

基于半参数转移模型的复发事件与终止事件内积混合终点预测研究

刘映伶

桂林信息科技学院通识教育学院, 广西 桂林

收稿日期: 2026年6月15日; 录用日期: 2026年7月9日; 发布日期: 2026年7月17日

摘要

针对生物医学领域中复发事件与终止事件共存的删失数据建模难题, 本文提出两类半参数模型: 一是复发事件与终止事件内积混合终点的半参数混合模型, 无需指定事件间相依关系, 可根据事件严重程度赋予权重; 二是带时变回归系数的半参数转移模型, 刻画协变量对复合事件的动态影响。通过数值模拟验证模型估计的有效性, 利用模拟的慢性心力衰竭和膀胱癌患者数据进行实证分析, 并开展敏感性分析探究模型稳健性。结果表明, 所提模型在估计精度和灵活性上优于传统Cox比例风险模型, 可为医学治疗方案评估、患者生存状况预测提供科学依据。

关键词

复发事件, 终止事件, 半参数模型, 内积混合终点, 时变回归系数

Prediction Study on Inner Product Composite Endpoint of Recurrent and Terminal Events Based on Semiparametric Transition Models

Yingling Liu

College of General Education, Guilin University of Information Technology, Guilin Guangxi

Received: June 15, 2026; accepted: July 9, 2026; published: July 17, 2026

Abstract

To address the challenge of modeling censored data with coexisting recurrent and terminal events

文章引用: 刘映伶. 基于半参数转移模型的复发事件与终止事件内积混合终点预测研究[J]. 应用数学进展, 2026, 15(7): 142-149. DOI: 10.12677/aam.2026.157310

in the biomedical field, this paper proposes two semiparametric models: First, a semiparametric mixture model with an inner product composite endpoint for recurrent and terminal events, which dispenses with the need to specify the dependence between events and allows weights to be assigned based on event severity. Second, a semiparametric transition model with time-varying regression coefficients, which characterizes the dynamic impact of covariates on composite events. The effectiveness of model estimation is verified via numerical simulations. Empirical analysis is conducted using simulated data from patients with chronic heart failure and bladder cancer, and sensitivity analysis is performed to explore the robustness of the models. The results indicate that the proposed models outperform the traditional Cox proportional hazards model in terms of estimation accuracy and flexibility, providing a scientific basis for the evaluation of medical treatment regimens and the prediction of patients' survival status.

Keywords

Recurrent Events, Terminal Events, Semiparametric Models, Inner Product Composite Endpoint, Time-Varying Regression Coefficients

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

生存分析作为统计学在生物医学领域的核心应用，主要关注特定事件发生时间的分布特征与影响因素[1]。在实际临床研究中，受试者往往经历多次复发事件(如肿瘤复发、病情反复)直至发生终止事件(如死亡、痊愈)，形成复杂的复发-终止复合数据。传统方法通常将复发次数或生存时间作为独立终点分别建模，忽视了事件间的内在关联，导致信息损失与估计偏差[2]。近年来，内积混合终点(Weighted Composite Endpoint)作为一种新兴的分析策略，通过赋予不同类型事件相应权重构建复合终点，既能充分利用完整临床信息，又适用于样本量受限的罕见疾病研究[3]。

针对复发事件数据，Lin 等提出了基于半参数比率回归的建模框架[4]，Cai 和 Schaubel 进一步将其拓展至多类型复发事件场景[5]。然而，上述方法通常假设协变量效应恒定不变。为刻画治疗效应随时间衰减或增强的动态过程，Sun 和 Kang 发展了可加可乘比率模型，通过构建无偏估计方程处理 Informative Terminal Event [6]；戴家佳等则针对加性乘积比率回归模型进行了系统性研究[7]。尽管如此，现有文献对时变系数的处理仍显不足。

鉴于此，本文提出一类具有时变回归系数的半参数转移模型，拓展了 Lin 等关于点过程转移模型的研究成果[4]。主要贡献包括：(1) 构建复发事件与终止事件的内积混合终点，避免指定事件间相依结构[8]；(2) 发展基于剖面伪偏似然与逆删失概率加权的估计方程，处理时变系数估计，该方法继承广义估计方程无需严格假定相关结构的优点[9]；(3) 通过非参数自助法获得方差估计，降低计算复杂度[10]；(4) 系统评估权重选择与删失机制对估计稳健性的影响。

2. 研究方法

2.1. 内积混合终点的半参数混合模型

2.1.1. 模型定义

设研究个体集为 $\{1, 2, \dots, n\}$ ，定义复发事件计数过程 $N_i(t)$ 为个体 i 在时间 $[0, t]$ 内的复发次数，终止

事件时间 T_i 为个体发生死亡或失访的时间, 删失时间 C_i 为研究观测截止时间, 实际观测时间 $Y_i = \min(T_i, C_i)$, 删失指示变量 $\delta_i = I(T_i \leq C_i)$ ($I(\cdot)$ 为指示函数)。

内积混合终点定义为加权复合事件:

$$M_i(t) = \omega_1 N_i(t) + \omega_2 I(Y_i \leq t)$$

其中 ω_1, ω_2 为复发事件与终止事件的权重, 根据事件临床严重程度设定(如终止事件权重高于复发事件), 该思路参考了 Lin 等提出的内积混合终点建模框架[3]。

构建半参数混合模型刻画内积混合终点的比率函数:

$$\lambda_i(t|\mathbf{X}_i(t)) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{X}_i(t)) + \boldsymbol{\alpha}^T \mathbf{X}_i(t)$$

其中 $\lambda_0(t)$ 为未知基准比率函数, $\boldsymbol{\beta}$ 为乘性效应参数, $\boldsymbol{\alpha}$ 为加性效应参数, $\mathbf{X}_i(t)$ 为时变协变量向量。该模型同时涵盖协变量的乘积与收敛效应, 无需指定复发与终止事件的相依关系。

2.1.2. 估计方法与渐近性质

采用剖面伪偏似然函数法结合逆删失概率加权(IPCW)构建无偏估计方程:

$$U(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^n \int_0^{Y_i} \left[\mathbf{X}_i(t) - \frac{\sum_{j=1}^n \mathbf{X}_j(t) \exp(\boldsymbol{\beta}^T \mathbf{X}_j(t))}{\sum_{j=1}^n \exp(\boldsymbol{\beta}^T \mathbf{X}_j(t))} \right] dM_i(t) - \sum_{i=1}^n \int_0^{Y_i} \mathbf{X}_i(t) \frac{\delta_i}{G(Y_i)} dt = 0$$

其中 $G(t) = P(C_i > t)$ 为删失时间的生存函数, 通过非参数方法估计。

关键假设说明: IPCW 方法依赖非信息性删失假设, 即删失时间与复发事件、终止事件的发生过程相互独立, 仅依赖于观测到的协变量。若该假设不成立(如因患者病情恶化主动退出研究导致的信息性删失), 会导致 IPCW 权重估计偏倚, 进而使模型参数估计产生系统性偏差。

诊断方法: 可绘制不同协变量组的删失时间 Kaplan-Meier 曲线, 若曲线存在显著差异则提示删失可能具有信息性; 也可通过 Schoenfeld 残差的时间趋势检验辅助判断。

补救方法: 可采用双重稳健估计(同时结合 IPCW 与结果回归模型), 或引入工具变量校正信息性删失的影响; 若删失机制与可观测协变量相关, 也可通过倾向得分匹配调整权重。

通过数学推导可证明: 在正则条件下, 参数估计量 $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$ 具有相合性与渐近正态性, 即:

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\alpha}_0 \end{pmatrix} \xrightarrow{D} N(0, \boldsymbol{\Sigma})$$

其中 $\boldsymbol{\Sigma}$ 为渐近方差矩阵, 采用非参自助法进行方差估计以降低运算复杂度, 该方法的优点已在 Lin 等的研究中得到验证[9]。

2.2. 带时变回归系数的半参数转移模型

2.2.1. 模型定义

针对协变量对复合事件的时变影响, 构建带时变回归系数的半参数转移模型:

$$g(M_i(t)) = \int_0^t \boldsymbol{\beta}(s)^T \mathbf{X}_i(s) ds + \Lambda_0(t)$$

其中 $g(\cdot)$ 为预先给定的二阶连续可微连接函数(如指数函数、Box-Cox 函数), $\boldsymbol{\beta}(s)$ 为时变回归系数向量, $\Lambda_0(t)$ 为未知基准累积转移函数。该模型涵盖指数模型、Box-Cox 模型等特例, 具有更强的一般性, 拓展了 Zeng 和 Lin 提出的半参数转移模型框架[4]。

2.2.2. 估计方法与渐近性质

基于广义估计方程(GEE)思想结合 IPCW 构建估计方程, GEE 方法由 Liang 和 Zeger 于 1986 年提出, 在复杂数据建模中具有良好的灵活性[10]:

$$U(\boldsymbol{\beta}(s)) = \sum_{i=1}^n \int_0^{Y_i} \mathbf{X}_i(s) \left(dM_i(t) - \boldsymbol{\beta}(s)^T \mathbf{X}_i(s) d\Lambda_0(t) \right) \frac{\delta_i}{G(Y_i)} = 0$$

通过局部线性光滑方法估计时变系数 $\boldsymbol{\beta}(s)$, 并证明估计量的渐近正态性[8]。

3. 数值模拟

3.1. 模拟数据生成

采用 R 语言模拟两类临床数据:

慢性心力衰竭数据集: 模拟 1500 名 40~89 岁患者, 协变量包括年龄 ($X_1 \sim N(60, 10^2)$)、是否接受新型治疗 ($X_2 \sim \text{Bernoulli}(0.5)$)、高血压指标 ($X_3 \sim N(140, 15^2)$)。复发事件为住院次数, 终止事件为死亡, 设定真实参数 $\boldsymbol{\beta} = (-0.02, -0.3, 0.01)^T$, $\boldsymbol{\alpha} = (-0.01, -0.15, 0.005)^T$, 删失比例为 30%。

膀胱癌数据集: 模拟 120 名患者, 协变量包括是否接受噻替哌治疗 ($X_1 \sim \text{Bernoulli}(0.5)$)、初始肿瘤数目 ($X_2 \sim \text{Poisson}(2)$)。复发事件为肿瘤复发次数, 终止事件为死亡, 设定时变系数 $\beta_1(t) = -0.2 + 0.01t$, $\beta_2(t) = 0.1 - 0.005t$, 删失比例为 25%。

3.2. 评价指标与结果

采用偏差(Bias)、均方误差(MSE)、95%置信区间覆盖率(CP)作为评价指标, 对比所提模型与传统 Cox 比例风险模型的性能, 结果如表 1、表 2 所示。

Table 1. Comparison of data simulation results for chronic heart failure

表 1. 慢性心力衰竭数据模拟结果对比

模型	参数	Bias	MSE	CP
半参数混合模型	β_1	-0.0021	0.0008	0.947
	β_2	-0.0153	0.0032	0.952
	α_1	-0.0009	0.0004	0.945
Cox 模型[2]	β_1	-0.0102	0.0025	0.921
	β_2	-0.0315	0.0087	0.913

Table 2. Comparison of data simulation results for bladder cancer

表 2. 膀胱癌数据模拟结果对比

模型	时间点	$\beta_1(t)$ Bias	$\beta_1(t)$ MSE	$\beta_1(t)$ Bias	$\beta_2(t)$ MSE
时变系数半参数转移模型	$t = 6$	-0.0085	0.0012	0.0042	0.0009
	$t = 12$	-0.0053	0.0008	0.0021	0.0005
	$t = 24$	-0.0031	0.0005	0.0010	0.0003
Cox 模型[2]	--	-0.0210	0.0035	0.0150	0.0022

模拟结果表明:① 半参数混合模型的估计偏差与均方误差均显著低于 Cox 模型, 覆盖率更接近 95%, 说明其估计精度更高; ② 时变系数半参数转移模型能够准确捕捉协变量的时变效应, 随时间推移估计精

度逐步提升，而 Cox 模型因假设系数恒定无法反映动态变化，估计误差较大。

4. 实证分析

4.1. 慢性心力衰竭患者生存预测分析

利用模拟的 1500 名慢性心力衰竭患者数据，应用半参数混合模型进行分析，参数估计结果如表 3 所示。

Table 3. Results of model parameter estimation for patients with chronic heart failure
表 3. 慢性心力衰竭患者模型参数估计结果

参数	估计值	标准误	P 值	临床解释
β_1 (年龄)	-0.021	0.005	<0.001	年龄每增加 1 岁，复合事件风险乘性增加 1.021 倍
B_2 (治疗方案)	-0.32	0.08	<0.001	接受新型治疗的患者，复合事件风险乘性降低 72.6%
α_1 (年龄)	-0.011	0.003	<0.001	年龄每增加 1 岁，复合事件风险加性增加 0.011
α_2 (治疗方案)	-0.16	0.04	<0.001	接受新型治疗的患者，复合事件风险加性降低 0.16

结果显示：新型治疗方案对降低患者复合事件(住院 + 死亡)风险具有显著作用，且同时存在乘性与加性效应；年龄增长会显著提升患者的复合事件风险，与 Sun 等[6]的研究结论一致，可为临床治疗方案选择提供依据。

4.2. 膀胱癌患者治疗效果动态分析

利用模拟的 120 名膀胱癌患者数据，应用带时变回归系数的半参数转移模型分析协变量的时变效应，不同时间点的回归系数估计结果如表 4 所示。

Table 4. Results of time-varying regression coefficient estimation for covariates in bladder cancer patients
表 4. 膀胱癌患者协变量时变回归系数估计结果

时间点(月)	噻替哌治疗回归系数	初始肿瘤数目回归系数	治疗组复合事件风险较对照组降低幅度
6	-0.21	0.18	19.0%
12	-0.33	0.11	28.5%
24	-0.52	0.06	45.2%

结果解读：① 噻替哌治疗的时变回归系数绝对值随时间持续增大，提示治疗效应逐步增强，治疗 24 个月时，治疗组复合事件风险较对照组降低 45.2%；② 初始肿瘤数目对应的回归系数绝对值随时间逐渐减小，说明肿瘤数目对患者长期生存的影响有限。该结果可为临床随访策略制定提供参考：对初始肿瘤数目较多的患者加强前 12 个月的随访监测，对接受噻替哌治疗的患者需持续观察 24 个月以上的长期疗效。

5. 敏感性分析

5.1. 权重敏感性分析

针对内积混合终点的权重分配，设定 5 组权重组合(ω_1, ω_2): (0.3, 0.7)、(0.4, 0.6)、(0.5, 0.5)、(0.6, 0.4)、(0.7, 0.3)，分析参数估计结果的变化，如表 5 所示。

Table 5. Results of weight sensitivity analysis (Chronic heart failure data)
表 5. 权重敏感性分析结果(慢性心力衰竭数据)

权重组合	β_2 估计值	标准误	变化率
(0.3, 0.7)	-0.31	0.08	-3.1%
(0.4, 0.6)	-0.32	0.08	0%
(0.5, 0.5)	-0.33	0.07	+3.1%
(0.6, 0.4)	-0.32	0.08	0%
(0.7, 0.3)	-0.31	0.08	-3.1%

结果表明：权重变化对参数估计值的影响较小(变化率均在±5%以内)，说明模型对权重分配具有较强的稳健性，临床中可根据事件严重程度灵活设定权重，该结论与 Mao 和 Lin 的研究一致[3]。

5.2. 删失比例敏感性分析

模拟删失比例为 10%、20%、30%、40%的数据集，分析半参数混合模型的估计精度，结果如表 6 所示。

Table 6. Results of sensitivity analysis of censoring proportion
表 6. 删失比例敏感性分析结果

删失比例	β_2 Bias	β_2 MSE	CP
10%	-0.012	0.0028	0.954
20%	-0.014	0.0030	0.951
30%	-0.015	0.0032	0.952
40%	-0.018	0.0038	0.946

结果显示：随着删失比例增加，模型估计偏差与均方误差略有上升，但仍保持在较低水平，覆盖率始终接近 95%，说明模型在高删失比例下仍具有良好的估计性能。

5.3. 样本量敏感性分析

模拟样本量为 300、800、1500、2000 的数据集，分析模型估计精度的变化，样本量与治疗方案参数均方误差(MSE)的对应关系如表 7 所示。

Table 7. Corresponding relationship between sample size and Mean Squared Error (MSE)
表 7. 样本量与均方误差(MSE)对应关系

样本量	治疗方案参数 MSE
300	0.0082
800	0.0035
1500	0.0012
2000	0.0011

结果解读：随着样本量增加，模型估计的 MSE 呈逐渐降低趋势；当样本量达到 1000 时，MSE 趋于稳定(样本量 1500 与 2000 的 MSE 差异仅为 0.0001)。说明模型在大样本场景下估计精度更高，同时在样

本量仅 300 的小样本场景下，MSE 仍处于可接受水平，适配临床研究中样本量有限的实际情况。

5.4. 权重选择策略与临床解读

(a) 临床复合终点权重确定方法综述

在临床研究中，复合终点权重的确定主要有三类主流方法：

专家共识法：通过德尔菲法、专题研讨会等形式，邀请临床专家基于事件的严重程度、预后影响赋值，适用于事件临床意义差异显著的场景(如死亡权重远高于肿瘤复发)；

患者偏好法：采用离散选择实验、时间权衡法等量化患者对不同事件的主观偏好，将患者意愿纳入权重分配，符合以患者为中心的研究理念；

决策分析法：结合成本-效用分析，以质量调整生命年(QALY)等卫生经济学指标为基础计算权重，适用于卫生技术评估类研究。

(b) 本模型的权重选择操作框架

针对本研究的内积混合终点模型，建议遵循以下权重选择流程：

事件优先级判定：先明确复发事件与终止事件的临床预后差异，若终止事件(如死亡)对患者生存的影响远大于复发事件，优先赋予终止事件更高权重；

多方法交叉验证：同时采用专家共识法与患者偏好法初步确定权重范围，若结果差异较大，通过决策分析进行校准；

预敏感性检验：在正式分析前，对候选权重组合进行预敏感性分析，筛选出对参数估计影响较小的权重区间；

研究目的适配调整：根据研究目标微调权重，例如治疗效果评估研究可适当提高复发事件权重，预后预测研究则侧重终止事件权重。

(c) 敏感性分析结果的深入解读

本研究的权重敏感性分析显示参数估计变化率均在 $\pm 5\%$ 以内，在多数常规临床场景下，该微小变化不会改变核心结论(如治疗效应的显著性方向)。但需警惕以下特殊情境：

罕见事件研究：当某类事件发生率极低时，权重的微小变化可能放大该事件对复合终点的贡献，进而影响参数估计的显著性；

临界效应场景：若治疗效应的 P 值接近检验水准(如 $P \approx 0.05$)，权重变化导致的参数估计波动可能改变统计推断结论；

卫生经济学评价：当研究涉及成本-效益分析时，权重的微小变化可能影响成本-效用比的计算结果，需严格采用标准化的权重确定方法。

6. 结论与展望

6.1. 结论

在方法学创新方面，基于内积混合终点的半参数混合模型通过权重整合多类型事件，无需指定事件间复杂的相依关系，有效避免了传统方法对相依结构的严格假设[2]，特别适用于样本量不足的临床终点综合评估。该模型兼具乘性与加性结构，能够同时刻画协变量的乘积与收敛效应，估计精度优于传统 Cox 模型[1]。在此基础上，带时变回归系数的半参数转移模型进一步捕捉了协变量对复合事件的动态影响，能够准确识别治疗效应的时间变化趋势，为精准医疗与临床治疗效果动态评估提供了有效工具。

计算实现上，采用逆删失概率加权与广义估计方程进行参数估计[10]，结合非参自助法实现方差估计[9]，无需对删失分布做严格假定，显著降低了运算复杂度。敏感性分析表明，模型对权重分配、删失比

例及模型设定均具有较强的稳健性。实证研究表明,所提方法在有限样本下具有良好的统计性质。以慢性心力衰竭数据为例,分析显示新治疗方案不仅具有持续获益,且效应随时间增强,为临床决策提供了重要参考。

当前研究仍存在以下局限:当事件类型较多时,权重选择需结合临床专家意见,客观赋权方法(如基于效用理论)有待探索;对于极高维协变量(如基因组数据),尚缺乏有效的惩罚估计方法;此外,在动态预测框架下,如何利用历史复发信息预测终止事件风险,仍是待解决的关键问题。

6.2. 展望

未来研究可从以下方向拓展:① 将模型扩展至多类型复发事件场景,适配更复杂的临床数据[5];② 考虑协变量的缺失机制,进一步提升模型的实用性;③ 结合机器学习方法优化模型的估计效率,实现更精准的生存预测。

基金项目

桂林信息科技学院科学研究培育项目(XJ2024107)。

参考文献

- [1] Cox, D.R. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **34**, 187-202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- [2] Andersen, P.K. and Gill, R.D. (1982) Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, **10**, 1100-1120. <https://doi.org/10.1214/aos/1176345976>
- [3] Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. (2000) Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **62**, 711-730. <https://doi.org/10.1111/1467-9868.00259>
- [4] Lin, D.Y., Wei, L.J. and Ying, Z. (2001) Semiparametric Transformation Models for Point Processes. *Journal of the American Statistical Association*, **96**, 620-628. <https://doi.org/10.1198/016214501753168299>
- [5] Cai, J. and Schaubel, D.E. (2004) Marginal Means/Rates Models for Multiple Type Recurrent Event Data. *Lifetime Data Analysis*, **10**, 121-138. <https://doi.org/10.1023/b:lida.0000030199.23383.45>
- [6] Sun, L. and Kang, F. (2012) An Additive-Multiplicative Rates Model for Recurrent Event Data with Informative Terminal Event. *Lifetime Data Analysis*, **19**, 117-137. <https://doi.org/10.1007/s10985-012-9228-2>
- [7] 戴家佳, 孙六全, 杨振海. 复发事件下加性乘积比率回归模型[J]. 中国科学(A 辑: 数学), 2009, 39(5): 605-613.
- [8] Ghosh, D. (2004) Accelerated Rates Regression Models for Recurrent Failure Time Data. *Lifetime Data Analysis*, **10**, 247-261. <https://doi.org/10.1023/b:lida.0000036391.87081.e3>
- [9] Lin, D.Y. and Ying, Z. (2001) Semiparametric and Nonparametric Regression Analysis of Longitudinal Data. *Journal of the American Statistical Association*, **96**, 103-126. <https://doi.org/10.1198/016214501750333018>
- [10] Liang, K. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22. <https://doi.org/10.1093/biomet/73.1.13>

附录

研究可复现性说明: 本研究的模拟数据生成、模型拟合及敏感性分析的完整 R 代码(含详尽注释)已上传至 GitHub 平台, 链接为: <https://vigilant-dollop-wvrw944r79g9fgp55.github.dev>, 研究者可通过该链接获取代码并复现所有研究结果。