状态序列分析在慢性病进展轨迹研究中的应用*

刘 洋[1,2] 李春霞[1,2] 曹 雯[1,2] 吴秉杰[1,2] 张 涛[1,2,3△]

【提 要】 **目的** 介绍状态序列分析(state sequence analysis, SSA)的基本原理,并通过实例阐述其在慢性病进展轨迹研究中的应用。**方法** SSA将个体随访期间经历的事件视为由不同状态构成的有序列表,通过计算序列间的不相似性并结合聚类分析,识别、描述和分析复杂的纵向模式。实例基于多中心健康体检项目数据,采用SSA识别典型慢性病进展轨迹,并应用logistic回归分析探索与慢性病进展轨迹组相关的潜在影响因素。**结果** 本研究共纳入8952名研究对象,识别出四种不同的慢性病进展轨迹。logistic回归分析结果显示,甘油三酯-葡萄糖指数、低密度脂蛋白/高密度脂蛋白比值、谷草转氨酶/谷丙转氨酶比值为健康体检人群慢性病进展轨迹的影响因素(P<0.05)。**结论** SSA是慢性病进展轨迹研究的有效工具,可为制定针对性的干预措施提供科学依据。

【关键词】 状态序列分析 纵向数据 健康体检队列 慢性病
【中图分类号】 R195.1 【文献标志码】 A **DOI** 10.11783/j.issn.1002-3674.2026.04.003

Application of State Sequence Analysis in Studying Chronic Disease Progression Trajectories

Liu Yang, Li Chunxia, Cao Wen, et al(*Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan* 250012)

【**Abstract**】 **Objective** To introduce the basic principles of state sequence analysis (SSA) and demonstrate its application in chronic disease progression trajectories research through an example. **Methods** SSA considers the events experienced by individuals during follow-up as an ordered list of different states. By calculating sequence dissimilarity and applying cluster analysis, SSA can identify, describe, and analyze complex longitudinal patterns. In this study, SSA was applied to multicenter health examination data to identify typical chronic disease progression trajectories, and logistic regression analysis was used to explore potential factors associated with these trajectories. **Results** A total of 8952 participants were included in the study, and four distinct chronic disease progression trajectories were identified. Logistic regression analysis revealed that the triglyceride-glucose index, the low-density lipoprotein/high-density lipoprotein ratio, and the aspartate aminotransferase/alanine aminotransferase ratio were factors influencing chronic disease progression trajectories in the health examination population (P<0.05). **Conclusion** SSA is an effective tool for studying chronic disease progression trajectories and can provide a scientific basis for developing targeted interventions.

【**Key words**】 State sequence analysis; Longitudinal data; Health examination cohort; Chronic disease

随着我国人口老龄化加剧,慢性病发病率和死亡率迅速上升,识别其进展轨迹并把握关键干预时间点已成为慢性病防控的热点问题[1]。目前分析慢性病进展轨迹的方法主要有两类:第一类,如潜类别分析[2]、聚类分析[3]和多状态马尔科夫模型[4],前两者主要聚焦于慢性病并发症的聚类,而后者则关注慢性病之间的转移强度。这些方法均有助于识别始发和继发疾病,但无法展示个体的完整慢性病进展过程。第二类,如组基轨迹模型[5]和有监督机器学习[6],通过分析患病数量或共病得分反映个体的完整慢性病进展过程,但难以明确区分其中的具体疾病。因此,现有方法在进行慢性病进展轨迹研究时会损失部分统计信息,而这种信息偏倚限制了对慢性病进展的深入理解。

状态序列分析(state sequence analysis, SSA)由Abbott和Forrest提出[7],并首次应用于社会科学领域[8]。该方法基于个体随访期间事件由一系列离散状态组成的思想,将慢性病进展过程视为由不同慢性病状态构成的序列。因此,在分析时能够同时考虑慢性病进展的动态变化及各慢性病的发生顺序。此外,SSA结合聚类分析能够识别典型的轨迹组,再结合回归分析方法可进一步比较各组之间的特征差异。本文将介绍SSA的基本原理,并采用该方法分析多中心健康体检人群的慢性病进展轨迹模式。

基本原理

1. 状态序列分析基本概念

SSA是一种探索性分析方法,该方法通过将个体在随访期间内所经历的事件描述为由不同类别状态构成的有序列表。随后量化序列间的成对不相似性,并构建不相似矩阵。基于该不相似性矩阵对序列进行聚类,从而识别出具有代表性的轨迹模式。

与其他方法直接使用原始数据进行分析不同,

*基金项目:国家自然科学基金(82222064; 82473730)
1.山东大学齐鲁医学院公共卫生学院生物统计学系(250012)
2.山东大学健康医疗大数据研究院
3.天津医科大学公共卫生学院
△通信作者:张涛,E-mail:taozhang@sdu.edu.cn

SSA首先需将这些数据转换为适合其分析的状态序列格式（如图1所示）。具体来说，研究时间区间 $[0, T]$ 需划分为 l 个等间隔的离散时间段，因此个体在此期间内经历的事件可映射为由 l 个有序状态构成的序列。假设有限状态集合 $S = s_1, s_2, s_3, \cdots, s_m$ 表示个体所有可能存在的状态，进而个体在时间段 t 上的状态 s_m 可表示为 $X(t) = s_m,\ s_m \in S$。此外，在SSA中，时间段称为时间单位（time unit）或时间点（time point）；有限集合 S 被称为字母表（alphabet）；有限集合内的状态被称为元素（element）。综上所述，在时间区间 $[0, T]$ 内序列长度为 l 的状态序列可被定义为从字母表 S 中按照所个体经历的事件顺序依次选择的 l 个元素的有序列表：

$$X = (X_1,\ X_2,\ X_3,\ \ldots,\ X_l), \quad X(t) \in S$$

个体的有序列表 X 可以视为一个整体，即一个概念单元（concept unit）。而在群体层面，多个个体的有序列表共同构成一个序列集（sequential set），每个概念单元在其中占据一行。

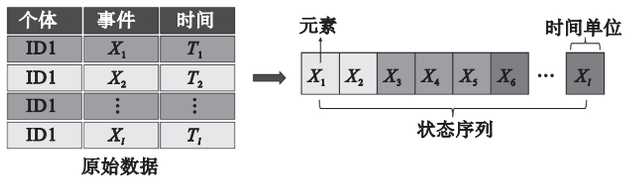


图1 SSA格式转换及其示例

2. 状态序列分析实施步骤

基于SSA的基本概念，该方法遵循以下步骤：序列编码、确定时间框架、构建状态序列、计算序列间的不相似性、寻找相似的状态序列，并对簇特征进行描述和可视化（如图2所示）。

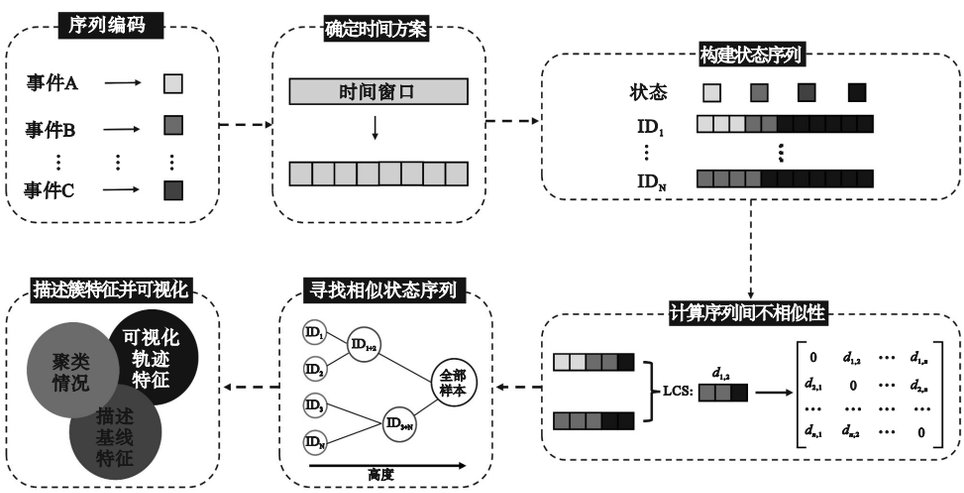


图2 SSA实现步骤

（1）序列编码

在序列编码时，需对所有可能的状态进行赋值并优化其数量，以简化序列的复杂性并提升结果的可解释性。由于这一过程需要对状态进行合理简化，它在很大程度上依赖于领域知识的支持。

（2）确定时间方案

在确定时间方案时，需明确观察期和时间单位。这两个因素决定了序列的起点、终点及其长度，并直接影响最终结果的解释。虽然时间方案的制定没有统一标准，但应根据具体的研究问题和分析目标进行合理设定。

（3）构建状态序列

在确定字母表和时间方案后，需为每个研究对象根据其时间窗口内的事件创建一个状态序列。

（4）计算序列间不相似性

在计算序列间不相似性时，需先选择适合研究目标的测量方法。常用的不相似性测量方法包括：对状态转换敏感的最优匹配（optimal matching）[9-11]、对时序不匹配敏感的汉明距离（Hamming distance）[12]以及对状态顺序敏感的最长公共子序列（longest common subsequence，LCS）[13-15]。为更清晰地说明不相似性的计算过程，以下以最长公共子序列为例介绍。首先，需要计算每对参与者状态序列之间的距离或不相似性：

$$d(y_1, y_2) = A(y_1, y_1) + A(y_2, y_2) - 2A(y_1, y_2)$$

其中 $d(y_1, y_2)$ 状态序列 y_1 和 y_2 之间的距离，$A(y_1, y_2)$ 表示 y_1 和 y_2 之间以相同顺序出现相同状态的数量。通过序列集 $Y = (y_1, y_2, \cdots, y_n)$ 生成不相似性矩阵 D：

$$D = \begin{bmatrix} 0 & d(y_1, y_2) & \cdots & d(y_1, y_n) \\ d(y_2, y_1) & 0 & \cdots & d(y_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ d(y_n, y_1) & d(y_n, y_2) & \cdots & 0 \end{bmatrix}$$

（5）寻找相似状态序列

在寻找相似状态序列时，需基于不相似性矩阵进行聚类分析，以识别具有代表性的簇。SSA中常使用Ward准则进行层次聚类分析[16]。在确定最佳聚类数量时，可以综合考虑树状图、聚类结果的简洁性和可

解释性，此外还可以通过一系列指标来评估聚类性能[17]。这些指标包括点二列相关系数（point biserial correlation，PBC）、Hubert’s Gamma（Hubert’s Gamma，HG）指数、Hubert’s Somers’D（Hubert’s Somers’D，HGSD）指数、Hubert’s C（Hubert’s C index，HC）指数、平均轮廓宽度（average silhouette width，ASW）、Calinski-Harabasz（Calinski-Harabasz，CH）指数和伪决定系数R^2（Pseudo R^2，R^2）等。其中，ASW值接近1表示聚类方案一致性较好，HC值接近0则表示聚类效果优异。对于其他指标，数值越高，聚类效果越佳。

（6）描述簇特征并可视化

SSA的结果可以与统计分析方法相结合，用以描述各簇的基线特征，或将这些簇作为解释变量或响应变量，评估特定轨迹组与健康结局或暴露之间的关联[18]。此外，各簇的纵向模式可通过可视化分析进行直观展示[19]。

3. 缺失处理

在实际调查中，由于测量时间点差异或失访，转换后的状态序列中可能出现缺失状态或序列长度不一致[20]。为应对这种情况，若缺失比例较低，在构建状态序列时，可将缺失视为字母表中的一种状态进行后续分析；或者将缺失视为无效状态，SSA则会忽略具有无效状态的时间单位并缩短序列进行分析。此外，对于序列长度不一致的情况，可以选择允许时间扭曲的测量方法（如最优匹配及其变体）来计算序列间的不相似性[21]。

4. 软件实现

在R统计软件中，可以使用“TraMineR”程序包构建状态序列，利用“WeightedCluster”程序包进行聚类分析及聚类性能评估。对于状态序列分析的可视化，可以使用“TraMineR”、“seqhandbook”和“TraMineRextras”程序包。有关上述程序包的使用代码以及SSA代码，已在其他资料中进行了详细介绍[22]。

研究实例

1. 数据来源及统计分析

（1）数据来源

数据来源于多中心健康体检项目，该项目自2007年开展，涵盖山东大学齐鲁医院、山东大学第二医院、济宁医学院附属医院等12家医疗机构，共纳入271394人（男性150993人，女性120401人）。体检数据由各医院按项目要求定期上传，共包含21708项检测项目。疾病数据来自病案首页和医保记录，涵盖60501个疾病类型。死亡数据来源于山东省死因监测数据库，涉及853种死因。本研究已获山东大学公共卫生学院伦理委员会批准。

本研究以2012—2013年作为基线，以2014—2023年作为观察期。排除基线年龄小于18岁、调查期间失访或死亡的个体，以及存在同一时间点诊断的多种新发慢性病的个体和ICD-10编码不全或错误的个体。最终，共纳入8952名研究对象，收集17970条慢性病记录。

（2）研究变量

研究参考一项针对中国成年人慢性病的meta分析，选取了以下23类慢性病[23]：癌症、青光眼与白内障、病毒性肝炎、胆囊、胆道及胰腺疾病、高脂血症、骨关节炎与类风湿关节炎、骨质疏松症、冠心病、甲状腺疾病、焦虑症、慢性肝病、慢性肾病、慢性胃肠炎、慢性阻塞性肺疾病、脑血管疾病、前列腺增生、痛风、哮喘、心力衰竭、抑郁症、高血压、糖尿病和周围血管病。

研究纳入了研究对象的一般信息，包括性别、年龄以及测量了BMI、收缩压、舒张压和甘油三酯等体检指标，并进一步计算了甘油三酯-葡萄糖（triglyceride-glucose，TyG）指数、低密度脂蛋白胆固醇（low density lipoprotein cholesterol，LDL-C）/高密度脂蛋白胆固醇（highdensitylipoprotein cholesterol，HDL-C）比值（LDL-C/HDL-C）、谷草转氨酶（aspartate aminotransferase，AST）与谷丙转氨酶（alanine aminotransferase，ALT）比值（AST/ALT）。其中，TyG指数计算公式为：

$$TyG=\ln\left[s\frac{甘油三酯(mg/dL)\times 空腹血糖(mg/dL)}{2}\right]$$

（3）统计分析

采用SSA对健康体检人群的慢性病进展轨迹进行分析，时间单位设置为半年。为准确识别慢性病轨迹，在构建状态序列时，采取以下处理方式：①若个体在时间单位 i 首次诊断慢性病 h，在时间单位 t 首次诊断慢性病 j，则在时间单位 i 到 t 之间的状态为 h，在时间单位 t 之后状态为 j。②若在观察期前已诊断慢性病 h，则在状态序列的首个时间单位记录状态 h。如果在观察期前有多个慢性病诊断，则记录距离观察期最接近的慢性病。③若个体在时间单位 s 前未诊断任何慢性病，则该时间单位之前的状态记录为“健康”。研究使用最长公共子序列法计算不相似性。

研究选择TyG指数、LDL-C/HDL-C和AST/ALT作为自变量，以慢性病进展轨迹组作为因变量。通过logistic回归分析，识别与慢性病进展轨迹模式相关的潜在影响因素。此外，研究将样本随机分为两个子样本，分别进行SSA以验证结果的稳健性[24]。计数资料以例（%）表示，组间比较采用χ^2检验，服从正态分布的计量资料采用$\bar{x}\pm s$表示，组间比较采用方差分析；不服从正态分布的计量资料采用$M(Q_1,Q_3)$表示，组间比较采用Kruskal-Wallis检验。双侧检验，检验水准α=0.05。

2. 结果

（1）最佳聚类数目选择

聚类性能衡量指标与聚类数 k 的关系如表1所示。随着 k 的增加，ASW值显著提升，并在 $k=4$ 时达到最大值。同时，$k=4$ 时的其他指标值也相对较优。此外，对树状图进行剪枝操作时，$k=4$ 也显示出良好聚类效果。因此，本研究的最佳聚类数确定为4。

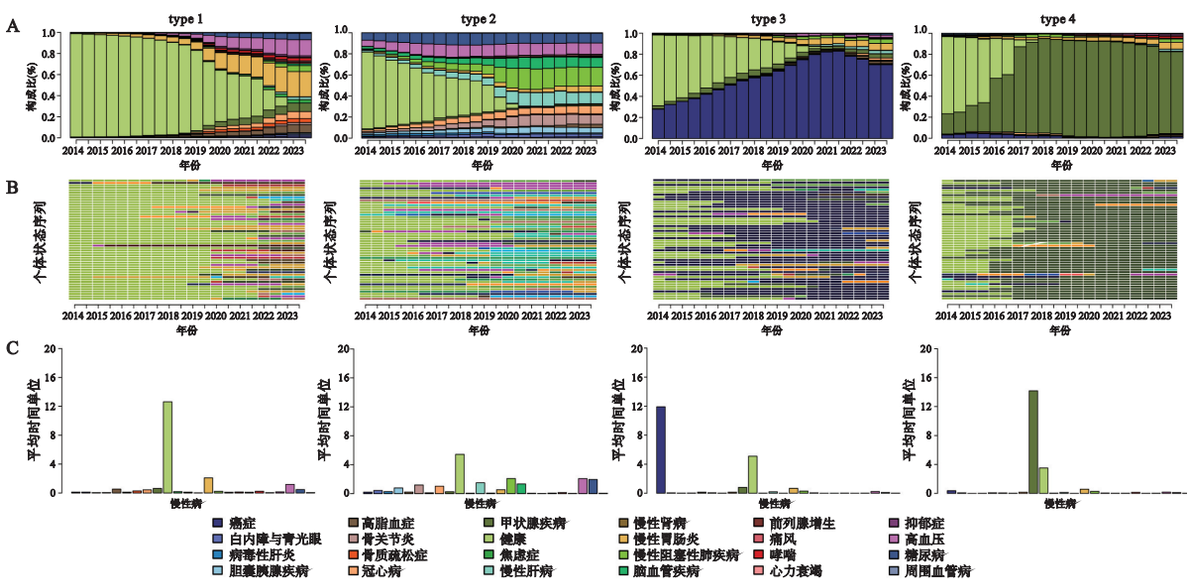
表1　聚类性能衡量指标

聚类数	PBC	HG	HGSD	ASW	CH	R^2	HC
2	0.402	0.493	0.477	0.231	868.750	0.088	0.270
3	0.470	0.571	0.554	0.248	745.292	0.143	0.229
4*	0.522	0.633	0.615	0.270	665.462	0.182	0.196
5	0.464	0.581	0.565	0.198	701.438	0.239	0.202
6	0.515	0.652	0.635	0.217	657.407	0.269	0.163
7	0.520	0.658	0.641	0.225	610.146	0.290	0.159
8	0.376	0.548	0.533	0.180	640.235	0.334	0.216
9	0.399	0.596	0.581	0.202	619.471	0.357	0.196
10	0.413	0.623	0.608	0.214	587.711	0.372	0.185

*：最佳聚类数目

（2）慢性病进展轨迹组的轨迹特征

图3A展示了各慢性病进展轨迹组在每个时间单位的状态分布，图3B展示了各组成员的状态序列，而图3C则显示了各状态在整个观察期所占的时长。具体而言，“type 1”组在研究后期健康状态比例逐渐下降，同时慢性胃肠炎、高血压和高脂血症等慢性疾病的比例逐渐上升。相比之下，“type 2”组在研究早期就出现健康状态比例下降的趋势，并伴随循环系统和内分泌与代谢系统疾病比例的上升。在“type 3”组中，癌症这一慢性疾病在研究早期的占比迅速上升，并在之后的时间段中保持较高水平。而“type 4”组则表现出甲状腺疾病在研究早期的占比快速上升，并在随后的时间段中维持较高占比。通过序列指数图可以进一步看出，“type 2”组的慢性病状态转移更频繁，而“type 4”组的慢性病状态转移则最为缓慢。



A为健康体检人群慢性病进展轨迹的状态序列分布图，B为健康体检人群慢性病进展轨迹的序列指数图，C为健康体检人群慢性病进展轨迹的平均时间图。

图3　健康体检人群慢性病进展轨迹特征

（3）慢性病进展轨迹组的基线特征

表2展示了不同慢性病进展轨迹组的基线特征。结果显示，四个轨迹组分别包含5749人、2050人、585人和568人，且轨迹组在各基线特征上存在显著差异（$P<0.01$）。

（4）慢性病进展轨迹组的影响因素

表3展示了慢性病进展轨迹组的logistic回归分析结果。相较于“type 1”组，TyG指数升高的个体处于“type 2”组的风险较高（OR=3.78，95% CI：2.47~5.79）；AST/ALT比值升高的个体处于“type 3”组风险较高（OR=1.36，95% CI：1.05~1.76）。此外，LDL-C/HDL-C比值升高的个体更易处于“type 1”组。

讨　论

本文介绍了SSA的基本原理，并通过实例展示其识别慢性病进展轨迹的能力。作为纵向数据分析方法，SSA在保持个体数据完整性的基础上，揭示个体在观察期内的动态变化，因此在真实世界研究中具有重要意义[11]。其次，与事件历史分析不同，SSA无须对数据进行先验假设[13]，能够更准确地描述和评估事件，并有效识别复杂数据结构中的模式，从而提供更全面且有价值的结果。此外，SSA提供了丰富的可视化工具，清晰且直观地呈现分析结果。

本研究识别出健康体检人群的四组慢性病进展轨迹。“type 1”组成员主要发生的慢性病包括慢性胃

表2 健康体检人群慢性病进展轨迹的基线特征

变量	type 1(*n*=5749)	type 2(*n*=2050)	type 3(*n*=585)	type 4(*n*=568)	统计量	*P*值
年龄,岁	41.90±10.53	45.50±12.29	41.16±10.80	34.25±10.63	164.36	<0.001
男性,频数(%)	3011(52.37)	1005(49.02)	133(22.74)	53(9.33)	531.92	<0.001
BMI,kg/m^2	24.53±3.47	24.82±3.58	23.73±3.46	22.70±3.30	54.23	<0.001
收缩压,mmHg	133.40±18.52	137.38±20.44	131.49±17.73	126.97±16.70	24.50	<0.001
舒张压,mmHg	83.53±12.59	85.29±12.85	81.41±12.32	77.77±11.57	20.90	<0.001
甘油三酯,mmol/L	1.15(0.78, 1.77)	1.19(0.80, 1.78)	0.99(0.69, 1.54)	0.91(0.67, 1.30)	76.75	<0.001
总胆固醇,mmol/L	5.04±0.93	5.06±0.96	4.91±0.86	4.81±0.95	9.22	<0.001
HDL-C,mmol/L	1.21±0.31	1.41±0.35	1.28±0.34	1.39±0.33	74.91	<0.001
LDL-C,mmol/L	2.91±0.78	2.63±0.74	2.75±0.72	2.61±0.68	39.87	<0.001
空腹血糖,mmol/L	5.15±1.01	5.49±1.65	5.01±0.68	4.96±0.77	32.81	<0.001
谷丙转氨酶,U/L	20.00(14.00, 29.00)	20.00(14.00, 29.00)	16.00(13.00, 23.25)	17.00(13.00, 24.75)	41.95	<0.001
谷草转氨酶,U/L	20.00(17.00, 25.00)	21.00(17.00, 25.00)	19.00(17.00, 23.00)	20.00(17.00, 23.00)	12.49	0.006
TyG指数	2.53±0.26	2.57±0.32	2.49±0.23	2.49±0.26	9.04	<0.001
LDL-C/HDL-C	2.43(1.83, 3.18)	2.17(1.65, 2.82)	1.88(1.47, 2.43)	1.84(1.42, 2.37)	287.97	<0.001
AST/ALT	1.05(0.79, 1.36)	1.00(0.79, 1.33)	1.17(0.92, 1.44)	1.16(0.91, 1.45)	42.04	<0.001
间隔时间,年	6.63±2.04	2.97±2.45	2.76±2.48	1.86±1.30	2326.32	<0.001
所患慢性病数量	1.89±1.32	2.46±1.70	1.96±1.30	1.65±1.16	96.32	<0.001

注:HDL-C为高密度脂蛋白胆固醇,LDL-C为低密度脂蛋白胆固醇,TyG指数为甘油三酯-葡萄糖指数,LDL-C/HDL-C为低密度脂蛋白胆固醇与高密度脂蛋白胆固醇比值,AST/ALT为谷草转氨酶与谷丙转氨酶比值,间隔时间为始发慢性病至观察期起点平均时间。

表3 慢性病进展轨迹组的logistic回归分析

	模型1		模型2	
	*OR*值(95%*CI*)	*P*值	*OR*值(95%*CI*)	*P*值
TyG指数				
type 1	1.00		1.00	
type 2	1.62(1.27~2.06)	<0.001	3.78(2.47~5.79)	<0.001
type 3	0.55(0.34~0.89)	0.015	1.56(0.55~4.40)	0.405
type 4	0.55(0.28~1.11)	0.098	0.59(0.15~2.23)	0.434
LDL-C/HDL-C				
type 1	1.00		1.00	
type 2	0.74(0.69~0.79)	<0.001	0.65(0.59~0.70)	<0.001
type 3	0.48(0.42~0.54)	<0.001	0.53(0.44~0.64)	<0.001
type 4	0.45(0.38~0.53)	<0.001	0.58(0.44~0.76)	<0.001
AST/ALT				
type 1	1.00		1.00	
type 2	0.95(0.83~1.08)	0.415	0.87(0.74~1.02)	0.094
type 3	1.76(1.43~2.17)	<0.001	1.36(1.05~1.76)	0.019
type 4	1.71(1.27~2.31)	<0.001	1.00(0.65~1.53)	0.989

注:模型1:未调整任何协变量,模型2:调整基线年龄、性别、体重指数、收缩压、舒张压、总胆固醇。TyG指数为甘油三酯-葡萄糖指数,LDL-C/HDL-C为低密度脂蛋白胆固醇与高密度脂蛋白胆固醇比值,AST/ALT为谷草转氨酶与谷丙转氨酶比值。

肠炎、高脂血症和高血压,logistic回归分析显示较高的LDL-C/HDL-C比值与该组显著相关,提示高脂血症可能在该组聚类过程中起关键作用[25]。“type 2”组在观察早期就已呈现出较高的慢性病发病率,序列指数图进一步显示该组疾病进展较快。此外,TyG指数与该组呈显著相关性,提示该组患者可能存在胰岛素抵抗和心脑血管疾病的风险[26-27],而这些疾病通常具有许多共同的危险因素。鉴于此,应通过积极的生活方式对该组人群实施干预,例如体育锻炼、戒烟限酒和合理膳食。“type 3”组与AST/ALT比值显著相关,这可能与该组中存在一定比例的肝癌患者有关。“type 4”组呈现出最为缓慢的疾病进展。一项针对奥地利住院患者的研究发现,在所识别的1260种不同疾病轨迹中,涉及甲状腺疾病向其他疾病转变的仅占不到0.1%[28],这一结果进一步支持了研究的结论。

SSA因其独特优势和广泛应用潜力,近年来受到研究者高度关注,并不断发展。为减少因填补非随机缺失引入的偏倚,研究人员开发了R程序包“PST”利用概率树来预测和填补缺失的状态[29]。为整合不同维度数据,研究人员提出了多通道SSA方法[11]。马尔科夫链的引入以及R程序包“SeqHMM”的开发,进一步提升了SSA在解释状态转移概率和构建共病网络方面的能力[30]。此外,SSA在医学领域的应用潜力也逐渐显现,如应用于患者护理轨迹分析[10]、药物治疗依从性分析[18]、双相情感障碍病程分析[31]等。本研究通过慢性病进展轨迹的实例分析,进一步拓展了其在医学研究的应用。

另一方面,SSA存在一定的局限性,在实际应用时应注意以下几点:①作为一种探索性方法,在解释结果时需谨慎,避免过度推断。②应用SSA时需保证观察期不可太短或者过宽的时间窗口,较短的序列长度可能无法充分展现复杂的动态变化,从而限制了对分析结果的解读。③随着状态序列数量的增加,计算复杂度呈平方级增长,因此,数据集过大时计算时间显著增加,且聚类结果可能存在偏差。④SSA为无监督学习方法,尽管其在揭示潜在模式方面具有一定优势,但未来仍需大量模拟研究以提升其在多样化应用场景中的解释力和准确性。

参 考 文 献

［1］王霁一, 冯星淋. 我国中老年人慢性病共病患病率的变化趋势及影响因素分析[J]. 医学与社会, 2024, 37(7): 41-48.
［2］Shen E, Lee JS, Mularski RA, et al. COPD comorbidity profiles and 2-year trajectory of acute and postacute care use [J]. Chest, 2021, 159(6): 2233-2243.
［3］范文婷. 无症状高尿酸血症患者尿酸结晶沉积及合并症的聚类分析[C]//中国超声医学工程学会. 中国超声医学工程学会第七届全国肌肉骨骼超声医学学术会议论文汇编, 2019:2.
［4］Wu G, Cai M, Wang C, et al. Ambient air pollution and incidence, progression to multimorbidity and death of hypertension, diabetes, and chronic kidney disease: A national prospective cohort [J]. Sci Total Environ, 2023, 881: 163406.
［5］Newman MG, Porucznik CA, Date AP, et al. Generating older adult multimorbidity trajectories using various comorbidity indices and calculation methods [J]. Innov Aging, 2023, 7(3): igad023.
［6］Ko J, Furby H, Ma X, et al. Clustering and prediction of disease progression trajectories in Huntington's disease: An analysis of Enroll-HD data using a machine learning approach [J]. Front Neurol, 2022, 13: 1034269.
［7］Abbott A, Forrest J. Optimal matching methods for historical sequences [J]. J Interdiscipl Hist, 1986, 16(3): 471-494.
［8］Bras H, Liefbroer AC, Elzinga CH. Standardization of pathways to adulthood? An analysis of dutch cohorts born between 1850 and 1900 [J]. Demography, 2010, 47(4): 1013-1034.
［9］Leblanc S, Roux J, Tillaut H, et al. Disease-modifying therapy usage in patients with multiple sclerosis in France: A 6-year population-based study [J]. Rev Neurol(Paris), 2021, 177(10): 1250-1261.
［10］Xu H, Bayless TM, Østbye T, et al. Care sequences leading to the diagnosis of Alzheimer's disease and related dementias: An analysis of electronic health records [J]. Alzheimers Dement, 2024, 20 (3): 2155-2164.
［11］Vanasse A, Courteau J, Courteau M, et al. Healthcare utilization after a first hospitalization for COPD: A new approach of state sequence analysis based on the '6w' multidimensional model of care trajectories [J]. BMC Health Serv Res, 2020, 20(1): 177.
［12］Biggin F, Ashcroft Q, Howcroft T, et al. Discovering patterns in outpatient neurology appointments using state sequence analysis [J]. BMC Health Serv Res, 2023, 23(1): 1208.
［13］Savaré L, Ieva F, Corrao G, et al. Capturing the variety of clinical pathways in patients with schizophrenic disorders through state sequences analysis [J]. BMC Med Res Methodol, 2023, 23(1): 174.
［14］Vanoli J, Nava CR, Airoldi C, et al. Use of state sequence analysis in pharmacoepidemiology: A tutorial [J]. Int J Environ Res Public Health, 2021, 18(24): 13398.
［15］Martin GL, Atramont A, Mazars M, et al. Days spent at home and mortality after critical illness: A cluster analysis using nationwide data [J]. Chest, 2023, 163(4): 826-842.
［16］Ward Jr JH. Hierarchical grouping to optimize an objective function [J]. J Am Stat Assoc, 1963, 58(301): 236-244.
［17］Studer M. Weightedcluster library manual. A practical guide to creating typologies of trajectories in the social sciences with R[R]. Technical Report, LIVES Working Papers, 2013: 24.
［18］Laforgue EJ, Istvan M, Chaslerie A, et al. Characterization of antipsychotic utilization before clozapine initiation for individuals with schizophrenia: An innovative visualization of trajectories using French National Health Insurance data [J]. Epidemiol Psychiatr Sci, 2023, 32: e59.
［19］Gabadinho A, Ritschard G, Müller NS, et al. Analyzing and visualizing state sequences in R with TraMineR [J]. Journal of Statistical Software, 2011, 40: 1-37.
［20］汤宁, 宋秋月, 易东, 等. 医学纵向数据建模方法及其统计分析策略[J]. 中国卫生统计, 2019, 36(3): 441-444.
［21］Studer M, Ritschard G. What matters in differences between life trajectories: A comparative r eview of sequence dissimilarity measures [J]. Journal of the Royal Statistical Society Series A: Statistics in Society, 2015, 179(2): 481-511.
［22］Mohammed S, Sonsoles LP. Learning analytics methods and tutorials [M]. Switzerland: Springer Nature, 2024.
［23］耿叶. 我国成年人慢性病共病患病现状及共病模式研究[D]. 沈阳:中国医科大学, 2023.
［24］McVicar D, Anyadike-Danes M. Predicting successful and unsuccessful transitions from school to work by using sequence methods [J]. Journal of the Royal Statistical Society Series A: Statistics in Society, 2002, 165(2): 317-334.
［25］朱柏林, 梁丽婷, 李秀琴, 等. LDL-C/HDL-C比值在血脂治疗监测中的应用价值分析[J]. 中国医学创新, 2021, 18(4): 148-151.
［26］孙文宇, 乔晶, 孟敏敏, 等. 甘油三酯葡萄糖指数与2型糖尿病胰岛素抵抗及血管并发症的相关性[J]. 临床内科杂志, 2023, 40 (10): 681-685.
［27］方奎, 张晶, 朱晨, 等. 三酰甘油-葡萄糖指数(TyG)与中老年人新发脑卒中风险的关联性研究[J]. 公共卫生与预防医学, 2024, 35 (4): 37-40.
［28］Dervić E, Sorger J, Yang L, et al. Unraveling cradle-to-grave disease trajectories from multilayer comorbidity networks [J]. NPJ Digit Med, 2024, 7(1): 56.
［29］Gabadinho A, Ritschard G. Analyzing state sequences with probabilistic suffix trees: The PST R package [J]. Journal of Statistical Software, 2016, 72(3): 1-39.
［30］Helske S, Helske J, Eerola M. Combining sequence analysis and hidden markov models in the analysis of complex [M]// Ahmed SE. Sequence Analysis and Related Approaches. Switzerland: Springer Nature, 2018: 185-200.
［31］Stevens A, Draisma S, Goossens PJJ, et al. The course of bipolar disorder in pregnant versus non-pregnant women [J]. Int J Bipolar Disord, 2021, 9(1): 35.

（责任编辑:邓　妍）