临床研究 DOI：10.13406/j.cnki.cyxb.003572

基于机器学习算法的静脉用药风险预测模型构建及验证

杨 洋[1]，王红梅[2]，单雪峰[3]，肖明朝[4]

（1.重庆医科大学附属第一医院麻醉科，重庆 400016；2.重庆医科大学附属第一医院药学部，重庆 400016；
3.重庆医科大学附属璧山医院药学部，重庆 402760；4.重庆医科大学附属第一医院泌尿外科，重庆 400016）

【摘 要】目的：收集患者临床信息，采用机器学习算法构建患者静脉用药风险预测模型。**方法：**回顾性纳入静脉用药患者（建模组 1 302 例和验证组 281 例），采用药学监护联盟协会提出的药物相关问题 V 9.09 分类标准分析患者存在的药物相关问题，采用 logistics 回归、神经网络、CHAID 决策树、贝叶斯网络、支持向量机等机器学习算法构建静脉用药风险预测模型，并采用混淆矩阵格式对各预测模型进行评价。通过准确率、召回率、精确率、F1 值以及生成验证受试者工作特征曲线下面积（area under curve，AUC）评价模型的预测性能。**结果：**患者药物相关问题发生率为 26.9%。患者药物相关问题主要集中在治疗安全性方面（n=556，94.9%），其次是治疗有效性方面（n=30，5.1%）。构建的模型中支持向量机的预测效能最好，AUC 为 0.826。**结论：**机器学习算法构建的静脉用药风险预测模型预测效能良好，可为静脉用药安全管理提供新思路和新方法。

【关键词】静脉用药；用药安全；机器学习；预测模型

【中图分类号】R979.9 **【文献标志码】**A **【收稿日期】**2024-03-05

Construction and validation of a machine learning-based model for predicting the risk of intravenous medication

Yang Yang[1], Wang Hongmei[2], Shan Xuefeng[3], Xiao Mingzhao[4]

(1. Department of Anesthesiology, The First Affiliated Hospital of Chongqing Medical University;
2. Department of Pharmacy, The First Affiliated Hospital of Chongqing Medical University;
3.Department of Pharmacy, Bishan Hospital of Chongqing Medical University; 4.Department of Urology,
The First Affiliated Hospital of Chongqing Medical University)

【Abstract】*Objective*: To construct a predictive model for the risk of intravenous medication using machine learning algorithms based on the clinical information of patients. ***Methods*:** A retrospective analysis was performed for the patients receiving intravenous medication, with 1302 patients in the modeling group and 281 in the validation group. The drug-related problem classification system V9.09 proposed by the European Society of Clinical Pharmacy was used to analyze the drug-related problems in patients. Machine learning algorithms, including logistic regression, neural network, CHAID decision tree, Bayesian network, and support vector machine, were used to construct risk predictive models for intravenous medication, and confusion matrices were used to evaluate the performance of each predictive model. Accuracy, recall rate, precision, and the area under the ROC curve (AUC) for the subjects in the validation group were used to evaluate the predictive performance of the model. ***Results*:** The incidence rate of drug-related problems was 26.9% among these patients. These drug-related problems mainly involved treatment safety (n=556, 94.9%), followed by treatment effectiveness (n=30, 5.1%). Among the models constructed, support vector machine algorithm showed the best predictive performance, with an AUC of 0.826. ***Conclusion*:** The predictive model for the risk of intravenous medication constructed using machine learning algorithms has good predictive performance, which can provide new insights and methods for the management of intravenous medication safety.

【Key words】intravenous medication; medication safety; machine learning; predictive model

静脉给药是一种将药物直接注射或输入血液的快速给药途径，相比于其他途径给药，静脉用药相关的药物伤害发生率高。我国国家药品不良反应监测年度报告显示，因静脉注射给药途径导致的不良反应/不良事件发生率高达 60%，且近年来一直

作者介绍：杨 洋，Email：382040924@qq.com，研究方向：患者安全，流行病学。

通信作者：肖明朝，Email：xmz.2004@163.com。

基金项目：重庆市科卫联合医学资助项目（编号：2020MSXM120）。

优先出版：https://link.cnki.net/urlid/50.1046.R.20240822.1409.002（2024-08-23）

位列第一并且呈逐年上升趋[1-2]。此外，1项研究分析表明，在我国因用药导致的死亡的病例中，也是通过静脉途径给药的药物导致死亡的事件较多[3]。因此，静脉用药安全不容忽视，开展早期静脉用药风险预测有助于预防药物相关伤害的发生，提高患者用药安全。

随着现代信息技术的发展，数据挖掘的方法已相当成熟，目前常见的机器学习方法包括支持向量机、随机森林、神经网络等。这些机器学习算法具有强大的非线性建模能力、模型泛化能力等特点，这使得机器学习算法成为许多实际问题中广泛使用的高效预测模型构建方法。在医学领域，机器学习算法已经得到广泛应用，并表现出不同的预测性能[4-6]。本研究拟通过机器学习算法，构建静脉用药安全风险早期预测模型，以期为静脉用药安全管理提供新思路和新方法。

1 资料与方法

1.1 研究对象

收集2019年3月至2022年12月就诊于重庆医科大学附属第一医院三甲综合医院的心血管及内分泌系统疾病的住院患者。患者纳入标准：①年龄≥18的患者；②使用治疗药物；③未接受任何药学干预。患者排除标准：①出院前非药物导致死亡的；②接受了或者正在参与药物的临床试验。

1.2 资料收集

通过医院电子病历系统收集患者一般情况和临床相关资料。由2名临床药师同时提取患者相关信息。①患者的一般信息：病历号、年龄、性别、婚姻状况、医保类型、体质量、身高、血压、合并症指数（charlson comorbidity index，CCI）、住院时长、饮酒史、吸烟史、药物过敏史等；②临床实验室检查指标：尿素、血肌酐、丙氨酰氨基转移酶、门冬氨酸氨基转移酶、糖化血红蛋白（hemoglobin A1C，HbA1C）、血钙、血钾、血钠、血氯、甘油三酯、总胆固醇、高密度脂蛋白等；③药物信息：用药数量、中药注射剂、药物相关问题等。

1.3 药物相关问题评价

采用根据欧洲药学监护联盟协会提出的药物相关问题V 9.09分类标准（Pharmaceutical Care Network Europe DRP V9.09 classification，PCNE-DRPs V 9.09）评价患者是否存在药物相关问题[7]。

1.4 模型构建

将患者是否存在静脉用药风险作为模型构建所需要的因变量。“存在静脉用药风险”定义为：存在中等或重度临床意义的药物相关问题，“不存在静脉用药风险”定义为：不存在中等或重度临床意义的药物相关问题。对该变量的确定由2名药师共同评估，当存在异议时，咨询第三人（临床医生）。

将收集到的数据导入IBM SPSS Statistics 23.0软件，并采用SPSS modeler 18.0进行模型的构建。采用的建模方法有神经网络、logistics回归、CHAID决策树、贝叶斯网络、支持向量机。

1.5 统计学方法

采用Excel表收集患者资料，应IBM SPSS Statistics 23.0软件对数据进行统计分析。计量资料用范围、均值和标准差表示，计数资料用频数和百分比表示。2组间计量资料比较采用t检验，计数资料比较采用卡方检验。单因素及多因素分析均采用logistic回归分析。采用混淆矩阵格式对优化后的各预测模型进行评价。通过准确率、召回率、精确率、F1值以及生成验证受试者工作特征曲线下面积（area under curve，AUC）评价模型的预测性能。检验水准α=0.05。

2 结 果

2.1 患者一般资料

共收集到住院患者1 583例，采用IBM SPSS Statistics 23.0软件构建住院患者静脉用药安全数据库。根据时间顺序，将患者分为建模组和验证组，建模组1 302例，验证组281例。其中建模组中，男性724（55.6%）例，平均年龄（66.4 ± 14.2）岁，年龄在60岁及以上的老年患者936（71.9%）例。高血压（47.2%）是最常见的诊断。患者平均住院时间为（10.3 ± 6.1）d，平均用药（3.0 ± 4.0）种（表1）。与验证组相比，建模组体质指数（body mass index，BMI）、冠心病、心力衰竭、肝功能、肾功能、白细胞、Ca^{2+}、K^{+}、Na^{+}、Cl^{-}、糖化血红蛋白、静脉用药数量及使用的中成药注射剂差异具有统计学意义（P<0.05）。

2.2 静脉用药药物相关问题

静脉用药相关问题根据PCNE-DRP V9.0进行分类。1 302例住院患者中，存在药物相关问题的患者数为350例。药物相关问题发生率为26.9%。药物相关问题主要集中在医嘱环节，其中仅有1个药物相关问题的患者有206例（占15.8%），存在2个药物相关问题的患者有88例（占6.8%），存在3个及其以上药物相关问题的患者有55例（占4.2%）。共发现药物相关问题586个，主要集中在：①治疗安全性方面（n=556，94.9%）。如2种中成药（注射剂）使用可能增加不良反应风险和增加患者经济负担。在药物相关问题中，药物相互作用是最常见的问题，如氨茶碱与利尿剂联用，可能影响氨茶碱的血药浓度，导致不良事件风险增加；胺碘酮注射液与阿托伐他汀钙片合用可升高其血药浓度，增加发生肌肉毒性的风险等。②治疗有效性方面（n=30，5.1%）。如静脉滴注哌拉西林他唑巴坦剂量和给药频次不合理，导致抗感染效

表1 患者信息（n,%;$\bar{x} \pm s$）

建模组变量	建模组患者数（n=1 302）	验证组患者数（n=281）	χ^2/t值	P值
一般信息				
年龄,≥60岁	936(71.9)	186(66.2)	3.634	0.057
均值±标准差	66.4±14.2	64.5±16.0	1.980	0.048
性别,男	724(55.6)	148(52.7)	0.806	0.369
BMI			7.566	0.023
BMI<18.5	147(11.2)	17(6.0)		
24≤BMI	537(41.2)	143(50.9)		
临床诊断				
高血压	615(47.2)	124(44.1)	1.849	0.604
冠心病	519(39.9)	89(31.7)	6.693	0.010
心力衰竭	472(36.3)	124(44.1)	5.971	0.015
房颤	202(15.5)	46(16.4)	0.117	0.732
糖尿病	490(37.6)	93(33.1)	2.123	0.145
住院时间,≥10 d	624(47.9)	131(46.6)	0.158	0.691
CCI,≥2分	667(51.3)	156(55.5)	3.523	0.172
肝功能,异常	296(22.7)	83(29.5)	5.874	0.015
肾功能,异常	727(55.9)	177(63.0)	11.791	0.003
实验室结果				
白细胞,异常	251(19.3)	48(17.1)	6.207	0.045
血红蛋白,异常	207(15.9)	42(14.9)	0.037	0.848
血小板,异常	176(13.5)	34(12.1)	2.310	0.315
白蛋白,异常	113(8.7)	21(7.5)	0.314	0.575
胆固醇,异常	166(12.7)	42(14.9)	1.547	0.214
甘油三酯,异常	290(22.3)	66(23.5)	0.676	0.411
Ca^{2+},异常	484(37.2)	121(43.1)	13.346	0.001
K^+,异常	157(12.1)	60(21.3)	17.704	<0.010
Na^+,异常	171(13.2)	47(16.7)	6.807	0.033
Cl,异常	331(25.4)	99(35.3)	18.011	<0.010
HbA1c>6.0%	585(44.8)	136(48.4)	4.907	0.027
用药情况				
静脉用药数	3.0±4.0	2.4±3.7	2.436	0.015
静脉用药总数≥3种	483(37.1)	85(30.2)	25.811	<0.001
使用中药注射剂,是	275(21.1)	13(4.6)	42.251	<0.010
使用抗菌药物,是	421(32.3)	79(28.1)	4.636	0.098
使用利尿剂,是	430(33.0)	95(33.8)	0.478	0.788

注：BMI,体质指数；CCI,合并症指数；HbA1c,糖化血红蛋白

果不佳；氟喹诺酮类药物与含钙或铁离子的药物联用，导致氟喹诺酮类药物疗效降低。药物相关问题分类见表2。由表2可见，不适当的组合占比最高，占66.6%，主要为药物与药物之间存在潜在的药物相互作用。

2.3 模型构建

构建各模型的效能参数见表3。支持向量机的预测效能最好（AUC=0.976），其次是CHAID决策树（AUC=0.925）。支持向量机模型、CHAID决策树模型的预测变量重要性的前十位因素分别见图1、图2。由图1和图2可见，在2个模型中，静脉用药数量、利尿剂的使用是较重要的变量。在后续的模型验证中，使用支持向量机、CHAID决策树同时进行验证。

2.4 静脉用药风险预测模型验证

采用支持向量机、CHAID决策树进行验证。281例患者中实际存在静脉用药风险的患者为80例，无静脉用药风险的患者为201例。经支持向量机预测为存在静脉用药风险的患者为64例，预测为无静脉用药风险的患者为271例。模型的预测效能AUC为0.826，支持向量机的预测效能更好，其他指标见表4。

表2 建模组DRPs的问题类别、原因分类（n,%）

项目	患者数	比例（%）
问题类别		
P1治疗有效性	19	3.2
P2治疗安全性	486	82.9
P3其他	81	13.8
具体问题		
P1.1药物治疗无效	2	0.3
P1.2治疗效果不佳	17	2.9
P1.3有未治疗的症状或适应症	0	0.0
P2.1（可能）发生药物不良事件	486	82.9
P3.1治疗成本效益问题	51	8.7
P3.2不必要的药物治疗	30	5.1
P3.3不确定的问题或抱怨，需要进一步说明	0	0.0
原因类别		
C2药物剂型	9	1.5
C3剂量选择	5	0.9
C4治疗疗程	3	0.5
C5调剂	0	0.0
C6药物使用过程	2	0.3
C7患者相关	0	0.0
C8患者治疗地点转换相关	0	0.0
具体原因		
C1.1选药不适宜（因未依据指南或处方集）	0	0.0
C1.2选药不适宜（指南推荐，但对此病人无使用禁忌）	15	2.6
C1.3无指针用药	27	4.6
C1.4不适当的组合（药物与药物或药物与草药或药物与保健品）	390	66.6
C1.5药物重复使用（药理作用相同或活性成分相同）	86	14.7
C1.6尽管存在适应证，未给予药物治疗或没有给与完整的药物治疗	0	0.0
C1.7同一疾病使用太多种药物	43	7.3
C2.1剂型不适宜（对该患者而言）	9	1.5
C3.1药物剂量过低	1	0.2
C3.2药物剂量过高	3	0.5
C3.3给药频次不足	1	0.2
C3.4给药频次过多	0	0.0
C3.5用药时间的指示错误，不清晰或遗漏	0	0.0
C4.1疗程过短	0	0.0
C4.2疗程过长	3	0.5
C6.1不恰当的给药时间或给药间隔	2	0.3
C8.1患者转换治疗地点时未进行药物重整	0	0.0
C9.2其他原因：溶媒不适宜、输注浓度不适宜、用法不适宜	6	1.0

表3 各模型的效能参数

建模方法	准确率	精确率	召回率	F1	AUC
logistics回归	0.855	0.731	0.766	0.748	0.720
神经网络	0.865	0.769	0.772	0.770	0.717
CHAID决策树	0.864	0.723	0.760	0.741	0.925
贝叶斯网络	0.849	0.789	0.723	0.755	0.742
支持向量机	0.945	0.906	0.890	0.898	0.976

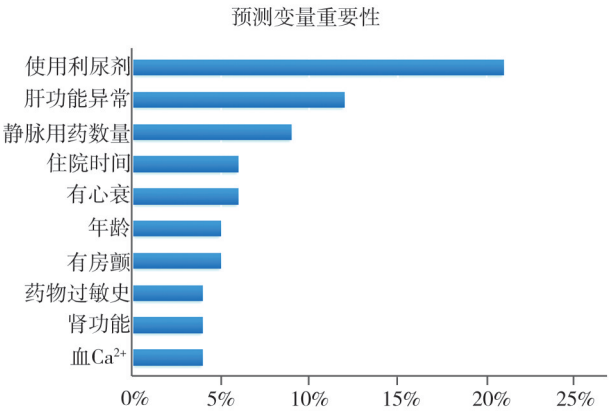


图1　支持向量机模型的预测变量重要性

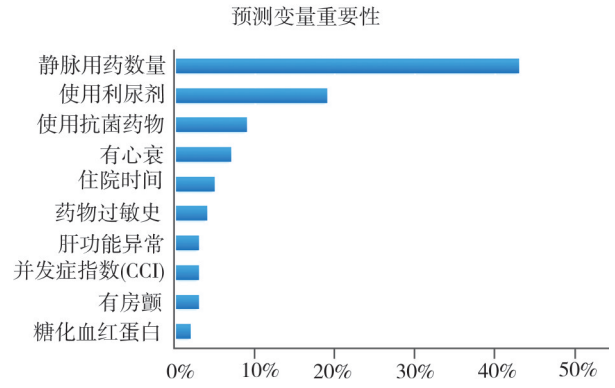


图2　CHAID 决策树的预测变量重要性

表4　支持向量机、CHAID 决策树模型验证的效能参数

建模方法	准确率	精确率	召回率	F1	AUC
CHAID 决策树	0.794	0.500	0.690	0.580	0.788
支持向量机（最优模型）	0.801	0.550	0.688	0.611	0.826

3　讨　论

静脉给药是住院患者最常用，甚至是首选的给药方式。虽然我国静脉用药安全令人担忧，但关注静脉用药风险预测的研究较少。本研究结果发现静脉用药的药物相关问题发生率为 26.9%，该结果低于以往文献报道结果的，如在我国老年人群中 34.5% 的患者存在至少 1 种药物相关问题[8]；慢性肾病患者中药物相关问题发生率高达 45.95%[9]。本研究的药物相关问题发生率低于文献报道的可能原因是因为研究的人群和药物的给药途径不同，之前的研究对象未区别静脉给药和非静脉途径给药，计算的药物相关问题发生率为总的发生率，而本研究重点关注的是静脉用药的药物相关问题，未考虑口服给药等其他非静脉途径给药。本研究结果发现药物相关问题主要涉及的类型是治疗安全性，其次才是治疗有效性，而发生药物相关问题的主要原因是药物选择。1 项研究分析了神经内科患者存在的药物相关问题，也发现了相似的趋势[10]。因此，也提示本课题组需要关注药物选择不当引起的用药安全问题。

本研究采用机器学习的方法包括神经网络、CHAID 决策树、贝叶斯网络、支持向量机等。通过比较各算法的预测效果，结果发现 CHAID 决策树和支持向量机的预测效能相对较好。比较 CHAID 决策树和支持向量机模型汇总的预测变量重要性发现，静脉用药数量、使用利尿剂、有心衰、肝功异常、住院时间、过敏等因素是 2 种机器学习算法筛选出的患者存在用药风险的共同预测因素。目前尚未见静脉用药风险的影响因素研究，但有研究报道多药联用、肥胖、肾功能损害和过敏与药物相关问题的高风险相关[11]。另有研究发现使用≥5 种药物的患者发生药物相关问题的风险明显高于使用少于 5 种药物的患者（OR=1.96；95%CI=1.31%~2.93%）[12]。另 1 项前瞻性研究发现，药品数量每增加 1 种，药物相关问题增加约 8.6%[13]。使用的药物数量越多，药物相互作用等药物相关问题增加，还可能增加患者住院时间延长的风险[14-15]。由此可见，使用的药物数量是患者用药安全中尤为重要的因素。在合理用药过程中，有必要进行处方精简。

另外，目前国内外报道的关于患者用药安全方面的预测研究多为不良反应预测、用药不依从性预测模型等[16-17]。还有些研究为药物相关问题预测或评分模型，这些研究采用 logistics 回归构建预测模型，但预测模型的预测效能欠佳，AUC 仅为 0.6-0.78[18-19]。本研究也使用了 logistics 回归构建预测模型，其 AUC 为 0.720，与文献报道结果相似，但与其他机器学习算法相比，logistics 回归构建的预测模型并不具有优势。

本研究仍存在一些局限性。首先，本研究纳入的人群仅涉及心血管疾病和内分泌疾病患者，预测模型中找到的影响因素可能不适用于其他疾病患者。其次，本研究未考虑口服药物等非静脉途径给药的药物，因此构建的预测模型可能无法准确预测非静脉途径用药的风险。另外，本研究收集的数据

为回顾性研究，可能存在部分患者资料不全，导致预测模型中未纳入资料不全的因素，因此，还需要大样本、前瞻性的多中心临床研究进一步验证。

总之，本研究基于患者的临床信息，成功构建了一个患者静脉用药风险的预测模型，具有良好的预测性能，对探索患者用药安全管理策略具有重要意义。

参 考 文 献

[1] 国家药品监督管理总局药品评价中心，国家药品不良反应监测中心．国家药品不良反应监测年度报告（2022年）[EB/OL]．（2023-01-12）[2024-03-03]. https://www.cdr-adr.org.cn/drug_1/aqjs_1/drug_aqjs_sjbg/202303/t20230324_50019.html.

Center for drug reevaluation, National center for ADR nonitoring, China. National adverse drug reaction monitoring annual report (2022) [EB/OL]. (2023-01-12)[2024-03-03]. https://www.cdr-adr.org.cn/drug_1/aqjs_1/drug_aqjs_sjbg/202303/t20230324_50019.html.

[2] 国家药品监督管理总局药品评价中心，国家药品不良反应监测中心．国家药品不良反应监测年度报告（2021年）[EB/OL]．（2022-01-10）[2023-02-03]. https://www.cdr-adr.org.cn/drug_1/aqjs_1/drug_aqjs_sjbg/202203/t20220330_49586.html.

Center for drug reevaluation, National center for ADR nonitoring, China. National adverse drug reaction monitoring annual report (2021) [EB/OL]. (2022-01-10)[2023-02-03]. https://www.cdr-adr.org.cn/drug_1/aqjs_1/drug_aqjs_sjbg/202203/t20220330_49586.html.

[3] Li HN, Deng JX, Yu PM, et al. Drug-related deaths in China: an analysis of a spontaneous reporting system[J]. Front Pharmacol, 2022, 13:771953.

[4] Wu YF, Fang Y. Stroke prediction with machine learning methods among older Chinese[J]. Int J Environ Res Public Health, 2020, 17(6): 1828.

[5] Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone [J]. BMC Med Inform Decis Mak, 2020, 20(1): 16.

[6] Zhang JL, Tang WM. Building a prediction model for iron deficiency anemia among infants in Shanghai, China[J]. Food Sci Nutr, 2019, 8(1): 265-272.

[7] Pharmaceutical Care Network Europe Association Classification for drug related problems V 9.1. The PCNE Classification V 9.1 2020

[8] Ma Z, Sun SS, Zhang CX, et al. Characteristics of drug-related problems and pharmacists′ interventions in a geriatric unit in China[J]. Int J Clin Pharm, 2021, 43(1): 270-274.

[9] 谢雅琼，林孝怡．血清游离轻链在鉴别诊断不同病因肾病的应用价值及其与患者肾功能分期的相关性分析[J]. 诊断学理论与实践，2023，22(2)：166-171.

Xie YQ, Lin XY. Value of serum-free light chain assay in differential diagnosis and staging of nephropathy of various etiologies[J]. J Diagn Concepts Pract, 2023, 22(2): 166-171.

[10] Liu PP, Li GY, Han M, et al. Identification and solution of drug-related problems in the neurology unit of a tertiary hospital in China[J]. BMC Pharmacol Toxicol, 2021, 22(1): 65.

[11] Garin N, Sole N, Lucas B, et al. Drug related problems in clinical practice: a cross-sectional study on their prevalence, risk factors and associated pharmaceutical interventions[J]. Sci Rep, 2021, 11(1): 883.

[12] 张咪雪，刘 莎．抑郁症的药物治疗新进展[J]. 重庆医科大学学报，2024，49(5)：631-637.

Zhang MX, Liu S. New progress in drug therapy for depression[J]. Journal of Chongqing Medical University, 2024, 49(5): 631-637.

[13] Viktil KK, Blix HS, Moger TA, et al. Polypharmacy as commonly defined is an indicator of limited value in the assessment of drug-related problems[J]. Br J Clin Pharmacol, 2007, 63(2): 187-195.

[14] 幸晓琼，王红梅，杨佳丹，等．某院心内科住院患者严重和禁忌的潜在药物相互作用横断面研究[J]. 中国药学杂志，2021，56(8)：694-698.

Xing XQ, Wang HM, Yang JD, et al. Cross-sectional study of major and contraindicated potential drug-drug interactions in hospitalized cardiac patients in our hospital[J]. Chin Pharm J, 2021, 56(8): 694-698.

[15] 席田兰，幸晓琼，杨佳丹，等．基于潜在药物相互作用的心血管疾病患者住院时间延长风险预测模型的构建[J]. 中国医院用药评价与分析，2023，23(4)：400-403.

Xi TL, Xing XQ, Yang JD, et al. Construction of prediction model for prolonged length of stay in patients with cardiovascular disease based on potential drug interactions[J]. Eval Anal Drug Use Hosp China, 2023, 23(4): 400-403.

[16] Chung AY, Anand S, Wong IC, et al. Improving medication safety and diabetes management inHong Kong: a multidisciplinary approach [J]. Xianggang Yi Xue Za Zhi, 2017, 23(2): 158-167.

[17] Onder G, Petrovic M, Tangiisuran B, et al. Development and validation of a score to assess risk of adverse drug reactions among in-hospital patients 65 years or older: the GerontoNet ADR risk score[J]. Arch Intern Med, 2010, 170(13): 1142-1148.

[18] Geeson C, Wei L, Franklin BD. Development and performance evaluation of the Medicines Optimisation Assessment Tool (MOAT): a prognostic model to target hospital pharmacists' input to prevent medication-related problems[J]. BMJ Qual Saf, 2019, 28(8): 645-656.

[19] Urbina O, Ferrández O, Grau S, et al. Design of a score to identify hospitalized patients at risk of drug-related problems[J]. Pharmacoepidemiol Drug Saf, 2014, 23(9): 923-932.

（责任编辑：周一青）