

多状态 Markov 模型在阿尔茨海默病自然史中的应用*

秦瑶¹ 韩红娟^{1,2} 刘龙¹ 陈杜荣¹ 王浩基¹ 崔靖¹ 白文琳^{1,3} 余红梅^{1,4,Δ}

【摘要】目的 将多状态 Markov 模型应用于阿尔茨海默病 (Alzheimer's disease, AD) 自然史研究,为慢性病自然史研究提供方法学借鉴。**方法** 基于美国公共数据库 ADNI,构建正常认知(normal cognition, NC)-轻度认知障碍(mild cognitive impairment, MCI)-AD 多状态 Markov 模型,估计不同状态间的转移强度、转移概率、影响因素和各状态的逗留时间,绘制生存曲线,并进行模型拟合优度评价。**结果** 性别、年龄、FAQ、MMSE、CDRSB、ADAS13、TMT-B-Time 是 AD 自然史中重要的影响因素;NC 与 MCI 状态的平均逗留时间分别为 7.502 年,11.621 年。多状态 Markov 模型拟合结果较好。**结论** 多状态 Markov 模型在 AD 等多状态慢性病转归研究中具有很好的应用价值。

【关键词】 多状态 Markov 模型 阿尔茨海默病 慢性病 自然史

【中图分类号】 R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.02.014

Application of Multi-state Markov Model in the Natural History of Alzheimer's Disease

Qin Yao, Han Hongjuan, Liu Long, et al (Department of Health Statistics, School of Public Health, Shanxi Medical University (030001), Taiyuan)

【Abstract】Objective The aim of this study was to explore the natural history of Alzheimer's disease (AD) based on the multi-state Markov model and to provide methodological reference for the chronic diseases. **Methods** Based on the American public database ADNI, we constructed a three states Markov model to estimate the transition intensity, transition probability, influencing factors and sojourn time of different states. Survival curves and goodness of fit was obtained. **Results** Gender, age, FAQ, MMSE, CDRSB, ADAS13 and TMT-B-Time were important influencing factors in the natural history of AD. The mean sojourn time of NC and MCI was 7.502 and 11.621 years, respectively. Model fit can be accepted. **Conclusion** Multi-state Markov model has a good application value for the outcome prediction of multi-state chronic diseases such as AD.

【Key words】 Multi-state Markov model; Alzheimer's disease; Chronic disease; Natural history

阿尔茨海默病 (Alzheimer's disease, AD) 是一种进行性中枢神经系统的退行性病变,起病隐匿,典型症状表现为认知能力损害,致残率高。一般认为轻度认知障碍 (mild cognitive impairment, MCI) 是介于正常认知 (normal cognition, NC) 和 AD 之间的一种中间过渡状态。以往的 AD 研究多采用传统的统计学方法,如 logistic 回归、Cox 回归等,无法提供疾病状态转移概率如何随时间变化的重要信息^[1-3]。多状态 Markov 模型作为处理随访资料的有效工具,可同时考虑多种状态、状态间转移的时间信息以及可能的影响因素,动态研究疾病的进展^[4-5]。本研究将多状态 Markov 模型应用于 AD 自然史研究,分析不同状态转移的影响因素及其转归规律,为动态生命学过程的慢性病研究提供方法学借鉴。

资料来源

研究对象来自美国公共数据库阿尔茨海默病神经影像学计划 (Alzheimer's disease neuroimaging initiative, ADNI)。ADNI 是一项多中心、跨学科的纵向观察性研究,多年来致力于为 AD 寻找更敏感、更准确的方法来检测早期 AD,并通过生物标志物来标记其进展^[6]。有关 ADNI 的详细信息,包括参与者纳入和排除标准以及完整的研究方案可以在 <http://adni.loni.usc.edu/> 上找到。所有受试者在进入研究时均获得书面知情同意书,并且该研究已经得到所有参与地点的当地机构审查委员会批准。

基本信息包括年龄、性别、受教育年限、婚姻状况和 APOE4。日常功能评估采用功能活动问卷 (Functional activities questionnaire, FAQ); 认知功能评估包括简易精神状态量表 (mini-mental state examination, MMSE), 和临床痴呆评定量表 (clinical dementia rating sum of boxes, CDRSB); 阿尔茨海默病评定量表-13 条目 (alzheimer's disease assessment scale, ADAS13), 评估认知领域中的记忆、语言和实践; 连线测验 B 部分花费时间 (trail making test-part B Time, TMT-B-Time), 评估执行功能中的定势转移能力,同时也反映

* 基金项目:国家自然科学基金面上项目 (81973154); 山西省应用基础研究计划青年项目 (201801D221399)

1. 山西医科大学公共卫生学院卫生统计学教研室 (030001)

2. 山西医科大学基础医学院数学教研室

3. 山西医科大学公共卫生学院儿少卫生学教研室

4. 重大疾病风险评估山西省重点实验室

Δ通信作者:余红梅, E-mail: yu@sxmu.edu.cn

手-眼的协调能力、注意能力和空间知觉;简短认知测试包含数字符号替换测验 (digit symbol substitution test, DSST);逻辑记忆的延迟回忆(logical memory-delayed recall, LM-DR)。

原理与方法

多状态 Markov 模型 (muti-Markov model, MSM) 是探讨慢性病不同发展阶段影响因素及转归规律较有效的统计方法^[7]。根据 AD 自然史,特确定一个三状态模型,即:状态 1(NC),状态 2(MCI),状态 3(AD)。首先,模型假定 AD 进程是 Markov 过程,NC 和 MCI 可以相互转化,但 MCI 进展为 AD 后将不再可逆,即将 NC 和 MCI 作为暂态,AD 作为吸收态;从现状态转移到下一状态仅仅取决于目前状态,而不直接受以前各状态的影响,即“无后效性”特征^[8]。由此构造的多状态结构见图 1,图中每个圆圈表示不同的状态,箭头表示状态转移的方向。

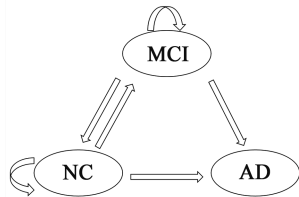


图 1 AD 自然史中状态结构图

MSM 模型假定某个体从 t_j 时刻的状态 r 在 t_{j+1} 时刻转移到状态 s 是依据概率随机发生的,采用转移强度 q 描述状态间的瞬时转移危险,即在 t 时刻处于状态 r 的个体,将在很短的时间区间内转移到状态 s 的概率^[9]。 q_{rs} 为转移强度矩阵中的元素; $q_{rs}^{(0)}$ 表示状态 r

转移到状态 s 的基线转移强度; $z(t)$ 表示状态 r 转移到状态 s 所对应的协变量向量; β_{rs} 表示协变量所对应的回归系数。根据转移强度矩阵各行之和等于 0,及对角线的转移强度等于该行中其他转移强度之和的相反数,本研究中模型的转移强度矩阵 Q 如下,共有 4 个待估转移强度: $q_{NC \rightarrow MCI}$ 、 $q_{NC \rightarrow AD}$ 、 $q_{MCI \rightarrow NC}$ 和 $q_{MCI \rightarrow AD}$ 。

$$Q = \begin{pmatrix} -(q_{NC \rightarrow MCI} + q_{NC \rightarrow AD}) & q_{NC \rightarrow MCI} & q_{NC \rightarrow AD} \\ q_{MCI \rightarrow NC} & -(q_{MCI \rightarrow NC} + q_{MCI \rightarrow AD}) & q_{MCI \rightarrow AD} \\ 0 & 0 & 0 \end{pmatrix}$$

则 MSM 模型可表示为 $q_{rs}(z(t)) = q_{rs}^{(0)} \exp(\beta_{rs}^T z(t))$ 。

个体的某次观察对似然函数的贡献为: $L_{i,j} = p_{rs}(t_{j+1} - t_j, z(t))$,则整个似然函数为每个个体每次观察对似然函数贡献的乘积,即 $L = \prod_{i=1}^N \prod_{j=1}^{m_i-1} L_{i,j}$ 。采用 Quasi-Newton 迭代法可获得模型参数的极大似然估计值,同时通过构造信息矩阵可获得这些参数标准误的渐近估计值^[10]。采用似然比检验进行变量的筛选, $G = -2(\ln L_0 - \ln L_1) = (-2\ln L_0) - (-2\ln L_1)$ 。

结果

1. 基本情况

本研究共纳入 1359 人,男性 751 (55.26%) 人,女性 608 (44.74%) 人,年龄平均 73.49 ± 6.96 岁,最长随访次数为 14 次 (15.14 年)。基线资料的详细描述见表 1。

2. MSM 模型的构建

依据各观察时刻研究对象所处的状态,得出各状态相互转移出现的频数,见表 2。

表 1 1359 名研究对象的基线资料

变量	总人群	基线状态		$\chi^2/t/z$ 值	P 值
		NC (n=516)	MCI (n=843)		
性别					
男	751 (55.26)	251 (48.64)	500 (59.31)	14.736	<0.001
女	608 (44.74)	265 (51.36)	343 (40.69)		
婚姻状况					
已婚	1010 (74.32)	354 (68.60)	656 (77.82)	14.234	<0.001
独身	349 (25.68)	162 (31.40)	187 (22.18)		
年龄 (岁)	73.49±6.96	74.11±5.89	73.11±7.52	2.745	0.006
受教育年限 (年)	16.09±2.79	16.39±2.67	15.92±2.85	3.014	0.003
APOE					
未携带	780 (57.40)	369 (71.51)	411 (48.75)	67.789	<0.001
携带	579 (42.60)	147 (28.49)	432 (51.25)		
FAQ	0 (0,3)	0 (0,0)	2 (0,5)	19.788	<0.001
MMSE	29 (27,30)	29 (29,30)	28 (26,29)	-15.465	<0.001
CDRSB	0.5 (0,1.50)	0 (0,0)	1.5 (1,2)	31.283	<0.001
ADAS13	13 (8.33,18.33)	9 (6,12)	16.33 (11,21)	19.688	<0.001
TMT-B-Time	86 (66,120)	76 (59,97)	96 (71,137)	9.858	<0.001
DSST	43 (35,51)	48 (41,55)	40 (33,48)	-11.622	<0.001
LM-DR	8 (4,12)	13 (11,15)	6 (3,9)	-27.750	<0.001

表 2 AD 各状态间转移频数

后一状态	前一状态		
	NC	MCI	AD
NC	2295	144	7
MCI	91	3179	358

根据各状态相互转移频数,经过多次迭代计算,最后设定初始转移强度矩阵为:

$$Q = \begin{pmatrix} 0 & 0.055 & 0 \\ 0.031 & 0 & 0.122 \\ 0 & 0 & 0 \end{pmatrix}$$

(1) 单因素三状态 Markov 模型

按 $\alpha=0.05$ 的水准,对性别、年龄、婚姻状况等 12 个因素采用似然比检验进行单因素分析。结果显示:性别、年龄、受教育年限、APOE4、FAQ、MMSE、CDRSB、ADAS13、TMT-B-Time、DSST 和 LW-DR 11 个因素具有统计学意义,结果见表 3。

表 3 单因素三状态 Markov 模型似然比检验

变量	-2logLR	G	df	P
无协变量模型	4199.999	-	-	-
性别	4190.337	9.663	3	0.022
婚姻状况	4195.865	4.134	3	0.247
年龄(岁)	4175.234	24.765	3	<0.001
受教育年限(年)	4191.736	8.263	3	0.041
APOE4	4125.308	74.692	3	<0.001
FAQ	3839.587	360.412	3	<0.001
MMSE	4007.230	192.770	3	<0.001
CDRSB	3917.051	282.948	3	<0.001
ADAS13	3721.223	478.776	3	<0.001
TMT-B-Time	4014.861	185.138	3	<0.001
DSST	4024.100	175.900	3	<0.001
LW-DR	3816.485	383.514	3	<0.001
多变量模型	3473.835	726.464	33	<0.001

(2) 多因素三状态 Markov 模型

将上述单因素分析中有意义的 11 个影响因素构建多因素多状态 Markov 模型,估计的转移强度矩阵见表 4,NC 进展为 MCI 的强度最大,MCI 恶化为 AD 的强度是逆转为 NC 的 1.29 倍。图 2 绘制了 15 年内的状态间的转移概率,可见 6 年内 NC 和 MCI 状态稳定率逐年降低,但仍高于其他状态间转移概率;NC 与 MCI 向 AD 的转移概率逐年升高,MCI 恶化率更高;NC 向 MCI 的进展率在第 11 年达到最高 48.97%,之后趋于稳定;MCI 向 NC 的逆转率在 15 年内趋于稳定。NC 与 MCI 状态的平均逗留时间分别为 7.502 (4.606, 12.219)年,11.621(9.908, 13.631)年。从基线到追踪观察 15 年内各状态的总逗留时间见图 3,可见每年内处于 NC 状态的时间最长,处于 MCI 和 AD 的时间逐年增高;追踪观察第 15 年时,处于 NC 的时间约为 7.331 年,处于 MCI 的时间约为 5.877 年,处于 AD 的时间约为 1.792 年。

表 4 转移强度矩阵

to	from		
	NC	MCI	AD
NC	-0.133 (-0.217, -0.082)	0.133 (0.082, 0.217)	0
MCI	0.038 (0.029, 0.049)	-0.086 (-0.101, -0.073)	0.049 (0.040, 0.059)

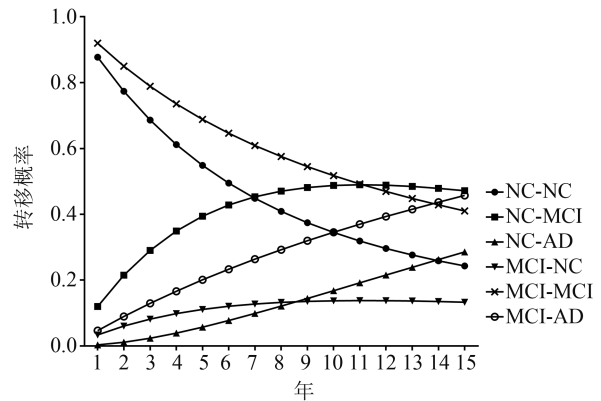


图 2 不同状态间 15 年内的转移概率

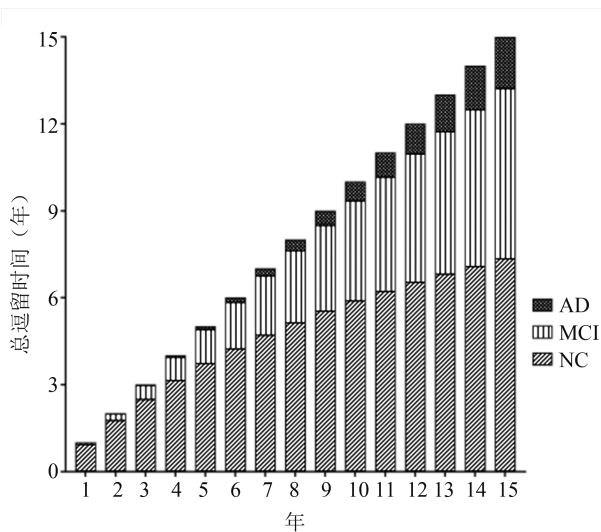


图 3 不同状态 15 年内的总逗留时间

不同影响因素在状态间转移的风险比(hazard ratios, HR)值见图 4。将 AD 作为生存分析的终点,NC 和 MCI 分别作为起点,绘制不同认知状态 15 年内的生存曲线,见图 5。图 6 分别给出了每个认知状态实际频率(实线)和理论频率(虚线)变化曲线,其横轴为时间,纵轴为频率,通过观察两条曲线的吻合程度,我们可以对多状态模型拟合优度进行评价,模型拟合较好。

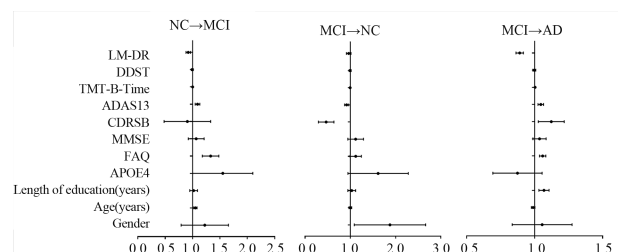


图 4 不同影响因素在状态间转移的森林图

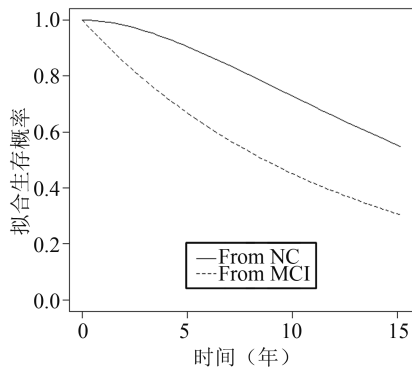


图5 不同认知状态 15 年内的生存概率曲线

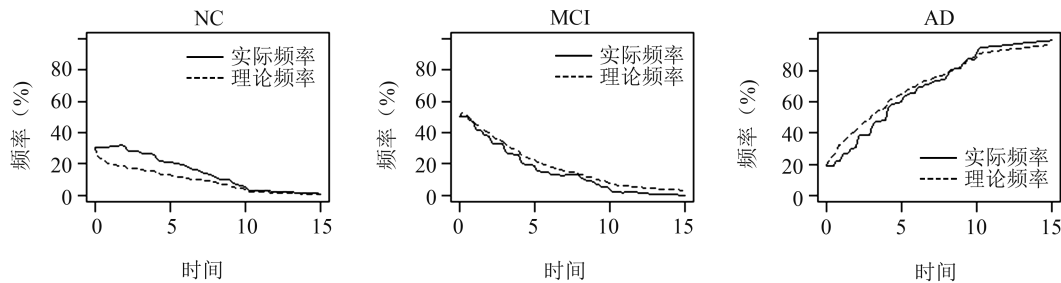


图6 不同认知状态实际频率与理论频率的变化曲线

应尽早预防 NC 人群认知障碍的发生^[11]。作为 NC 与 AD 的中间状态, MCI 进展为 AD 的研究已经成为神经病学研究的热点, 本研究发现 MCI 进展为 AD 的转移概率逐年上升, 且教育年月较长, 较高的 FAQ、CDRSB、ADAS13 得分与 TMT-B-Time 是 MCI 进展为 AD 的危险因素, 提示应关注日常生活能力障碍和文化程度较高的 MCI 患者, 对其进行日常生活活动的康复训练, 减少其未来进展为 AD 的风险。此外, MCI 也可以逆转为 NC, 其异质性经常被忽视^[12]。本研究发现较低的 CDRSB 和 ADAS13 得分和男性患者是 MCI 逆转为 NC 的保护因素, 表明较好的认知与执行功能是利于 MCI 患者的逆转, 与国外既往研究一致^[13-14]; MCI 向 NC 逆转的转移概率在 15 年之内稳定于 1%~2%, 尽管远远低于其他状态间的转移概率, 但临床医生也应重视 MCI 的逆转, 使具有良好认知轨迹的 MCI 患者避免接受多余的检查与昂贵的抗痴呆治疗, 进而有助于优化医疗资源分配和临床决策。

2. MSM 模型对传统生存分析的有益补充

许多慢性病呈现多状态、渐进性的特点, 大多数研究采用传统的疾病转归分析方法 (logistic 回归、Cox 回归等) 均不能提供疾病状态转移概率如何随时间变化的重要信息, 如 logistic 回归可处理多个结局但无法分析不同状态间的转移; Cox 回归只可分析观察到的一个结局, 无法处理具有多个结局的纵向数据^[15]。MSM 模型对数据要求及其严格, 需要个体明确的状态间转移时间, 否则只能进行传统的生存分析^[16]。MSM 模型可对涉及到多种不同结局的纵向资料构建

讨论

1. AD 自然史的动态研究

随着我国人口老龄化, 老年人群的认知障碍问题愈发受到广泛的关注。如能降低 NC 向 MCI 进展的风险, 对老年人的生活质量具有重大意义。本研究发现 NC 向 MCI 的转移概率 11 年之内逐年升高, 之后认知状态趋于稳定, 转移概率波动较小; 高龄、携带 APOE4 与较高的 FAQ 和 ADAS13 得分是 NC 进展为 MCI 的危险因素, 提示老年人群应经常进行认知训练, 尤其是对携带 APOE4 的高龄人群采取干预措施,

疾病转移模型, 考察不同状态转移的影响因素及协变量对其的影响, 估计不同状态间的转移概率和各个状态的逗留时间, 并依据现有数据对个体进行预测, 更深入地了解慢性病过程^[17]。

本研究以 AD 自然史为例, 讨论了 MSM 模型的应用, 同样可以用来分析其他慢性病的疾病进程。本研究尚存在些局限性。首先 MSM 模型需假定 AD 进展仅仅取决于目前状态, 而与过去的各状态无关^[1]。但在实际的临床工作中, 疾病转归可能还会与某些既往史相关。其次, 本研究将转移强度设定为常数, 假定所有转移危险在时间上是恒定的, 没有考虑转移强度随时间的变化, 接下来可以构建非齐次 MSM 模型, 更为全面地讨论 AD 自然史的相关影响因素及其变化规律。

总之, 对于多状态多阶段的慢性病纵向资料, 多状态 Markov 模型具有很好的应用价值, 在认识疾病进程的影响因素, 分析状态转移概率以及这些概率如何随着影响因素和时间的变化, 动态地评价疾病进程等方面具有得天独厚的优势。

参 考 文 献

[1] 安小妹, 凌莉. Markov 模型在生命统计中的研究进展. 中国卫生统计, 2007, 24(4):436-439.
 [2] 高建伟. 多状态 Markov 模型在轻度认知障碍向阿尔茨海默病转归研究中的应用. 太原: 山西医科大学, 2011.
 [3] 杨珊珊. 多状态 Markov 模型在轻度认知损害转归研究中的应用. 太原: 山西医科大学, 2010.
 [4] 余红梅. 多状态模型简介. 中国卫生统计, 2005, 22(3):177-180.
 [5] Liangliang Z, Chae YL, Tapabrata M, et al. Analysis of conversion of Alzheimer's disease using a multi-state Markov model. Statistical Methods in Medical Research, 2019, 28(9):2801-2819.