

载脂蛋白 E 联合临床相关指标预测阿尔兹海默病模型的建立与验证*

吴天晨¹ 杨卉^{2△} 梁艳¹ 杨永刚¹

【摘要】目的 以载脂蛋白 E(ApoE)为基础,结合阿尔兹海默病发病危险因素及临床常见指标,构建阿尔兹海默病发病风险的临床预测模型。**方法** 从南京市中医院大数据平台中,检索时间为 2019 年 1 月至 2021 年 1 月两年时间内,阿尔兹海默病患者 61 例,模糊匹配健康体检者 111 例。利用 LASSO 回归筛选危险因素,构建 logistic 回归预测模型,10 折交叉进行区分度验证,bootstrap 法进行校准度验证,临床决策曲线评判预测模型的临床指导性,最后以列线图可视化呈现临床预测模型。**结果** 筛选出 12 个变量,最终纳入 4 个危险因素:年龄、游离三碘甲状腺素(FT₃)、性别、ApoE。全样本 ROC 曲线下面积为 0.879,10 折交叉 9 次验证后 ROC 曲线下平均面积为 0.864;采用 Bootstrap 法抽样及 Hosmer-Lemeshow 校准度检验,结果 $\chi^2=6.496, P=0.592>0.05$ 。临床决策曲线阈值概率区间为 1%~88.6%。**结论** 利用年龄、游离三碘甲状腺素、性别、ApoE 构建的临床预测模型对患者进行个体化评估,可以预警阿尔兹海默病,开展早期预防干预,减缓该病的发生发展。

【关键词】 痴呆 阿尔兹海默病 预测模型 载脂蛋白 E

【中图分类号】 R74 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.01.005

Apolipoprotein ApoE Combined with Clinical Related Indices to Predict and Verify a Model for Alzheimer's Disease

Wu Tianchen, Yang Hui, Liang Yan, et al(*Nanjing Hospital of Traditional Chinese Medicine(210022), Nanjing*)

【Abstract】Objective To construct a clinical prediction model for the risk of Alzheimer's disease based on ApoE, combined with risk factors and common clinical indicators. **Methods** There were 61 cases of Alzheimer's disease patients and 111 cases of fuzzy matching healthy physical examination from Nanjing Hospital of Chinese Medicine data platform from January 2019 to January 2021. Using LASSO regression screening of risk factors, constructing logistic regression forecasting model, 10 fold cross verifies the degree of differentiation, validation the calibration of the bootstrap method. The clinical guidance of the prediction model was evaluated by the clinical decision curve, and finally, the clinical prediction model was visualized by nomogram. **Results** 12 variables were screened out and four risk factors were included, which are age, free triiodothyroxine(FT₃), gender and ApoE. The AUC of ROC of the whole sample was 0.879, and the average AUC of ROC after 10 folded and 9 crossed training sets verification was 0.864. Bootstrap method and Hosmer-Lemeshow were used to test the calibration degree. Results $\chi^2=6.496, P=0.592>0.05$. The threshold probability of clinical decision curve ranged from 1% to 88.6%. **Conclusion** Individualized evaluation of patients using clinical prediction models constructed by age, FT₃, gender and ApoE can provide early warning of Alzheimer's disease, carry out early prevention intervention and slow down the development of the disease.

【Key words】 Dementia; Alzheimer's disease; Prediction model; Apolipoprotein E

随着全球人口老龄化,认知障碍成为一项严重威胁全球人口健康和生活质量的世界性公共卫生难题,2021 阿尔兹海默病协会国际会议(AAIC)指出,截止到当时,全球约有痴呆患者 5000 万例,预计到 2050 年将上升到 1.52 亿^[1]。全球医疗卫生每年投入大量的资金,约占用全球生产总值的 1.1%^[2]。阿尔兹海默病(Alzheimer's disease, AD)导致的认知障碍占比约为 60%~70%^[3]。由于 AD 早期发现率较低,导致患病率、发病率和死亡率逐年上升。轻

度认知障碍最终会有 75% 发展成 AD^[4]。AD 临床前期诊断至关重要。早预测,早干预,能有效减缓病程进展。

载脂蛋白 E(aplipoprtein E, ApoE)是 AD 的风险基因,由三个等位基因($\epsilon 2$ 、 $\epsilon 3$ 、 $\epsilon 4$)相互组合成 6 种基因型,产生 3 种蛋白表达亚型(ApoE2、ApoE3、ApoE4)。ApoE 能产生 β -淀粉样肽来维持神经功能,同时参与血脂代谢。胆固醇及甘油三脂等脂质代谢的紊乱与 AD 发生密切相关^[5]。垂体-甲状腺轴功能与认知障碍也存在相关性^[6]。

本研究希望构建以 ApoE 为基础,结合 AD 危险因素及临床潜在关联性指标,构建一个预测 AD 发病风险的临床模型,成为其量化工具,为临床诊疗提供理性直观的信息,为患者个体化评估提供参考^[7]。

* 基金项目:国家自然科学基金(81904112);江苏省自然科学基金资助项目(BK20190136);南京市中医药青年人才培养项目(ZYQ20047)

1.南京市中医院(210001)

2.南京中医药大学

△通信作者:杨卉, E-mail: yanghuiyoung@126.com

资料与方法

1.一般资料

本研究以南京市中医院临床大数据库为平台。检索时间为2019年1月至2021年1月两年时间内,南京市中医院脑病科门诊及住院明确诊断阿尔兹海默病患者61例,将具有ApoE4血检结果的健康体检人群,按照1:2比例,以年龄和性别为人口学特征进行模糊匹配,去除临床观察指标缺失项超过30%的个案11例,共纳入111例作为对照组,建立临床预测模型。

纳入标准参考美国国家衰老研究院-阿尔茨海默协会(NIA-AA)对AD的诊断标准:①临床依据简易精神状态检查(MMSE)、Blessed痴呆评定量表(BDRS)或类似检查诊断为痴呆;②认知缺陷至少累积2个认知领域,没有意识障碍;③发病年龄在40~100岁之间。

排除标准:①严重脑血管病造成的认知障碍或痴呆,意识障碍,谵妄,假性痴呆(抑郁等导致)以及短暂性意识混乱和因药物、毒物引起的认知障碍。②入院72h内相关血检指标缺失过多,数据完整性无法保证。流程图见图1。

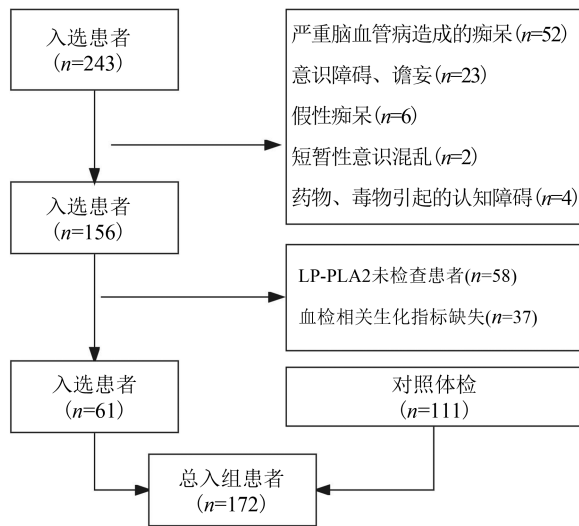


图1 数据纳入流程图

2.方法

纳入的相关危险因素及临床观察指标,包括性别(gender)、年龄(age)、高血压(hypertension)、糖尿病(diabetes)、冠心病(coronary heart disease)、载脂蛋白E(ApoE)、甘油三酯(TG)、总胆固醇(TC)、低密度脂蛋白(LDL-C)、高密度脂蛋白(HDL-C)、脂蛋白a(LPa)、载脂蛋白A1(ApoA1)、载脂蛋白B(ApoB)、促甲状腺激素(TSH)、游离甲状腺素(FT4)、游离三碘甲状腺素(FT3)、总三碘甲状腺素(TT3)、总甲状腺素(TT4)。以分类资料及数值资料形式进行前期收集,后期数据整理转化。转化方法:

①年龄以10岁为1个等级进入模型;其余数值指标以临床正常参考值分为低中高3个等级,赋值为1、2、3;②分类变量赋值如下:ApoE2($\epsilon 2/\epsilon 2$ 、 $\epsilon 2/\epsilon 3$)、ApoE3($\epsilon 3/\epsilon 3$ 、 $\epsilon 2/\epsilon 4$)、ApoE4($\epsilon 3/\epsilon 4$ 、 $\epsilon 4/\epsilon 4$)分别赋值为2、3、4;③性别赋值为男=1、女=0,是否有冠心病、高血压、糖尿病等分类变量赋值为有=1、无=0。

3.统计学方法

采用SPSS 26.0和STATA 17统计软件进行数据处理及统计分析,两组之间差异性比较,等级数据使用多样本秩和检验(Kruskal Wallis),分类数据使用卡方检验。选择LASSO回归^[8]进行自变量的筛选。无序多分类变量以独热变量形式进入LASSO回归。采用最小误差(最小EBIC)值为截断点,筛选自变量,进行单因素分析后进行多因素分析,构建logistic回归模型。因样本量数量限制,不进行验证集及训练集的划分,选用k折交叉验证法进行区分度验证^[9],绘制ROC曲线,计算曲线下面积(AUC)。选择bootstrap重复抽样法进行拟合优度验证(校准度)^[10],重复抽样(n=1000次)。临床决策曲线(decision curve analysis, DCA)评判预测模型的临床指导性^[11]。列线图可视化呈现临床预测模型^[12]。相关模型构建要求及结果报告参考TRIPOD声明清单^[13]。

结果

1.痴呆组和健康对照组基本特征及相关实验室指标比较

总样本量为172例,痴呆组与健康对照组在高血压、冠心病、载脂蛋白A1、载脂蛋白B、总胆固醇、高密度脂蛋白、低密度脂蛋白、促甲状腺激素、总三碘甲状腺素、总甲状腺素、游离甲状腺素进行 χ^2 检验,差异无统计学意义(P>0.05)。痴呆组与对照组在性别、年龄、纤维蛋白原、ApoE、脂蛋白a、甘油三酯、高密度脂蛋白、游离三碘甲状腺素进行 χ^2 检验,差异有统计学意义(P<0.05),见表1。

2. LASSO 回归筛选变量及构建预测模型

建立LASSO回归模型,将全数据集18个变量纳入模型,构建LASSO回归路径图及交叉验证图(图2~3)。以EBIC为截断点的 $\lambda = 3.922, \ln(3.922) = 1.37$ (垂直于X轴黑色细短虚线),筛选出age、TG、LDL-C、Lpa、ApoA1、ApoB、FT3、TT3、gender、hypertension、diabetes、ApoE共12个变量,构建logistic回归模型。最终纳入4个变量, Age、FT3、gender、ApoE建立预测模型(表2)。

3.验证预测模型的区分度及校准度

根据样本量,选择k折交叉验证法对构建的模型进行模型区分度的验证(本研究选取10重交叉验证,

9 轮训练集验证后根据损失函数计算 AUC 的平均值) (图 4), 全样本 ROC 曲线下面积为 0.879, 10 折交叉 9 次验证后 ROC 曲线下面积平均为 0.864, 区分度较高。采用 Bootstrap 法进行重复抽样及 Hosmer-

Lemeshow 校准度检验(图 5), 重复抽样 1000 次, 种子设定值(123), 预测值: 观察值比=0.969, 斜率值=0.933, 截距项=-0.030, $\chi^2 = 6.496, P = 0.592 > 0.05$ 构建的模型的校准度较高, 一致性良好。

表 1 健康对照组和痴呆组基本特征及相关实验室指标比较[例数(%)]

因素	对照组	痴呆组	P	因素	对照组	痴呆组	P
性别			0.001	总胆固醇 (mmol/L)			0.055
女	68(0.40)	20(0.12)		≤ 2.84	7(0.04)	9(0.05)	
男	43(0.25)	41(0.24)		2.85~5.69	81(0.47)	46(0.27)	
高血压			0.251	≥ 5.70	23(0.13)	6(0.04)	
无	29(0.17)	21(0.12)		甘油三酯 (mmol/L)			0.021
有	82(0.48)	40(0.23)		≤ 0.44	0(0.00)	3(0.02)	
糖尿病			0.031	0.45~1.70	76(0.44)	46(0.27)	
无	75(0.44)	31(0.18)		≥ 1.71	35(0.20)	12(0.07)	
有	36(0.21)	30(0.17)		高密度脂蛋白 (mmol/L)			0.158
冠心病			0.123	≤ 1.03	37(0.22)	27(0.16)	
无	75(0.44)	34(0.20)		1.04~1.55	63(0.37)	32(0.19)	
有	36(0.21)	27(0.16)		≥ 1.56	11(0.06)	2(0.01)	
年龄(岁)			0.001	低密度脂蛋白 (mmol/L)			0.216
≤ 48	8(0.05)	2(0.01)		≤ 1.49	17(0.10)	12(0.07)	
49~58	18(0.11)	2(0.01)		1.50~3.12	65(0.38)	40(0.23)	
59~68	32(0.19)	7(0.04)		≥ 3.13	29(0.17)	9(0.05)	
69~78	23(0.13)	20(0.12)		促甲状腺激素 (uiU/L)			0.549
79~88	25(0.15)	21(0.12)		≤ 0.54	7(0.04)	5(0.03)	
≥ 89	5(0.03)	9(0.05)		0.55~4.78	96(0.56)	49(0.29)	
ApoE			0.001	≥ 4.79	8(0.05)	7(0.04)	
ApoE2	26(0.15)	2(0.01)		总三碘甲状腺素 (nmol/L)			0.077
ApoE3	76(0.44)	30(0.17)		≤ 0.91	6(0.03)	8(0.05)	
ApoE4	9(0.05)	29(0.17)		0.92~2.79	105(0.61)	53(0.31)	
脂蛋白 a(mg/L)			0.024	总甲状腺素 (nmol/L)			0.536
0~300	90(0.52)	40(0.23)		≤ 58.0	5(0.03)	4(0.02)	
≥ 301	21(0.12)	21(0.12)		58.1~140.6	101(0.59)	56(0.33)	
载脂蛋白 A1(g/L)			0.083	≥ 140.7	5(0.03)	1(0.01)	
≤ 0.9	5(0.03)	11(0.06)		游离甲状腺素 (pmol/L)			0.991
1.0~1.6	87(0.51)	40(0.23)		≤ 11.4	4(0.02)	2(0.01)	
≥ 1.7	19(0.11)	10(0.06)		11.5~22.7	105(0.61)	58(0.34)	
载脂蛋白 B(g/L)			0.178	≥ 22.8	2(0.01)	1(0.01)	
≤ 0.5	7(0.04)	7(0.04)		游离三碘甲状腺素 (pmol/L)			0.001
0.6~1.1	76(0.44)	45(0.26)		≤ 3.4	1(0.01)	14(0.08)	
≥ 1.2	28(0.16)	9(0.05)		3.5~6.5	106(0.62)	47(0.27)	
				≥ 6.6	4(0.02)	0(0.00)	

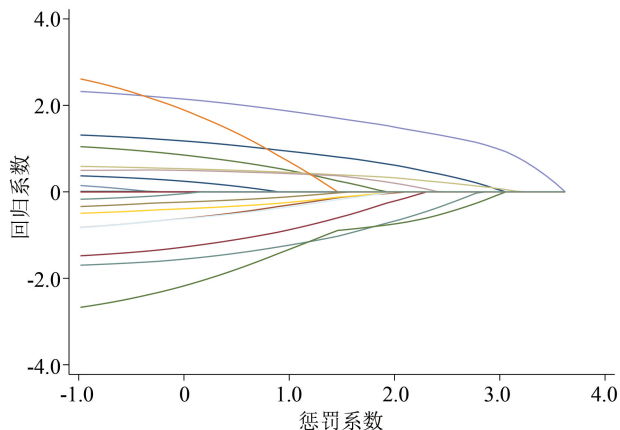


图 2 LASSO 回归路径图

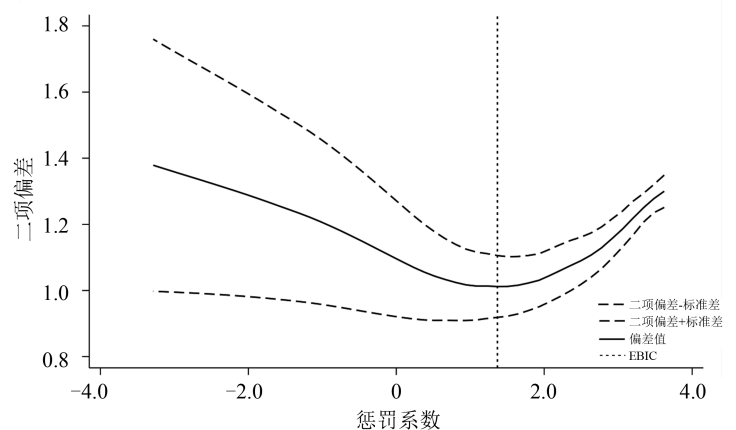


图 3 LASSO 回归交叉验证图

4. 预测模型的临床应用及可视化呈现

临床决策曲线阈值概率区间为 1%~88.6%，其中 0.886 后曲线波动较大，对预测概率干扰较大，作为

截断值，见图 6。列线图进行预测模型的可视化展示和应用，见图 7。

表 2 临床预测模型单因素及多因素 logistic 回归分析

变量名	单因素分析				多因素分析			
	β	Wald	P	OR(95%CI)	β	Wald	P	OR(95%CI)
年龄	0.56	15.58	0.001	1.75(1.33~2.31)	0.586	9.745	0.002	1.8(1.24~2.60)
甘油三酯	-0.79	4.75	0.029	0.45(0.22~0.92)	—	—	—	—
低密度脂蛋白	-0.41	2.48	0.115	0.66(0.40~1.11)	—	—	—	—
脂蛋白 a	0.81	5.01	0.025	2.25(1.11~4.58)	—	—	—	—
载脂蛋白 A1	-0.57	3.05	0.081	0.56(0.30~1.07)	—	—	—	—
载脂蛋白 B	-0.58	3.37	0.066	0.56(0.30~1.04)	—	—	—	—
游离三碘甲状腺素	-3.51	11.33	0.001	0.03(0.00~0.23)	-3.272	7.381	0.007	0.04(0.00~0.40)
总三碘甲状腺素	-0.97	2.95	0.086	0.38(0.12~1.15)	—	—	—	—
性别	1.18	12.31	0.001	3.24(1.68~6.23)	1.075	5.957	0.015	2.93(1.24~6.94)
高血压	-0.40	1.31	0.253	0.68(0.34~1.33)	—	—	—	—
糖尿病	0.70	4.61	0.032	2.02(1.06~3.82)	—	—	—	—
ApoE2*	—	30.96	0.001	—	—	—	—	—
ApoE3	1.64	4.57	0.001	5.13(1.15~22.78)	2.298	5.88	0.015	9.96(1.55~63.80)
ApoE4	-2.10	22.95	0.001	41.89(8.28~211.89)	4.254	17.076	<0.001	70.42(9.36~529.74)

* : ApoE 以哑变量纳入模型, ApoE2 为参照。

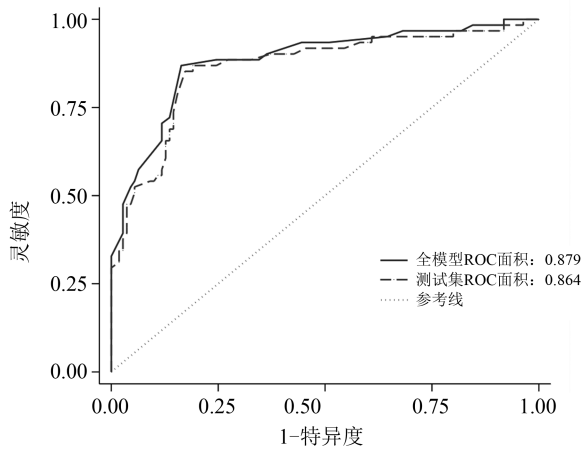


图 4 模型区分度

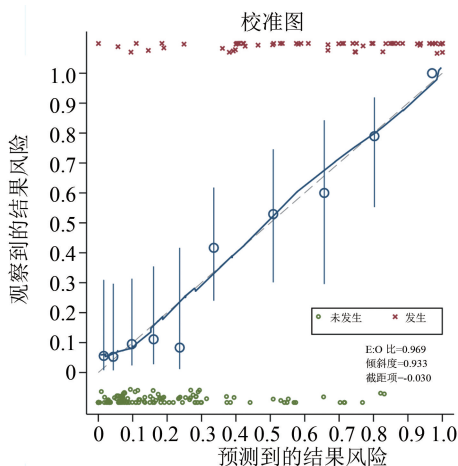


图 5 模型校准度

讨论

AD 是造成全球老年人群痴呆的主要原因，目前全球约有 4400 万患者^[14]。早发现、早干预是减缓病

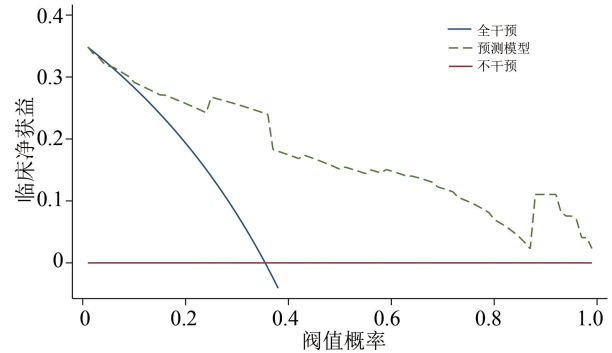


图 6 模型决策曲线图

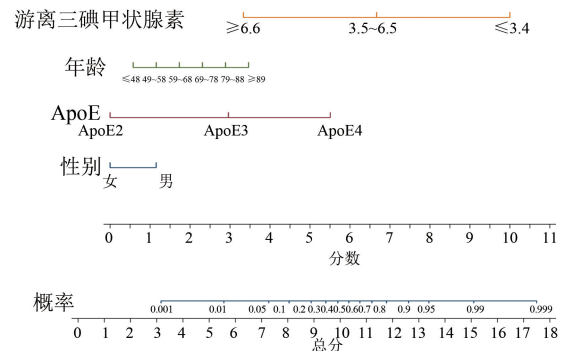


图 7 模型列线图

情发展的关键。结合特异性强、灵敏性高的生物学标志及临床常见生化指标，提高 AD 诊断率，进行早期干预。本研究利用医院大数据平台，通过前期数据分析整理及后期模型的建立与验证，提炼出四个指标。

通过列线图(nomogram)展示可见，ApoE 中的亚型 ApoE4 单项分值最高，提示 ApoE4 是 AD 的重要危险因素。A β -淀粉样蛋白异常沉积及 tau 蛋白形成的神经纤维缠结是造成 AD 两个重要病理标志。研究显示 ApoE4 能够促进 AD 患者脑内 A β -淀粉样蛋白

沉积,降低其清除率^[15]。此外,ApoE4 能够诱导触发内质网应激,诱导 tau 过度磷酸化和轴突病变^[16]。在多因素 logistic 回归方程中,以 ApoE2 为参照,携带 ApoE4 的人群发生 AD 发病风险是携带 ApoE2 人群的 70.42 倍左右(95% CI: 9.36 ~ 529.74, $P < 0.001$), Yamazaki 等^[17]通过人脑器官模型的建立,发现携带 ApoE4 的 AD 患者,脑神经元细胞凋亡时间的提前和突触完整性的降低明显高于其他两种基因型。

列线图提示 FT3 水平与预测 AD 的评分呈负相关,在多因素回归方程中可以发现随着 FT3 上升,AD 的风险呈下降趋势(0.04,95% CI: 0.00~0.40, $P = 0.007$)。在 Quinlan 的研究中^[18],与健康人群相比,AD 患者血清中 FT3 水平明显降低,FT3 水平的下降会影响 AD 患者脑内血供、能量代谢及信号传导。FT3 对海马生理作用起到调节功能,AD 人群中,较高血清浓度的 FT3 可以减少海马体积出现萎缩的概率。

根据回归方程,男性的发病风险是女性的 2.93 倍左右(95% CI: 1.24~6.94, $P = 0.015$)。Liesinger 的研究^[19]指出,女性 AD 患者约占总数的 2/3。但在神经病理学的诊断中,通过死亡后的病理解剖分析,男女在发病率上没有显著差异,并且男性出现认知障碍较女性提前。其特点为病程短,死亡率高,非典型临床症状隐匿,故男性 AD 发病率低可能与其隐匿性高,报告较少有关。年龄每上升 10 岁,AD 风险增加 1.8 倍(95% CI: 1.24~2.60, $P = 0.002$)。Lemprière 的研究中^[20]提出 A β -淀粉样蛋白异常沉积是 AD 的病理基础,年龄的增长会促进 A β -淀粉样蛋白活化小胶质细胞群,从而增加 AD 的发病风险。

通过构建合理的模型,以列线图进行展示,利用 4 个临床较为常见的实验室指标和风险因素对 AD 进行预测分析,从而更好地发现 AD 临床前期患者,进行早期的干预和决策。

AD 的影响因素较多,考虑到样本量限制,选用 AD 发病危险因素及临床指标作为变量纳入模型,忽略了患者的受教育程度及生活影响因素,如抑郁、压力、不良嗜好等。后期可通过增大样本量、增加可疑因素(变量)数量,进行多中心外部验证、多时间点前后验证,从而建立更加稳健的模型,提高模型的预测度。

本研究收集本院 AD 患者数据,建立临床预测模型。找到 4 个指标对 AD 发病风险进行预测分析,列线图实行个体化评估,旨在早期预警 AD 发病风险,开展预防干预,减缓 AD 的发生发展。

参 考 文 献

[1] Nichols E, Steinmetz JD, Vollset SE, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health*, 2022, 7(2) : e105-e125.

- [2] Nichols E, Szeke C, Vollset SE, et al. Global, regional and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*, 2019, 18(1) : 88-106.
- [3] Leng F, Edison P. Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nat Rev Neurol*, 2021, 17(3) : 157-172.
- [4] Tahami Monfared AA, Byrnes MJ, White LA, et al. Alzheimer's Disease: Epidemiology and Clinical Progression. *Neurol Ther*, 2022, 11(2) : 553-569.
- [5] Qin W, Li W, Wang Q, et al. Race-Related Association between APOE Genotype and Alzheimer's Disease: A Systematic Review and Meta-Analysis. *Alzheimers Dis*, 2021, 83(2) : 897-906.
- [6] Choi BW, Kim S, Kang S, et al. Relationship Between Thyroid Hormone Levels and the Pathology of Alzheimer's Disease in Euthyroid Subjects. *Thyroid*, 2020, 30(11) : 1547-1555.
- [7] 谷鸿秋, 周支瑞, 章仲恒, 等. 临床预测模型: 基本概念、应用场景及研究思路. *中国循证心血管医学杂志*, 2018, 10(12) : 1454-1456+1462.
- [8] McEligot AJ, Poyner V, Sharma R, et al. Logistic LASSO regression for dietary intakes and breast cancer. *Nutrients*, 2020, 12(9) : 2652.
- [9] Dankers FJWM, Traverso A, Wee L, et al. Prediction Modeling Methodology. *Fundamentals of Clinical Data Science*. Cham (CH) : Springer, 2019: 101-120.
- [10] Noma H, Shinozaki T, Iba K, et al. Confidence intervals of prediction accuracy measures for multivariable prediction models based on the bootstrap-based optimism correction methods. *Stat Med*, 2021, 40(26) : 5691-5701.
- [11] Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J*, 2021, 21(10) : 1643-1648.
- [12] Zuo Z, Zhang G, Song P, et al. Survival nomogram for stage IB non-small-cell lung cancer patients, based on the SEER database and an external validation cohort. *Ann Surg Oncol*, 2020, 28(7) : 3941-3950.
- [13] Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med*, 2015, 162(1) : w1-w74.
- [14] Dumurgier J, Sabia S. Epidemiology of Alzheimer's disease: latest trends. *Rev Prat*, 2020, 70(2) : 149-151.
- [15] Yamazaki Y, Zhao N, Caulfield TR, et al. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nat Rev Neurol*, 2019, 15(9) : 501-518.
- [16] Liang T, Xue F, Hang W, et al. Neuron-Specific Apolipoprotein E4 (1-272) Fragment Induces Tau Hyperphosphorylation and Axonopathy via Triggering Endoplasmic Reticulum Stress. *Journal of Alzheimer's Disease*, 2019, 71(2) : 597-611.
- [17] Zhao J, Fu Y, Yamazaki Y, et al. APOE4 exacerbates synapse loss and neurodegeneration in Alzheimer's disease patient iPSC-derived cerebral organoids. *Nat Commun*, 2020, 11(1) : 5540
- [18] Quinlan P, Horvath A, Eckerström C, et al. Higher thyroid function is associated with accelerated hippocampal volume loss in Alzheimer's disease. *Psychoneuroendocrinology*, 2022, 139: 105710.
- [19] Liesinger AM, Graff-Radford NR, Duara R, et al. Sex and age interact to determine clinicopathologic differences in Alzheimer's disease. *Acta Neuropathol*, 2018, 136(6) : 873-885.
- [20] Lemprière S. Age-related microglial activation accelerated in AD. *Nat Rev Neurol*, 2019, 15(7) : 369.