Cox模型中变量选择后的统计推断：一项模拟研究

Lena Schemet医学数学统计与人工智能，奥格斯堡大学，德国奥格斯堡
lena.schemet@uni-a.de

Sarah Friedrich-Welz医学数学统计与人工智能，奥格斯堡大学；高级分析与预测科学中心（CAAPS），奥格斯堡大学，德国奥格斯堡sarah.friedrich@uni-a.de

arXiv:2602.07477v1 [stat.ME] 7 Feb 2026

摘要

选择相关预测变量是生物医学生存时间数据分析的核心。然而，经典频率学派推断假定协变量集合预先固定，并未考虑数据驱动的变量选择。因此，朴素的选择后推断可能存在偏倚并产生误导。在右删失生存情形下，删失带来的额外不确定性可能会进一步加剧这些问题。我们研究了在Cox模型背景下，对Lasso及其扩展方法自适应Lasso进行变量选择后，应用于其系数的若干推断程序。所考察的方法包括样本拆分、精确选择后推断和去偏Lasso。我们通过一项中性模拟研究考察这些方法的性能，该研究反映了生物医学应用中常见的现实协变量结构和删失率。为补充模拟结果，我们使用一个公开可用的生存数据集，通过一个应用实例展示这些程序在实际中的表现。

关键词 生存分析；Cox模型；Lasso；变量选择后的推断；选择后推断；去偏Lasso

1 引言

变量选择已成为现代回归分析不可或缺的组成部分，在协变量结构复杂且样本量有限的生物医学研究中尤为如此。Lasso [1] 及其变体 [2] 等方法被广泛用于识别相关预测变量，同时通过正则化稳定估计。然而，传统推断程序并未针对变量选择的数据驱动特性进行设计。因此，模型选择后计算出的朴素置信区间和 p 值可能存在严重偏倚、过于乐观，最终造成误导 [3]。

这个问题在线性模型和广义线性模型中广为人知，越来越多的研究强调，需要采用明确考虑选择步骤的推断程序 [4]。对于生存时间结局，这一问题尤为突出。右删失会带来额外的不确定性，而事件时间同时依赖于协变量和删失机制，这使得标准渐近论证的应用更加复杂 [5]。尽管惩罚Cox模型在生物医学应用中被广泛使用[6, 7] ，但在变量选择后对其系数进行严谨推断仍颇具挑战，且相关研究相对不足。

为解决这些局限，研究者提出了多种方法。样本拆分是一种概念性的玩具策略，将模型选择与后续推断分开进行。其有效性通常依赖于足够大的样本量，因为这两个步骤都只使用部分可用数据，这可能会降低统计效率 [8]。选择后推断会明确以选定模型为条件，并可在回归情形中获得有效的选择后置信区间。关于有效且精确的选择后推断，已有一般性框架得到提出并在理论上得到论证，但主要针对线性回归模型 [3, 4, 9]。这些方法在Cox模型中的扩展目前仍有限 [10, 11]。具体而言，Cox回归的精确选择后推断程序迄今仅在固定收缩且不进行随机化的条件下得到研究。研究表明，随机化选择后推断能够提高线性模型中的检验效能和稳定性 [12, 13], ，但Cox模型的相应理论与实现尚未建立。

除了明确以选择事件为条件的选择性推断方法外，还有用于变量选择后推断的去偏方法。这些方法对惩罚估计量进行近似单步校正，以恢复渐近正态性。因此，无需明确以选定模型为条件，也可以在变量选择后进行推断。这些方法的理论性质和实证表现主要针对线性回归模型进行了研究[14, 15]。因此，尽管低维和高维线性模型领域已取得显著进展，但关于生存分析中变量选择后推断的有效性和实际表现的现有证据仍相对有限 [10, 11]。

受这些考量的启发，我们系统评估了Cox模型中Lasso型估计量在变量选择后所采用的推断程序。基于Kammer等人 [16], 在高斯情形下评估变量选择后推断的框架，我们将其模拟设计扩展至右删失生存数据，并考察删失、协变量依赖和模型稀疏性如何影响不同方法的有效性和效率。考察的方法包括样本拆分、精确选择后推断和去偏Lasso。

我们的主要贡献如下：

- 我们将Kammer等人 [16] 的模拟研究调整并扩展至Cox模型，并纳入与生物医学应用相关的现实生存和删失结构。
- 我们对多种变量选择后推断方法进行了比较评估，重点考察广泛情景下的覆盖率、置信区间宽度和选择性检验功效。
- 我们通过一个使用公开生存数据的应用示例，说明这些方法在实践中的影响。

本文其余部分安排如下。第2节介绍Cox模型设定、Lasso和自适应Lasso估计量，以及所考察的推断程序，包括样本拆分、选择性推断和去偏方法。第3节描述模拟研究设计，包括数据生成机制、估计目标和性能指标。第4节展示模拟结果。第5节通过一个真实数据示例说明所提出的方法。第6节总结主要发现及其意义，并展开讨论。

总体而言，我们的结果为在生存分析中寻求基于Lasso的变量选择后可靠推断的实践者提供了指导，并指出了值得进一步研究的方法学方面。

2 方法

本节介绍本研究的方法学框架。我们首先介绍模型设定和记号，随后阐述所考虑的变量选择程序和选择性推断方法。

2.1 设定

令 $X = (X^{(1)}, \ldots, X^{(p)})^\top$ 表示一个 p 维协变量向量，其中 $(\cdot)^\top$ 表示任意相关向量的转置，并令 T 表示所关注的失效时间。

在右删失情况下，删失时间为 C，我们观测到生存时间 $Y = \min(T, C)$ 和事件指标 $\delta = \mathbb{1}(T \leq C)$。我们按照 Andersen等人 [5], 的定义，假设删失是非信息性的，即在给定观测到的协变量后，删失机制与事件过程相互独立，且不依赖于未知的回归参数。

对于 n 个独立个体，我们观测到独立同分布数据 (Y_i, δ_i, X_i)，其中 $i = 1, \ldots, n$，$X_i = (X_i^{(1)}, \ldots, X_i^{(p)})^\top$ 表示 p 维协变量向量。

Cox模型 [17] 将时间 t 的条件风险率指定为

$$h(t \mid X) = h_0(t) \exp(X^\top \beta^0),$$

其中，$h_0(t)$ 是未知的基准风险函数，$\beta^0 = (\beta_1^0、\ldots\ldots、\beta_p^0)^\top$ 表示未知的真实回归系数向量。

给定观测数据，对数部分似然可视为参数 $\beta \in \mathbb{R}^p$的函数，其表达式为

$$\ell(\beta) = \sum_{i=1}^{n} \delta_i \left(X_i^\top \beta - \log \sum_{j: Y_j \geq Y_i} \exp(X_j^\top \beta) \right). \tag{1}$$

真实回归参数 β^0 的估计基于最大化对数部分似然 (1)，由此得到估计量 β。在协变量较多的情况下，Lasso 及其变体等惩罚方法已成为标准工具 [6, 18, 7]。

2.2 Lasso

适配于Cox模型的Lasso估计量 [1]，，通过最大化 ℓ_1惩罚偏似然得到，

$$L_P(\beta) = \ell(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|,$$

其中 $\lambda > 0$ 为正则化参数。

Lasso具有若干吸引人的性质 [7]。特别是，ℓ_1惩罚通过将某些系数精确收缩至零来实现变量选择。该方法已在线性回归中得到广泛应用，并越来越多地用于Cox模型，尤其是在涉及大规模协变量数据的生物医学应用中 [19, 20]。

然而，这些优势也伴随着众所周知的局限性：

1. Lasso倾向于纳入过多变量，从而导致变量选择中的假阳性率偏高 [21]。2. 估计系数偏向于零，对于真实效应较大的变量尤为如此 [14, 22]。

缓解这种收缩偏差的一种方法是自适应Lasso [2]，，它将数据依赖权重 w_j，$j \in \{1, \ldots, p\}$纳入惩罚项。一种常见做法是将这些权重定义为初步系数估计的反函数，例如在 $\gamma \in \mathbb{N}$条件下的 $w_j = 1/|\beta_j|^\gamma$。该加权方案可减少较大系数的收缩，同时保持较小系数的稀疏性。在此选择下，自适应Lasso估计量达到 $\sqrt{n}$相合性，并满足预言机性质 [7]。

尽管有所改进，自适应 Lasso 程序在线性回归变量选择后的推断中，并不总能达到令人满意的覆盖性质 [3]。受此启发，我们在 Cox模型中考察并比较标准Lasso与自适应Lasso。

2.3 概念框架：变量选择后的选择性推断

选择性推断旨在回答这样一类推断问题：在观测数据之前，假设、模型或推断目标尚未确定，而是在数据分析过程中产生的[3, 9]。变量选择问题中自然会出现这种情形：首先利用数据选择一组协变量，随后仅针对选定变量报告推断结果。

在变量选择的背景下，利用数据选择一个子模型 $M \subseteq M_F$，其中 M_F 表示候选预测变量的完整集合，例如通过 Lasso 或自适应Lasso进行选择。随后对与选定变量相关的回归系数进行推断。然而，经典频率学派推断构造（$1 - \alpha$）全模型参数 β_j^F的置信区间，使得

$$\mathbb{P}\big(\beta_j^F \in CI_j\big) = 1 - \alpha,$$

在推断程序独立于任何数据驱动的模型选择这一假设下。一旦引入变量选择，这些保证通常就会失效，因为选择步骤会改变估计量和检验统计量的分布 [23, 24]。

2.3.1 从全模型世界到子模型世界

变量选择从根本上改变了推断视角。经过数据驱动的选择步骤后，推断不再是在全模型参数空间中进行，而是在由选择程序所诱导的子模型世界中进行。在这个子模型世界中，回归系数仅相对于选定变量集定义，通常与其在完整模型中的对应系数不同。

对于给定的子模型 $M \subseteq M_F$，相应的总体层面系数向量为

$$\beta_M = (\beta_{j,M})_{j \in M} = \beta_M(\beta^F),$$

它可以被视为将完整数据参数 β^F 投影到与 M[12, 4] 对应的子模型上。在线性回归中，这些子模型系数有闭式表达式；而在 Cox模型中则没有这样的解析表示 [5, 25]。相反，它们通过偏似然隐式定义，并通过数值方法求得，例如使用 Newton–Raphson型算法 [17, 5]。

2.3.2 选择性置信区间作为主要估计目标

在本文中，我们并不将这些子模型系数本身视为主要推断目标。相反，我们关注的是变量选择后报告的不确定性陈述。

我们使用术语选择性置信区间（SCI）来指代在数据驱动的变量选择步骤之后，为选定变量报告的置信区间。在本文中，这些选择性置信区间构成主要估计目标。我们通过不同推断程序所生成的选择性置信区间来评估和比较这些程序，尽管这些区间可能依赖于不同的条件化原则，也可能对应不同的底层参照系数。这种操作性视角使我们能够将各种方法作为选择后量化不确定性的实用工具进行比较，尽管它们的形式化推断目标存在差异。

对于基于样本拆分和精确条件后选择推断的方法，选择性置信区间的构造 $\widehat{\ }$ 以选定模型 M 为条件，并以相应的子模型特定系数为目标 $\beta_{j,\,\widehat{M}}$。对于这些方法，选择性覆盖率是以选择事件为条件定义的，其形式为

$$\mathbb{P}\Big(\beta_{j,\widehat{M}} \in SCI_{j,\widehat{M}} \;\Big|\; \widehat{M}\Big) = 1-\alpha, \qquad j \in \widehat{M}. \tag{2}$$

相比之下，去偏Lasso为全模型系数构造选择性置信区间 β_j^0，且不以选择事件为条件。因此，其选择性覆盖率是无条件的，具有渐近性质，并不对应于条件子模型推断。

尽管本研究考察的推断程序在其底层估计目标和覆盖保证方面各不相同，但它们都针对一个共同的实际问题：如何量化数据驱动的变量选择程序所选回归系数的不确定性。因此，本文始终采用对选择性推断 的广义操作性定义：在变量选择后进行、并针对选定变量报告的推断，其中选择性置信区间是比较的核心对象。

2.4 推断程序

下面介绍本研究考察的具体推断程序。所有方法都在 Lasso 变量选择后应用，用于构建选择性置信区间，但它们在推断目标、条件化策略和覆盖保证方面各不相同。

2.4.1 样本拆分

Cox [8]最早讨论了模型选择与统计推断的分离。样本拆分将这一思路转化为一种实用程序，即把数据分成两个不相交的部分。

$\widehat{\ }$ 首先，使用一部分数据选择一个子模型 M。然后，使用剩余数据估计相应的子模型系数并构建选择性置信区间。由于推断使用的数据未参与选择步骤，所得选择性置信区间满足条件选择性覆盖性质 (2)。具体而言，使用推断子样本对选定变量拟合无惩罚 Cox 模型，并基于偏似然构建标准 Wald 型区间，从而获得选择性置信区间。

在应用研究中，将数据简单地按 50/50 拆分通常被用作默认选择。线性回归情形下的模拟研究支持这一做法，表明这种拆分能够在稳定的模型选择与高效的推断之间实现

合理的折中 [12, 26]。不过，目前对于 Cox模型 还没有类似的指导原则，因为有效信息量取决于删失机制和样本量 [10, 11]。通过选择拆分比例，可以平衡选择准确性和推断精度。为简单起见，我们全程采用 50/50 拆分。

从实现角度来看，样本拆分简单直接，且在计算上可行。除标准未惩罚Cox模型拟合中所需的操作外，它不需要约束优化或矩阵求逆。

2.4.2 精确条件后选择推断

Lee等人 [4] 提出的精确条件后选择推断框架（精确PSI）通过对选择事件进行条件化，在Lasso选择后提供有限样本有效推断。
推断以选定子模型为条件进行，目标是相应的子模型特定估计目标。因此，所得选择性置信区间满足条件选择性覆盖性质（2）。
对于Cox模型，由于偏似然的结构，需要进行相应调整，但其基本的条件推断原理在概念上保持不变。现有软件实现支持在将惩罚参数 λ 视为固定值时进行精确条件选择性推断。

因此，当正则化参数预先指定，而非以依赖数据或预测导向的方式选择时，精确PSI适用。对于线性回归模型，已有研究提出精确PSI的随机化扩展，以适应数据驱动的调参选择。目前，Cox模型尚无此类随机化程序；即使在线性场景中，这类程序的计算开销也很大 [16]。

2.4.3 去偏Lasso

去偏Lasso最初是为线性回归模型提出的，通过校正 ℓ_1 惩罚所导致的收缩偏差，实现有效的统计推断 [14, 15]。随后，Lu 和 Xia [27]将该方法扩展至Cox比例风险模型。

其核心思想是在Lasso估计量上加入单步校正，以消除主要偏差项，从而恢复渐近正态性。对于Cox模型，具体做法是将最小二乘得分函数和信息矩阵替换为由偏似然推导出的对应量。随后，基于Cox偏似然的得分函数以及逆Fisher信息矩阵的近似进行推断，并在该近似中考虑删失。

具体而言，系数 j 的去偏估计量为

$$\tilde{\beta}_j = \widehat{\beta}_j + \widehat{\Theta}_j^\top U(\widehat{\beta}),$$

$\widehat{}$ $\widehat{}$其中， β_j 是Lasso估计量的第 j个分量， $U(\beta)$表示在 β处计算的Cox部分$\widehat{}$ $\widehat{}$似然的得分函数，而 Θ_j 表示逆Fisher信息矩阵[27]估计量的第 j行。

$\tilde{}$在满足适当的正则条件时，去偏估计量 β_j 渐近服从正态分布，因此可以构造如下形式的选择性置信区间：

$$SCI_j^{\mathrm{DB}} = \tilde{\beta}_j \pm z_{1-\alpha/2}\,\widehat{\sigma}_j/\sqrt{n},$$

其中， $z_{1-\alpha/2}$ 表示标准正态分布的$(1-\alpha/2)$分位数， $\widehat{\sigma}_j$ 是 β_j渐近标准差的相合估计量$\tilde{}$。对于Cox模型， $\widehat{\sigma}_j$ 是基于节点回归对逆Fisher信息矩阵 Θ_j 进行近似得到的，该方法考虑了删失，并避免在高维情况下直接求逆 [15, 27]。

与样本拆分和精确PSI不同，去偏Lasso的目标是全模型估计目标 β_j^F（或其部分似然投影），且不以选择事件为条件。因此，它在变量选择后进行推断，而不显式考虑选择步骤。

从计算角度来看，去偏Lasso比样本拆分和精确PSI的计算要求更高，因为它需要拟合多个辅助回归模型，以近似逆Fisher信息矩阵 [15]。Lu 和 Xia 提供了Cox模型中去偏推断的R实现，作为研究代码 [27]。

3 模拟研究

我们开展了一项全面的模拟研究，用于比较在Cox模型中进行变量选择后所应用的选择性推断程序。该研究旨在系统评估不同选择性推断方法在不同样本量、协变量维度、相关结构、删失水平和调优策略下的表现。研究特别关注选择性推断有效性、估计效率，以及推断性能与预测性能之间的关系。

模拟设计和报告遵循ADEMP结构 [28]。

3.1 研究目标 (A)

本模拟研究旨在评估数据驱动的变量选择后，针对高维Cox回归的选择性推断程序。我们考察了不同选择后推断方法所产生的选择性置信区间（SCIs）的选择性覆盖率，以及相关选择性假设的有效性、选定子模型的正确性和稳定性，以及由此得到的预测性能。我们在广泛的数据生成机制下比较了所有程序，这些机制反映了典型的生物医学生存分析场景。

3.2 数据生成机制 (D)

我们开展了一项蒙特卡罗模拟研究，每个情景重复 $n_{\text{sim}} = 1000$ 次。我们将模拟迭代集合记为 $S = \{1, \ldots, n_{\text{sim}}\}$。样本量 n 介于75至800之间，从 $n = 75$ 开始，此后以100为步长递增（表2）。

3.2.1 玩具模拟

对每个个体，我们生成了一个 p维协变量向量，其维数取 $p \in \{10,20,50\}$。协变量抽取自均值为零且具有相关结构的多元正态分布

$$\Sigma_{ij} = \rho^{|i-j|}, \qquad \rho \in \{0.0, 0.3\},$$

从而得到相互独立或中度相关的预测变量。在旨在模拟混合数据类型的情景中，将一部分协变量二值化，以得到近似服从伯努利(0.5)分布的变量。

真实系数向量 β^0 按照四种预定义模式之一生成（表1）：全一模式，即所有协变量具有相等的非零效应；高对比模式，即前四个系数呈现交替出现的大效应和小效应；真实模式，即少数协变量具有中等效应；以及稀疏模式，即仅有一个或两个非零系数。其余所有系数均设为零，具体取决于维数 p。

表1：用于生成数据向量 β^0 的系数模式。

	系数模式	描述
全为一	(1, 1, ..., 1)	对所有协变量具有相同效应。
高对比	(0.3, 1.0, 0.3, 1.0, 0, . . . , 0)	交替出现的大效应和小效应。
现实	(0.8, 0.7, 0.5, 0.8, 0, . . . , 0)	少数协变量上存在中等效应。
稀疏	(1, 1, 0, . . ., 0)	两个非零系数。

事件时间由 基于Ramos等人 [29]的Cox模型线性预测值 $X^\top \beta^0$ 生成。Two个基线风险分布被考虑。对于指数基线，

$$T = -\frac{\log U}{\exp(X^\top \beta^0)}, \qquad U \sim \text{Unif}(0,1),$$

这对应于恒定的基线风险。对于 Weibull基线，生存时间通过逆变换抽样生成，如下所示

$$T = \left(\frac{-\log U}{s \, \exp(X^\top \beta^0)} \right)^{1/k}, \qquad U \sim \text{Unif}(0,1),$$

这对应于累积基线风险为 $H_0(t) = s\,t^k$ 且基线风险为$h_0(t) = s\,k\,t^{k-1}$的 Cox模型。此处， $k > 0$ 表示 Weibull形状参数，当$k > 1$ 时基线风险递增，当 $k < 1$时基线风险递减，而 $s > 0$ 是控制基线风险总体时间尺度的尺度参数。

在所有威布尔分布情景中，我们将尺度参数固定为 $s = 1$ ，以简化模拟设计。由于基线尺度在Cox模型中不可识别，且不会影响回归系数的偏似然，因此预计这一选择不会影响对 β 的推断，并使我们能够专注于选择性推断的性质 [30, 25]。此外，形状参数固定为 $k = 2$，对应于适度递增的基线风险。本研究并未重点考察基线风险形状的变化，因为Cox模型中变量选择后的推断预计主要取决于回归结构，而非基线风险的具体参数形式。

我们通过一种简单的行政删失机制施加独立删失。对于每次重复模拟，我们首先生成事件时间$\{T_i\}_{i=1}^n$ ，然后将删失截点 $c^\star$ 定义为$\{T_i\}_{i=1}^n$的第（$1 - \pi_C$）分位数，其中 $\pi_C \in \{0,\ 0.1,\ 0.3\}$ 表示目标删失比例。观测时间定义为$\min(T,\ c^\star)$，事件指标为 $\delta = \mathbf{1}(T \le c^\star)$。按此构造，删失比例在设计上接近目标水平 [29]。

为避免并列值，仅对重复的事件时间（以及适用情况下的删失时间）添加了微小的确定性抖动。具体而言，对于每组相同的时间，我们分别为事件时间添加量级为 $\varepsilon_T = 10^{-8}$ 的偏移量，为删失时间添加量级为 $\varepsilon_C = 5 \cdot 10^{-9}$ 的偏移量，并根据相应时间变量的取值范围进行缩放。

3.2.2 现实模拟

除了玩具设置外，我们还考虑了一个经过乳腺癌分子分类国际联盟（METABRIC）研究校准的现实模拟场景 [31], 。该研究是一个具有长期临床随访的大型乳腺癌队列。第 5 节的真实数据示例也使用了同一队列。此模拟设计旨在全面控制数据生成机制，同时纳入受实证结果启发的协变量分布、效应大小和依赖结构。这种方法遵循了 Sauer等人 [32]讨论的现实参数模拟设计原则。

我们构建了一组固定的临床协变量，以反映第 5 节真实数据分析中使用的变量，包括诊断年龄、肿瘤分期、ER/HER2状态、孕激素受体状态、化疗、激素治疗、放射治疗、组织学分级、肿瘤大小、阳性淋巴结数量以及诺丁汉预后指数。分类预测变量通过显式的无截距虚拟变量设计矩阵进行编码。为确保不同模拟重复之间的设计稳定且可复现，我们删除了常量列和非常罕见的 $0/1$ 虚拟变量，并对其余协变量进行中心化和标准化。所得的稳定化设计矩阵在整个模拟研究中保持不变。

事件时间通过逆变换抽样生成，采用具有威布尔基线风险的Cox模型，并遵循与玩具模拟相同的数据生成机制。选择威布尔基线是因为它能提供灵活的单调风险函数；此外，在初步比较多种参数化基线分布（包括指数分布、对数正态分布和对数逻辑斯蒂分布）的实验中，它在时间依赖性AUC方面表现出最佳区分度。有关这些模型比较的详情，请参见补充材料S1.1。

我们在稳定化METABRIC设计矩阵上拟合了一次威布尔模型，并将拟合回归系数视为模拟真实值。在每次模拟重复中，我们从METABRIC队列中有放回地重新抽取协变量特征，并根据校准的Weibull基线模型生成事件时间。采用与玩具模拟相同的行政删失机制设定目标删失比例，并在不同情景中改变删失水平。

总体而言，METABRIC模拟是对玩具设计的结构化扩展：生存时间生成、基线风险构造和删失机制均保持一致，而协变量分布和回归系数则以真实临床数据为基础。这使我们能够在受控且经过现实校准的条件下研究选择性推断程序。

3.2.3 设计因素

模拟网格由表 2 中列出的设计因素定义。在玩具模拟中，每种情景均在两种基线风险分布（指数分布和威布尔分布）以及四种系数模式（全一、高对比、现实和稀疏）下进行评估。相比之下，METABRIC模拟基于单一的

威布尔基线风险，以及根据METABRIC数据推导并经经验校准的系数模式。其余所有设计因素，包括样本量和目标删失比例，均按照玩具设置以类似方式变化。

表2：模拟研究中变化的设计因素。

因素	玩具模拟	现实（METABRIC）	模拟
样本量 n	75–800	75–800	
协变量数量 p	10，20，50	10	
相关性 ρ	0。0，0。3	—	
目标删失比例	0,0.10，0.30	0,0.10，0.30	
基线分布	指数分布、威布尔分布	威布尔分布	
系数模式	四种不同的设定（表1）	基于真实数据的模式	

3.3 估计目标（E）

模拟研究的主要推断目标是针对由数据驱动的变量选择程序选出的回归系数所报告的选择性置信区间（SCIs）集合。对于$\widehat{\ }$$\widehat{\ }$选定子模型 $M \subseteq M_F$和一个系数索引 $j \in M$，我们将相应的选择性置信区间记为 $SCI_{j,\ \widehat{M}}$。在$\widehat{\ }$模拟研究中，我们将第 $s \in S$ 次迭代中选定的子模型记为 M_s，将该次迭代中针对变量 j 报告的选择性置信区间记为 $SCI_{j,\ \widehat{M}_s}$$\widehat{\ }$，前提是 $j \in M_s$。

次要估计目标包括相关的选择性零假设 $H_{0,\ j}$： $\beta_j = 0$，以及由此产生的选择后模型。

3.4 方法 (M)

所有考察的方法均采用两阶段结构。第一步，使用标准Lasso或自适应Lasso选出一部分协变量。第二步，应用专门的推断程序，为选定变量对应的回归系数构造选择性置信区间（SCIs）。表3对所考察的方法进行了系统概述；这些方法在如何处理变量选择步骤方面有所不同。

3.4.1 基准与朴素比较方法

为建立比较基准，我们纳入了一个包含所有可用协变量的完整 Cox模型，以及一个仅包含真实非零系数的预言模型。

此外，我们还考虑了两种忽略选择诱导偏差的朴素选择后重新拟合策略。第一种是标准的重新拟合方法，即对 Lasso 选出的变量拟合无惩罚 Cox 模型，并基于对偏似然进行完全重新优化后得到的常规 Wald置信区间进行推断。第二种变体记为 *refit0*，它在变量选择后不再进行额外的牛顿–拉夫森更新，而是使用惩罚拟合程序结束时得到的单步估计量。

纳入 refit0 变体是出于方法学方面的考虑。检查 `selectiveInference` package [33] 中 Cox模型精确选择后推断的实现后发现，其底层估计量对应的是这种单步重新拟合，而非对偏似然进行完全重新优化后得到的拟合。因此，纳入 refit0 使我们能够区分估计量本身的影响与精确PSI中采用的额外条件化步骤的影响，并评估朴素单步重新拟合与理论上有依据的选择性推断程序所依赖的估计量有多接近。

3.4.2 选择性推断方法

除了这些基准和朴素方法外，我们还考察了几种旨在提供变量选择后有效推断的方法。这些方法包括样本拆分，即通过独立的数据分割，将模型选择与推断分开；去偏Lasso，则通过基于投影的校正来

减轻收缩偏差；以及精确PSI，它通过对选择事件进行条件化，实现精确的选择后有效性。

所有基于Lasso的方法都依赖于正则化参数 λ 的选择。调参策略与选择后推断有效性之间的相互作用，是我们实证研究的一个核心方面。

3.4.3 调参规则的选择

标准Lasso和自适应Lasso的性能都在很大程度上取决于正则化参数 λ 的选择。为进行全面且具有实践意义的比较，我们考察了五种调参规则，汇总于表4。这些规则可分为预测导向、模型选择导向和固定方法。

首先，我们采用了两种基于10折交叉验证（交叉验证）的预测导向调参规则。调参参数 $\lambda_{\mathrm{CV},\ \min}$ 选择使交叉验证部分似然风险达到最小值的 λ，而$\lambda_{\mathrm{CV},\ 1\mathrm{SE}}$ 规则则选择最大的 λ，使其交叉验证风险处于最小值的一个标准误范围内。后一种规则倾向于选择更简约的模型，同时保持相近的预测性能 [34, 第7.10节1]。

其次，采用了两种信息论调优规则。AIC型选择 λ_{AIC} 通过最小化赤池型信息准则来定义，主要针对预测准确性；而BIC型选择 λ_{BIC} 则对模型复杂度施加更强的惩罚，因此倾向于选择更稀疏的模型。在适当的正则条件下，已知BIC类准则能够实现一致变量选择[35, 36, 37]。

最后，将一个固定的正则化参数 λ_{fix} 作为基准。该值根据一个大型外部数据集（$N = 100,\ 000$）得出，在该数据集上进行了 1,000 次重复 Lasso 拟合，以获得稳定且针对特定情景的适当惩罚水平估计。有关该固定调参参数的构建以及赤池信息准则和贝叶斯信息准则的定义，技术细节见补充材料S1.2。

总而言之，这一系列调参规则反映了应用型 Lasso 分析中常用的选择，并有助于系统评估不同的变量选择策略如何与后续推断相互作用，以及表3所列各方法的变量选择性质。

表3：本研究所考察方法概览。

方法	变量选择	调参	推断
full	none	–	Wald置信区间
预言机	none	–	Wald置信区间
重新拟合	Lasso	依赖于λ	重新拟合所选模型后的Wald置信区间
refit0	Lasso	依赖于λ	不重新拟合的 Wald 置信区间
拆分	Lasso	λ依赖	基于样本拆分的 Wald 置信区间
去偏	Lasso	依赖于λ	去偏 Wald 同时置信区间
精确PSI	Lasso	依赖于λ	精确条件 SCI

表4：本研究中用于选择惩罚参数 λ 的调参规则概述。

调参规则	调参类型	描述
min	基于交叉验证	λ使交叉验证风险最小化
1se	基于交叉验证	最小交叉验证风险一个标准误范围内的最大 λ
fix	固定	通过在大型模拟总体数-据集上拟合 Lasso 得到的固定惩罚参数 $\lambda_{\min}$ 据集
aic	基于信息	λ由赤池信息准则选出
bic	基于信息的	λ 由贝叶斯信息准则选出

3.5 性能指标（P）

使用 选择性覆盖率、*SCI*宽度、选择性检验功效 和 选择性第一类错误 评估性能。此外，还记录了各方法的运行时间（以秒为单位），并报告了模型性能指标。

选择性覆盖率定义为：在选定变量中，包含真实目标系数的选择性置信区间所占的比例；评估时以名义90%水平为基准，这与我们的主要软件包默认设置一致。覆盖率越接近名义水平，说明对变量选择步骤的调整效果越好。

SCI宽度对应选择性置信区间的平均长度，可作为估计精度的衡量指标；区间越短，精度越高。

当相应系数实际上不为零时，选择性检验功效表示正确拒绝选择性原假设的概率；而选择性第一类错误则表示真实系数为零时错误拒绝原假设的概率。

所有主要性能指标的详细定义和公式见补充材料 S1.3。

除推断性能外，我们还将选定模型的预测性能作为次要结果进行评估。尽管一致性指数（C指数）被广泛用于评估生存分析中的区分度，但其可解释性有限且对删失敏感，因此受到批评 [38]。因此，我们主要使用积分布里尔评分（IBS）评估预测准确性，该指标汇总整个随访期内时间依赖的预测误差平方，同时反映区分度和校准度。IBS采用删失逆概率权重计算 [39]。IBS值越低，预测性能越好。为保证结果完整，并便于与以往研究比较，我们也报告了C指数。

为量化变量选择造成的效率损失，我们在评估预测性能的同时，也评估了模型规模和变量选择准确性。利用真实活跃集信息拟合的预言模型作为理论基准。相应的预言模型结果见补充分析。

3.6 软件与实现细节

所有模拟和分析均使用R（版本4.4.2）完成。使用 `survival` package [40, 25] 拟合Cox模型并处理生存时间数据。使用 `glmnet` package [41, 42, 43]通过Lasso和自适应Lasso进行惩罚估计。

精确条件后选择推断通过 `selectiveInference` 包 [33]实现。尽管该框架在固定惩罚条件下支持Cox模型，但随机或数据驱动的惩罚选择程序目前仅限于线性回归设定。

对于Cox模型尚无成熟、现成软件支持的方法，均根据文献中的方法学描述手动实现。具体而言，Cox模型的去偏推断按照Lu 和 Xia [27], 实现，并以其研究代码作为参考。所有定制实现和模拟代码均在补充材料中提供。

4 结果

本节呈现模拟研究的结果。我们首先报告主要选择性估计目标的推断性能，然后总结与预测和变量选择相关的非选择性性能指标。

4.1 主要估计目标

第3节中定义的主要估计目标，其结果通过选择性覆盖率、SCI宽度、选择性检验功效和选择性第一类错误来报告。对选择性覆盖率、检验功效和第一类错误的定性评估，依据其与名义水平的接近程度、在样本量逐渐增大时的稳健性，以及相对于预言机程序的一致性。不设定严格的数值阈值，因为解读侧重于系统性模式，而非个别数值。

4.1.1 选择性覆盖率

选择性覆盖率按照第3.4节中的定义进行评估。覆盖概率越接近名义90%水平，表明对变量选择步骤的调整越有效。当选择性覆盖率始终接近名义90%水平，且在不同样本量下波动很小时，称其为“高”；当出现明显但非系统性的偏差时，称其为“中等”；当观察到显著或持续的覆盖不足时，称其为“低”。

图 1 展示了一个具有现实系数模式且满足 $p = 20$ 的代表性情形下的选择性覆盖概率。在采用显式选择调整时，各方法和调优策略的选择性覆盖率总体上接近名义水平。

具体而言，在大多数模拟情景中，对于中到大样本量（$n \gtrsim 300$），样本拆分和去偏Lasso的覆盖率都接近目标水平。一个例外是 $\lambda_{\mathrm{CV,\ 1SE}}$ 调参规则，此时模型选择的不确定性增加，导致覆盖率略微偏离名义水平。

相比之下，精确PSI与基于交叉验证的调参相结合时，会出现明显的覆盖不足，尤其是在 $\lambda_{\mathrm{CV,\ min}}$ 和$\lambda_{\mathrm{CV,\ 1SE}}$ 的情况下。这种现象反映了精确条件后选择推断对数据驱动的正则化参数调优较为敏感，而这种调优违反了该方法所依赖的固定-λ 假设。

这些定性模式在其他样本量、调优策略（λ_{fix}、 λ_{AIC}、 λ_{BIC}、 $\lambda_{\mathrm{CV,\ min}}$、$\lambda_{\mathrm{CV,\ 1SE}}$）以及目标删失比例下依然成立，详见补充材料 S2.1。

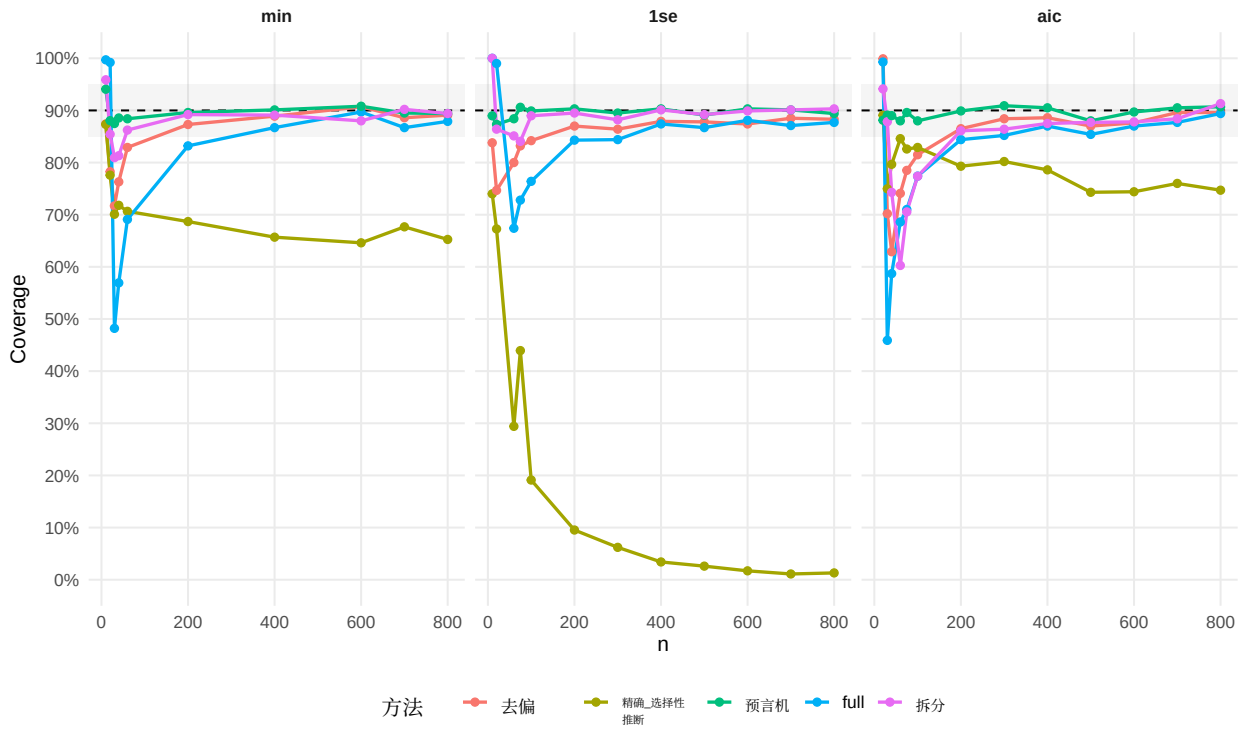

图 1：在现实系数模式下，使用 $p = 20$、威布尔基线风险、无删失且相关性为 $\rho = 0.3$ 时的选择性覆盖率。结果展示了使用非自适应Lasso、采用调参选项 $\lambda_{\mathrm{CV,min}}$,$\lambda_{\mathrm{CV,1se}}$,$\lambda_{\mathrm{AIC}}$ 时系数 X_1 的情况。

4.1.2 选择性置信区间宽度

SCI宽度用于衡量推断精度，区间越短表示效率越高。

图 2 总结了一个具有代表性的简单设定下，不同推断方法所得选择性置信区间长度的分布。SCI宽度从另一个角度展现了选择性推断的表现，突显了不同调整策略在实际应用中的影响。

在各种情景下，去偏Lasso得到的选择性置信区间相对较短且稳定，体现了偏差校正与方差膨胀之间较好的平衡。样本拆分得到的选择性置信区间更宽，变异性也更大；在基于 METABRIC 校准的设计中尤为明显，这与其有效样本量较低相符。

精确PSI总体上产生最宽的选择性置信区间，尤其是在结合基于交叉验证的调参时。这反映了对复杂选择事件进行条件化所导致的显著不确定性膨胀。增加样本量会一致地缩短所有方法的选择性置信区间宽度，但不会改变它们的相对排序。

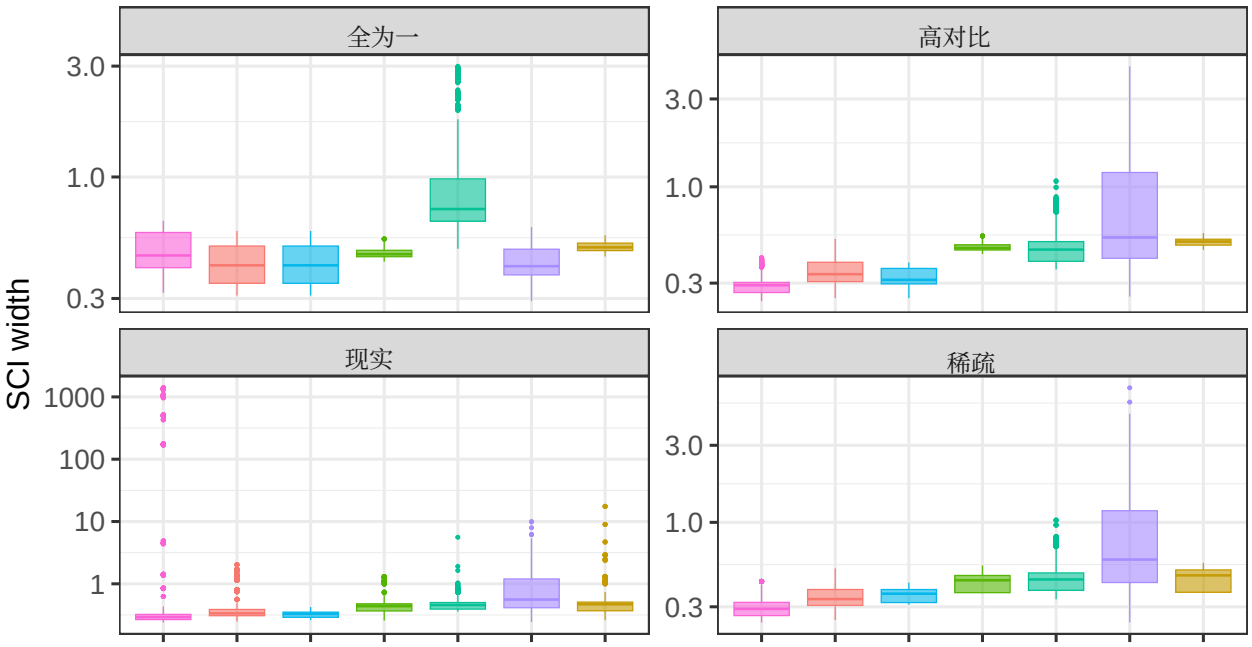

图 2：选择性置信区间长度的分布，针对样本量为 $n = 200$ 时的主要估计目标，跨不同推断方法。结果 e 以对数刻度展示了玩具设置和 METABRIC校准设置下的结果。

4.1.3 选择性检验功效与选择性第一类错误

按照第3.4节的定义评估选择性检验功效和选择性第一类错误，以量化在变量选择的情况下检测非零效应并控制错误拒绝的能力。

总体而言，明确考虑选择步骤的方法，在选择性检验功效与第一类错误控制之间实现了显著更好的平衡，相较于朴素选择后方法表现更优。特别是，去偏Lasso在保持较高检验效能的同时，使选择性I型错误率接近名义水平；而样本拆分尽管具有合理的检验效能，却表现出膨胀的I型错误率。相比之下，朴素重新拟合和非选择性程序往往较为宽松。

图3报告了在样本量为 $n = 75$时，针对玩具设计和METABRIC设计，在多个模拟情景下取平均得到的选择性检验功效和选择性I型错误率。跨情景取平均可减少特定设置所造成的变异，并突出不同推断程序之间的系统性差异。

样本量更大（$n = 100$ 和 $n = 200$）、采用替代调优规则（$\lambda_{\mathrm{CV,min}}$、 $\lambda_{\mathrm{CV,1se}}$、 λ_{fix}、 λ_{AIC}、λ_{BIC}）以及不同目标删失比例下的结果，见补充材料S2.3。在这一更广泛的设置范围内，定性结论仍然保持稳定。

增大样本量可提高选择性检验功效，并降低I类错误率的变异性，同时不改变各方法的相对排名。正则化程度较低的调优策略，例如 $\lambda_{\mathrm{CV,\ min}}$ 或 λ_{fix}规则，可带来更高的选择性检验功效，但其第一类错误控制的变异性也更大，尤其是在

METABRIC设计下。更保守的选择能以降低检验效能为代价提高稳定性。删失程度越高，选择性检验功效就越低，且不同方法之间的差异也越明显。

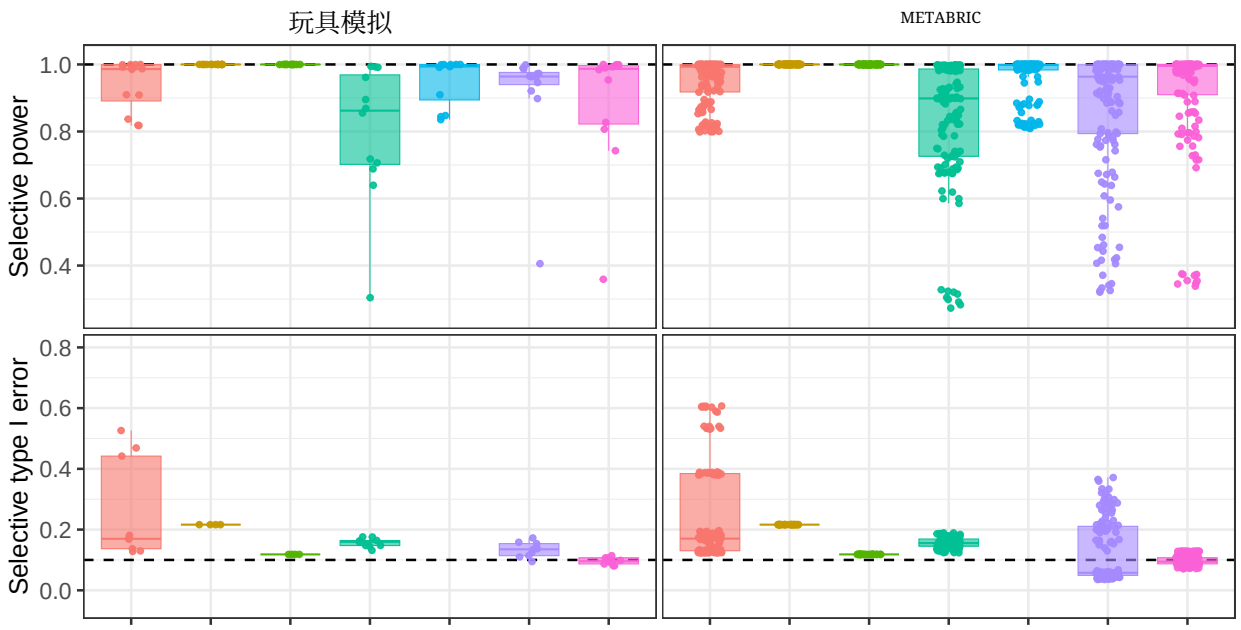

图3：样本量为 $n = 75$ 时，玩具模拟和 METABRIC 设置下的选择性检验功效（上排）和选择性I型错误率（下排）。虚线表示名义I型错误水平。

4.2 性能指标

选定模型的预测性能作为次要结局进行了评估。结果使用积分布里尔评分（IBS）进行汇总，数值越低表示预测准确性越高。与选择相关的指标（如平均模型规模，以及选定集合中真正活跃变量的比例（P_{true}））与 IBS 一并报告。一致性指数（ C 指数）的相应结果见补充材料 S2.5。

在各种模拟设置下，预测导向的调参规则，尤其是AIC型准则和$\lambda_{\text{CV, min}}$，最常获得最低的积分布里尔评分。然而，这些规则选出的模型规模显著更大，且 P_{true}明显更低，尤其是在 $p = 50$方面。

较为简约的调参规则，例如$\lambda_{\text{CV, 1SE}}$ 和BIC型准则 λ_{BIC}，选出的模型规模明显更小，且 P_{true}的值更高，同时仅以适度损失预测性能为代价。总体而言，这些结果体现了预测准确性与稀疏性之间预期的权衡：以预测为优化目标的调优策略往往倾向于稠密模型，而以选择为导向的规则则能产生更易解释且选择质量更高的解。

4.3 主要结果总结

在所有模拟情景中，明确考虑变量选择步骤的推断程序，相较于朴素选择后方法，对选择性覆盖率和选择性第一类错误的控制有显著改善。去偏Lasso和样本拆分在广泛的样本量、调优策略和删失比例范围内，覆盖率接近名义水平。相比之下，精确PSI对正则化参数的数据驱动调优表现出明显敏感性，尤其是在交叉验证下。

就推断效率而言，去偏Lasso始终产生比样本拆分和精确PSI更短的选择性置信区间，而精确PSI因以复杂选择事件为条件而导致选择性置信区间大幅膨胀。各方法之间的差异随样本量增加而缩小，但在中等维数设定中仍然可见。

因此，当经验错误率接近名义水平时，称I类错误“得到控制”；当观察到系统性过度拒绝时，称其“膨胀”；当拒绝率始终低于名义水平时，称其“保守”。选择性检验功效和选择性第一类错误之间存在明显且因方法而异的权衡，这些权衡在表5中作了定性总结。在该总结中，检验功效相对于预言机基准被描述为高或中等，而I类错误的表现则根据其相对于名义水平的系统性偏差，归类为得到控制、膨胀或保守。

表5：各方法和调优策略下选择性推断性能的定性总结。

方法	调优类别	选择性覆盖率	检验效能	I类错误
拆分	固定	接近名义水平	中等	膨胀
	CV	接近名义水平	中等	膨胀
去偏	固定	接近名义水平	high	受控
	CV	接近名义水平	high	受控
精确PSI	固定	低于名义水平	low	保守
	CV	低于名义水平	low	保守

以积分布里尔评分（IBS）衡量的预测性能，对推断程序和调参策略的选择相对不敏感。不同方法的预测准确性总体保持相对稳定，即使推断性质存在显著差异亦是如此；与预言模型相比，效率损失也较为适中。总体而言，结果凸显了变量选择后 Cox模型中推断有效性与预测性能之间的明显区别。

5 真实数据示例

我们使用乳腺癌分子分类国际联盟（METABRIC）研究的数据来阐释所提出的选择性推断框架 [31], ，该研究包含一个具有长期临床随访的大型乳腺癌队列。临床数据取自 cBioPortal。为确保分析的完整可复现性，仅使用了公开可获取的临床变量。未访问任何受控访问权限的分子数据或原始基因组数据。

5.1 研究问题

METABRIC数据示例旨在说明临床相关的右删失生存情境中的变量选择与选择后推断。该分析的重点在于方法论而非病因学：其目标不是进行因果解释或生物标志物发现，而是基于常规可获取的临床协变量开展预后建模。

具体而言，我们旨在考察：(i) 不同基于 Lasso 的方法和调优策略下变量选择的稳定性；(ii) 针对所选协变量得到的选择性置信区间（SCIs）；以及 (iii) 不同选择后推断方法如何在数据自适应建模流程中量化不确定性。

总生存时间（以月为单位）定义为 `overall_survival_months`，事件指标根据 `patients_vital_status`（死亡或删失）得出。我们纳入了常规可获得的临床预测变量，包括诊断年龄、肿瘤分期、ER状态、HER2状态、PR状态、化疗、激素治疗、放射治疗、组织学分级、肿瘤大小、阳性淋巴结数量以及诺丁汉预后指数（NPI）。

5.2 分析

采用与模拟研究相同的推断程序对METABRIC数据集进行了分析，包括fullCox模型、样本拆分、去偏Lasso和精确PSI。

为了评估变量选择的稳定性，我们采用了类似于Kammer等人 [16]的方法。具体而言，我们进行了100次子抽样重复。在每次重复中，重新运行变量选择步骤，并将选择频率计算为每个协变量在被选中的子样本中所占的比例。

此外，我们针对每个子样本和每种推断方法重新计算了置信区间。这样便可以综合评估不同方法下选择稳定性和不确定性量化，而不必依赖选定模型的一次单独实现。

除非另有说明，结果均针对采用交叉验证调参选择的非自适应Lasso报告$\lambda_{\mathrm{CV,\ min}}$，这体现了一种常见的实际分析流程。其他调参规则（$\lambda_{\mathrm{AIC}}$, $\lambda_{\mathrm{CV,\ 1SE}}$）呈现出定性上相似的模式，相关结果见补充材料。

5.3 结果

图4汇总了METABRIC数据的回归系数估计值及相应的90%选择性置信区间。图示布局沿用了Kammer等人 [16]提出的格式。结果展示的是交叉验证调参选择$\lambda_{\mathrm{CV,\ min}}$；对应的$\lambda_{\mathrm{CV,\ 1SE}}$ 和基于AIC的调参规则图示在定性上非常相似，因此为简洁起见不予列出。

不同推断方法的点估计总体上较为接近，但选择性置信区间的宽度和稳定性差异显著。样本拆分和精确PSI通常会产生较宽的区间，这反映了对变量选择步骤的明确调整。相比之下，去偏Lasso通常得到相对较窄的区间，但其在不同子样本间的变异性也更大。

更深入的考察表明，这些模式很大程度上取决于协变量类型。对于连续预测变量，去偏Lasso通常能提供更精确的选择性推断，从而得到相对较窄的置信区间。然而，对于二元或有序协变量，去偏Lasso会出现明显的区间膨胀，导致置信区间显著变宽。精确PSI则表现相反：与去偏Lasso相比，它对连续协变量给出的区间更宽，但对二元和有序预测变量则能得到更稳定、更易解释的选择性置信区间。

选择频率进一步印证了这一区别。肿瘤大小和肿瘤分期等核心临床预测变量会在不同子样本和推断程序中持续被选中。相比之下，化疗或放射治疗等治疗相关协变量的表现出明显的方法依赖性变异，这既反映了数据中的相关结构，也体现了不同方法在将选择不确定性传递至推断时的差异。

总体而言，METABRIC示例说明，在现实的临床场景中，如何运用选择性推断区分稳定与不稳定的预后信号。与孤立地解读各个置信区间相比，综合考量选择频率和选择性置信区间，能够透明地概括数据自适应建模流程中的变量重要性和推断不确定性。

6 讨论

在本研究中，我们对应用于 Cox比例风险模型中 Lasso 类型变量选择后的选择性推断程序进行了中立的比较研究。我们的目标并非找出普遍最优的方法，而是阐明在生存分析中，将数据驱动的变量选择与调参结合起来进行选择后不确定性量化时所涉及的推断保证、效率权衡和实际敏感性。

模拟情景中的一项核心发现是，变量选择后的有效不确定性量化，关键在于使推断目标与预期的分析目标相一致。在许多实际应用场景中，尤其是在探索性或预后建模中，关注的重点是选定子模型的系数，而非恢复某个唯一的潜在“真实”完整模型。从这种以子模型为导向的视角来看，选择性推断颇具吸引力，因为它明确考虑了变量筛选所带来的乐观偏差，并有助于以更可重复的方式报告不确定性。

我们的模拟结果表明，与朴素的选择后重新拟合相比，明确调整选择步骤的推断程序能够更有效地控制选择性覆盖率和选择性第一类错误。具体而言，在广泛的样本量、删失水平和调优策略下，去偏Lasso和样本拆分所实现的选择性覆盖率都接近名义水平。相比之下，精确PSI对数据驱动的调优选择表现出明显的敏感性，尤其是在正则化

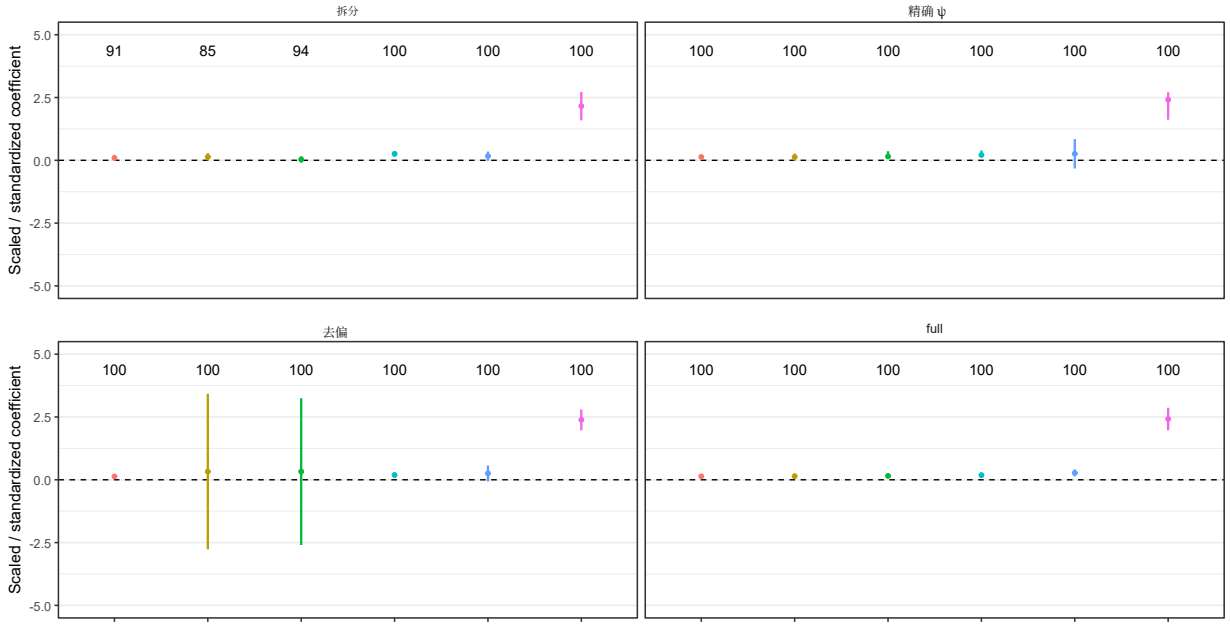

图4：真实数据示例 METABRIC：采用不同推断方法得到的回归系数点估计和90%选择性置信区间。结果针对交叉验证调参选择$\lambda_{\mathrm{CV,\ min}}$展示。系数按原始尺度显示，并按标准化效应大小递增排序。各面板上方的数字表示100个子样本中的选择频率（%）。

参数通过交叉验证选定时。这一现象反映出，精确PSI依赖于对固定选择事件进行条件化；当 λ 是通过自适应方式选定时，这一假设便不再成立。

与此同时，我们的结果也表明，选择调整并非没有代价。对选择事件进行强条件化的方法往往会产生较宽的选择性置信区间，这反映出显著的不确定性膨胀。尤其是，精确选择性推断在大多数情景下得到的区间最宽；而去偏Lasso在偏差校正与方差之间取得了更好的平衡，因此得到的区间相对较短且稳定。样本拆分在概念上简单且稳健，但会降低有效样本量，进而导致区间变宽，在中度至高度删失的情景下尤为明显。

正则化参数 λ 的调参策略是影响实际表现的关键因素。预测导向的规则（如 $\lambda_{\min}$ ）或 AIC 类准则 λ_{AIC} 通常会选择更大的模型，从而提高选择性检验功效，但也会增加选择性 I 类错误控制的变异性。较保守的选择（如 1-SE 规则 $\lambda_{1\mathrm{SE}}$ 或 BIC 类准则 λ_{BIC}）会倾向于选择更稀疏的模型，且往往能稳定错误率，但代价是检验效能降低，并且在某些情况下会出现轻微的覆盖不足。重要的是，这些权衡在不同样本量和删失比例下在定性上保持稳定，说明它们反映的是这些方法的系统性特征，而非特定模拟设置造成的假象。值得注意的是，这些推断行为上的显著差异在预测性能上仅有微弱体现，因为不同调参策略下的 IBS 值相对稳定。

删失会进一步放大这些差异。正如预期，随着删失比例增加，有效信息量减少，所有方法的选择性检验功效都会降低。同时，删失会加剧方法特异性敏感性，尤其是对于依赖强条件化或近似方法的程序

方差估计。这一观察凸显了在现实的删失机制下评估选择性推断方法的重要性，因为这类机制在生物医学生存研究中十分常见。
基于 METABRIC队列的真实数据分析，在具有临床相关性的场景中展示了这些方法学发现。该分析并不关注因果解释，而是强调如何利用选择性推断区分稳定的预后信号与较弱或不稳定的关联。将选择频率与选择性置信区间结合起来，可以透明地概括变量重要性和不确定性，并避免对数据自适应建模流程所产生的效应作出过度自信的解读。在此背景下，推断方法之间的差异最好理解为反映了不同的使用场景：当控制假发现至关重要时，保守方法可能更为可取；而在检验效能和区间宽度十分关键的探索性场景中，更有效率的方法则可能更具吸引力。值得注意的是，真实数据分析还表明，选择性推断方法的相对表现可能取决于协变量类型；连续预测变量与二元或有序预测变量之间存在差异。

我们的研究存在若干局限。首先，模拟设计仅涵盖有限的一组系数模式和基线风险函数设定。虽然我们改变了样本量、删失、相关结构和调优策略，但未考虑时变效应、交互作用或非比例风险等更复杂的特征。其次，我们采用子模型视角来评估推断。其他估计目标，例如完整模型参数或模型设定错误时的目标总体参数，可能会导致对这些方法相对优劣得出不同结论。第三，实施细节也很重要：为确保可比性，我们采用了软件默认选项，但实际分析可能会受益于针对具体问题对方差估计、数值容差或调优策略进行校准。

尽管存在这些局限性，定性结论在广泛情景中的一致性仍令人确信，主要见解可以推广到现实分析流程。总体而言，我们的结果强调，选择性推断可以大幅提高Cox模型中选择后结论的可信度，但应将其视为更广泛建模策略的一部分，该策略包括仔细调参、敏感性分析以及对预期推断目标的明确说明。

利益冲突

作者声明不存在利益冲突。

资助

衷心感谢德国研究联合会（项目编号 FR 4121/2-2）的资助。

数据可用性

本研究分析的所有数据均可从 cBioPortal [31] 公开获取。

代码可用性

用于模拟研究和数据分析的完整代码已在 `https://github.com/lena-222/selective-inference-cox` 公开。

稿件准备过程中生成式人工智能和人工智能辅助技术的声明

在准备本文期间，作者仅使用 ChatGPT 来改进稿件的语言和清晰度。使用该工具后，作者根据需要审阅并编辑了内容，并对已发表文章的内容承担全部责任。

作者贡献

L.S. 构思了这项研究，实施了所有方法，完成了分析并撰写了稿件。S.F.-W. 参与了方法学开发，并提供了指导、同行评审和批判性反馈。所有作者均阅读并批准了最终稿件。

参考文献

[1] Robert Tibshirani. 通过Lasso进行回归收缩与变量选择。皇家统计学会会刊*B*辑：统计方法，58(1):267–288，1996。[2] Hui Zou. 自适应Lasso及其预言机性质。美国统计学会会刊，101(476):1418–1429，2006。[3] Richard Berk、Lawrence Brown、Andreas Buja、Kai Zhang和Linda Zhao. 有效的选择后推断。统计年鉴，41(2):802–837，2013。[4] Jason D Lee、Dennis L Sun、Yuekai Sun和Jonathan E Taylor. 精确选择后推断及其在Lasso中的应用。统计年鉴，44(3):907–927，2016。[5] Per Kragh Andersen、Ørnulf Borgan、Richard D Gill和Niels Keiding. 模型设定与删失。载于基于计数过程的统计模型，第3章，第121–175页。施普林格，纽约，1993。[6] Robert Tibshirani. Cox模型中用于变量选择的Lasso方法。医学统计学，16(4):385–395，1997。[7] Hao Helen Zhang和Wenbin Lu. Cox比例风险模型的自适应Lasso。生物统计学，94(3):691–703，2007。[8] David R Cox. 关于用于评估显著性水平的数据拆分的说明。生物统计学，62(2):441–444，1975。[9] Jonathan Taylor和Robert J Tibshirani. 统计学习与选择性推断。美国国家科学院院刊，112(25):7629–7634，2015。[10] Yi Yu、Jelena Bradic和Richard J Samworth. 高维Cox模型的置信区间。中华统计学志，31(1):243–267，2021。[11] Shengchun Kong、Zhuqing Yu、Xianyang Zhang和Guang Cheng. 使用去稀疏化Lasso对Cox回归模型进行高维稳健推断。斯堪的纳维亚统计学杂志，48(3):1068–1095，2021。[12] William Fithian、Dennis Sun和Jonathan Taylor. 模型选择后的最优推断，2017。[13] Jonathan Taylor和Robert Tibshirani. ℓ_1惩罚似然模型的选择后推断。加拿大统计学杂志，46(1):41–61，2018。[14] Cun-Hui Zhang和Stephanie S Zhang. 高维线性模型中低维参数的置信区间。皇家统计学会会刊*B*辑：统计方法，76(1):217–242，2014。[15] Sara van de Geer、Peter Bühlmann、Ya'acov Ritov和Ruben Dezeure. 关于高维模型的渐近最优置信域与检验。统计年鉴，42(3):1166–1202，2014。[16] Michael Kammer、Daniela Dunkler、Stefan Michiels和Georg Heinze. 评估生物医学研究中Lasso选择性推断方法：一项比较模拟研究。*BMC*医学研究方法学，22(1):206，2022。[17] David R Cox. 回归模型与生命表。皇家统计学会会刊：*B*辑（方法论），34(2):187–220，1972。[18] Jianqing Fan和Runze Li. Cox比例风险模型与脆弱性模型的变量选择。统计年鉴，30(1):74–99，2002。[19] Zaixiang Tang、Yueping Shen、Xinyan Zhang和Nengjun Yi. 用于生存预测和相关基因检测的尖峰与板块Lasso Cox模型。生物信息学，33(18):2799–2807，2017。

[20] Liwei Wang、Yu Chang、Jinfeng Ma、Wenqing Qu和Yifan Li. 针对受“双重侵犯”影响的原发性胃癌，识别可延长无进展生存期的高风险候选者：一种利用Lasso-Cox回归的分析方法。BMC Cancer，25(1):381，2025。[21] Martin J. Wainwright. 使用 ℓ_1约束二次规划（Lasso）实现高维及含噪稀疏性恢复的精确阈值。IEEE信息论汇刊，55(5):2183–2202，2009。[22] Trevor Hastie、Robert Tibshirani和Martin Wainwright.《稀疏统计学习：Lasso及其推广》。Chapman & Hall/CRC，博卡拉顿，佛罗里达州，2015。Chapman & Hall/CRC统计学与应用概率专著丛书。[23] Hannes Leeb和Benedikt M. Pötscher. 模型选择与推断：事实与虚构。计量经济学理论，21(1):21–59，2005。[24] Hannes Leeb和Benedikt M. Pötscher. 能否估计模型选择后估计量的条件分布？统计年鉴，34(5):2554–2591，2006。[25] Terry M. Therneau和Patricia M. Grambsch.《生存数据建模：拓展Cox模型》。施普林格，纽约，2000。[26] Xiaoying Tian和Jonathan Taylor. 随机化响应下的选择性推断。统计年鉴，46(2):679–710，2018。[27] Lu Xia、Bin Nan和Yi Li. 对协变量数量不断增长的Cox比例风险模型进行统计推断。斯堪的纳维亚统计学杂志，50(2):550–571，2023。[28] Tim P. Morris、Ian R. White和Michael J. Crowther. 使用模拟研究评估统计方法。医学统计学，38(11):2074–2102，2019。[29] P. L. Ramos、D. C. F. Guzman、A. L. Mota、D. A. Saavedra、F. A. Rodrigues和F. Louzada. 删失数据抽样：实用指南。统计计算与模拟杂志，94(18):4072–4106，2024。[30] P. K. Andersen和R. D. Gill. 计数过程的Cox回归模型：大样本研究。统计年鉴，10(4):1100–1120，1982。[31] Christina Curtis、Sohrab P. Shah、Suet-Feung Chin、Gulisa Turashvili、Oscar M. Rueda、Mark J. Dunning、Doug Speed、Andy G. Lynch、Shamith Samarajiwa、Yinyin Yuan、Stefan Gräf、Gavin Ha、Gholamreza Haffari、Ali Bashashati、Roslin Russell、Steven McKinney、Anita Langerød、Andrew Green、Elena Provenzano、Gordon Wishart、Sarah Pinder、Peter Watson、Florian Markowetz、Leigh Murphy、Ian Ellis、Arnie Purushotham、Anne-Lise Børresen-Dale、James D. Brenton、Simon Tavaré、Samuel Aparicio和Carlos Caldas. 2,000例乳腺肿瘤的基因组与转录组结构揭示了新的亚组。《自然》，486(7403):346–352，2012。[32] Christina Sauer、F. Julian D. Lange、Maria Thurow、Ina Dormuth和Anne-Laure Boulesteix. 基于真实数据的统计参数模拟研究，2025。[33] Ryan Tibshirani、Rob Tibshirani、Jonathan Taylor、Joshua Loftus、Stephen Reid和Jelena Markovic. selectiveInference：选择后推断工具，2019。R软件包版本1.2.5。[34] Trevor Hastie、Robert Tibshirani和Jerome Friedman.《统计学习基础》。施普林格，纽约，NY，美国，第二版，2009。[35] Hirotugu Akaike. 统计模型识别的新视角。IEEE自动控制汇刊，19(6):716–723，1974。[36] Gideon Schwarz. 估计模型的维数。统计年鉴，6(2):461–464，1978。[37] Jianqing Fan和Runze Li. 通过非凹惩罚似然进行变量选择及其预言机性质。美国统计学会会刊，96(456):1348–1360，2001。[38] Paul Blanche、Michael W. Kattan和Thomas A. Gerds. C指数不适用于评估t年预测风险。生物统计学，20(2):347–357，2019。[39] Thomas A. Gerds和Martin Schumacher. 在含右删失事件时间的一般生存模型中，对期望布里尔评分进行一致估计。生物计量学杂志，48(6):1029–1040，2006。[40] Terry M Therneau. R中的生存分析软件包，2024。R软件包版本3.7-0。[41] Jerome Friedman、Trevor Hastie、Robert Tibshirani、Balasubramanian Narasimhan、Kenneth Tay、Noah Simon和Junyang Qian. glmnet：Lasso和弹性网正则化广义线性模型，2021。R软件包版本4.1-1。

[42] Jerome Friedman、Trevor Hastie 和 Robert Tibshirani。通过坐标下降求解广义线性模型的正则化路径。统计软件杂志，33(1):1–22，2010。[43] Noah Simon、Jerome Friedman、Trevor Hastie 和 Robert Tibshirani。通过坐标下降求解 Cox比例风险模型的正则化路径。统计软件杂志，39(5):1–13，2011。

参考文献

[1] Robert Tibshirani。《通过Lasso进行回归收缩与变量选择》。皇家统计学会会刊*B*辑：统计方法，58(1):267–288，1996。[2] Hui Zou。《自适应Lasso及其预言机性质》。美国统计学会会刊，101(476):1418–1429，2006。[3] Richard Berk、Lawrence Brown、Andreas Buja、Kai Zhang和Linda Zhao。《有效的选择后推断》。统计年鉴，41(2):802–837，2013。[4] Jason D Lee、Dennis L Sun、Yuekai Sun和Jonathan E Taylor。《精确选择后推断及其在Lasso中的应用》。统计年鉴，44(3):907–927，2016。[5] Per Kragh Andersen、Ørnulf Borgan、Richard D Gill和Niels Keiding。《模型设定与删失》。载于基于计数过程的统计模型，第3章，第121–175页。施普林格，纽约，1993。[6] Robert Tibshirani。《用于Cox模型变量选择的Lasso方法》。医学统计学，16(4):385–395，1997。[7] Hao Helen Zhang和Wenbin Lu。《Cox比例风险模型的自适应Lasso》。生物统计学，94(3):691–703，2007。[8] David R Cox。《关于用于评估显著性水平的数据拆分的说明》。生物统计学，62(2):441–444，1975。[9] Jonathan Taylor和Robert J Tibshirani。《统计学习与选择性推断》。美国国家科学院院刊，112(25):7629–7634，2015。[10] Yi Yu、Jelena Bradic和Richard J Samworth。《高维Cox模型的置信区间》。中华统计学志，31(1):243–267，2021。[11] Shengchun Kong、Zhuqing Yu、Xianyang Zhang和Guang Cheng。《使用去稀疏化Lasso对Cox回归模型进行高维稳健推断》。斯堪的纳维亚统计学杂志，48(3):1068–1095，2021。[12] William Fithian、Dennis Sun和Jonathan Taylor。《模型选择后的最优推断》，2017。[13] Jonathan Taylor和Robert Tibshirani。《 ℓ_1惩罚似然模型的选择后推断》。加拿大统计学杂志，46(1):41–61，2018。[14] Cun-Hui Zhang和Stephanie S Zhang。《高维线性模型中低维参数的置信区间》。皇家统计学会会刊*B*辑：统计方法，76(1):217–242，2014。[15] Sara van de Geer、Peter Bühlmann、Ya'acov Ritov和Ruben Dezeure。《关于高维模型的渐近最优置信域与检验》。统计年鉴，42(3):1166–1202，2014。[16] Michael Kammer、Daniela Dunkler、Stefan Michiels和Georg Heinze。《评估生物医学研究中Lasso选择性推断方法：一项比较模拟研究》。*BMC*医学研究方法学，22(1):206，2022。[17] David R Cox。《回归模型与生命表》。皇家统计学会会刊：*B*辑（方法学），34(2):187–220，1972。[18] Jianqing Fan和Runze Li。《Cox比例风险模型与脆弱性模型的变量选择》。统计年鉴，30(1):74–99，2002。[19] Zaixiang Tang、Yueping Shen、Xinyan Zhang和Nengjun Yi。《用于生存预测和相关基因检测的尖峰与板块Lasso Cox模型》。生物信息学，33(18):2799–2807，2017。[20] Liwei Wang、Yu Chang、Jinfeng Ma、Wenqing Qu和Yifan Li。《针对"双重侵犯"的原发性胃癌：采用Lasso-Cox回归的分析方法识别可延长无进展生存期的高风险候选者》。*BMC Cancer*，25(1):381，2025。

[21] Martin J. Wainwright。《利用 ℓ_1约束二次规划（Lasso）实现高维含噪稀疏恢复的精确阈值》。IEEE信息论汇刊，55(5):2183–2202，2009。[22] Trevor Hastie、Robert Tibshirani和Martin Wainwright。《稀疏统计学习：Lasso与推广》。Chapman & Hall/CRC，博卡拉顿，佛罗里达州，2015。Chapman & Hall/CRC统计学与应用概率专著丛书。[23] Hannes Leeb和Benedikt M. Pötscher。《模型选择与推断：事实与虚构》。计量经济学理论，21(1):21–59，2005。[24] Hannes Leeb和Benedikt M. Pötscher。《能否估计模型选择后估计量的条件分布？》统计年鉴，34(5):2554–2591，2006。[25] Terry M. Therneau和Patricia M. Grambsch。《生存数据建模：扩展Cox模型》。施普林格，纽约，2000。[26] Xiaoying Tian和Jonathan Taylor。《随机化响应下的选择性推断》。统计年鉴，46(2):679–710，2018。[27] Lu Xia、Bin Nan和Yi Li。《对协变量数量不断增加的Cox比例风险模型进行统计推断》。斯堪的纳维亚统计学杂志，50(2):550–571，2023。[28] Tim P. Morris、Ian R. White和Michael J. Crowther。《使用模拟研究评估统计方法》。医学统计学，38(11):2074–2102，2019。[29] P. L. Ramos、D. C. F. Guzman、A. L. Mota、D. A. Saavedra、F. A. Rodrigues和F. Louzada。《删失数据抽样：实用指南》。统计计算与模拟杂志，94(18):4072–4106，2024。[30] P. K. Andersen和R. D. Gill。《用于计数过程的Cox回归模型：大样本研究》。统计年鉴，10(4):1100–1120，1982。[31] Christina Curtis、Sohrab P. Shah、Suet-Feung Chin、Gulisa Turashvili、Oscar M. Rueda、Mark J. Dunning、Doug Speed、Andy G. Lynch、Shamith Samarajiwa、Yinyin Yuan、Stefan Gräf、Gavin Ha、Gholamreza Haffari、Ali Bashashati、Roslin Russell、Steven McKinney、Anita Langerød、Andrew Green、Elena Provenzano、Gordon Wishart、Sarah Pinder、Peter Watson、Florian Markowetz、Leigh Murphy、Ian Ellis、Arnie Purushotham、Anne-Lise Børresen-Dale、James D. Brenton、Simon Tavaré、Samuel Aparicio和Carlos Caldas。《2000例乳腺肿瘤的基因组与转录组结构揭示了新的亚组》。《自然》，486(7403):346–352，2012。[32] Christina Sauer、F. Julian D. Lange、Maria Thurow、Ina Dormuth和Anne-Laure Boulesteix。《基于真实数据的统计参数模拟研究》，2025。[33] Ryan Tibshirani、Rob Tibshirani、Jonathan Taylor、Joshua Loftus、Stephen Reid和Jelena Markovic。《selectiveInference：选择后推断工具》，2019。R软件包版本1.2.5。[34] Trevor Hastie、Robert Tibshirani和Jerome Friedman。《统计学习基础》。施普林格，纽约，纽约州，美国，第2版，2009。[35] Hirotugu Akaike。《重新审视统计模型识别》。IEEE自动控制汇刊，19(6):716–723，1974。[36] Gideon Schwarz。《估计模型的维数》。统计年鉴，6(2):461–464，1978。[37] Jianqing Fan和Runze Li。《通过非凹惩罚似然进行变量选择及其预言机性质》。美国统计学会会刊，96(456):1348–1360，2001。[38] Paul Blanche、Michael W. Kattan和Thomas A. Gerds。《C指数并不适用于评估t年预测风险》。生物统计学，20(2):347–357，2019。[39] Thomas A. Gerds和Martin Schumacher。《在右删失事件时间的一般生存模型中一致估计期望布里尔评分》。生物计量学杂志，48(6):1029–1040，2006。[40] Terry M Therneau。《R中的生存分析软件包》，2024。R软件包版本3.7-0。[41] Jerome Friedman、Trevor Hastie、Robert Tibshirani、Balasubramanian Narasimhan、Kenneth Tay、Noah Simon和Junyang Qian。《glmnet：正则化广义线性模型的Lasso与弹性网》，2021。R软件包版本4.1-1。[42] Jerome Friedman、Trevor Hastie和Robert Tibshirani。《通过坐标下降求解广义线性模型的正则化路径》。统计软件杂志，33(1):1–22，2010。[43] Noah Simon、Jerome Friedman、Trevor Hastie和Robert Tibshirani。《通过坐标下降求解Cox比例风险模型的正则化路径》。统计软件杂志，39(5):1–13，2011。