基于对数累积风险比轨迹的获益时间估计方法研究*

樊　静[1]　蔡嘉鑫[2]　索珊珊[1]　查景丽[1]　李　超[1△]　邵　方[3△]

【提　要】 **目的**　传统获益时间（time to benefit，TTB）估计多基于序贯截断（sequential truncation）的 Cox 比例风险模型风险比（hazard ratio，HR），但存在反复截断数据导致统计性质不明确、结果易被误解为时变效应及计算量大等问题。本研究旨在建立一种方法学更严谨、计算效率更高的 TTB 估计框架。**方法**　提出基于对数累积风险比时间轨迹的 TTB 估计新框架。整合 Weibull 回归模型与 Kaplan-Meier（KM）估计，并通过模拟研究及真实临床试验数据，系统比较新框架与传统序贯截断 Cox 模型在 TTB 估计中的性能。**结果**　模拟研究表明，新框架中 Weibull 回归模型综合性能最优，估计精度高且稳定，均方根误差（root mean square error，RMSE）始终小于 1。相比之下，序贯截断 Cox 模型在部分情景中对 TTB 持续显著时间点估计的 RMSE 超过 10。此外，Weibull 回归模型计算速度较序贯截断 Cox 模型提升约 35~250 倍。**结论**　基于对数累积风险比时间轨迹的 TTB 估计框架在避免序贯截断方法缺陷的同时，为治疗效应时间演变的稳健评估提供了统一且可靠的方法。

【关键词】 获益时间　生存分析　Weibull 回归模型　Kaplan-Meier 估计　Cox 比例风险模型
【中图分类号】 R195.1　　【文献标志码】 A　　**DOI**　10.11783/j.issn.1002-3674.2026.05.004

Study on Time to Benefit Estimation Method based on Log Cumulative Hazard Ratio Trajectory

Fan Jing, Cai Jiaxin, Suo Shanshan, et al (*Department of Epidemiology and Health Statistics, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an* 710061)

【**Abstract**】 **Objective**　Conventional estimation of the Time to Benefit (TTB) typically relies on Hazard Ratios (HR) derived from sequentially truncated Cox proportional hazards models. However, this approach is limited by poorly defined statistical properties due to repetitive data truncation, potential misinterpretation of results as time-varying effects, and high computational intensity. This study aims to establish a more methodologically rigorous and computationally efficient framework for TTB estimation. **Methods**　We propose a novel TTB estimation framework based on the temporal trajectory of the log-cumulative hazard ratio. By integrating Weibull regression models with Kaplan-Meier (KM) estimation, we systematically compared the performance of the proposed framework against the traditional sequentially truncated Cox model using both simulation studies and real-world clinical trial data. **Results**　Simulation results demonstrate that the Weibull regression model within the new framework yields superior overall performance, characterized by high estimation precision and stability; its root mean square error (RMSE) remained consistently below 1. In contrast, the RMSE of the sequentially truncated Cox model exceeded 10 in certain scenarios when estimating the time point of sustained significance for TTB. Furthermore, the computational speed of the Weibull regression model was approximately 35 to 250 times faster than that of the sequentially truncated Cox model. **Conclusion**　The TTB estimation framework based on the temporal trajectory of the log-cumulative hazard ratio effectively addresses the inherent flaws of sequential truncation methods. It provides a unified and reliable methodology for the robust assessment of the temporal evolution of treatment effects.

【**Key words**】 Time to benefit; Survival analysis; Weibull regression model; Kaplan-Meier estimation; Cox proportional hazards model

获益时间（time to benefit，TTB），其定义为患者接受干预后出现统计学显著获益的时间点，是生存分析中评价治疗效应的重要指标[1]。与传统的单一终点效应量相比，TTB 能够直观刻画干预效应随时间演变的动态特征，具有更强的临床可解释性。在肿瘤学、心血管疾病等需要长期随访的慢性病研究领域，TTB 对于精准评估治疗价值、优化临床决策具有重要意义。

尽管 TTB 的临床应用价值已得到广泛认可，但其统计估计方法仍缺乏统一标准[1-4]，在模型选择、效应量度量等方面存在显著差异[5-7]。目前一种常用方法是在 Cox 比例风险模型框架下，通过对生存数据进行序贯截断（sequential truncation），迭代计算各时间点的风险比（hazard ratio，HR）以估计 TTB[5]。然而，该方法存在两个核心的方法学缺陷：第一，逐点估计的 HR 容易被误解为时变效应（time-varying effect）[8]，但其本质并非基于时变系数模型的标准估计；第二，该方法对数据的反复截断与重复建模缺乏严格的统计理论基础，估计结果的无偏性、一致性等统计性质难以保证，稳健性存疑。因此，迫切需要建立方法学上更

* 基金项目：国家自然科学基金（82473732；82322060）；四大慢病重大专项（2024ZD0522700）；南京医科大学南京公共卫生研究院科研创新项目（NCX2405）；煤炭环境致病与防治教育部重点实验室开放课题基金项目（MEKLCEPP/SXMU-2025MEKLCEPP/SX MU-202501）
1. 西安交通大学医学部公共卫生学院流行病与卫生统计学教研室（710061）
2. 山西大学复杂系统研究所
3. 南京医科大学公共卫生学院生物统计学系
△ 通信作者：邵方，E-mail：shaofang@njmu. edu. cn；李超，E-mail：lcxjtu_research@163. com

为严谨、理论基础更为扎实的TTB估计框架。

针对上述方法学缺陷，本研究提出了一种基于效应轨迹的TTB估计新框架。该框架整合成熟的统计模型，包括参数化的Weibull回归模型[9]和非参数的Kaplan-Meier估计[10]，通过一次性整体建模估计两组间效应的差异函数（如对数累积风险比，即对数累积风险差异）。在此框架下，TTB的点估计及显著性区间边界估计可由差异函数的估计曲线及其逐点置信区间与零效应基线的交点直接确定[5]。该方法的核心优势在于：它基于完整的生存数据集进行统一的统计推断，从根本上规避了传统方法中对数据反复截断、多重建模带来的方法学问题。由于估计过程依托成熟的统计理论，相较于传统序贯截断Cox模型，本框架在处理大规模临床试验数据时，可显著提升计算效率，并在高删失率情况下仍能保持统计推断的稳健性。

基于上述新框架，本研究将系统性比较三种生存分析方法在TTB估计中的性能表现：参数化Weibull回归模型、非参数Kaplan-Meier（KM）估计以及传统框架下的半参数Cox比例风险模型。通过模拟研究，在不同样本量和删失率组合的场景设置下，基于点估计和置信区间边界估计方法，系统评估各模型估计TTB的统计性质。此外，采用两项慢性病临床试验的真实数据对新框架的有效性和实用性进行实证分析。本研究旨在为TTB的定量评估提供一套方法学严谨、统计稳健且易于实施的分析工具，以促进其在慢性病长期干预效果评价中的规范化应用。

模型与方法

1. 获益时间及其估计框架

本研究将TTB定义为：干预效应量的点估计值或其置信区间上界首次或持续由上至下穿越零效应基线（即无效阈值0）的最早时间点。为克服传统TTB估计方法的缺陷并系统比较不同模型的估计性能，本研究构建了统一的TTB估计框架。首先，根据试验设计将研究对象分为两组，其中Z=0表示对照组，Z=1表示干预组。对于任意组别$z\in\{0,1\}$，其生存函数和累积风险函数分别记为$S_z(t)=P(T>t|Z=z)$和$H_z(t)=-\log S_z(t)$。在本框架下，我们选择两组的对数累计风险比作为核心效应度量：

$$D(t)=\Delta\log H(t)=\log\left[\frac{H_1(t)}{H_0(t)}\right]=\log H_1(t)-\log H_0(t) \tag{1}$$

在此定义下，$D(t)$<0表明在时间点t时，干预组相较于对照组具有更低的累积风险，即呈现有益效应。尽管效应量亦可采用生存函数或累积风险函数之差，但本研究选择对数累积风险比基于以下考量：该度量在数值解释和变化趋势上，与传统TTB估计方法中迭代计算的对数风险比更为接近，便于方法学比较。特别地，当比例风险（proportional hazards，PH）假设成立时，效应函数$D(t)$在理论上等同于恒定的对数风险比。因此，效应函数$D(t)$的引入不仅为整合不同生存模型（参数、非参数、半参数）提供了一个统一的比较框架，也为后续推导其估计量的抽样分布和构建置信区间奠定了理论基础。

2. 获益时间估计方法

基于上述框架，本研究从点估计和显著性区间边界估计两个层面比较新框架下Weibull回归模型、KM估计以及传统框架下Cox比例风险模型在TTB估计中的表现。点估计直接依据效应函数估计$\hat{D}(t)$首次或持续穿越零效应基线的时间来确定TTB。显著性区间边界估计则通过分析效应函数$D(t)$的逐点置信区间边界曲线与零效应基线的交点推断TTB的统计显著性范围。需要说明的是，该方法并非直接为TTB构建置信区间，而是利用效应函数的置信区间信息推断TTB的不确定性。考虑到研究早期数据较为稀疏，置信区间通常较宽，其下界可能持续低于零值，而难以获得稳定估计，本研究沿用已有文献中的保守策略[5]，将TTB的显著性区间边界估计定义为：效应函数$D(t)$的置信区间上界首次或持续性地由上至下穿越零效应基线的时间点[5]。综上，我们采用以下四个TTB估计量：

① 点估计：

$$\begin{cases} TTB_{\text{首次}}=\inf\left\{t:\hat{D}(t)<0\right\} \\ TTB_{\text{持续}}=\inf\left\{t:\hat{D}(s)<0,\forall s\geqslant t\right\} \end{cases} \tag{2}$$

② 显著性区间边界估计：

$$\begin{cases} TTB_{\text{区间首次}}=\inf\left\{t:CI_L(t)<0,CI_U(t)<0\right\} \\ TTB_{\text{区间持续}}=\inf\left\{t:CI_L(s)<0,CI_U(s)<0,\forall s\geqslant t\right\} \end{cases} \tag{3}$$

其中，“首次”与“持续”分别对应首次观察到获益及持续获益的最早时间点。$\hat{D}(t)$表示时间点t的对数风险比或对数累积风险比估计；$CI_L(t)$和$CI_U(t)$分别为t时刻对数风险比或对数累积风险比的95%置信区间的下界和上界。$D(t)$具体可由下述模型进行估计。

（1）Weibull回归模型

在上述框架下，可采用基于Weibull分布[11]的参数化生存分析回归模型进行效应量估计。将治疗组别Z作为分层（strata）因素纳入回归模型从而对不同组别分别设定Weibull分布的尺度参数，并通过极大似然估计（maximum likelihood estimation，MLE）[12]获得组别特异的形状参数（$\hat{k}_z$）与尺度参数（$\hat{\lambda}_z$），进而计

算累积风险函数$\hat{H}_z(t)$和效应差异函数$\hat{D}(t)$。具体公式如下：

$$\hat{H}_z(t) = \left(\frac{t}{\hat{\lambda}_z}\right)^{\hat{k}_z} \tag{4}$$

$$\hat{D}(t) = \widehat{\Delta\log H}(t) = \hat{k}_1(\log(t) - \log(\hat{\lambda}_1)) - \hat{k}_0(\log(t) - \log(\hat{\lambda}_0)) \tag{5}$$

为量化$\hat{D}(t)$的不确定性，引入Delta方法[13]进行方差估计。利用一阶泰勒展开将非线性函数变换$\widehat{\Delta\log H}(t)$的渐进抽样分布方差估计转化为其关于模型参数的梯度向量与协方差矩阵的矩阵计算，最终构建$D(t)$的逐点95%置信区间，以实现TTB的显著性区间边界估计。记Weibull回归模型参数向量为$\theta = (\beta_0, \beta_1, -\log(k_0), -\log(k_1))^T$，其中$\hat{k}_0 = \exp(-\gamma_0)$，$\hat{\lambda}_0 = \exp(\hat{\beta}_0)$，$\hat{k}_1 = \exp(-\gamma_1)$，$\hat{\lambda}_1 = \exp(\hat{\beta}_0 + \hat{\beta}_1)$。$\theta$极大似然估计记为$\hat{\theta}$，对应的渐进协方差矩阵为$\hat{\Sigma}_\theta = \left[-\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\Big|_{\theta=\hat{\theta}} \right]^{-1}$。则：

$$\hat{D}(t) = \widehat{\Delta\log H}(t) = \hat{k}_1\left[\log(t) - (\hat{\beta}_0 + \hat{\beta}_1)\right] - \hat{k}_0\left[\log(t) - \hat{\beta}_0\right] \tag{6}$$

$$\frac{\partial\hat{D}(t)}{\partial\theta} = \frac{\partial\Delta\log H(t)}{\partial\theta} = \left(\hat{k}_0 - \hat{k}_1, -\hat{k}_1, \hat{k}_0(\log(t) - \hat{\beta}_0), -\hat{k}_1(\log(t) - (\hat{\beta}_0 + \hat{\beta}_1))\right)^{\mathrm{T}} \tag{7}$$

$$\widehat{\mathrm{Var}}\left(\widehat{\Delta\log H}(t)\right) \approx \left[\frac{\partial\Delta\log H(t)}{\partial\theta}\Big|_{\theta=\hat{\theta}}\right]^{\mathrm{T}} \hat{\Sigma}_{\hat{\theta}} \left[\frac{\partial\Delta\log H(t)}{\partial\theta}\Big|_{\theta=\hat{\theta}}\right] \tag{8}$$

$$95\%CI(t) = \widehat{\Delta\log H}(t) \pm 1.96 \times \sqrt{\widehat{\mathrm{Var}}\left(\widehat{\Delta\log H}(t)\right)} \tag{9}$$

公式中$\hat{k}_0$和$\hat{k}_1$分别为对照组和干预组的Weibull形状参数的估计值，$\hat{\lambda}_0$和$\hat{\lambda}_1$为对应尺度参数的估计值，$\hat{\beta}_0$是截距项估计值，$\hat{\beta}_1$是治疗效应系数的估计值。log(·)和exp(·)分别表示自然对数函数和指数函数。参数向量θ由β_0、β_1及形状参数的负对数项构成，其极大似然估计$\hat{\theta}$的协方差矩阵用$\hat{\Sigma}_{\hat{\theta}}$表示，$\ell(\theta)$表示对数似然函数，梯度向量$\frac{\partial\Delta\log H(t)}{\partial\theta}$表示$\Delta\log H$对$\theta$的梯度（偏导数），$\widehat{\mathrm{Var}}\left(\widehat{\Delta\log H}(t)\right)$表示其渐近方差估计。

（2）Kaplan-Meier估计

Kaplan-Meier（KM）估计[10]是一种非参数生存分析方法，不依赖生存时间的分布假设，直接依据观测事件时间和删失状态进行估计。本研究中采用KM乘积限估计量先分别计算两组生存函数$\hat{S}(t)$，通过$-\log\hat{S}(t)$转化为对数累积风险函数$\widehat{\log H}(t)$并作差构建效应估计$\hat{D}(t)$。为进行统计推断，采用Greenwood公式[14]估计生存函数的方差$\widehat{\mathrm{Var}}(\hat{S}(t))$，并结合Delta方法经由log变换得到$\widehat{\log H}(t)$的方差估计，最终构建$D(t)$的95%置信区间，实现TTB的显著性区间边界估计。具体计算公式如下：

$$\hat{S}(t) = \prod_{i:t_i \leqslant t}\left(1 - \frac{d_i}{n_i}\right) \tag{10}$$

$$\hat{D}(t) = \widehat{\Delta\log H}(t) = \log(-\log(\hat{S}_1(t))) - \log(-\log(\hat{S}_0(t))) \tag{11}$$

$$\widehat{\mathrm{Var}}\left(\widehat{\Delta\log H}(t)\right) = \frac{\widehat{\mathrm{Var}}(\hat{S}_1(t))}{[\hat{S}_1(t)\cdot\log(\hat{S}_1(t))]^2} + \frac{\widehat{\mathrm{Var}}(\hat{S}_0(t))}{[\hat{S}_0(t)\cdot\log(\hat{S}_0(t))]^2} \tag{12}$$

$$95\%CI(t) = \widehat{\Delta\log H}(t) \pm 1.96 \times \sqrt{\widehat{\mathrm{Var}}\left(\widehat{\Delta\log H}(t)\right)} \tag{13}$$

式中，d_i是在时间t_i发生的事件数，n_i为在时间t_i之前仍处于风险中的个体数。

（3）Cox比例风险模型

传统TTB估计方法基于时间序贯截断的Cox比例风险模型[5]，通过评估*HR*估计值随时间变化的动态特征进行TTB估计。该方法遍历每个观测时间点t，构建包含所有生存时间变量值$T_i \leqslant t$个体的截断子数据集$D_a(t)$，在该截断数据集上拟合Cox模型，估计t时刻对应的Cox模型回归干预效应系数$\hat{\beta}_t$及对数风险比$\hat{D}(t) = \widehat{\log HR}(t)$值。计算公式如下：

$$D_a(t) = \{(T_i, \delta_i, X_i) \mid T_i \leqslant t, i = 1, \cdots, n\} \tag{14}$$

$$h(s \mid X, D_a(t)) = h_0(s \mid D_a(t))\exp(X\beta_t), s \leqslant t \tag{15}$$

$$\hat{D}(t) = \widehat{\log HR}(t) = \hat{\beta}_t \tag{16}$$

$$95\%CI(t) = \hat{\beta}_t \pm 1.96 \times \widehat{\mathrm{SE}}\left(\hat{\beta}_t\right) \tag{17}$$

式中$D_a(t)$表示截断至时间t的数据集，T_i、δ_i、X_i分别为个体i的生存时间、事件状态指示变量（1为事件发生，0为删失）和二元分组变量。$h(s|X, D_a(t))$是基于数据集$D_a(t)$的风险函数，而$h_0(s|D_a(t))$是基准风险函数。$\hat{\beta}_t$为估计的回归系数，其标准误是$\widehat{\mathrm{SE}}(\hat{\beta}_t)$。

数据来源

本研究采用两项公开发表的大型随机对照试验数据进行实例分析，分别为SPRINT研究（systolic blood pressure intervention trial）[15]和SOLOIST-WHF研究（sodium-glucose cotransporter 2 inhibitor in patients hospitalized for worsening heart failure）[16]。

SPRINT旨在评估强化降压治疗相较于标准降压对高血压患者心血管结局的影响。该研究共纳入9361例高血压患者，按1∶1比例随机分配至强化降压组（目标收缩压<120 mmHg）与标准降压组（目标收缩

压<140 mmHg），中位随访时间为3.26年，累积人年数约30500人年。由于大部分患者在研究结束时未发生结局事件，该研究具有较高的删失率（约94%）。

SOLOIST-WHF研究旨在评估SGLT2抑制剂索格列净在因心力衰竭恶化住院且射血分数降低患者中的疗效。该研究共纳入1222例患者，按1∶1比例随机分配至索格列净组和安慰剂组。中位随访时间为9.0个月，累积人年数约625人年，删失率为50.9%。该研究因新冠疫情影响而提前终止，但仍观察到明确的治疗效应。

模拟实验设置

为系统评估各模型估计TTB的统计性质，本研究设计了基于真实临床试验参数的模拟实验。模拟参数取值基于SPRINT与SOLOIST-WHF两项试验数据的Weibull分布拟合结果，分别构建两个独立模拟场景（场景1和场景2），以反映不同的治疗效应模式和生存时间分布特征。在每个场景下，生存时间T由Weibull分布生成，删失时间C由均匀分布$U(0, C_{max})$独立生成，其中C_{max}为可调参数，用于控制目标删失率。观测时间定义为$Y=\min(T, C)$，删失指示变量为$\delta=I(T\leqslant C)$（当$\delta=1$时，表示事件发生；$\delta=0$时表示发生右删失）。基于参数设置值，可直接计算各场景下TTB的真值作为评估基准。构造不同删失率（0%，30%，60%，90%）和样本量（N=5000，10000）组合的实验条件，并对每个实验条件重复模拟1000次。估计方法包括新框架下Weibull回归模型、KM估计以及传统框架下Cox模型的TTB估计。

本研究从TTB点估计和显著性区间边界估计两方面评估各模型方法的性能。采用偏倚（Bias）、平均绝对误差（mean absolute error，MAE）和均方根误差（root mean square error，RMSE）进行点估计评价。偏倚反映系统性偏差，MAE反映平均绝对偏离程度，RMSE综合反映偏差和变异性。显著性区间边界估计参照置信区间的常规评价准则，采用以下两个指标进行比较：一是计算TTB点估计值与显著性区间边界估计值之间的平均距离，反映边界估计的精确程度；二是计算覆盖率，即显著性区间边界估计值大于TTB真值的比例，参照值为单侧置信区间名义置信水平（97.5%）。计算效率的评估指标为各模型完成模拟所需的实际运行时间（wall-clock time）。

本研究的模拟实验均在高性能计算服务器上完成。硬件配置为搭载双路Intel Xeon Gold 6346 CPU @ 3.10 GHz（共计32物理核心，64线程），并配备1TB DDR4内存。为了提高计算效率，针对不同样本量和删失率组合采用并行计算策略：通过R软件包doFuture与foreach进行任务调度管理，利用20个CPU核心线程并行执行模拟任务。

所有统计分析均使用R软件v4.4.1版本实现。主要使用的R软件包为“survival” v3.8.3、“survminer” v0.5.0、“ggplot2” v3.5.1、“foreach” v1.5.2、“doRNG” v1.8.6.1、“future” v1.58.0以及“doFuture” v1.1.2。

结　果

1. 模拟研究

场景1基于SPRINT数据集，不同模型在各样本量和删失率下的TTB估计性能存在明显差异。从点估计结果来看（表1），Weibull回归模型表现出最小的TTB估计偏倚、MAE和RMSE，且在不同条件下保持较好的稳定性。整体上，随样本量增加，各指标均逐步改善；随删失率升高，各指标有不同程度的恶化。相比之下，KM估计的TTB误差高于Weibull回归模型，且首次和持续穿越TTB的估计存在较大差异。Cox模型表现最差，无论是Bias、MAE还是RMSE，均远高于其他两种方法，且波动较大。

表1　SPRINT数据三种模型TTB点估计性能比较

删失率	模型	计算方法	类型	N=5000			N=10000		
				Bias	MAE	RMSE	Bias	MAE	RMSE
0%	Weibull	$\Delta\log H$	首次	0.108	0.288	0.423	0.050	0.196	0.268
			持续	0.108	0.288	0.423	0.050	0.196	0.268
	KM	$\Delta\log H$	首次	−0.065	0.370	0.516	−0.145	0.337	0.405
			持续	0.577	0.682	0.991	0.348	0.461	0.630
	Cox	$\log(HR)$	首次	−0.081	0.323	0.735	−0.035	0.386	0.988
			持续	42.167	42.167	43.050	38.973	38.973	39.567
30%	Weibull	$\Delta\log H$	首次	0.142	0.332	0.497	0.060	0.221	0.297
			持续	0.142	0.332	0.497	0.060	0.221	0.297
	KM	$\Delta\log H$	首次	−0.136	0.356	0.474	−0.190	0.335	0.382
			持续	0.541	0.662	0.968	0.328	0.454	0.622
	Cox	$\log(HR)$	首次	−0.080	0.327	0.571	−0.058	0.349	0.598
			持续	12.484	12.484	13.723	9.262	9.262	10.178

续表 1

删失率	模型	计算方法	类型	*N*=5000			*N*=10000		
				Bias	MAE	RMSE	Bias	MAE	RMSE
60%	Weibull	Δlog*H*	首次	0. 183	0. 405	0. 614	0. 072	0. 264	0. 365
			持续	0. 183	0. 405	0. 614	0. 072	0. 264	0. 365
	KM	Δlog*H*	首次	−0. 167	0. 358	0. 451	−0. 207	0. 343	0. 386
			持续	0. 555	0. 672	0. 994	0. 338	0. 458	0. 627
	Cox	log(*HR*)	首次	−0. 005	0. 388	0. 715	−0. 042	0. 346	0. 524
			持续	3. 527	3. 547	4. 352	2. 199	2. 227	2. 738
90%	Weibull	Δlog*H*	首次	0. 233	0. 502	0. 736	0. 104	0. 347	0. 470
			持续	0. 233	0. 502	0. 736	0. 104	0. 347	0. 470
	KM	Δlog*H*	首次	−0. 194	0. 364	0. 457	−0. 228	0. 349	0. 393
			持续	0. 600	0. 714	1. 110	0. 349	0. 469	0. 656
	Cox	log(*HR*)	首次	0. 038	0. 408	0. 636	−0. 030	0. 339	0. 469
			持续	0. 922	1. 032	1. 512	0. 510	0. 646	0. 890

注：①Δlog*H*，对数累积风险比；log(*HR*)，对数风险比；②首次，Weibull、KM 和 Cox 模型差异曲线自上而下首次穿越阈值 0；③持续，Weibull、KM 和 Cox 模型差异曲线自上而下持续穿越阈值 0。

表 2 展示了各方法 TTB 显著性区间边界估计的性能。Weibull 回归模型的平均距离较小且随样本量上升而下降，但随删失率升高而有所增大；覆盖率则相对稳定，接近理论水平 97. 5%。KM 估计的平均距离大于 Weibull 回归模型，且首次 TTB 的平均距离在高删失率下反而降低，其覆盖率也较低，在 90% 高删失情形中不足 0. 6。而 Cox 模型的平均距离明显偏高，但覆盖率普遍接近 1，表现较为保守。

表 2　SPRINT 数据三种模型 TTB 显著性区间边界估计性能比较

删失率	模型	计算方法	类型	*N*=5000		*N*=10000	
				距离	覆盖率	距离	覆盖率
0%	Weibull	Δlog*H*	首次	1. 151	0. 972	0. 707	0. 977
			持续	1. 151	0. 972	0. 707	0. 977
	KM	Δlog*H*	首次	1. 823	0. 793	1. 228	0. 796
			持续	2. 294	0. 996	1. 410	0. 993
	Cox	log(*HR*)	首次	35. 589	0. 956	36. 736	0. 957
			持续	18. 646	1. 000	13. 881	1. 000
30%	Weibull	Δlog*H*	首次	1. 346	0. 971	0. 826	0. 977
			持续	1. 346	0. 971	0. 826	0. 977
	KM	Δlog*H*	首次	1. 664	0. 710	1. 146	0. 717
			持续	2. 359	0. 996	1. 443	0. 993
	Cox	log(*HR*)	首次	9. 536	0. 951	7. 574	0. 957
			持续	12. 968	1. 000	10. 039	1. 000
60%	Weibull	Δlog*H*	首次	1. 582	0. 980	0. 993	0. 978
			持续	1. 582	0. 980	0. 993	0. 978
	KM	Δlog*H*	首次	1. 568	0. 649	1. 032	0. 638
			持续	2. 407	0. 996	1. 450	0. 993
	Cox	log(*HR*)	首次	4. 297	0. 958	3. 195	0. 955
			持续	6. 395	1. 000	4. 409	1. 000
90%	Weibull	Δlog*H*	首次	1. 881	0. 977	1. 207	0. 975
			持续	1. 880	0. 977	1. 207	0. 975
	KM	Δlog*H*	首次	1. 496	0. 588	0. 976	0. 573
			持续	3. 034	0. 998	1. 615	0. 994
	Cox	log(*HR*)	首次	2. 715	0. 965	1. 838	0. 962
			持续	3. 275	0. 998	2. 034	0. 995

注：①首次，显著性区间边界估计上界自上而下首次穿越阈值 0；持续，显著性区间边界估计上界自上而下持续穿越阈值 0；②距离：平均距离，即各方法对应的点估计值与显著性区间边界估计值差值的绝对值；③覆盖率，各方法的对应的显著性区间边界估计值大于真值的次数占模拟次数的比例。

计算效率方面（表3），Weibull回归模型和KM估计耗时较短，均在10秒内，且对样本量和删失率变化不敏感。而序贯截断Cox模型则需耗时十几分钟，远高于其他两种方法。

表3 SPRINT数据三种模型TTB估计耗时比较

删失率	模型	消耗时间（*s*）	
		N=5000	*N*=10000
0%	Weibull	6. 454	7. 099
	KM	2. 951	3. 225
	Cox	1067. 666	1336. 894
30%	Weibull	5. 084	6. 311
	KM	2. 316	2. 870
	Cox	982. 420	1322. 482
60%	Weibull	5. 254	11. 505
	KM	2. 261	4. 927
	Cox	599. 493	1476. 773
90%	Weibull	5. 396	12. 218
	KM	2. 334	5. 048
	Cox	181. 352	515. 606

场景2基于SOLOIST-WHF数据集，不同方法的表现趋势与场景1结果基本一致。表4和表5分别呈现了点估计和显著性区间边界估计的结果。Weibull回归模型在各项指标中均占优势，KM估计次之，Cox模型表现最差。计算效率差异（表6）也保持相同，Weibull回归模型和KM估计耗时短，Cox模型耗时长。

2. 实例分析

图1展示了SPRINT和SOLOIST-WHF两组真实数据集采用不同方法估计TTB的差异。以SPRINT数据为例（图1A、C、E），Weibull回归模型首次和持续TTB点估计均为0. 36年；KM估计首次TTB值为0. 05年，持续TTB为0. 84年；Cox模型首次和持续TTB分别为0. 08年和1. 96年。根据显著性区间边界估计，Weibull回归模型首次和持续TTB上限均为1. 70年；KM估计首次和持续TTB上限分别为2. 18年和2. 29年；Cox模型上限则更晚，分别为2. 46年和2. 64年。总体而言，Cox模型估计的TTB最晚，Weibull回归模型和KM估计较早。

类似地，SOLOIST-WHF数据结果（图1B、D、F）也表明，Weibull回归模型的TTB估计最早，Cox模型

表4 SOLOIST-WHF数据三种模型TTB点估计性能比较

删失率	模型	计算方法	类型	*N*=5000			*N*=10000		
				Bias	MAE	RMSE	Bias	MAE	RMSE
0%	Weibull	Δlog*H*	首次	0. 436	0. 436	0. 611	0. 375	0. 375	0. 464
			持续	0. 436	0. 436	0. 611	0. 375	0. 375	0. 464
	KM	Δlog*H*	首次	0. 277	0. 277	0. 550	0. 193	0. 193	0. 395
			持续	0. 992	0. 992	1. 310	0. 720	0. 720	0. 905
	Cox	log(*HR*)	首次	0. 213	0. 213	0. 383	0. 329	0. 329	1. 209
			持续	32. 111	32. 111	32. 850	29. 675	29. 675	30. 151
30%	Weibull	Δlog*H*	首次	0. 466	0. 466	0. 673	0. 391	0. 391	0. 500
			持续	0. 466	0. 466	0. 673	0. 391	0. 391	0. 500
	KM	Δlog*H*	首次	0. 178	0. 178	0. 403	0. 140	0. 140	0. 325
			持续	0. 961	0. 961	1. 287	0. 699	0. 699	0. 888
	Cox	log(*HR*)	首次	0. 269	0. 269	0. 563	0. 301	0. 301	0. 586
			持续	10. 066	10. 066	11. 182	7. 564	7. 564	8. 465
60%	Weibull	Δlog*H*	首次	0. 527	0. 527	0. 785	0. 425	0. 425	0. 574
			持续	0. 527	0. 527	0. 785	0. 425	0. 425	0. 574
	KM	Δlog*H*	首次	0. 154	0. 154	0. 387	0. 118	0. 118	0. 296
			持续	0. 969	0. 969	1. 303	0. 702	0. 702	0. 894
	Cox	log(*HR*)	首次	0. 307	0. 307	0. 637	0. 302	0. 302	0. 532
			持续	3. 343	3. 343	4. 104	2. 263	2. 263	2. 759
90%	Weibull	Δlog*H*	首次	0. 568	0. 568	0. 895	0. 477	0. 477	0. 682
			持续	0. 568	0. 568	0. 895	0. 477	0. 477	0. 682
	KM	Δlog*H*	首次	0. 139	0. 139	0. 355	0. 101	0. 101	0. 275
			持续	1. 019	1. 019	1. 389	0. 730	0. 730	0. 938
	Cox	log(*HR*)	首次	0. 368	0. 368	0. 720	0. 344	0. 344	0. 593
			持续	1. 290	1. 290	1. 776	0. 884	0. 884	1. 169

注：①Δlog*H*，对数累积风险比；log(*HR*)，对数风险比；②首次，Weibull、KM和Cox模型差异曲线自上而下首次穿越阈值0；③持续，Weibull、KM和Cox模型差异曲线自上而下持续穿越阈值0。

表 5　SOLOIST-WHF 数据三种模型 TTB 显著性区间边界估计性能比较

删失率	模型	计算方法	类型	*N*=5000		*N*=10000	
				距离	覆盖率	距离	覆盖率
0%	Weibull	Δlog*H*	首次	1. 235	1. 000	0. 760	1. 000
			持续	1. 235	1. 000	0. 760	1. 000
	KM	Δlog*H*	首次	1. 813	1. 000	1. 325	1. 000
			持续	2. 336	1. 000	1. 533	1. 000
	Cox	log(*HR*)	首次	24. 815	1. 000	25. 956	1. 000
			持续	14. 548	1. 000	11. 184	1. 000
30%	Weibull	Δlog*H*	首次	1. 413	1. 000	0. 882	1. 000
			持续	1. 413	1. 000	0. 882	1. 000
	KM	Δlog*H*	首次	1. 571	1. 000	1. 127	1. 000
			持续	2. 406	1. 000	1. 573	1. 000
	Cox	log(*HR*)	首次	7. 388	1. 000	5. 926	1. 000
			持续	11. 572	1. 000	8. 886	1. 000
60%	Weibull	Δlog*H*	首次	1. 673	1. 000	1. 067	1. 000
			持续	1. 673	1. 000	1. 067	1. 000
	KM	Δlog*H*	首次	1. 436	1. 000	1. 015	1. 000
			持续	2. 488	1. 000	1. 591	1. 000
	Cox	log(*HR*)	首次	3. 935	1. 000	2. 829	1. 000
			持续	6. 018	1. 000	4. 221	1. 000
90%	Weibull	Δlog*H*	首次	1. 964	0. 999	1. 287	0. 999
			持续	1. 960	1. 000	1. 287	1. 000
	KM	Δlog*H*	首次	1. 446	1. 000	1. 022	1. 000
			持续	3. 541	1. 000	1. 872	1. 000
	Cox	log(*HR*)	首次	2. 681	1. 000	1. 917	1. 000
			持续	3. 344	1. 000	2. 166	1. 000

注：①首次，显著性区间边界估计上界自上而下首次穿越阈值 0；持续，显著性区间边界估计上界自上而下持续穿越阈值 0；②距离：平均距离，即各方法对应的点估计值与显著性区间边界估计值差值的绝对值；③覆盖率，各方法的对应的显著性区间边界估计值大于真值的次数占模拟次数的比例。

表 6　SOLOIST-WHF 数据三种模型 TTB 估计耗时比较

删失率	模型	耗时（s）	
		N=5000	*N*=10000
0%	Weibull	9. 880	13. 198
	KM	4. 026	6. 069
	Cox	1782. 117	2986. 484
30%	Weibull	8. 010	13. 261
	KM	3. 686	6. 498
	Cox	1780. 189	3390. 668
60%	Weibull	8. 089	16. 010
	KM	3. 428	7. 241
	Cox	721. 055	1755. 052
90%	Weibull	7. 983	16. 755
	KM	3. 409	7. 949
	Cox	280. 484	833. 360

最晚。KM 估计点估计居中，但显著性区间边界估计值明显较晚。Weibull 回归模型的 TTB 点估计首次和持续 TTB 值均为 0 年，显著性区间边界估计均为 0. 1 年；KM 估计点估计首次和持续 TTB 值分别为 0 年和 0. 02 年，显著性区间边界估计值首次和持续 TTB 值分别为 0. 1 年和 0. 85 年；Cox 模型点估计首次和持续 TTB 值分别为 0. 02 年和 0. 20 年，显著性区间边界估计为 0. 30 年和 0. 46 年。

总体而言，三种方法估计的效应差异变化趋势基本一致，均呈现早期变化剧烈，中后期趋于平稳的特点。Weibull 回归模型曲线在初期下降最快，且置信区间较窄；KM 估计曲线在开始阶段波动较大，置信区间相对较宽；Cox 模型曲线的波动幅度最大，置信区间最宽。KM 估计和 Cox 模型首次和持续 TTB 估计差异较大。实证结果与模拟研究结果吻合。

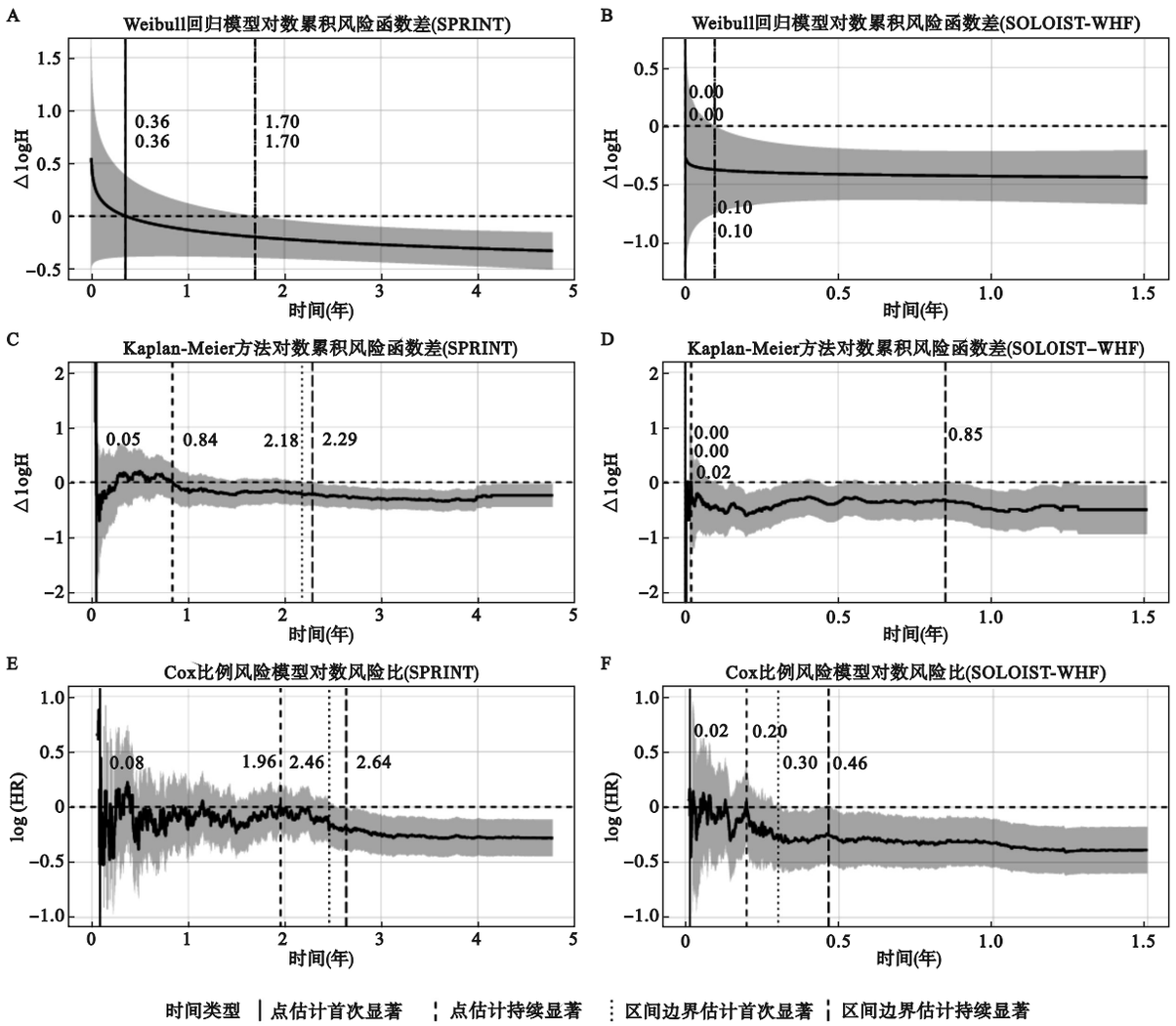


图1 基于SPRINT和SOLOIST-WHF试验数据的三类生存模型（Weibull、Kaplan-Meier、Cox）在对数累积风险比/对数风险比尺度下的获益时间估计

讨 论

本研究提出并评估了一个用于估计TTB的新方法学框架。该框架的核心创新在于采用对数累积风险比作为效应差异指标，并以其差异曲线首次或持续穿越无效阈值的时间点作为TTB估计。这种方法不仅能有效捕捉风险比随时间变化的动态趋势，还能基于成熟的统计理论对完整数据集进行一次性高效估计，避免了传统方法中数据反复截断并重复估计风险比所带来的局限。在不确定性量化方面，本研究采用显著性区间边界估计[5]来界定TTB的统计学显著区间，这种方法不同于直接构建指标置信区间[17-18]的传统做法。由于TTB的抽样分布难以直接推导，我们通过寻找差异函数的逐点置信区间上界与无效阈值的交点，来确定TTB的显著性边界。这种方法直观地反映了在考虑抽样误差后，治疗获益达到统计学显著的时间范围。为全面评估获益起始，我们沿用传统方法[5]，引入首次和持续两种TTB估计，分别对应于效应函数估计或其置信区间上界首次和持续穿越无效阈值的时间点，共同提供了治疗获益的证据。

模拟研究结果表明，新框架下的Weibull回归模型在TTB点估计精度、显著性区间边界估计的稳健性和计算效率等方面均占优势，覆盖率也较为理想。此外，结果显示Weibull回归模型首次与持续TTB估计值并无差异，这体现了参数化模型在效应轨迹拟合上的平滑性优势，该模型通过整体参数建模约束了效应演变的连续性。这种平滑性不仅提升了计算速度，还从根本上解决了效应轨迹在基线附近反复震荡导致的判定偏倚，为临床提供了一个更为稳健的获益判定标准。KM估计作为非参数替代，在某些条件下性能有所降低，但仍优于传统的Cox模型方法。其首次TTB的点估计表现优于持续TTB估计值，但显著性区间边界估计的性能则相对较差，反映了非参数方法在有限样本下进行区间估计的不稳定性。值得注意的是，在高删失率下，KM估计的首次TTB显著性区间边界估计反而出现较好的表现，这可能源于其标准误估

计的稳定性改善，但伴随的覆盖率下降提示该现象可能并非性能的可靠提升。与此同时，序贯截断的Cox模型结果反映了其方法学存在的问题。该模型在TTB显著性区间边界估计中的持续显著时间存在明显的高估偏倚，尤其在低删失率时尤为明显，其主要原因在于Cox模型在各截断点是通过截断数据估计HR的，删失时间点减少，而事件发生时间点增多且取值较大，使得中后期局部HR置信区间上界在基线附近持续波动，从而延迟了持续显著判定的时间。这种缺陷在长随访研究中会被放大，造成估计值严重偏离，且计算量大，进一步证实了该方法在理论和实际应用中的局限[19]。

实例分析进一步验证了模拟研究的发现。Weibull回归模型能最早识别出治疗获益时间点，无论是点估计还是显著性区间边界估计。相比之下，序贯截断Cox模型的TTB估计最晚。比较对数累积风险比与对数风险比曲线，两者在方向趋势上确实表现出相似性，这为新框架在理论解释上与传统方法的一致性提供了经验证据。差异曲线的比较也表明，Weibull回归模型在随访早期变化更为明显和平滑，KM估计和Cox模型存在初期波动剧烈的问题。在随访后期，KM估计差异曲线及其置信区间出现了轻微上升的波动，这种不稳定性可能源于风险集个体数量下降所致的估计方差增大，尽管存在波动，其效应估计并未跨越无效阈值。Weibull回归模型因受到其分布假设的约束，整体表现出更高的平滑度，而传统的序贯截断Cox模型方法的对数风险比置信区间也随着事件数的累积而逐步收窄，最终趋于稳定收敛，但计算效率远低于新框架下的两种方法。

综上所述，基于对数累积风险比轨迹的TTB估计新框架，在方法学上更为合理严谨，计算效率也明显优于传统序贯截断Cox模型方法。在此框架下，Weibull回归模型可作为首选方法，为临床研究中时间相关治疗效应的精准评估提供更可靠的统计工具。未来研究可进一步探讨该框架在其他生存分析场景（如竞争风险、复合终点）下的应用前景，从而为临床实践和公共卫生研究中的时间效应评价提供更加丰富可靠的工具支持。

参 考 文 献

[1] Lee SJ, Leipzig RM, Walter LC. “When will it help?” incorporating lagtime to benefit into prevention decisions for older adults[J]. JAMA, 2013, 310(24): 2609.

[2] Krishnaswami A, Peterson ED, Goyal P, et al. Time to benefit and harm of intensive blood pressure treatment: insights from SPRINT [J]. European Heart Journal-Quality of Care and Clinical Outcomes, 2021, 7(4): e1-e2.

[3] Yourman LC, Cenzer IS, Boscardin WJ, et al. Evaluation of time to benefit of statins for the primary prevention of cardiovascular events in adults aged 50 to 75 years: a meta-analysis [J]. JAMA Internal Medicine, 2021, 181 (2): 179-185.

[4] Lu M, Wu Z, Zhao Y, et al. Time to benefit estimation in multicenter studies using flexible hazard shared frailty models [J]. BMC Medical Research Methodology, 2026, 26 (1): 84.

[5] Chen K, Nie Z, Shi R, et al. Time to benefit of sodium-glucose cotransporter-2 inhibitors among patients with heart failure[J]. JAMA Network Open, 2023, 6(8): e2330754-e2330754.

[6] Ho VS, Cenzer IS, Nguyen BT, et al. Time to benefit for stroke reduction after blood pressure treatment in older adults: a meta-analysis[J]. Journal of the American Geriatrics Society, 2022, 70 (5): 1558-1568.

[7] Horiguchi M, Tian L, Kehl KL, et al. Assessing delayed treatment benefits of immunotherapy using long-term average hazard: a novel test/estimation approach [J]. Lifetime Data Analysis, 2025, 31 (4): 784-809.

[8] Bellera CA, MacGrogan G, Debled M, et al. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer[J]. BMC Medical Research Methodology, 2010, 10(1): 20.

[9] Carroll KJ. On the use and utility of the Weibull model in the analysis of survival data[J]. Controlled Clinical Trials, 2003, 24(6): 682-701.

[10] Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate[J]. International Journal of Ayurveda Research, 2010, 1(4): 274.

[11] Zhang Z. Parametric regression model for survival data: Weibull regression model as an example[J]. Annals of Translational Medicine, 2016, 4(24): 484.

[12] Odell PM, Anderson KM, D’Agostino RB. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model[J]. Biometrics, 1992, 48(3): 951-959.

[13] Pawitan Y. In all likelihood: statistical modelling and inference using likelihood[M]. Oxford: Oxford University Press, 2001.

[14] Rodenkirchen J, Hoyer A. An Extension of Greenwood’s Formulato Variances [PP/OL]. (2024-06-12) [2025-12-05]. https://doi.org/10. 48550/arXiv. 2406. 07994.

[15] SPRINT Research Group. A randomized trial of intensive versus standard blood-pressure control[J]. New England Journal of Medicine, 2015, 373(22): 2103-2116.

[16] Bhatt DL, Szarek M, Steg PG, et al. Sotagliflozin in patients with diabetes and recent worsening heart failure[J]. New England journal of medicine, 2021, 384(2): 117-128.

[17] Greenwood M. A report on the natural duration of cancer [R]. Reports on Public Health and Medical Subjects, No. 33. London: His Majesty’s Stationery Office, 1926.

[18] Neyman J. Outline of a theory of statistical estimation based on the classical theory of probability[J]. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1937, 236(767): 333-380.

[19] Hernán MA. The hazards of hazard ratios[J]. Epidemiology, 2010, 21(1): 13-15.

（责任编辑：林美艺）