.0169	0.016	0.0159	0.0168	0.0159	0.0159	0.0169	0.016	10000	2	0.05	(0.9,0.8)	-3.793
0.011	0.011	0.0145	0.0154	0.0146	0.0145	0.0154	0.0146	0.0145	0.0154	0.0146	10000	2	0.05	(0.98,0.9)	-3.793
0.0108	0.011	0.0139	0.0147	0.0144	0.0138	0.0147	0.0143	0.0138	0.0147	0.0144	10000	4	0.05	(0.9,0.8)	-3.793
0.0108	0.011	0.0127	0.0135	0.0131	0.0127	0.0134	0.0131	0.0127	0.0135	0.0131	10000	4	0.05	(0.98,0.9)	-3.793
0.0108	0.011	0.0148	0.016	0.0153	0.0147	0.0159	0.0152	0.0147	0.016	0.0153	5000	2	0.1	(0.9,0.8)	-3.111
0.0108	0.011	0.0135	0.0146	0.014	0.0135	0.0146	0.0139	0.0135	0.0146	0.014	5000	2	0.1	(0.98,0.9)	-3.111
0.0108	0.011	0.0137	0.014	0.0137	0.0137	0.0139	0.0137	0.0137	0.014	0.0137	5000	4	0.1	(0.9,0.8)	-3.111
0.0108	0.011	0.0123	0.0128	0.0126	0.0123	0.0128	0.0125	0.0123	0.0128	0.0126	5000	4	0.1	(0.98,0.9)	-3.111

网络表52-纯风险参数$\log\{\pi(\tau_1,\tau_2;\boldsymbol{x})\}$的经验方差和估计方差的均值，其中$\boldsymbol{x}=(1,1,0.6)'$, 采用非分层抽样，通过不同的分析方法和方差估计方法，在不同的协变量数据缺失概率下，对5,000个模拟队列进行分析。

网络附录 J. 用于在主文档表 3 中获取 SCC 和 SCC.Calib 的 $\hat{V}$ 的逐步伪代码

加载病例队列数据集 加载队列数据集 ---------- A. 使用带设计权重的分层病例队列进行估计

1. 使用设计权重，对病例队列数据运行 coxph 模型，以估计对数相对风险

2. 使用设计权重和步骤 A.2 中获得的对数相对风险估计值，从队列数据中以非参数方式估计基线风险

3. 根据步骤 A.1 和 A.2 中获得的相对风险估计值和基线风险估计值，估计纯风险

4. 计算病例队列个体对参数估计值的影响函数

5. 根据步骤 A.4 中获得的影响函数，估计参数估计值的方差

---------- B. 使用校准权重对分层病例队列进行估计 -------

3. 运行 coxph 模型，使用步骤 B.2 得到的插补后的队列数据估计对数相对风险

4. 计算队列个体对步骤 B.3 中得到的对数相对风险估计值的影响函数，以构建 Breslow 等人（2009）提出的辅助变量

5. 根据步骤 B.4 中得到的影响函数校准设计权重

6. 使用步骤 5 中得到的校准权重，对病例队列数据运行 coxph 模型，以估计对数相对风险

7. 计算整个队列在纯风险时间区间内的总随访时间

8. 使用步骤 6 中得到的对数相对风险估计值，计算插补后的队列数据中的相对风险

9. 将步骤 B.8 中得到的估计相对风险乘以步骤 7 中得到的总随访时间，以构建 Shin 等人（2020）提出的一个额外辅助变量

10. 根据步骤 B.4 和 B.9 中获得的影响函数对设计权重进行校准

11. 使用步骤 B.10 中获得的校准权重，对病例-队列数据运行 coxph 模型，以估计对数相对风险

12. 使用在步骤 B.10 中获得的校准权重和在步骤 B.11 中获得的对数相对风险估计值，从队列数据中以非参数方式估计基线风险

12. 根据在步骤 B.11 和 B.12 中获得的相对风险估计值和基线风险估计值，估计纯风险

12. 计算校准所涉及的病例队列个体对参数估计值的影响函数

13. 根据步骤 B.12 中得到的影响函数，估计参数估计值的方差

数据可用性声明

用于网络附录 E 和 I 中模拟的 R 代码及函数可在本文的补充材料中以及 GitHub 上找到，网址为 https://github.com/Etievant/CaseCohort。我们未获授权发布网络附录 F 中使用的临床数据。

参考文献

Barlow, W.E.（1994年）病例队列设计的稳健方差估计。*Biometrics*，**50**，1064–1072。

Borgan,O.、Langholz,B.、Samuelsen,S.O.、Goldstein, L. 和 Pogoda,J.（2000年）暴露分层病例队列设计。*Lifetime Data Analysis*，**6**，39–58。

Breslow, N.E. 和 Lumley, T.（2013）半参数模型与两阶段抽样：Cox 回归的应用。从概率到统计再回归：高维模型与过程——为纪念 *Jon A. Wellner* 而作的文集，**9**，65–78。

Breslow, N.E.、Lumley, T.、Ballantyne, C.M.、Chambless, L.E. 和 Kulich, M.（2009）《两阶段分层抽样中模型参数的改进 Horvitz–Thompson 估计：流行病学中的应用》。生物科学统计学，**1**，32–49。Deville, J.-C. 和 Sarndal, C.-E.（1992）《调查抽样中的校准估计量》。*Journal of theAmerican Statistical Association*，**87**，376–382。Graubard, B.I. 和 Fears, T.R.（2005）《简单与复杂抽样设计下归因风险的标准误》。*Biometrics*，**61**，847–855。Gray, R.J.（2009）《队列抽样设计的加权分析》。*Lifetime Data Analysis*，**15**，24–40。Jiao, J.（2002）《病例队列研究中方差估计量的比较》。博士学位论文，南加州大学。Pfeiffer, R.M. 和 Gail, M.H.（2017）《绝对风险：临床管理与公共卫生中的方法和应用》。Chapman and Hall/CRC，纽约。Reid, N. 和 Crépeau, H.（1985）《比例风险回归的影响函数》。*Biometrika*，**72**，1–9。

Samuelsen, S.O., Ånestad, H. 和 Skrondal, A.（2007年）一般队列抽样设计的分层病例队列分析。*Scandinavian Journal of Statistics*，**34**，103–119。

Shin, Y.E., Pfeiffer, R.M., Graubard, B.I. 和 Gail, M.H.（2020）通过权重校准提高队列中嵌套病例对照样本的纯风险估计效率。*Biometrics*，**76**，1087–1097。