



【专家简介】王炳顺,医学博士,上海交通大学医学院教授、博士生导师,现任临床医学研究院副院长、生物统计教研室主任。主要从事临床研究中的医学统计方法与妇幼保健流行病学研究,同时积极推动医学科研文化建设与临床研究体系规范化发展。在 *Lancet*, *PNAS*, *Blood*, *eClinicalMedicine* 及 *J Stat Soft* 等医学、统计学及综合性期刊发表论文近百篇。主编《医学统计与临床研究》《SAS统计软件应用》《医学统计学及SAS应用》, 副主编《临床试验设计与统计分析》等多部教材。作为项目负责人,主持国家自然科学基金、科技部“十三五”“十四五”国家重点研发计划课题及上海市软科学重点项目;担任全国医疗器械临床评价标准化技术归口单位专家组成员、国家药品监督管理局药品审评和医疗器械审评统计学专家,深度参与新药研发与创新医疗器械的临床评价方法学体系建设。

贝叶斯联合模型在纵向观测和生存数据整合分析中的应用

倪长宇^{1,2} 王筱金² 王炳顺^{2Δ}

(¹上海交通大学公共卫生学院 上海 200025; ²上海交通大学医学院临床医学研究院生物统计教研室 上海 200025)

【摘要】 在医学随访研究中,纵向观测数据(如重复测量的生物标志物或症状评分)与生存时间数据(如疾病进展或死亡事件)存在密切关联。传统的独立分析方法因忽视二者内在关联及测量误差,易导致统计推断偏差。联合模型通过共享随机效应关联纵向子模型与生存子模型,可纠正重复测量中的测量误差,提升参数估计效率和统计检验效能。传统频率学派的联合模型在简单场景下具有可行性,但在处理高维、非线性或复杂缺失机制处理时面临计算与推断挑战。贝叶斯联合模型基于马尔可夫链蒙特卡罗(Markov chain Monte Carlo, MCMC)方法,通过引入先验分布和后验抽样技术,在参数估计稳健性、模型扩展性和动态预测性能方面更具优势。本文简介贝叶斯联合模型的方法学框架,包括:(1)纵向子模型(如线性混合效应模型)与生存子模型(如Cox比例风险模型)的构建;(2)三类常见关联结构(当前值、当前斜率及累积面积);(3)基于MCMC的贝叶斯参数估计;(4)个体化动态预测与模型性能评估。以原发性胆汁性肝硬化为例,演示贝叶斯联合模型的实际应用流程:从临床预测指标筛选、单/多指标联合模型拟合与比较,到时间依赖性ROC曲线验证预测效能。实例分析显示,贝叶斯联合模型可有效整合纵向轨迹信息,动态更新个体生存概率,为临床精准决策提供量化依据。

【关键词】 纵向数据; 生存分析; 联合模型; 贝叶斯方法; 动态预测

【中图分类号】 R195.1 **【文献标志码】** A **doi:** 10.3969/j.issn.1672-8467.2026.02.021

Application of Bayesian joint models in the integrated analysis of longitudinal and survival data

NI Chang-yu^{1,2}, WANG Xiao-jin², WANG Bing-shun^{2Δ}

(¹School of Public Health, Shanghai Jiao Tong University, Shanghai 200025, China; ²Department of Biostatistics, Institute of Clinical Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China)

国家重点研发计划(2023YFC2411403)

^ΔCorresponding author E-mail: wangbingshun@sjtu.edu.cn

网络首发时间:2026-03-19 13:31:38 网络首发地址:https://link.cnki.net/urlid/31.1885.R.20260318.1419.002

【Abstract】 In medical follow-up studies, there is often an intrinsic association between longitudinal observations (e.g., repeatedly measured biomarkers or symptom scores) and time-to-event data (e.g., disease progression or mortality). Traditional approaches that analyze these two data types independently may lead to biased statistical inference, as they neglect this inherent association and measurement errors. Joint models (JMs) address this limitation by linking longitudinal and survival sub-models through shared random effects, thereby correcting measurement errors in repeated measurements and improving estimation efficiency and statistical power. While traditional frequentist approaches are feasible in simple scenarios, they face computational and theoretical challenges when handling high-dimensional, nonlinear, or complex missing data mechanisms. In contrast, Bayesian JMs leverage Markov chain Monte Carlo (MCMC) methods, incorporating prior distributions and posterior sampling techniques to enhance robustness in parameter estimation, model flexibility, and dynamic prediction performance. This article introduces the methodological framework of Bayesian joint models, including: (1) Specification of longitudinal sub-models (e.g., linear mixed-effects models) and survival sub-models (e.g., Cox proportional hazards models); (2) Three common association structures (current value, current slope, and cumulative area); (3) Bayesian parameter estimation via MCMC; (4) Personalized dynamic prediction and model performance evaluation. Using primary biliary cirrhosis (PBC) as a case study, we demonstrate the practical application of Bayesian JMs, ranging from the selection of predictive indicators, fitting and comparison of single/multi-indicator JMs to the predictive performance using time-dependent ROC curves. The case study indicates that Bayesian JMs can effectively integrate longitudinal trajectory information, dynamically update individual survival probabilities, and provide quantitative support for clinical decision-making.

【Key words】 longitudinal data; survival analysis; joint model; Bayesian method; dynamic prediction

* This work was supported by the National Key Research and Development Program of China (2023YFC2411403).

在临床研究及流行病学研究中,研究者常需同时分析两类密切相关的数据:重复测量的纵向指标(如生理参数或生物标志物)和记录终点事件的生存数据(如死亡或疾病复发)。传统分析方法通常将这两类数据割裂处理,不仅未能充分捕捉其内在关联性,也忽略了纵向测量中固有的误差,最终导致关联效应估计出现偏差^[1]。为解决这一问题,联合建模方法将纵向与生存数据整合在统一的分析框架中,通过建立共享的潜在结构形式,更准确地刻画纵向变化过程与生存风险间的动态关联,从而为疗效评估和疾病机制研究提供更可靠的统计依据^[2-3]。此外,该方法还能利用随访过程中持续更新的纵向数据,实现对个体生存概率的动态预测,展现出显著的临床应用价值^[4]。

联合模型的发展经历了不同阶段。早期 Tsiatis 等^[5]提出的两阶段方法首先拟合纵向数据的线性混合效应模型,再将预测的轨迹函数作为时变协变量纳入 Cox 比例风险模型。虽然该方法有所改进,但

其隐含假设(即生物标志物在两次测量间保持恒定)过于理想化,仍可能导致估计倚偏^[6]。随着研究深入,更稳健的共享参数模型逐渐发展成熟,该方法通过共享的随机效应直接关联纵向和生存子模型,有效克服了两阶段方法的固有缺陷,成为当前的主流建模策略。

在参数估计方面,早期研究主要采用频率学派的极大似然估计或偏似然估计^[5]。然而,当处理包含多结局指标、竞争风险事件或高维随机效应的复杂模型时,这类方法对分布假设较为敏感,容易出现估计不稳定的问题。基于马尔可夫链蒙特卡罗(Markov chain Monte Carlo, MCMC)算法的贝叶斯方法为此提供了更优的解决方案:通过引入参数的先验分布并结合数据得到后验分布进行统计推断,不仅显著提高了参数估计的稳健性和灵活性,还能自然处理多事件终点和复合纵向指标^[7],且无需对随机效应分布施加正态性假设^[8]。由此可见,贝叶斯联合建模方法有效突破了传统频率学派方法在

模型复杂度和分布假设方面的双重局限,为多维度、多类型临床数据的整合分析提供了更为完善的统计解决方案。

方法学框架 贝叶斯联合模型的核心是构建一个包含纵向子模型和生存子模型的一体化框架,并通过特定的关联结构将二者联系起来,在贝叶斯范式下完成统计推断与动态预测。

基本框架 联合模型的框架图清晰展示了自变量 X_i (如治疗组别)、 t 时刻测量的纵向数据 $Y_i(t)$ 以及生存结局 S 之间的相互关系(图 1)。其中, α 表示自变量对生存结局的直接效应, β 表示自变量对纵向数据的效应, γ 表示纵向数据对生存结局的效应。为了量化以上变量间的关联,联合模型由两个子模型组成:纵向子模型与生存子模型。

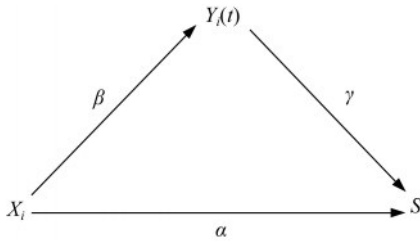


图 1 联合模型的框架示意图

Fig 1 Schematic diagram of the framework for joint models

纵向子模型 旨在描述个体内部重复测量指标的轨迹,并捕捉个体间的异质性,通常采用线性混合效应模型^[1]。结构可表示为:

$$Y_i(t) = \beta_1 X_i + \beta_2 t + a_i + b_i t + \varepsilon_{it} \quad (1)$$

其中 X_i 代表自变量(如治疗分组),其固定效应系数为 β_1 ;模型中同时纳入固定效应斜率为 β_2 的线性时间趋势项,也可用函数 $f(t)$ 替代线性函数 $\beta_2 t$,以表征非线性时间趋势; a_i 为随机截距,用于刻画个体间纵向数据初始水平的异质性; b_i 为随机斜率,用于刻画个体间纵向数据变化速率的异质性; ε_{it} 为测量误差,假定其独立于 a_i 、 b_i 和 X_i 。

生存子模型 用于描述结局事件发生的时间,通常为 Cox 比例风险模型或其拓展形式^[1]。个体 i 在时间点 t 的风险函数可表示为:

$$h_i(t) = h_0(t) \exp\{\alpha X_i + \gamma Y_i^*(t)\} \quad (2)$$

其中, $h_0(t)$ 为基线风险函数, $Y_i^*(t)$ 为纵向数据随时间连续变化的真实轨迹。值得注意的是,纵向子模型与生存子模型可以使用不同的协变量集合,两者的变量筛选可基于临床意义与模型拟合效果独立进行,随后再进行联合模型拟合^[9]。

关联结构 在联合建模框架中,纵向子模型与生存子模型通过特定的关联结构相连接,该结构定义了纵向数据轨迹如何影响事件发生的风险^[10]。常见的关联结构有三种形式^[8,11-12]:一是“当前值”关联,即某时刻的事件风险取决于当前时刻纵向指标的真实水平(如当前 CD4 计数影响 HIV 患者死亡风险);二是“时变斜率”关联,即风险同时受到该时刻纵向指标瞬时变化速率的影响(如 eGFR 下降速度影响肾替代治疗风险);三是“累积面积”关联,即风险与从基线至某时刻的纵向指标曲线下面积所反映的累积暴露效应相关(如血铅浓度增加即铅在体内累积暴露量增加神经系统不可逆损伤风险)。这些结构分别捕捉了纵向过程的即时状态、动态变化趋势与历史累积负荷对事件风险的不同影响机制。因而基于潜在生物学机制选择合适的关联结构对模型的科学解释和预测准确性至关重要。

参数估计 在贝叶斯框架下,联合模型基于 MCMC 进行参数估计^[13]。首先,通过设定先验分布表达对参数的先验知识,随后基于联合似然函数(通常假设在给定随机效应下,纵向与生存过程相互独立)计算后验分布。采用 MCMC 方法从后验分布中抽取样本,进而基于样本均值、标准差和分位数计算参数的点估计和可信区间。收敛性可通过 Gelman-Rubin 诊断和轨迹图进行评估^[14]。该方法无需依赖渐近近似或正态性假设,并且可灵活纳入先验信息,适用于复杂联合模型。

动态预测 贝叶斯联合模型的突出优势是能够进行个体化的动态预测,即基于个体截至某时间点的纵向观测值、基线协变量及生存状态,实时更新未来结局的预测。对于个体 i ,在已知其截至时刻 t 的纵向观测值 $Y_i(t)$ 、基线协变量 X_i 以及生存至时间 t (即 $T_i^* > t$) 的条件下,可对其未来某一时刻 $u(u > t)$ 的纵向观测值进行预测,公式如下:

$$Y_i(u|t) = E\{y_i(u) | T_i^* > t, Y_i(t), \mathcal{D}_n\} \quad (3)$$

其中 $y_i(u)$ 表示个体 i 在 u 时刻真实的纵向观测值, $Y_i(t)$ 表示在 t 时刻及之前已经观测到的所有纵向数据, \mathcal{D}_n 被定义为用于拟合联合模型的训练样本。在此基础上,可进一步预测该个体在未来时刻 u 的生存概率:

$$S_i(u|t) = P\{T_i^* \geq u | T_i^* > t, Y_i(t), X_i, \mathcal{D}_n\} \quad (4)$$

对个体 i 进行动态预测时,模型参数的后验分

布 $p(\theta | \mathcal{D}_n)$ 是基于训练样本 \mathcal{D}_n 得到的,并用于计算 $S_i(u|t)$ 。此外,模型还能动态更新:当个体 i 在 $t' > t$ 时刻获得新的观测信息时,可将 $Y_i(t')$ 与 $T_i^* > t'$ 代入上述公式,重新计算预测值,从而实现预测结果的实时迭代与优化^[15-16]。这种动态更新特性非常适合临床随访中的实时风险监控与决策辅助。

模型拟合效果与预测性能评估 在贝叶斯联合建模框架下,模型拟合效果通常通过信息准则进行评估,如偏差信息准则(deviance information criterion, DIC)和广泛适用信息准则(widely applicable information criterion, WAIC)。DIC 基于后验偏差与有效参数个数的权衡,值越小表示模型拟合越好且复杂度适中^[17];WAIC 通过计算对数预测密度评估模型泛化能力,同样以较小值为优^[18]。这两种准则均适用于复杂随机效应结构。

模型预测性能的评估主要关注区分度与校准度。区分度常用时间依赖性 ROC 曲线(time-dependent ROC)及其曲线下面积(area under curve, AUC)来评估模型区分度。动态 AUC 能贴切地反映预测模型在随访不同时间点的判别力。校准度反映预测风险与实际风险的一致性,可通过动态预期布里尔分数(Brier score, BS)或绘制校准图来评估^[19]。

建模工具 R 语言提供了多个拟合联合模型的软件包,经历了从经典频率学派的最大似然估计向贝叶斯框架的转变,同时提升了模型灵活性、功能范围和计算效率。

JM 包^[20]通过共享参数模型实现了对经典联合模型的稳定拟合与推断,为联合建模奠定了方法学基础。但其在处理不同数据类型和复杂模型方面存在局限性。

JMbayes 包^[13]利用 MCMC 算法进行后验推断,实现了向贝叶斯框架的重要转变。该包内置了动态预测及区分度、校准度评估工具,功能显著增强。灵活的关联结构可以支持更广泛的模型类型,使其能够处理二分类、删失数据等非连续纵向结局。而且支持基于个体截至某时间点的纵向测量数据,实时更新其未来的生存概率或生物标志物轨迹预测。

JMbayes2 包^[21]是 JMbayes 的扩展版本,进一步支持混合类型(连续/分类)或多个纵向结局,以及多重事件时间(如竞争风险及多状态过程)的联合建模。此外,通过底层算法的 C++ 优化和一系列辅助函数的集成,构建了一个从数据预处理、模型拟

合比较到预测验证的高效分析流程。

实例分析 为具体说明贝叶斯联合模型的应用流程,本文以 R 包 JMbayes2 内置的原发性胆汁性肝硬化(primary biliary cirrhosis, PBC)数据集 PBC2 作为分析实例。该数据集源自梅奥诊所 1974—1984 年开展的 PBC 患者长期随访研究,包含 312 名患者信息,共 1 945 条随访记录,涵盖基线特征、临床症状、多次随访测量的生物标志物水平及生存状态(结局为死亡或肝移植的复合事件)等关键信息。

研究人群按结局分为存活组(143 例,45.8%)和死亡与肝移植合并的复合结局组(169 例,54.2%)。两组在年龄、性别、腹水、肝肿大、蜘蛛痣、水肿程度及组织学期等方面的差异均有统计学意义($P < 0.05$),而用药情况的组间差异无统计学意义($P = 0.636$)(表 1)。

表 1 PBC2 数据集人群基线特征

Tab 1 Baseline characteristics of the study population

Variables	in the PBC2		P
	Alive	Dead/ Transplanted	
Cases	143 (45.8)	169 (54.2)	
Age (Mean ± SD)	48.4 ± 9.8	51.4 ± 11.0	0.012
Sex			0.001
Male	7 (4.9)	29 (17.2)	
Female	136 (95.1)	140 (82.8)	
Drug			0.636
D-penicillamine	75 (52.4)	83 (49.1)	
placebo	68 (47.6)	86 (50.9)	
Ascites			<0.001
Yes	1 (0.7)	23 (13.6)	
No	142 (99.3)	146 (86.4)	
Hepatomegaly			<0.001
Yes	46 (32.2)	114 (67.5)	
No	97 (67.8)	55 (32.5)	
Spiders			0.001
Yes	27 (18.9)	63 (37.3)	
No	116 (81.1)	106 (62.7)	
Edema			<0.001
No edema	124 (86.7)	123 (72.8)	
Edema no diuretics	18 (12.6)	26 (15.4)	
Edema despite diuretics	1 (0.7)	20 (11.8)	
Histologic stage			<0.001
1	15 (10.5)	1 (0.6)	
2	43 (30.1)	24 (14.2)	
3	58 (40.6)	62 (36.7)	
4	27 (18.9)	82 (48.5)	

模型首先以单一纵向指标——血清胆红素作为纵向子模型的因变量。为了满足拟合线性混合模型的因变量近似正态分布的要求,对血清胆红素进行对数变换得到 $\log(\text{serBilir})$ 。本研究对纵向子模型和生存子模型分别采用最优子集法筛选自变量,以赤池信息准则(Akaike information criterion, AIC)最低为标准。纵向子模型在基线特征中筛选出腹水和蜘蛛痣等4个变量;生存子模型使用相同方法筛选变量,并通过Schoenfeld残差法检验比例风险假设,筛选出性别、腹水和水肿等5个变量。同时,基于临床意义,将用药情况纳入两个子模型中进行分析。

本文先比较了联合模型中时间效应以及两个子模型关联结构组合出的不同模型,即联合模型中时间效应(线性 vs. 非线性)与关联结构(当前值、当前斜率、累积面积)的6种组合。当时间效应为非线性、连接函数形式为当前值时,模型表现最优(DIC=3 902.792, WAIC=3 881.708),故将该模型作为本研究的最优联合模型。其中贝叶斯参数估计的先验

分布设定采用JMbayes2包默认的弱信息先验。

基于最优联合模型的后验分布估计结果,对模型的具体参数进行分析。在纵向数据子模型部分,采用2个节点的自然样条函数 $\text{ns}(\text{year}, 2)$ 拟合观测时间对 $\log(\text{serBilir})$ 的影响,其中 $\text{ns}(\text{year}, 2)_1$ 和 $\text{ns}(\text{year}, 2)_2$ 两个分项协同捕捉观测时间与 $\log(\text{serBilir})$ 之间的非线性关系。

结果显示,时间的两个非线性分项对 $\log(\text{serBilir})$ 的变化具有统计学意义($\text{ns}(\text{year}, 2)_1, \beta=1.838, 95\% \text{CI}: 1.530 \sim 2.166$; $\text{ns}(\text{year}, 2)_2, \beta=1.709, 95\% \text{CI}: 1.246 \sim 2.226$),提示观测时间与 $\log(\text{serBilir})$ 呈显著非线性相关,随观测时间延长 $\log(\text{serBilir})$ 逐渐升高(表2)。此外,腹水、蜘蛛痣、水肿均与 $\log(\text{serBilir})$ 升高显著相关,而药物使用没有统计学意义。在生存分析子模型部分,纵向指标 $\log(\text{serBilir})$ 对患者发生死亡/移植风险的影响显著,结果表明 $\log(\text{serBilir})$ 每增加1个单位,结局发生风险上升3.884倍(HR=3.884, 95%CI: 3.208~4.785)(表2)。

表2 纵向子模型与生存子模型参数估计

Tab 2 Parameter estimates of the longitudinal submodel and survival submodel

Variables	Estimate (95% CI)	Std. Dev	P	HR (95% CI)
Longitudinal Outcome [$\log(\text{serBilir})$]				
(Intercept)	0.391 (0.190, 0.592)	0.102	<0.001	—
Ns(year, 2) ₁	1.838 (1.530, 2.166)	0.163	<0.001	—
Ns(year, 2) ₂	1.709 (1.246, 2.226)	0.249	<0.001	—
Drug	-0.134 (-0.346, 0.077)	0.107	0.211	—
Ascites	0.182 (0.112, 0.258)	0.037	<0.001	—
Spiders	0.132 (0.075, 0.189)	0.029	<0.001	—
Edema no diuretics	0.118 (0.057, 0.180)	0.031	<0.001	—
Edema despite diuretics	0.193 (0.089, 0.297)	0.052	<0.001	—
Histologic stage	0.036 (-0.005, 0.079)	0.022	0.084	—
Survival Outcome				
Sex	-0.403 (-0.845, 0.060)	0.227	0.080	0.668 (0.430, 1.062)
Drug	-0.048 (-0.459, 0.364)	0.214	0.830	0.953 (0.632, 1.439)
Ascites	0.574 (-0.067, 1.186)	0.320	0.075	1.775 (0.935, 3.274)
Edema no diuretics	0.780 (0.280, 1.247)	0.247	0.005	2.181 (1.324, 3.479)
Edema despite diuretics	0.940 (0.302, 1.544)	0.317	0.003	2.561 (1.352, 4.685)
Hepatomegaly	0.223 (-0.157, 0.596)	0.191	0.247	1.250 (0.854, 1.814)
Histologic stage	0.549 (0.311, 0.796)	0.126	<0.001	1.732 (1.365, 2.218)
Value[$\log(\text{serBilir})$]	1.357 (1.166, 1.566)	0.101	<0.001	3.884 (3.208, 4.785)

Std.Dev: Standard deviation.

在模型性能评价方面,考虑患者整体随访时间,设定预测终点为第10年,分别呈现随访第2、3、4、5、6年时的预测性能。模型在各时点的时间依赖

AUC值显示,在随访早期模型区分能力最强,随时间推移略有波动,但整体维持在较高水平,总体时间依赖AUC高达0.840(图2)。

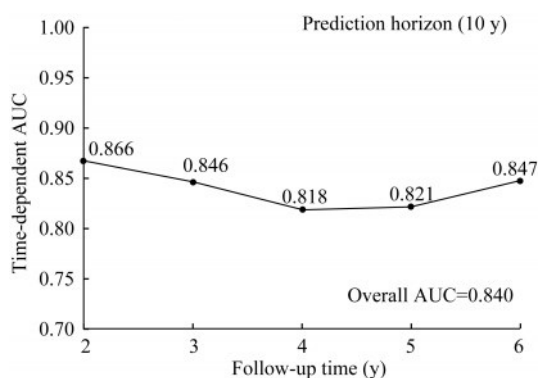


图2 最佳联合模型在不同随访时间的时间依赖AUC

Fig 2 Time-dependent AUC at different follow-up time for the optimal joint model

表3 多纵向指标与单纵向指标联合模型预测能力比较

Tab 3 Comparison of predictive performance among joint models with multiple longitudinal indicators and single longitudinal indicators

Longitudinal variables	Follow-up time (y)					Overall AUC
	2	3	4	5	6	
log (serBilir)	0.866	0.846	0.818	0.821	0.847	0.840
log (serBilir)+prothrombin	0.864	0.841	0.815	0.817	0.848	0.837
log (serBilir)+prothrombin+albumin	0.858	0.839	0.820	0.819	0.854	0.838

AUC: Area under curve.

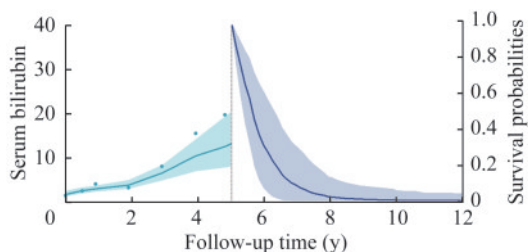


图3 联合模型个体生存概率预测

Fig 3 Individual survival probability prediction of joint model

结语 基于贝叶斯框架的联合模型为医学研究中纵向数据与生存数据的整合分析提供了强大而灵活的统计解决方案,突破了传统分析方法的局限。通过共享参数结构,贝叶斯联合模型有效纠正了传统分析中的偏倚,并借助MCMC算法解决了复杂模型的计算难题。其动态预测功能尤其适用于个性化医疗和临床实时决策支持。

需要注意的是,联合模型的应用前提是同时存在纵向与生存两类核心数据,且二者存在潜在关联^[22]。本文基于PBC2数据集的实证研究表明,贝叶斯联合模型在临床研究中展现出良好的拟合与预测性能。然而,当纳入多个纵向指标(如血清胆红素、凝血酶原时间和白蛋白)后,模型的预测效能

除了血清胆红素,进一步探索纳入多个纵向指标(如凝血酶原时间与白蛋白)共同构建的联合模型。结果显示,在此数据集中,加入额外纵向指标并未显著提升模型的整体预测区分度(overall AUC)(表3)。

采用纵向指标对目标患者的生存概率进行个体化动态预测。例如,整合93号患者随访前5年血清胆红素的纵向变化轨迹,来展示该患者至第12年的生存概率变化(图3)。结果显示,该患者的血清胆红素成上升趋势,其生存概率则随随访时间延长而持续下降,且前期下降速度较快,后期逐渐趋缓。

却未见显著提升,这可能与样本量、疾病类型或指标间的共线性有关^[10]。相比之下,一项针对肾移植患者的研究显示,同时整合肾小球滤过率和蛋白尿两个纵向指标后,模型的判别性能显著提高且更为稳定^[23],这印证了联合模型在多个纵向指标整合分析中的潜在价值。

目前,贝叶斯联合模型在医学领域的应用正逐步深化,但仍面临高维多模态数据整合、缺失数据处理及模型可解释性提升等关键挑战。未来研究可聚焦于以下几个方向:(1)拓展模型对高维多模态纵向数据(如影像组学、基因组学数据)的处理能力;(2)优化针对非随机缺失数据的联合建模方法;(3)结合贝叶斯模型平均(Bayesian model averaging, BMA)等技术,整合不同关联结构模型的预测结果以提升稳健性;(4)通过大规模临床试验和真实世界研究验证其应用价值,推动精准医学实践。随着计算工具的日益完善和个体化预后评估需求的增长,贝叶斯联合模型有望在更广泛的医学研究领域发挥重要作用。

作者贡献声明 倪长宇 数据整理,结果分析和解释,论文撰写。王筱金 论文构思。王炳顺

研究设计和指导,论文修订。

利益冲突声明 所有作者均声明不存在利益冲突。

参 考 文 献

- [1] ASAR Ö, RITCHIE J, KALRA PA, *et al.* Joint modelling of repeated measurement and time-to-event data: an introductory tutorial[J]. *Int J Epidemiol*, 2015, 44(1): 334-344.
- [2] IBRAHIM JG, CHU H, CHEN LM. Basic concepts and methods for joint models of longitudinal and survival data [J]. *J Clin Oncol*, 2010, 28(16): 2796-2801.
- [3] 翟映红, 陈琪, 韩贺东, 等. 联合模型介绍及在医学研究中的应用[J]. *中华流行病学杂志*, 2019, 40(11): 1456-1460.
- [4] RIZOPOULOS D. Joint modeling of longitudinal and time-to-event data: challenges and future directions [M/OL]// TORELLI N, PESARIN F, BAR-HEN A. *Advances in Theoretical and Applied Statistics*. Berlin, Heidelberg: Springer, 2013: 199-209. (2013-01-01) [2026-03-08]. https://doi.org/10.1007/978-3-642-35588-2_19.
- [5] TSIATIS AA, DEGRUTTOLA V, WULFSOHN MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS[J]. *J Am Stat Assoc*, 1995, 90(429): 27-37.
- [6] SWEETING MJ, THOMPSON SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture [J]. *Biom J*, 2011, 53(5): 750-763.
- [7] RIZOPOULOS D, GHOSH P. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event[J]. *Stat Med*, 2011, 30(12): 1366-1380.
- [8] GOULD AL, BOYE ME, CROWTHER MJ, *et al.* Joint modeling of survival and longitudinal non-survival data: Current methods and issues. Report of the DIA bayesian joint modeling working group[J]. *Stat Med*, 2015, 34(14): 2181-2195.
- [9] RIZOPOULOS D. *Joint models for longitudinal and time-to-event data: With applications in R* [M]. New York: Chapman and Hall/CRC, 2012.
- [10] RIZOPOULOS D, HATFIELD LA, CARLIN BP, *et al.* Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging[J]. *J Am Stat Assoc*, 2014, 109(508): 1385-1397.
- [11] YE W, LIN X, TAYLOR JMG. Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach [J]. *Biometrics*, 2008, 64(4): 1238-1246.
- [12] WOLBERS M, BABIKER A, SABIN C, *et al.* Pretreatment CD4 cell slope and progression to AIDS or death in HIV-infected patients initiating antiretroviral therapy—the CASCADE collaboration: a collaboration of 23 cohort studies[J]. *PLoS Med*, 2010, 7(2): e1000239.
- [13] RIZOPOULOS D. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC[J]. *J Stat Softw*, 2016, 72: 1-46.
- [14] GODANA AA, MOLLA BT, ABATIHUN D. Bayesian longitudinal modeling of blood pressure measurements of hypertensive patients at wachemo university nigist elleni mohamed memorial teaching and referral hospital hosanna, southern ethiopia[J]. *Heliyon*, 2023, 9(12): e22984.
- [15] YU M, TAYLOR JMG, SANDLER HM. Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model[J]. *J Am Stat Assoc*, 2008, 103(481): 178-187.
- [16] TAYLOR JMG, PARK Y, ANKERST DP, *et al.* Real-time individual predictions of prostate cancer recurrence using joint models[J]. *Biometrics*, 2013, 69(1): 206-213.
- [17] SPIEGELHALTER DJ, BEST NG, CARLIN BP, *et al.* Bayesian measures of model complexity and fit [J]. *J R Statist Soc B*, 2002, 64(4): 583-639.
- [18] GELMAN A, HWANG J, VEHTARI A. Understanding predictive information criteria for bayesian models[J]. *Stat Comput*, 2014, 24(6): 997-1016.
- [19] LI K, LUO S. Dynamic predictions in bayesian functional joint models for longitudinal and time-to-event data: An application to alzheimer's disease [J]. *Stat Methods Med Res*, 2019, 28(2): 327-342.
- [20] RIZOPOULOS D. JM: an R package for the joint modelling of longitudinal and time-to-event data [J]. *J Stat Softw*, 2010, 35: 1-33.
- [21] RIZOPOULOS D, MIRANDA-AFONSO P, PAPAGEORGIU G. JMbayes2: extended joint models for longitudinal and time-to-event data [CP/OL]. (2026-01-28) [2026-03-08]. <https://cran.r-project.org/web/packages/JMbayes2/index.html>.
- [22] LIU L, ZHENG C, KANG J. Exploring causality mechanism in the joint analysis of longitudinal and survival data[J]. *Stat Med*, 2018, 37(26): 3733-3744.
- [23] RAYNAUD M, AUBERT O, DIVARD G, *et al.* Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence: an observational, international, multicohort study [J]. *Lancet Digit Health*, 2021, 3(12): e795-e805.

(收稿日期:2026-03-14; 编辑:张秀峰)