

三种轨迹分析方法的比较研究*

王俊杰¹ 陈依娜¹ 陆梦依^{1,2} 尤东方¹ 赵 杨^{1,2,3△}

【摘要】 目的 介绍几种常见的轨迹分析方法,并对其使用效果和适用条件进行比较。方法 在模拟实验中对不同轨迹分析方法在各种情境下的效果以及适用条件进行比较研究,并在实例分析中进行验证。结果 增长混合模型(growth mixture modeling, GMM)在单轨迹分析中表现最好,基于群体的轨迹模型(group-based trajectory modeling, GBTM)总体表现较好且可以拓展对多轨迹进行研究,聚类分析表现尚可且能分析多轨迹,但对随访数据质量要求较高。结论 三种方法各有其优点,可根据分析数据的具体情况选择合适的方法。

【关键词】 轨迹分析 基于群体的轨迹模型 增长混合模型 聚类分析

【中图分类号】 R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.03.002

Comparative Study of Three Trajectory Analysis Methods

Wang Junjie, Chen Yina, Lu Mengyi, et al (Department of Biostatistics, School of Public Health, Nanjing Medical University (211166), Nanjing)

【Abstract】 **Objective** To introduce and compare different trajectory analysis methods. **Methods** Simulation experiments were conducted to compare the effect and applicable conditions of different trajectory analysis methods under different scenarios. Dataset of SPRINT was analyzed using these methods. **Results** Growth mixture modeling (GMM) performed best in single trajectory analysis while group-based trajectory modeling (GBTM) and cluster analysis were able to analyze multivariate trajectory. Cluster analysis had strict requirements on follow-up data. **Conclusion** Each of these three methods has its own advantages, and the appropriate method can be selected according to the specific situation of the dataset.

【Key words】 Trajectory analysis; Group-based trajectory modeling; Growth mixture modeling; Cluster analysis

疾病的发生发展是一种动态的过程,是某种暴露因素持续作用于机体的结果。而队列研究则是研究疾病动态发展过程的重要工具,在纵向队列研究中,可以在不同时间点测量个体的同一指标,研究暴露因素对疾病风险的长期效应。对于每个个体,这一指标的纵向数据可以看成一条轨迹(trajec-tory),对轨迹进行分析可以帮助我们理解暴露因素的纵向变化和疾病发生发展的关系。

对轨迹进行分析的方法可分为使用平均水平的方法和以个体为中心的方法^[1-4]。使用平均水平的方法主要包括方差分析(analysis of variance, ANOVA)以及多变量回归(multiple regression)。这类方法有一定的误导性,假设一组患者中一半的个体情况改善,另一半恶化时,那么平均值表明随着时间推移该组内部没有发生变化,但实际上该组内部的个体拥有两种相反的变化模式^[3,5]。由于这类方法的局限性,以个体为中心的方法日益得到重视,这类方法可找出总体内部发展趋势不同的亚组^[6],主要包括基于群体的轨迹模

型(group-based trajectory modeling, GBTM)、增长混合模型(growth mixture modeling, GMM)以及聚类分析(cluster analysis)^[3,7-9]。

针对轨迹分析方法进行比较的研究较少,在实际研究中关于这些方法如何选择尚无定论。本研究主要通过模拟实验比较不同轨迹分析方法对不同轨迹的区分能力,并在实例分析中进行应用。

基本原理与方法

1. 基于群体的轨迹模型

GBTM 假设了人群的离散分布且同一亚组中的个体之间没有差异(增长参数没有组内方差),因此可以在人群中区分出包括同质性个体的亚组,同一亚组内的个体具有相似的轨迹^[9-10]。用 $Y_i = \{y_{i1}, y_{i2}, \dots, y_{iT}\}$ 表示 T 周期内个体 i 的指标 Y 的纵向测量序列, $P(Y_i) = \sum_j \pi_j P^j(Y_i)$ 表示 Y_i 的概率(单一轨迹),当个体 i 属于第 j 组时, $y_{i1}, y_{i2}, \dots, y_{iT}$ 相互独立,因此 $P^j(Y_i) = \prod_t P^j(Y_{it})$ 可以表示个体 i 属于 j 组时 Y_i 的条件概率, π_j 表示该个体 i 属于第 j 组的条件概率。根据所研究的数据类型(连续、二分类、计数),可以用不同的概率分布函数来表示,利用最大似然估计可以得到不同轨迹组的参数。

设置不同的分组数量以及轨迹形状参数对不同模

* 基金项目: 国家自然科学基金(82173620, 82373690, 82204156); 国家重点研发计划(2022YFC3702702)

1. 南京医科大学公共卫生学院生物统计学系(211166)

2. 南京医科大学国家疫苗研发创新平台

3. 江苏省恶性肿瘤生物标志物与防治重点实验室, 肿瘤个性化医学省部共建协同创新中心

△通信作者: 赵杨, E-mail: yzhao@njmu.edu.cn

型进行分析,模型选择时的参考标准包括贝叶斯信息准则(Bayesian information criteria, BIC)、平均后验概率(average posterior probability, AvePP)、LMR-LRT 检验(Lo-Mendell-Rubin likelihood ratio test, LMR-LRT)以及基于 Bootstrap 似然比检验(Bootstrap likelihood ratio test, BLRT)等。该模型确定了每个个体属于一个亚组或另一个亚组的概率(后验组概率),每个个体最终被分配到后验概率最高的轨迹组。GBTM 单轨迹模型还可拓展为多轨迹模型以研究多个纵向指标间的相互作用以及它们的联合发展对疾病的影响。GBTM 是目前最常使用的轨迹分析方法,可使用 SAS 软件中的 PROC TRAJ 程序或者 R 软件中的 lcmm, gbmt 程序包实现。

2. 增长混合模型

与 GBTM 的原理类似,GMM 可以估计每个组的平均增长曲线,进行分析前也要假设潜在类别的数量以及每个类别的曲线形状,指定模型后,根据随时间变化的每条轨迹的形状设置参数。不同的是,GMM 通过在模型中引入随机效应,可以估计增长参数的方差,以表示个体参数和亚组平均参数的差异^[2,11-12],从而考虑到了组内每个个体之间的差异。例如第 k 组的个体 i 的第 j 次测量指标可以用

$$Y_{ij} = (\beta_0^k + b_{0i}^k) + (\beta_1^k + b_{1i}^k) X_{ij} + (\beta_2^k + b_{2i}^k) X_{ij}^2 + \varepsilon_{ij}$$

表示。模型参数的估计可以通过极大似然或贝叶斯方法实现,模型的选择标准包括 BIC, AvePP, 每个轨迹中样本比例以及 LMR-LRT 检验等。可使用 R 软件中的 lcmm 程序包实现。

3. 聚类分析

前述的两种方法要求研究的纵向指标为连续资料时服从正态分布,在使用时可能会受到限制。聚类分析可以作为替代的方法,它不需要任何的正态性或者参数假设。可以利用轨迹之间的距离来进行聚类分析,例如 R 软件中的 kml 程序包^[13],kml 包聚类方法的基本原理是 K-均值算法,先设置几个聚类中心,根据轨迹间的距离将每个个体分配到距离最近的类别中,然后更新每种类别的中心,重复这一过程直到每个类别不再改变,最终所有个体的轨迹都被分配到一定数量的轨迹组内。最佳分类数量可以使用 Calinski & Harabasz 准则, Ray & Turi 准则以及 Davies & Bouldin 准则等来选择^[14]。

4. 单轨迹情形下的模拟实验

笔者基于以下模型产生模拟数据:个体 i 的第 j 次测量指标可表示为与时间 t 相关的函数:

$$Y_{ij} |_{G=g} = (\mu_{0g} + \nu_{0g}) + (\mu_{1g} + \nu_{1g}) \times t_{ij} + (\mu_{2g} + \nu_{2g}) \times t_{ij}^2 + \varepsilon_{ij}$$

其中 $\mu_{0g}, \mu_{1g}, \mu_{2g}$ 为固定效应系数,在不同的轨迹亚组间不同, $\nu = [\nu_{0g}, \nu_{1g}, \nu_{2g}]^T$ 为随机效应系数,在每个个体中均不相同,用于描述每个个体偏离平均轨迹的情

$$\text{况, } \nu = [\nu_{0g}, \nu_{1g}, \nu_{2g}]^T \sim N(0, D), D = \begin{pmatrix} 0 & 3 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}; \varepsilon_{ij} \text{ 为符}$$

合正态分布的随机误差, $\varepsilon_{ij} \sim N(0, 1)$; 每组包含 100000 个体,每个个体测量 10 次(0.1~1.0 共 10 个时间点)。每次模拟从总样本中抽取 100 个个体,重复抽取 100 次。模拟基于总人群中存在两个或三个潜在类别进行,每个类别占总体的比例均相同。通过改变模型参数中的固定效应系数改变平均轨迹的形状,如逐渐上升、逐渐下降或者先升后降等,在平均轨迹的基础上为不同个体加入随机效应后生成一个轨迹组,设置的情境则为生成轨迹组的不同组合,研究不同形状的轨迹组混合在一起时不同方法对于轨迹组的识别能力。该模拟实验共生成 5 个轨迹组并将其组合以设置 6 个不同情境,轨迹组平均轨迹形状的参数设置如表 1,不同情境的平均轨迹组合情况如图 1、2。

5. 多轨迹情形下的模拟实验

设置两个测量指标 y 和 Y , 总体中共包含两个分组,设置两个组的平均轨迹的函数,第一组: $y = 5t^2 + 6t + 10, Y = -2t^2 - 3t + 10$, 第二组: $y = -16t^2 + 16t + 10, Y = 16t^2 - 16t + 10$ 。每个轨迹组内部两个指标间设置从 -1 到 1 的相关系数(-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1)。每组样本量设为 100000,每个个体测量 10 次(取 0.1~1.0 共 10 个时间点),根据设定情境的平均轨迹,代入该点时间取值计算两个指标在所有个体中的平均值,该点所有个体的两个指标满足二元正态分布,通过设置协方差矩阵控制指标间的相关系数以及指标内部的变异大小,并生成所有个体在该时间点 y 和 Y 的值。每次模拟从总样本中抽取 100 个个体,重复抽取 100 次。主要研究多轨迹模拟识别轨迹组的能力以及指标间相关系数的改变对不同方法效果的影响。

6. 模拟实验评价指标

包括正确分类率,相对偏倚和对轨迹形状的还原效果。正确分类率指被正确分类的个体占总样本量的百分比,越高则代表分类精度越高。相对偏倚指 GMM 和 GBTM 两种方法中得到的轨迹参数减去真实设置参数的差值与设置参数比值的绝对值。对每种方法得到的平均轨迹图像与设置情境的真实图像比较,如果形状不一致则判定为该方法错误分类并不再计算其他评价指标。

表 1 单轨迹模拟实验平均轨迹参数设置

轨迹组 (g)	截距 (u_0)	一次项 (u_1)	二次项 (u_2)
1	10	6	5
2	10	-3	-2
3	10	16	-16
4	10	-16	16
5	15	-3	-2

7. 实例分析

本研究利用 SPRINT(收缩压干预实验)数据库分析不同方法的效果^[15]。SPRINT 共招募了 9361 名参与者,参与者被分配不同的收缩压控制目标,即低于 140 mmHg(标准治疗组)或低于 120 mmHg(强化治疗组),在 3.26 年的随访中,强化治疗组和标准治疗组的平均收缩压分别为 121.5 mmHg 和 134.6 mmHg。

结 果

1. 单轨迹模拟实验结果

在包含两组轨迹的情境一、二、四以及包含三组轨迹的情境六中,三种方法都能正确识别出轨迹的形状,并且有较高的平均正确分类率,其中 GMM 的识别效果最好,对模拟轨迹的固定效应系数的估计精度也高于 GBTM。GBTM 和聚类分析的精度相似,但和 GMM 有一定差距。在情境三和情境五中,GMM 依然

识别出了正确的轨迹组形状,分类也较为准确,而其他两种方法得到的轨迹形状与模拟设置的不同,未能识别出正确的轨迹组。对 GMM 和 GBTM 所得轨迹参数的相对偏倚进行比较,可以发现 GMM 对于模拟设置参数的估计更加精确。具体的平均正确分类率和系数估计的相对偏倚的比较情况见表 2~4,初始模拟和三种方法得到的平均轨迹图像见图 1、2。

表 2 不同单轨迹分析方法的平均正确分类率

情境	GMM	GBTM	聚类
情境一	0.9977	0.9855	0.9856
情境二	0.9986	0.9024	0.9037
情境三	0.9840	错误分类	错误分类
情境四	0.9988	0.9508	0.9507
情境五	0.9797	错误分类	错误分类
情境六	0.9958	0.9216	0.9223

注:所得平均轨迹图像与设置情境不同时判定为错误分类,评价指标不计算。

表 3 GMM 和 GBTM 所得轨迹参数的相对偏倚(情境一到情境四)

情境		GMM		GBTM	
		第一组	第二组	第一组	第二组
情境一	二次项	-0.0422	0.0070	-0.0777	-0.0054
	一次项	0.0319	-0.0097	0.0196	-0.0146
	截距	0.0006	0.0015	-0.0086	0.0107
情境二	二次项	-0.0064	0.0044	-0.1084	-0.3427
	一次项	-0.0069	-0.0053	-0.0856	0.2111
	截距	0.0008	0.0013	-0.0386	0.0401
情境三	二次项	-0.0370	0.0015	错误分类	错误分类
	一次项	0.0363	0.0009	错误分类	错误分类
	截距	0.0002	0.0020	错误分类	错误分类
情境四	二次项	-0.0461	0.0065	-0.1214	-0.0267
	一次项	0.0330	-0.0079	-0.0305	-0.0433
	截距	0.0004	0.0017	-0.0289	0.0464

表 4 GMM 和 GBTM 所得轨迹参数的相对偏倚(情境五、情境六)

情境		GMM			GBTM		
		第一组	第二组	第三组	第一组	第二组	第三组
情境五	二次项	-0.0045	-0.0029	-0.0044	错误分类	错误分类	错误分类
	一次项	0.0055	-0.0285	-0.0110	错误分类	错误分类	错误分类
	截距	-0.0117	-0.0127	-0.0102	错误分类	错误分类	错误分类
情境六	二次项	-0.0005	0.0226	0.0017	0.0031	-0.7415	-0.0791
	一次项	0.0067	-0.0207	-0.0003	-0.0372	0.1666	-0.0508
	截距	-0.0025	-0.0047	-0.0004	0.0396	-0.0274	-0.0189

2. 多轨迹模拟实验结果

当轨迹指标内部变异较小时,两类方法的正确分类率都较高,所以增大指标内部方差进行比较。内部方差即每个轨迹组内同一时间点所有个体测量指标的方差,越大则组内变异越大。当内部方差为 8 时,两种方法的精度没有明显的差异,也未观察到分类效果随

着相关性的变化而变化。当内部方差增加到 32 和 64 时,两种方法的精度均下降,但 GBTM 要优于聚类分析。并且当两个指标间的相关性越强时,GBTM 的分类效果越好,但聚类分析的分类效果与指标间的相关性大小无关,见表 5 和图 3。

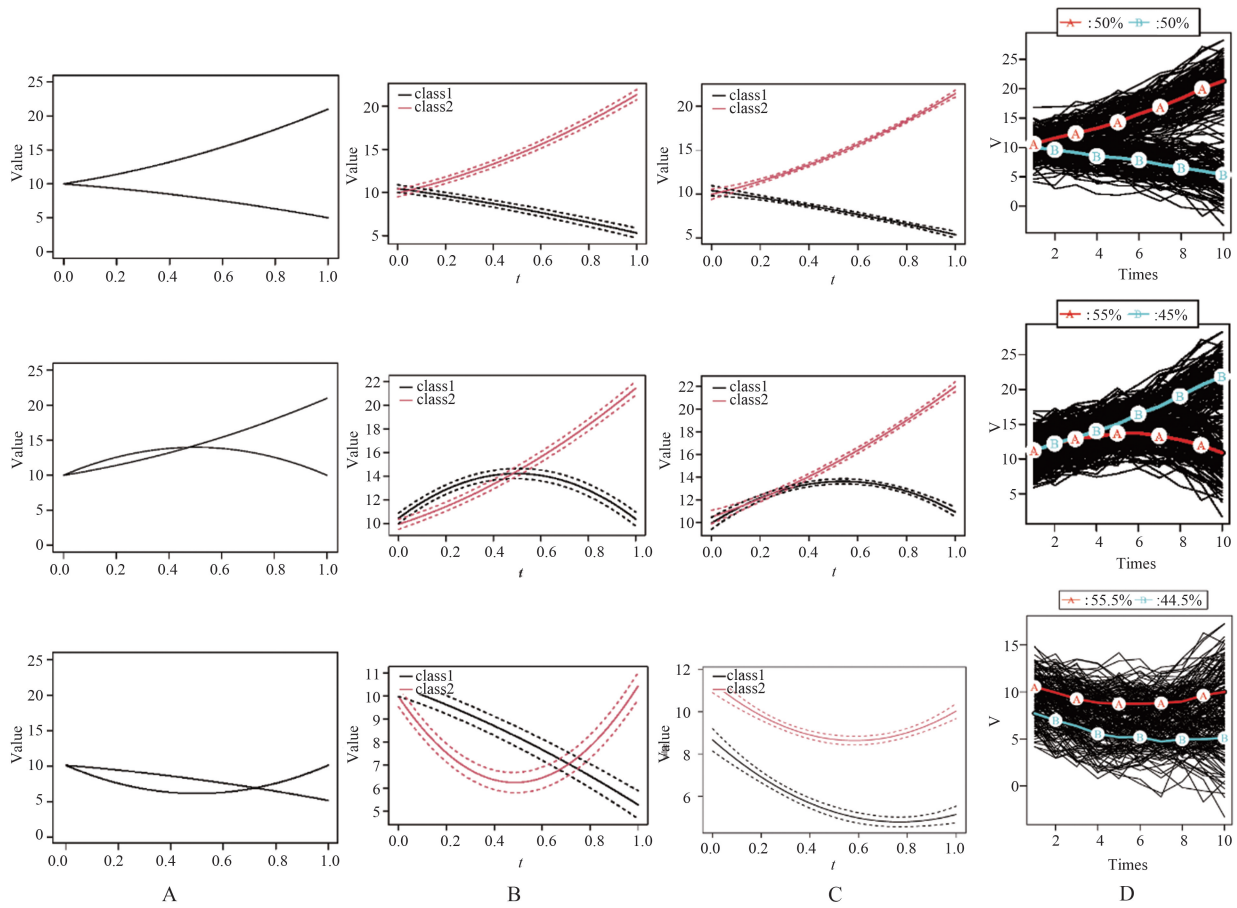


图1 情境一至三下(对应图片1~3行)初始模拟和三种方法的平均轨迹及分组情况 (A:模拟;B:GMM;C:GBTM;D:聚类分析)

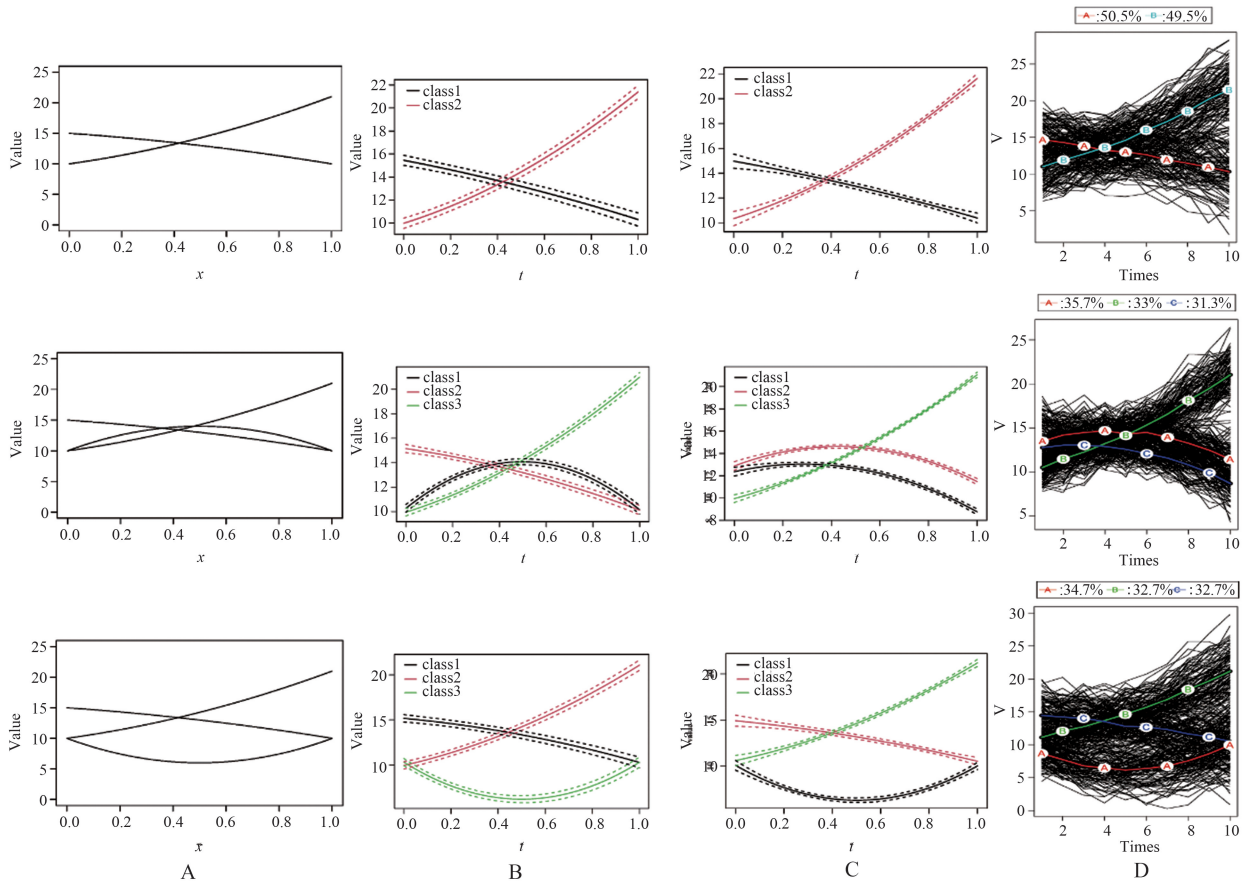


图2 情境四至六下(对应图片1~3行)初始模拟和三种方法的平均轨迹及分组情况 (A:模拟;B:GMM;C:GBTM;D:聚类分析)

表 5 不同指标内部方差(8,32,64)及不同指标间的相关性下两种方法的平均正确分类率

相关系数	GBTM			聚类		
	8	32	64	8	32	64
1	0.93500	0.97000	0.98000	1	0.91305	0.59540
0.75	1	0.99600	1	1	0.94940	0.71140
0.5	0.99995	0.97450	0.83450	0.99995	0.94890	0.82985
0.25	0.99950	0.94990	0.91265	0.99945	0.93920	0.71650
0	0.99865	0.92855	0.82415	0.99870	0.92650	0.77775
-0.25	0.99800	0.91785	0.86680	0.99660	0.91095	0.82055
-0.5	0.99835	0.92100	0.82095	0.99375	0.88915	0.80830
-0.75	0.99970	0.94975	0.97035	0.98970	0.86920	0.63080
-1	1	1	0.86775	0.98415	0.84505	0.74395

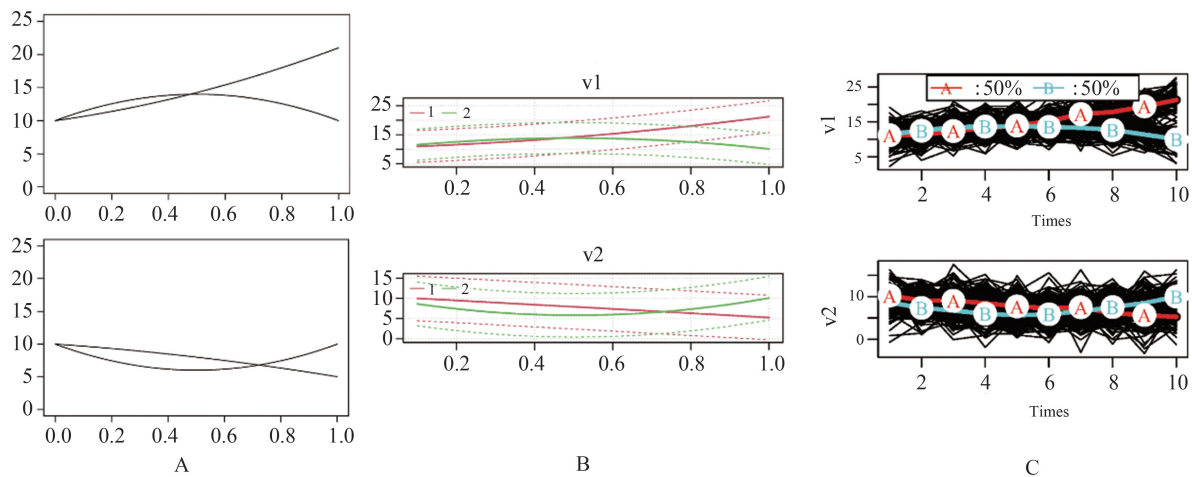


图 3 多轨迹模拟情境(两个变量)下初始模拟以及两种方法的平均轨迹及分组情况 (A:模拟;B:GBTM;C:聚类分析;)

单轨迹分析方法,设置分组数量为 2 组,三种方法得到的平均轨迹见图 4,大致符合上述参与者的实际情况,三种方法的正确分类率分别为 84.12%、82.89% 和 82.51%,效果较为理想。

表 6 SPRINT 数据中标准治疗组和强化治疗组参与者的基线基本情况

基线特征	强化治疗组 (n=4678)	标准治疗组 (n=4683)
年龄	67.9 ± 9.4	67.9 ± 9.5
女性 (%)	1684 (36.0)	1648 (35.2)
基线血压 (mmHg)		
收缩压	139.7 ± 15.8	139.7 ± 15.4
舒张压	78.2 ± 11.9	78.0 ± 12.0
收缩压分布		
≤ 132 mmHg	1583 (33.8)	1553 (33.2)
132 mmHg~145 mmHg	1489 (31.8)	1549 (33.1)
≥ 145 mmHg	1606 (34.3)	1581 (33.8)
吸烟状态		
从不	2050 (43.8)	2072 (44.2)
当前	639 (13.7)	601 (12.8)
过去	1977 (42.3)	1996 (42.6)
BMI(kg/m ²)	22.9 ± 5.8	29.8 ± 5.7

使用 GBTM 和聚类分析分别对 SPRINT 数据的

3.实例分析

2010 年 11 月至 2013 年 3 月期间共 9361 名参与者入组并接受了基线测量,参与者被随机分组至标准治疗组和强化治疗组接受不同的血压控制治疗,两组参与者的主要基线指标见表 6,主要基线指标在两个治疗组间无显著差异^[15]。

本研究先使用单轨迹分析的方法对 SPRINT 数据库参与者的收缩压轨迹进行分析,观察验证现有方法是否能够区分出两个血压控制组,并对其效果进行比较。SPRINT 数据库的参与者的真实收缩压变化轨迹大致可分为两组,一组从高血压水平下降至 135 mmHg 左右,另一组从高血压水平迅速降低至 120 mmHg 左右。使用 GMM,GBTM 和聚类分析三种

收缩压和舒张压以及尿素氮和碳酸氢盐两组指标进行多轨迹分析。GBTM 设置轨迹最大次项为 2 次,从 1 至 4 组拟合模型;聚类分析结果为使用 Calinski & Harabasz 准则自动输出的最佳分组数量。GBTM 模型选择情况见表 7、8,两种方法得到的平均轨迹见图 5、6。使用轨迹分析对所有参与者进行分组后,为探究不同轨迹分组对健康结局的影响,选择发生数量较多的全因死亡作为研究结局,调整年龄、性别、心血管疾病史,使用 logistic 回归分析不同轨迹组对全因死亡结局的影响。第一组分析中 GBTM 和聚类分析最佳分组数量分别为 4、2,GBTM 得到的四组轨迹中,第 1 组内的参与者的收缩压水平最低而舒张压水平最高,第 2 组内收缩压水平最高而舒张压水平最低,第 1、4、3、2 组之间呈现出收缩压水平逐渐提高而舒张压逐渐降低的趋势,第 1、3、4 组相比第 2 组死亡风险更低,提示脉压差较大的人群可能有更高的死亡风险;聚类分析得到的两组轨迹中,B 组收缩压和舒张压相比 A 组都更高,两组间死亡风险无明显差异;第二组分析中 GBTM 和聚类分析最佳分组数量分别为 4、2,GBTM

表 7 GBTM 对收缩压和舒张压多轨迹分析的模型选择

分组数量	BIC	成员比例	AvePP
1	2274267	100%	100%
2	2178712	43.9%/56.1%	98.3%/98.5%
3	2137202	28.8%/26.1%/42.1%	98.0%/98.0%/96.9%
4	2116733	22.9%/13.1%/ 28.7%/35.2%	97.5%/97.9%/ 96.1%/96.0%

表 8 GBTM 对尿素氮和碳酸氢盐多轨迹分析的模型选择

分组数量	BIC	成员比例	AvePP
1	980221	100%	100%
2	922549	27.1%/72.9%	96.5%/98.1%
3	899451	15.1%/41.6%/43.2%	96.6%/95.8%/95.0%
4	887655	30.2%/14.6%/ 27.8%/27.3%	93.9%/96.8%/ 91.5%/92.6%

的四组轨迹中,第 2 组尿素氮水平最高而碳酸氢盐水平最低,其余组相比第 2 组尿素氮水平较低而碳酸氢盐水平较高,与第 1 组比,第 2 组死亡风险更高而第 4 组死亡风险更低;聚类分析得到的两组轨迹的特征与

GBTM 相似,B 组相比 A 组尿素氮水平较高而碳酸氢盐水平较低,B 组相比 A 组死亡风险更高,虽然两种多轨迹分析方法分组结果并不一致,但都提示人群中尿素氮水平较高而碳酸氢盐水平较低的个体可能有更高的死亡风险。具体见表 9、10。

表 9 对收缩压和舒张压轨迹组进行 logistic 分析的结果

方法	参考组	比较组	OR	P
GBTM	2	1	0.58	0.003
		3	0.57	<0.001
		4	0.54	<0.001
聚类分析	A	B	0.99	0.910

表 10 对尿素氮和碳酸氢盐轨迹组进行 logistic 分析的结果

方法	参考组	比较组	OR	P
GBTM	1	2	2.03	<0.001
		3	0.83	0.165
		4	0.68	0.007
聚类分析	A	B	2.26	<0.001

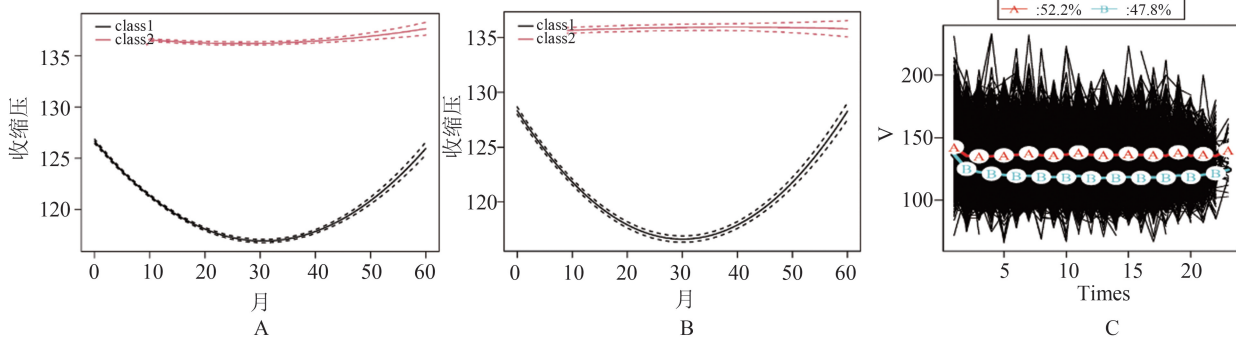


图 4 对 SPRINT 数据收缩压进行单轨迹分析的情况(A:GBTM; B:GMM; C:聚类分析)

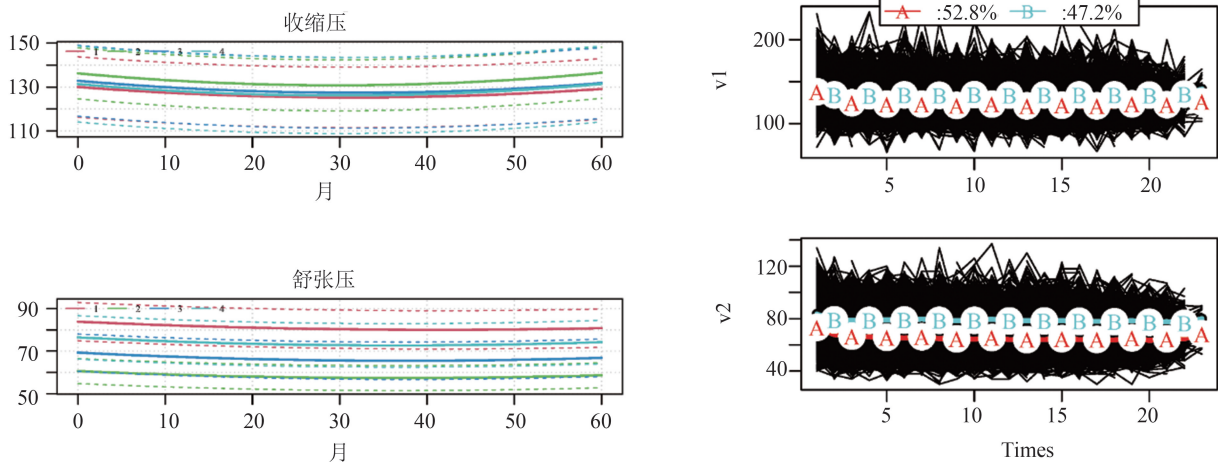


图 5 GBTM 和聚类分析对 SPRINT 收缩压及舒张压的多轨迹分析情况

讨论

轨迹分析方法最初多用于社会学、心理学等领域^[2,16-20],近年来,这些方法也被临床或流行病学相关的研究人员所接受^[19,21]。例如有研究通过分析体力

活动指标相关轨迹,发现增加体力活动和卒中后功能恢复相关^[22];还有研究分析了慢性阻塞性肺疾病患者的肺活量轨迹对死亡风险的影响^[23]。

本文对目前常用的几类轨迹分析方法进行比较,单轨迹模拟实验中,GMM 的效果相比 GBTM 和聚类

分析更好,多轨迹模拟实验中,GBTM 的效果相对聚类分析较好,并且当设置的指标间相关性越高,其分类也越好。聚类分析作为非参数方法对于数

据分布的要求宽松,还可以拓展对多轨迹进行分析,在通过多项式方法模拟的数据中也取得了和参数方法接近的效果。

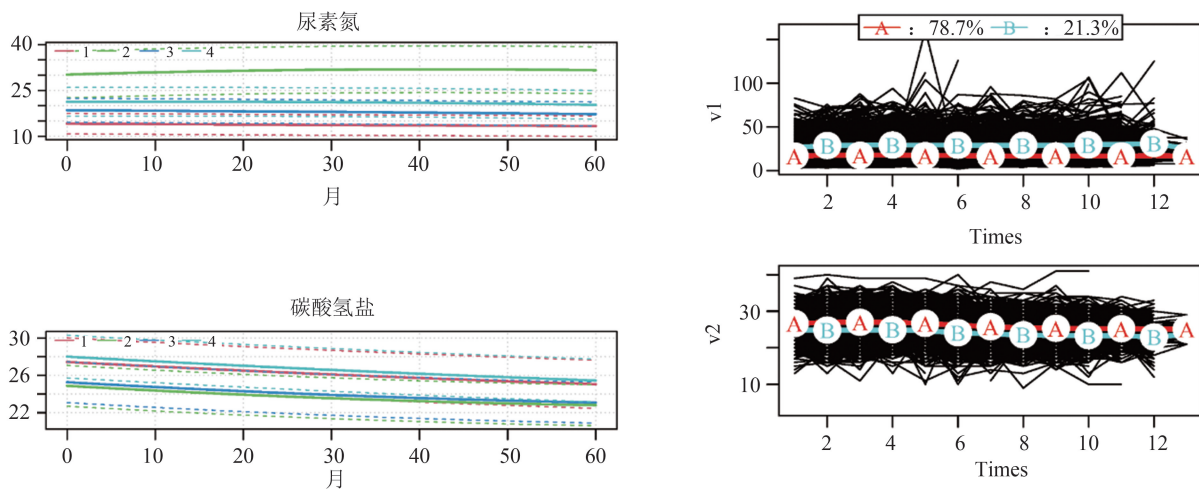


图 6 GBTM 和聚类分析对 SPRINT 尿素氮及碳酸氢盐的多轨迹分析情况

在实例分析中,GBTM、GMM 和聚类分析都能识别出不同的单轨迹分组且分类效果较好。聚类分析使用存在局限,因为它要求总体中每个参与者有相同的随访间隔以便于计算轨迹间的距离,而现实中数据对不同的参与者的随访并不都严格遵循相同的间隔进行。GBTM 和 GMM 对于个体随访时间没有严格的要求,所以在实际应用中更有广泛的适用性。轨迹分析方法在实际应用中最主要的问题就是对分组数量的选择,虽然不同的方法都提供了选择标准,包括 BIC、AIC,结合 AvePP(越高越好且不低于 70%)、轨迹组人群比例(最好不低于 5%)决定最终模型^[10],但是选择不同的判定标准可能会产生不同的结果,有时还需要相关领域的临床知识来下最终结论,所以最终模型的选择依然带有一定的主观性。在多轨迹分析中,GBTM 和聚类分析并未得到相似的分组且轨迹间的区分度不是很高,这可能是由于参与者的指标轨迹异质性和波动幅度较低,但得到的部分轨迹分组间死亡风险存在差异,依然有一定的指导意义。

本研究存在一些局限,首先模拟的情形中随访时间间隔在个体间一致,轨迹组的数量以及形状较为简单,实际应用中情况可能会更复杂;多轨迹模拟中只考虑了一种两个指标变量分析的情境,未设置更复杂的情境;在 SPRINT 数据单轨迹分析中对轨迹分类能力进行验证时,事先确定了轨迹组的数量,而在实际应用中不会出现这种情况。

综上,结合模拟实验和实例分析,三类方法中 GMM 在单轨迹分析中表现最好。GBTM 和聚类分析的表现较为稳定且可以拓展进行多轨迹分析以研究多种指标或者暴露对疾病进程的影响。聚类分析虽然对研究资料类型无限制且不需要设置参数,但是对随访

数据的质量又有较高要求。在实际研究中,我们应根据具体情况选择合适的方法。

参 考 文 献

- [1] Muthén B, Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcohol Clin Exp Res*, 2000, 24(6) : 882-891.
- [2] Jung T, Wickrama KaS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*, 2008, 2(1) : 302-317.
- [3] Frankfurt S, Frazier P, Syed M, et al. Using Group-Based Trajectory and Growth Mixture Modeling to Identify Classes of Change Trajectories. *The Counseling Psychologist*, 2016, 44(5) : 622-660.
- [4] Nguena Nguéack HL, Pagé MG, Katz J, et al. Trajectory Modeling Techniques Useful to Epidemiological Research: A Comparative Narrative Review of Approaches. *Clin Epidemiol*, 2020, 12: 1205-1222.
- [5] Westen D, Novotny CM, Thompson-Brenner H. The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychol Bull*, 2004, 130(4) : 631-663.
- [6] Curran PJ, Hussong AM. The use of latent trajectory models in psychopathology research. *J Abnorm Psychol*, 2003, 112(4) : 526-544.
- [7] Curran PJ, Muthén BO. The application of latent curve analysis to testing developmental theories in intervention research. *Am J Community Psychol*, 1999, 27(4) : 567-595.
- [8] Heggeseth BJD, Gradworks T. Longitudinal Cluster Analysis with Applications to Growth Trajectories. *Dissertations & Theses - Graduorks*, 2013:4-13.
- [9] Jones BL, Nagin DS. Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. *Sociological Methods & Research*, 2007, 35(4) : 542-571.
- [10] Nagin DS. Group-based modeling of development. Harvard University Press, 2005:23-27.

- [11] Kaplan D. The Sage handbook of quantitative methodology for the social sciences. Sage Publications, 2004; 345-367.
- [12] Van Der Nest G, Lima Passos V, Candel M, et al. An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Adv Life Course Res*, 2020, 43: 100323.
- [13] Genolini C, Falissard B. Kml: a package to cluster longitudinal data. *Comput Methods Programs Biomed*, 2011, 104(3): e112-121.
- [14] Genolini C, Falissard B. Kml: k-means for longitudinal data. *Computational Statistics*, 2010, 25(2): 317-328.
- [15] Wright JT Jr, Williamson JD, Whelton PK, et al. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med*, 2015, 373(22): 2103-2116.
- [16] Caspi A, Roberts B. Personality Development Across the Life Course: The Argument for Change and Continuity. *Psychological Inquiry*, 2001, 12: 49-66.
- [17] Rod NH, Bengtsson J, Budtz-Jørgensen E, et al. Trajectories of childhood adversity and mortality in early adulthood: a population-based cohort study. *Lancet*, 2020, 396(10249): 489-497.
- [18] Laurenceau JP, Hayes AM, Feldman GC. Some methodological and statistical issues in the study of change processes in psychotherapy. *Clin Psychol Rev*, 2007, 27(6): 682-95.
- [19] Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*, 2010, 6: 109-138.
- [20] Girard LC, Tremblay RE, Nagin D, et al. Development of Aggression Subtypes from Childhood to Adolescence: a Group-Based Multi-Trajectory Modelling Perspective. *J Abnorm Child Psychol*, 2019, 47(5): 825-838.
- [21] Burckhardt P, Nagin DS, Padman R. Multi-Trajectory Models of Chronic Kidney Disease Progression. *AMIA Annu Symp Proc*, 2016, 2016: 1737-1746.
- [22] Buvarp D, Viktorisson A, Axelsson F, et al. Physical Activity Trajectories and Functional Recovery After Acute Stroke Among Adults in Sweden. *JAMA Netw Open*, 2023, 6(5): e2310919.
- [23] Backman H, Blomberg A, Lundquist A, et al. Lung Function Trajectories and Associated Mortality among Adults with and without Airway Obstruction. *Am J Respir Crit Care Med*, 2023, 208(10): 1063-1074.

(责任编辑:郭海强)

(上接第 330 页)

究,后续研究中将会对不同资料类型,不同信息时间等情况加以探讨。

参 考 文 献

- [1] Wassmer G, Brannath W. Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. Cham: Springer International Publishing, 2016; 133-134.
- [2] Cui L, Hung HMJ, Wang SJ. Modification of Sample Size in Group Sequential Clinical Trials. *Biometrics*, 1999, 55(3): 853-857.
- [3] Shen Y, Fisher L. Statistical Inference for Self-Designing Clinical Trials with a One-Sided Hypothesis. *Biometrics*, 1999, 55(1): 190-197.
- [4] Denne JS. Sample size recalculation using conditional power. *Statistics in Medicine*, 2001, 20(17-18): 2645-2660.
- [5] Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Statistics in Medicine*, 2011, 30(28): 3267-3284.
- [6] Mehta C, Liu L. An objective re-evaluation of adaptive sample size re-estimation: commentary on 'Twenty-five years of confirmatory adaptive designs': COMMENTARY. *Statistics in Medicine*, 2016, 35(3): 350-358.
- [7] Jennison C, Turnbull BW. Adaptive sample size modification in clinical trials: start small then ask for more. *Statistics in Medicine*, 2015, 34(29): 3793-3810.
- [8] Jennison C, Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect.. *Statistics in Medicine*, 2003, 22(6): 971-993.
- [9] Mehta C, Bhingare A, Liu L, et al. Optimal adaptive promising zone designs. *Statistics in Medicine*, 2022, 41(11): 1950-1970.
- [10] Emerson SS, Levin GP, Emerson SC. Comments on 'Adaptive increase in sample size when interim results are promising: A practical guide with examples'. *Statistics in Medicine*, 2011, 30(28): 3285-3301.
- [11] Chen YJ, Li C, Lan KG. Sample size adjustment based on promising interim results and its application in confirmatory clinical trials. *Clinical Trials*, 2015, 12(6): 584-595.
- [12] Pilz M, Kieser M, Kunzmann K. Comments on "Adaptive sample size modification in clinical trials: Start small then ask for more". *Statistics in Medicine*, 2020, 39(1): 97-98.
- [13] Kieser M. Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials. Cham: Springer International Publishing, 2020; 270-271.
- [14] Broberg P. Sample size re-assessment leading to a raised sample size does not inflate type I error rate under mild conditions. *BMC Medical Research Methodology*, 2013, 13(1): 94.
- [15] Lan KKG, Wittes J. The B-Value: A Tool for Monitoring Data. *Biometrics*, 1988, 44(2): 579.
- [16] Gao P, Ware JH, Mehta C. Sample Size Re-Estimation for Adaptive Sequential Design in Clinical Trials. *Journal of Biopharmaceutical Statistics*, 2008, 18(6): 1184-1196.
- [17] Liu Y, Hu M. Testing multiple primary endpoints in clinical trials with sample size adaptation. *Pharmaceutical Statistics*, 2016, 15(1): 37-45.

(责任编辑:郭海强)