

· 临床研究 ·

基于 Logistic 回归模型及卡方自动交互探测决策树模型预测胎儿染色体异常的研究

阳 蓉 罗孝勇 李 陶

摘要 **目的** 应用 Logistic 回归模型及卡方自动交互探测 (CHAID) 决策树模型分析胎儿染色体异常的影响因素, 比较两种模型对其的预测价值。**方法** 回顾性分析 10 种常见超声软指标阳性并获得羊水穿刺结果的 642 例单胎孕妇资料, 以胎儿染色体结果为因变量, 超声软指标为自变量, 建立 Logistic 回归模型和 CHAID 决策树模型, 筛选胎儿染色体异常的影响因素; 绘制受试者工作特征 (ROC) 曲线分析两种模型对胎儿染色体异常的预测价值。**结果** 单因素 Logistic 回归模型显示, NT 增厚、鼻骨缺失、侧脑室增宽均为胎儿染色体异常的影响因素 ($OR=4.942$ 、 2.558 、 2.463 , 均 $P<0.05$); 多因素 Logistic 回归筛选 NT 增厚 ($OR=7.511$, $P<0.05$)、鼻骨缺失 ($OR=4.819$, $P<0.05$)、侧脑室增宽 ($OR=4.789$, $P<0.05$) 用于回归模型的拟合, 获得回归方程: $Y=-2.888+2.016\times NT \text{ 增厚}+1.572\times \text{鼻骨缺失}+1.566\times \text{侧脑室增宽}$ 。CHAID 决策树模型显示, NT 增厚、鼻骨缺失均为胎儿染色体异常的影响因素。ROC 曲线分析显示, Logistic 回归模型预测胎儿染色体异常的曲线下面积大于 CHAID 决策树模型 (0.712 vs. 0.675), 差异有统计学意义 ($Z=2.267$, $P<0.05$)。**结论** Logistic 回归模型和 CHAID 决策树模型均可预测胎儿染色体异常, Logistic 回归模型的预测价值优于 CHAID 决策树模型。

关键词 Logistic 回归模型; 卡方自动交互探测决策树模型; 染色体异常, 胎儿; 超声软指标
[中图分类号] R445.1; R714.5 [文献标识码] A

Prediction of fetal chromosomal abnormalities based on Logistic regression model and chi-squared automatic interaction detector decision tree model

YANG Rong, LUO Xiaoyong, LI Tao

Department of Ultrasound, Daping Hospital, Army Military Medical University, Chongqing 400042, China

ABSTRACT **Objective** To explore the predictive value of fetal chromosomal abnormalities by Logistic regression model and chi-squared automatic interaction detector (CHAID) decision tree model, and to analyze the influencing factors of fetal chromosomal abnormalities. **Methods** The data of 642 singleton pregnant women with 10 common positive ultrasound soft markers and amniocentesis results were retrospectively analyzed, fetal chromosomal abnormalities as the dependent variables and ultrasound soft marker as independent variables to establish Logistic regression model and CHAID decision tree model, the influence factors affecting fetal chromosomal abnormalities were screened. Receiver operating characteristic (ROC) curve was drawn to analyze the predictive value of the two models for fetal chromosomal abnormalities. **Results** The univariate Logistic regression model showed that the nuchal translucency (NT) thickening, nasal bone loss, and lateral ventricle widening were influence factors for fetal chromosomal abnormalities ($OR=4.942$, 2.558 , 2.463 , all $P<0.05$). Multivariate Logistic regression analysis were used to screen NT thickening ($OR=7.511$, $P<0.05$), nasal bone loss ($OR=4.819$, $P<0.05$) and lateral ventricle widening ($OR=4.789$, $P<0.05$) for fitting the regression model. The regression equation was obtained: $Y=-2.888+2.016\times NT \text{ thickening}+1.572\times \text{nasal bone loss}+1.566\times \text{lateral ventricle widening}$. The CHAID decision tree model showed that NT thickening and nasal bone loss were the influencing factors for fetal chromosomal abnormalities. ROC curve analysis showed that the area of

基金项目: 重庆市技术创新与应用示范社会民生类一般项目 (cstc2018jcsx-msybX0018); 重庆市影像医学与核医学临床医学研究中心 (CSTC2015YFPT-gcjsyjzx0175)

作者单位: 400042 重庆市, 陆军军医大学大坪医院超声科 (阳蓉、李陶); 四川省遂宁市中心医院超声科 (罗孝勇)

通讯作者: 李陶, Email: taoli39@163.com

the Logistic regression model was higher than that of the CHAID decision tree model (0.712 vs. 0.675), the difference was statistically significant ($Z=2.267, P<0.05$). **Conclusion** Logistic regression model and CHAID decision tree model have certain predictive value for fetal chromosomal abnormalities, and the Logistic regression model is better than the CHAID decision tree model.

KEY WORDS Logistic regression model; Chi-squared automatic interaction detector decision tree model; Chromosomal abnormality, fetus; Ultrasound soft marker

染色体异常可能导致胎儿发育畸形甚至死亡,活产儿中染色体异常的发生率约为 0.67%^[1]。目前有创性产前诊断方法是筛查胎儿染色体异常的金标准,其中以羊水穿刺最常用,但其可能引起早产、流产、胎盘早剥、宫内感染等风险。研究^[2-3]表明超声软指标与胎儿染色体异常存在一定关联,但对于超声软指标异常孕妇是否需行有创性产前诊断存在一定争议^[4]。本研究通过建立 Logistic 回归模型和卡方自动交互探测 (CHAID) 决策树模型筛选胎儿染色体异常的影响因素,并比较两种模型的预测价值,以期提高产前筛查的准确性。

资料与方法

一、研究对象

选取 2017 年 1 月至 2021 年 12 月在四川省遂宁市中心医院行羊水穿刺的 642 例孕妇,年龄 16~45 岁,平均(29.45±5.35)岁,孕 17~31 周,平均孕(20.83±2.88)周。纳入标准:①单胎妊娠;②10 种常见超声软指标中有一项为阳性;③获得羊水穿刺结果,孕期资料和产后资料完整。排除标准:①双胞胎或多胎妊娠;②有两项或多项超声软指标为阳性;③有先兆流产征兆;④胎儿存在严重结构畸形。本研究经四川省遂宁市中心医院医学伦理委员会批准,所有孕妇及家属均知情同意。

二、仪器与方法

1. 超声检查:使用 GE E 10 彩色多普勒超声诊断仪, C1-6 和 C2-9 凸阵探头,频率分别为 1~6 MHz 和 2~9 MHz。孕妇取仰卧位,经腹部进行胎儿标准化切面扫查,明确胎儿数目及其是否存活,估测胎儿孕周及体质量,观察胎儿是否存在结构畸形及阳性超声软指标。10 种常见超声软指标具体为^[5]:①NT 增厚,即孕 11~13⁶周时胎儿正中矢状切面上颈后部皮肤高回声区深部的无回声或带状低回声大小 ≥ 2.5 mm;②鼻骨缺失,即超声检查胎儿各个切面均未见鼻骨回声为双侧鼻骨缺失,超声检查双眼球横切面仅见一侧鼻骨回声为单侧鼻骨缺失;③侧脑室增宽,即侧脑室后角内径 10~15 mm;④肾盂分离,即双肾横切面显示肾盂

分离但不足以诊断肾盂积水,孕 20 周内肾盂分离 > 4 mm,孕 20~30 周肾盂分离 > 5 mm,孕 30 周以上肾盂分离 > 7 mm;⑤脉络丛囊肿,即超声示脉络丛内见直径 ≥ 3 mm 的圆形或椭圆形无回声结构,囊壁薄,边缘光滑、整齐,单双侧均可出现;⑥单脐动脉,即脐带纵切面仅见一条脐动脉,横切面为“吕”字结构代替正常的“品”字结构,CDFI 仅显示膀胱一侧的一条血管;⑦永存左上腔静脉,即三血管-气管切面示肺动脉左侧可见一条多余的静脉,四腔心切面示冠状静脉窦增宽;⑧永久性右脐静脉,即腹部横切面可见脐静脉与右门静脉相连,向胃泡左侧走行;⑨肠管回声增强,即肠管回声类似或强于骨骼回声;⑩心室强光点,即心室内见点状强回声。

2. 染色体检查:孕 16 周后于超声引导下羊膜腔穿刺术,常规抽取羊水 20 ml 进行细胞培养及染色体核型分析。

三、统计学处理

应用 SPSS 26.0 统计软件,计数资料以率或频数表示。以胎儿染色体结果为因变量($Y=1$ 为阳性, $Y=0$ 为阴性),对超声软指标进行赋值(有=1,无=0),先行单因素 Logistic 回归分析,再将差异有统计学意义的变量采用逐步向前法进行多因素 Logistic 回归分析,筛选胎儿染色体异常的影响因素。将超声软指标进行决策树交互检验分析,规则如下:①树的生长“枝条”分割显著性水准 $\alpha=0.05$;②设定决策树父节点、子节点最小样本量分别为 100 例和 50 例,并产生 CHAID 决策树模型的胎儿染色体异常预测概率。绘制受试者工作特征(ROC)曲线分析 Logistic 回归模型和 CHAID 决策树模型预测胎儿染色体异常的价值,曲线下面积比较采用 Z 检验。 $P<0.05$ 为差异有统计学意义。

结 果

一、一般情况

642 例孕妇阳性超声软指标中染色体异常发生情况见表 1,阳性超声软指标声像图见图 1。

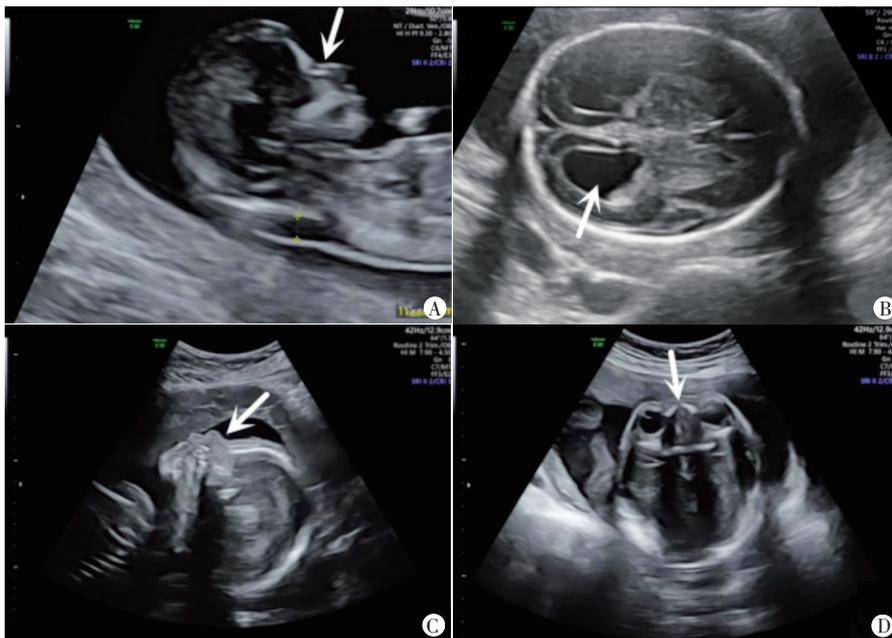
二、Logistic 回归分析结果

单因素 Logistic 回归分析显示,NT 增厚、鼻骨缺

表1 642例孕妇阳性超声软指标中染色体异常发生情况

例

超声软指标	总例数	染色体异常例数	染色体异常类型		
			三体	性染色体异常	结构异常
NT增厚	78	21	10	5	6
鼻骨缺失	52	11	7	2	2
侧脑室增宽	38	9	4	2	3
肾盂分离	90	4	3	1	0
脉络丛囊肿	137	8	2	2	4
单脐动脉	35	1	1	0	0
永久性右脐静脉	22	1	0	0	1
永存左上腔静脉	20	2	0	0	2
肠管回声增强	33	1	0	1	0
心室强光点	137	9	0	2	7
合计	642	67	27	15	25



A:NT增厚(箭头示);B:侧脑室增宽(箭头示);C,D:鼻骨缺失(箭头示)

图1 阳性超声软指标声像图

鼻、侧脑室增宽均为胎儿染色体异常的影响因素;多因素 Logistic 回归筛选 NT增厚($OR=7.511, P<0.001$)、

鼻骨缺失($OR=4.819, P<0.001$)、侧脑室增宽($OR=4.789, P<0.001$)用于回归模型的拟合,获得回归方程: $Y=-2.888+2.016\times NT增厚+1.572\times 鼻骨缺失+1.566\times 侧脑室增宽$ 。见表2,3。

三、CHAID 决策树模型分析结果将超声软指标用于建立 CHAID 决策树模型,决策树生长共2层,3个终端节点数。见图2。结果显示 NT增厚、鼻骨缺失均为胎儿染色体异常的影响因素,其中 NT增厚对染色体异常影响最大。CHAID 决策树模型中,第一层为 NT增厚,NT增厚时胎儿染色体异常的发生率为29.5%,NT正常时胎儿染色体异常的发生率为7.8%;第二层为鼻骨缺失,在 NT正常的基

表2 胎儿染色体异常影响因素的单因素 Logistic 回归分析

变量	回归系数	标准误	Wald χ^2 值	P值	OR值(95%可信区间)
NT增厚	1.598	0.294	29.579	<0.001	4.942(2.779~8.790)
鼻骨缺失	0.939	0.367	6.535	0.011	2.558(1.245~5.257)
侧脑室增宽	0.901	0.421	4.588	0.032	2.463(1.080~5.620)
单脐动脉	-1.423	1.023	1.934	0.164	0.241(0.032~1.790)
肾盂分离	-1.019	0.529	3.713	0.054	0.361(0.128~1.018)
脉络丛囊肿	-0.758	0.390	3.777	0.052	0.469(0.218~1.006)
心室强光点	-0.748	0.390	3.677	0.055	0.474(0.221~1.017)
永久性右脐静脉	0.917	1.032	0.790	0.374	0.400(0.053~3.020)
肠管回声增强	-1.391	1.023	1.847	0.174	0.249(0.033~1.849)
永存左上腔静脉	-0.049	0.757	0.004	0.948	0.952(0.216~4.196)

表3 胎儿染色体异常影响因素的多因素 Logistic 回归分析

变量	回归系数	标准误	Wald χ^2 值	P值	OR值(95%可信区间)
NT增厚	2.016	0.322	39.134	<0.001	7.511(3.993~14.126)
鼻骨缺失	1.572	0.397	15.697	<0.001	4.819(2.213~10.489)
侧脑室增宽	1.566	0.448	12.234	<0.001	4.789(1.991~11.521)
常数项	-2.888	0.205	197.536	<0.001	0.056(-)

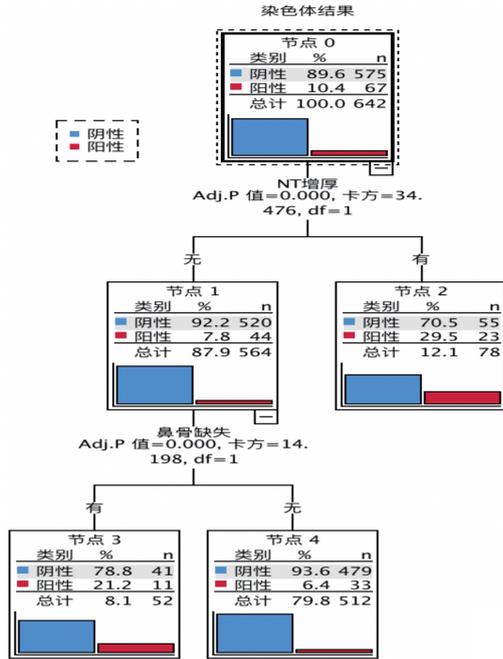


图2 CHAID决策树模型图

四、ROC 曲线分析

ROC 曲线分析显示, Logistic 回归模型预测胎儿染色体异常的曲线下面积为 0.712(95% 可信区间 0.675~0.746), 约登指数 40.77%, 敏感性 62.69%, 特异性 78.09%; CHAID 决策树模型预测胎儿染色体异常的曲线下面积为 0.675(95% 可信区间 0.637~0.711), 约登指数 34.05%, 敏感性 50.75%, 特异性 83.30%。Logistic 回归模型的曲线下面积大于 CHAID 决策树模型, 差异有统计学意义($Z=2.267, P<0.05$)。见图 3。

讨 论

染色体异常是导致胎儿发育畸形甚至死亡的主要原因之一, 产前诊断在提高筛查准确性的同时应遵循由无创到有创的顺序。超声软指标与胎儿染色体异常存在一定关联, 从而被越来越多的临床医师用作有创性产前诊断的指征之一, 但对所有超声软指标阳性者进行有创性产前诊断会造成不必要的资源浪费, 且可能增加宫内感染和流产的风险。本研究选取 642 例有一项超声软指标阳性并获得羊水穿刺结果的单胎妊娠孕妇, 应用 Logistic 回归模型和 CHAID 决策树模

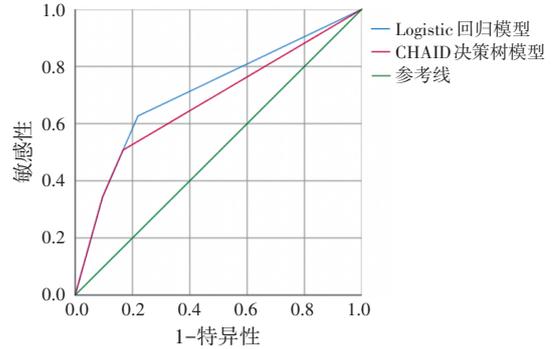


图3 Logistic 回归模型和 CHAID 决策树模型预测胎儿染色体异常的 ROC 曲线图

型分析 10 种常见超声软指标在预测胎儿染色体异常中的价值, 旨在为临床产前筛查提供参考。

本研究中胎儿染色体异常的多因素 Logistic 回归分析结果显示, NT 增厚胎儿发生染色体异常的风险是正常胎儿的 7.511 倍, CHAID 决策树模型的首层为 NT 增厚, 表明本研究纳入的 10 项超声软指标中, NT 增厚与胎儿染色体异常的关联最强。CHAID 决策树模型显示, NT 增厚胎儿染色体异常的发生率为 29.5%, 高于 NT 正常胎儿染色体异常的发生率(7.8%), 与研究^[6-8]报道 NT 增厚时胎儿染色体异常风险增加的观点一致。CHAID 决策树模型的第二层显示鼻骨缺失时胎儿染色体异常的发生率为 21.2%, Logistic 回归模型中鼻骨缺失胎儿发生染色体异常的风险是鼻骨正常胎儿的 4.819 倍。赵萍等^[9]研究发现鼻骨发育异常是胎儿染色体异常的危险因素($OR=6.771$), 本研究结论与其相似。目前胎儿 NT 增厚、鼻骨缺失在早孕期即可筛查, 因此当临床发现 NT 增厚、鼻骨缺失时, 建议及时行产前诊断及系统全面的超声筛查。

目前针对侧脑室增宽能否作为胎儿染色体异常的独立危险因素, 进而建议孕妇行有创性产前诊断仍存在争议。常清贤^[10]研究发现孕妇年龄和侧脑室增宽均是胎儿染色体异常的危险因素, 但年龄因素的 OR 值较侧脑室增宽更高(9.669 vs. 5.450), 说明年龄对于筛查胎儿染色体异常的价值更大。本研究 Logistic 回归分析显示, 侧脑室增宽是胎儿染色体异常的影响因素($OR=4.789, P<0.001$), 但该指标在 CHAID 决策树模

型中被剔除,分析原因可能是本研究纳入的侧脑室增宽的样本量相对较少,且未考虑孕妇年龄、宫内感染等因素,因此对于临床发现侧脑室增宽的胎儿可综合评估孕妇年龄及血清学检查结果后再进行有创性产前诊断。

本研究 ROC 曲线分析显示, Logistic 回归模型和 CHAID 决策树模型预测胎儿染色体异常的曲线下面积分别为 0.712、0.675, 差异有统计学意义 ($Z=2.267$, $P<0.05$), 表明 Logistic 回归模型的预测效能优于 CHAID 决策树模型。Logistic 回归模型的敏感性为 62.69%, 特异性为 78.09%; CHAID 决策树模型的敏感性为 50.75%, 特异性为 83.30%, CHAID 决策树模型的特异性高于 Logistic 回归模型, 即误诊率较低。在进行临床筛查时, 高特异性可以减少为了确定实际染色体正常而进行有创性产前诊断的数量; CHAID 决策树模型的敏感性低于 Logistic 回归模型, 即漏诊率较高。研究^[11-13]显示, Logistic 回归模型的优势在于分析因变量(胎儿染色体结果)与自变量(超声软指标)的依存关系时, 是在控制其他变量不变的情况下分析某变量的主效应, 但不能很好地反映各自变量间的交互作用; CHAID 决策树模型是在已知各独立变量(超声软指标)发生概率的基础上对数据进行分类, 其优势在于算法简单, 图形直观易于理解, 能反映超声软指标之间的交互作用, 但受子节点的数据影响结果相对不稳定。本研究中部分超声软指标的单因素 Logistic 回归分析差异无统计学意义, 如单脐动脉、脉络丛囊肿、永存左上腔静脉、永久性右脐静脉、肠管回声增强、心室强光点等, 因此当临床发现这些单一指征时, 应观察是否合并其他高危因素, 综合考虑是否需要行有创性产前诊断。

综上所述, 本研究通过对两种模型的比较、分析发现, Logistic 回归模型预测胎儿染色体异常的价值优于 CHAID 决策树模型。但本研究仅依据单一超声软指标进行染色体筛查, 敏感性较低, 且胎儿染色体异常的影响因素很多, 今后需结合孕妇年龄、血清学检查结果、胎儿结构畸形、多项软指标阳性等高危因素进一步综合评估。

参考文献

[1] Rose NC, Kaimal AJ, Dugoff L, et al. Screening for fetal chromosomal

abnormalities: ACOG practice bulletin, number 226[J]. *Obstet Gynecol*, 2020, 136(4):48-69.

- [2] Hu T, Tian T, Zhang Z, et al. Prenatal chromosomal microarray analysis in 2466 fetuses with ultrasonographic soft markers: a prospective cohort study[J]. *Am J Obstet Gynecol*, 2021, 224(5): 1-16.
- [3] 姚妍怡, 张成成, 宋婕萍. 中孕期超声软指标与染色体异常相关分析[J]. *现代妇产科进展*, 2021, 30(12):933-936.
- [4] Prabhu M, Kuller JA, Biggio JR, et al. Society for Maternal-Fetal Medicine Consult Series #57: Evaluation and management of isolated soft ultrasound markers for aneuploidy in the second trimester: (Replaces Consults #10, Single umbilical artery, October 2010; #16, Isolated echogenic bowel diagnosed on second-trimester ultrasound, August 2011; #17, Evaluation and management of isolated renal pelviectasis on second-trimester ultrasound, December 2011; #25, Isolated fetal choroid plexus cysts, April 2013; #27, Isolated echogenic intracardiac focus, August 2013)[J]. *Am J Obstet Gynecol*, 2021, 225(4):B2-B15.
- [5] 李胜利, 罗国阳. 胎儿畸形产前超声诊断学[M]. 北京: 科学出版社, 2017:899-919.
- [6] 谢思培, 陈锦珍, 林少梅, 等. 超声结构筛查对 16~18 周胎儿染色体核型正常而 NT 增厚的价值[J]. *中国超声医学杂志*, 2021, 37(3): 319-322.
- [7] Shakoor S, Dileep D, Tirmizi S, et al. Increased nuchal translucency and adverse pregnancy outcomes[J]. *J Matern Fetal Neonatal Med*, 2017, 30(14):1760-1763.
- [8] Sagi-Dain L, Singer A, Ben Shachar S, et al. Risk of clinically significant chromosomal microarray analysis findings in fetuses with nuchal translucency from 3.0 mm through 3.4 mm[J]. *Obstet Gynecol*, 2021, 137(1):126-131.
- [9] 赵萍, 宋勇, 崔丽清, 等. 超声软指标在胎儿染色体异常筛查中的应用价值[J]. *临床超声医学杂志*, 2021, 23(1):18-22.
- [10] 常清贤. 产前胎儿侧脑室扩张病因和预后研究及与婴幼儿神经系统发育关系的前瞻性研究[D]. 广州: 南方医科大学, 2013.
- [11] 焦莉萍, 郭晶晶, 杨云云, 等. 决策树模型与 Logistic 回归模型在生活饮用水水质影响因素分析中的应用[J]. *中国卫生统计*, 2020, 37(6):874-877, 882.
- [12] Feng Y, Wang J, Shao Z, et al. Predicting related factors of immunological response to hepatitis B vaccine in hemodialysis patients based on integration of decision tree classification and logistic regression[J]. *Hum Vaccin Immunother*, 2021, 17(9): 3214-3220.
- [13] Khazaei S, Najafi-GhOBADI S, Ramezani-Doroh V. Construction data mining methods in the prediction of death in hemodialysis patients using support vector machine, neural network, logistic regression and decision tree[J]. *J Prev Med Hyg*, 2021, 62(1): 222-230.

(收稿日期: 2022-04-24)