

一种简便的临床决策模型 ——注射用血栓通不良反应危险因素分析*

肖正国¹, 刘效栓^{1△}, 李喜香¹, 刘军刚¹, 张民¹, 王炯¹, 李莉¹, 豆金彦¹, 季文莹²

1 甘肃省中医院, 甘肃 兰州 730050; 2 兰州大学第二医院, 甘肃 兰州 730000

[摘要] 目的: 利用在兰州地区收集的数据, 探讨引起注射用血栓通(*Xueshuantong* injection, XSTI)不良反应的潜在危险因素, 并开发一种简单的工具以预测和控制其人为因素, 为临床合理用药提供依据。方法: 在兰州地区收集1 229例使用注射用血栓通的住院病例, 其中23例发生了不良反应。运用倾向性匹配(propensity score matching, PSM)对性别、年龄、民族、身高、体质量等混杂因素按照(1:4)匹配后, 得到共115个病例, 用决策树(decision tree, DT)分析建模, 用随机森林模型(random forest, RF)计算出各自变量的重要性。结果: 重要性排名前四的变量为合并用药种类(number of combined drugs, NOCD)、单位体表面积单次给药量(single dose per body surface area, DPBSA)、静滴速度(IV drip rate, IV rate)、溶媒种类(solution type, S type), 其平均Gini指数减少量>3.7, 是最重要的危险因素。最终临床预测模型中输出的变量只有NOCD、DPBSA, 是密切相关因素。结论: Decision Tree模型是一种简单的不良反应风险预测工具; 合并用药种类、单位体表面积给药剂量与注射用血栓通不良反应发生率密切相关; 注射用血栓通给药剂量应按体表面积法计算, 单位体表面积给药剂量应<289.2 mg/m²。

[关键词] 注射用血栓通; 倾向性匹配; 决策树; 随机森林; 单位体表面积给药剂量; 不良反应; 危险因素

[中图分类号] R285.6 **[文献标识码]** A **[文章编号]** 2096-9600(2025)03-0036-06

A Simple Clinical Decision Model: Analysis of Adverse Reaction Risk Factors of *Xueshuantong* Injection

XIAO Zhengguo¹, LIU Xiaoshuan^{1△}, LI Xixiang¹, LIU Jun'gang¹, ZHANG Min¹, WANG Jiong¹,
LI Li¹, DOU Jinyan¹, JI Wenyong²

1 Gansu Provincial Hospital of Traditional Chinese Medicine, Lanzhou 730050, China;

2 Lanzhou University Second Hospital, Lanzhou 730000, China

Abstract Objective: To discuss the potential risk factors of adverse reaction of *Xueshuantong* injection (XSTI) using the data collected in Lanzhou region, and to develop a simple tool for the prediction and management of its human factors, so as to provide the reference for clinical rational medication. Methods: A total of 1229 inpatient cases of applying XSTI were gathered in Lanzhou region, among them, 23 cases suffered adverse reactions. 115 medical cases were obtained after using propensity score matching (PSM) to match the confounder including gender, age, ethnic groups, body height and body mass in accordance with 1 : 4, and these medical cases were analyzed using decision tree (DT) for modeling, and the importance of the respective variables was calculated using the random forest (RF) model. Results: The top four variables in terms of importance were number of combined drugs (NOCD), single dose per body surface area (DPBSA), IV drip rate (IV rate) and solution type (S type), and the mean Gini index reduction>3.7, was the most important risk factor. The variables exported from the final clinical prediction model were NOCD and DPBSA, which were closely related factors. Conclusion: Decision Tree model is a simple tool for predicting the risk of adverse reactions; NOCD and DPBSA are closely related to the incidence of adverse reaction of XSTI; The dose of XSTI administered should be calculated according to the body surface area method, and the dose administered per unit of body surface area should be <289.2 mg/m².

Keywords *Xueshuantong* injection; propensity score matching; decision tree; random forest; single dose per body surface area; adverse reaction; risk factors

注射用血栓通(*Xueshuantong* injection, XSTI)是一种常用的中药注射剂,由三七总皂苷等组成^[1],含人参皂苷Rb1、Rd、Rg1,三七皂苷R1等主要成分和少量的未知成分。XSTI被广泛用于治疗中风偏瘫、胸痛、视网膜中央静脉阻塞等疾

病^[2]。倾向性评分匹配(propensity score matching, PSM)是一种均衡混杂因素的方法,近年来广泛用于大样本、非随机观察性研究^[3-4]。决策树(decision tree, DT)是数据挖掘领域常用的模型,它基于非线性判别方法,利用一组独立变量将

样本逐步分解成更小的子组。该过程在树的每个分支处迭代,根据特定的准则选择与因变量关联最强的自变量,从而构造一棵可用于预测新样本单元所属类别的树^[5]。随机森林(random forest, RF)通过多次随机抽样构建多棵树,通过平均 Gini 指数减少量排序给出各自变量用于预测的重要性评价。

有众多文献报道了 XSTI 在心脑血管及眼科疾病方面的有效性^[6-7]。但在合并用药种类、单位体表面积给药剂量、静滴速度等^[8]潜在的危险因素与不良反应之间的相关性方面未检索到相关报道。笔者通过 XSTI 上市后的安全性评价项目,利用在兰州地区收集的数据,对影响 XSTI 不良反应发生的影响因素进行了研究。

1 资料与方法

1.1 资料来源与数据描述 在不干预临床医生用药、不影响患者治疗、不进行诱导提问的情况下,对 2015 年 4 月至 2016 年 2 月间兰州地区使用

XSTI 的住院患者集中监测,对包括性别、年龄、体质指数、个人过敏史、家族过敏史、批号、配置放置时间、溶媒种类、用量、单次给药剂量、给药频次合并用药在内的观察单位进行详细记录。收集到兰州地区住院患者 1 229 人,其中男 630 例,女 599 例;平均年龄(51.52±15.33)岁;平均身高(167.20±7.46)cm;平均体质量(66.719±9.671)kg;平均体质指数(23.523±3.13)kg/m²,其中 23 例发生不良反应,数据代码字典见表 1,给药剂量统计描述见表 2,合并用药数统计描述见表 3。前期研究表明,注射用血栓通不良反应的发生与年龄等因素之间存在潜在联系,是混杂因素。发生不良反应的组(观察组 23 例)与未发生不良反应的组(对照组 1206 例),两组在年龄、性别、民族、身高、体质量方面存在差异,组间临床特征基线不均衡(匹配前),见表 4。本研究采用倾向性匹配,做了组间均衡性校正。

表1 数据编码字典

变量编码	标签	多分类变量值
hospital	医院编号	
date	入组日期	
mr	病历号	
sex	性别	1=man,2=female
nationality	民族	1=ethnic Han,2=else
age	年龄	Age in years
height	身高	cm
bw	体质量	kg
PHA	过敏史	0=Without,1=with
allergen	过敏原	0=no,1=food,2=drug,3=unknown Reason,4=else
NAD	过敏药物种类	1=PN,2=sulfonamides,3=else
AtP	过敏表现	1=shock,2=laryngeal Edema,3=bronchospasm,4=erythra,5=pruritus,6=stomachache,7=constipation,8=emesis,9=diarrhea,10=else
FHOA	家族过敏史	0=Without,1=with
PR	与本人关系	0=elas,1=parent,3=sibling
diagnose	用药主诊断	1=CHD,2=Silent Myocardial Ischemia,3=stenocardia,4=Acute Ventricular MyocardialIschemia,5=Ischemic Cardiomyopathy,6=cerebral Infarction,7=Cerebral Hemorrhage Convalescence,8=Transient Ischemic Attack,9=Obstruction Of Central Retinal Vein,10=else
HYUIB	是否首次使用	0=no,1=yes
DFA	给药时间(天)	
group	是否发生不良反应	0=no,1=yes
ARS	不良反应症状	0=no,1=damage to skin and accessories,2=Systemic damage,3=Central and peripheral nervous system damage,4=Gastrointestinal damage,5=Damage to the extracardial vessels,6=Cardiac toxicity,7=Respiratory damage,8=Visual damage,9=Damage to the musculoskeletal system,10=Metabolic dystrophy,11=Damage to the hepatobiliary system,12=Urinary system damage
CNA	过敏反应病例流水号	
BN	批号	-

续表 1

变量编码	标签	多分类变量值
CPT	配置放置时间(min)	-
BSA	人体体表面积	-
DF	给药频次(次/天)	-
SDD	单次给药剂量(mg)	-
Daily dosing	每日给药剂量	-
DPBSA	单位体表面积单次剂量(mg/m ²)	-
SD group	单次给药剂量分组	-
S type	溶媒种类	-
SolvV	单次溶媒体积(mL)	-
Drug C	给药浓度(mg/mL)	-
Drop Sp	静滴速度(滴/min)	-
NOCD	合并用药种类	-

注:-表示无变量值

表2 给药剂量统计描述

单次给药剂量(mg)	例数	构成比(%)	平均剂量(mg)	标准差
100	4	0.3		
200	12	1.0		
250	3	0.2		
300	50	4.1		
400	54	4.4	488.45	4.364
450	114	9.3		
500	901	73.3		
600	66	5.4		
750	25	2.0		
总计	1229	100.0		

表3 合并用药数分组统计描述

合并用药种类(种)	例数	最低值	最高值	平均值
≤5	657			
5<种类<10	303			
有效 ≤15	165	3	30	6.70
>15	102			
总计	1227			
缺失	2			
总计	1229			

1.2 统计学方法 使用 IBM SPSS Statistics 22.0 及 R 3.4.3 软件,基于信息熵算法的 Decision Tree 模型设定随机种子数(36 902),从 115 例中随机抽取 100 例为训练集建模,其余 15 例为测试集,对模型进行评价;Random Forest 模型基于平均 Gini 指数减少量对涉及的自变量进行排序(ntree=5 000);连续变量年龄、身高、体质量采

用 Independent-Samples T-Test 检验,分类变量性别、民族采用 χ^2 检验,当最小期望频数<5 时,用 Fisher 确切概率检验,静滴速度组间比较用 t 检验(滴注速度并非严格意义上的连续变量)。

2 结果

2.1 协变量与倾向性匹配 将所有的患者分为两组,发生不良反应为观察组 23 例,未发生不良反应为对照组 1 206 例。倾向评分的协变量是性别、年龄、民族、身高、体质量,用 PSM 法按 1:4 匹配,用以调整基线临床特征的差异,得到观察组 23 例,对照组 92 例,共 115 例。患者匹配后性别、年龄、民族、身高、体质量在两组间无差异,匹配后结果见表 4。

a 连续型变量使用 Independent-Samples T-Test 检验,b 分类变量使用 χ^2 检验,c 有一个期望频数小于 1 采用 Fisher 确切概率检验。

2.2 模型建立及评价 设随机种子数为 36902,从 115 例中随机抽取 100 例作为训练集,构建 Decision Tree 模型,输出图 1,剩余 15 例作为测试集测试。

随机森林模型算法:rf<-random Forest (group~HYUIB+PHA+DF+SDD+daily dosing+DPBSA+S Type+Solv V+drug C+IV rate+NOCD,data=a,n tree=5000,importance=TRUE,proximity=TRUE)

输出 Cleveland,见图 2,得到变量重要性排序见表 5。

根据 Decision Tree 模型(图 1)分析提示,当合并用药≥7 种,单位体表面积单次给药剂量≥289.2 mg/m²时,就判断为有发生不良反应可能(准确度 66.7%、灵敏度 90%、特异度 20%)。该临床预测模型简便提示医务人员,该患者合并用药数是

否超过控制线(7种),单位体表面积给药量是否 超过警戒线($\geq 289.2\text{ mg/m}^2$)。

表4 倾向性评分匹配前后两组间临床特征及P值变化

	基线资料	匹配前倾向性评分($n=1229$)			P	基线资料	匹配后倾向性评分($n=115$)			P
		发生不良反应	没有发生不良反应				发生不良反应	没有发生不良反应		
平均年龄(岁, $\bar{x}\pm s$)	51.52 \pm 15.334	41.65 \pm 15.361	51.71 \pm 15.289	0.002	43.58 \pm 12.926	41.65 \pm 15.361	44.07 \pm 12.292	0.490		
性别				0.451					0.851	
男[n(%)]	51.30%(630/1229)	43.5%(10/23)	51.4%(620/1206)		45.20%(52/115)	43.5%(10/23)	45.7%(42/92)			
女[n(%)]	48.7%(599/1229)	56.5%(13/23)	48.6%(586/1206)		54.80%(63/115)	56.5%(13/23)	54.3%(50/92)			
民族				0.396					1	
汉族[n(%)]	93.30%(1147/1229)	23(100.0%)	93.2%(1124/1206)		115(100%)	23(100%)	92(100%)			
其他民族[n(%)]	6.70%(82/1229)	0(0.0%)	6.8%(82/1206)		0	0	0			
平均身高(cm, $\bar{x}\pm s$)	167.20 \pm 7.649	167.88 \pm 6.823	167.18 \pm 7.483	0.659	167.85 \pm 7.362	167.87 \pm 6.824	167.85 \pm 7.526	0.990		
平均体质量(kg, $\bar{x}\pm s$)	65.72 \pm 9.374	66.931 \pm 9.0637	65.696 \pm 9.3792	0.531	66.77 \pm 9.503	66.93 \pm 9.063	66.73 \pm 9.657	0.926		

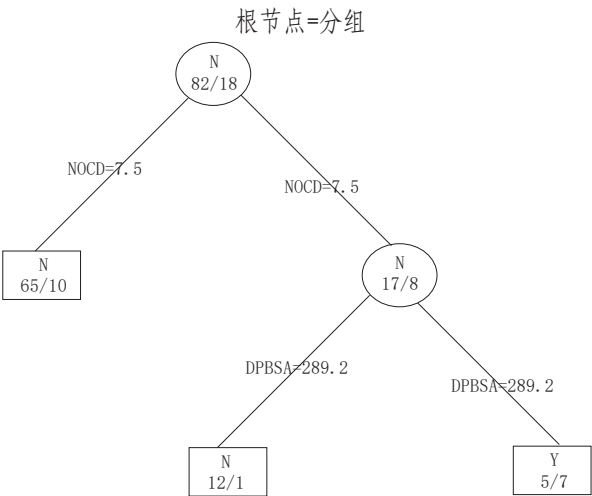


图1 基于信息熵算法的决策树模型

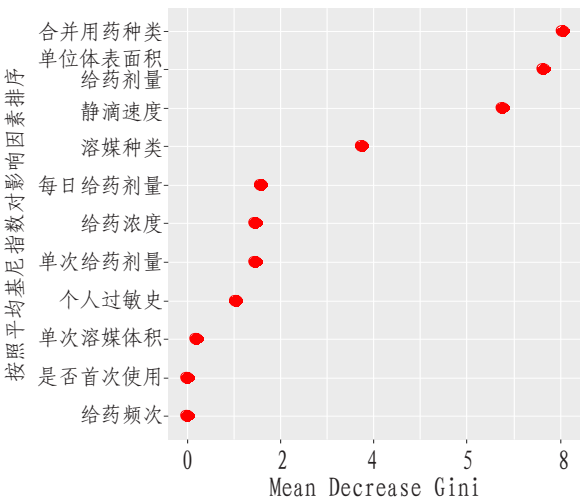


图2 MDG数据Cleveland图

2.3 自变量重要性分析 由图2及表5结果分析可知,重要性排名前四的变量为合并用药种类(number of combined drugs,NOCD)、单位体表面积单次给药剂量(DBPSA)、静滴速度(IV rate)、溶媒种类(S Type),其平均Gini指数减少量 ≥ 3.7 ,

是最重要的危险因素。但Decision Tree模型中只输出的变量只有NOCD、DPBSA,因此将静滴速度(Iv rate)、溶媒种类(S Type)进一步分析,做Independent-Samples T-Test检验,结果显示静滴速度在两组间存在差异但无统计学意义(见表6),因此即使在变量重要性排名第三,最终模型中还是没有输出该变量。

对溶媒种类分析,发现发生不良反应的一组中有1例使用了葡萄糖氯化钠注射液作为溶媒,分析结果无统计学意义,也无临床实际意义。

为评价表5中MDG排序第五与第四的变量之间是否有差别,本研究做了ROC曲线下[灵敏度(1~0.9)、特异度(1~0.8)]偏AUC分析[roc1<-roc(a\$group,a\$SType),roc2<-roc(a\$group,a\$daily dosing)],结果灵敏度pAUC分别为0.064、0.006;特异度pAUC分别为0.121/0.017; $P<0.001$;见图3—6。因此对MDG $<1.592\ 243$ 的变量不再分析。

表5 各自变量重要性排序

变量代码	中文标签	平均减少基尼系数
NOCD	合并用药种类	8.046139
DPBSA	单位体表面积给药剂量(mg/m^2)	7.640889
Iv rate	静滴速度(滴/min)	6.765919
S Type	溶媒种类(种)	3.743778
daily dosing	每日给药剂量(mg)	1.592243
drug C	给药浓度(mg/mL)	1.472925
SDD	单次给药剂量(mg)	1.468894
PHA	个人过敏史	1.044482
Solv V	单次溶媒体积(mL)	0.197339
HYUIB	是否首次使用	0.019245
DF	给药频次(次)	0.001254

表6 静滴速度组间统计

组别(是否发生不良反应)	数量	平均值 (滴/min)	标准 偏差 (滴/min)	标准误差 平均值 (滴/min)	t	p
否	92	61.68	13.505	1.408	0.754	0.452
是	23	64.09	14.321	2.986		

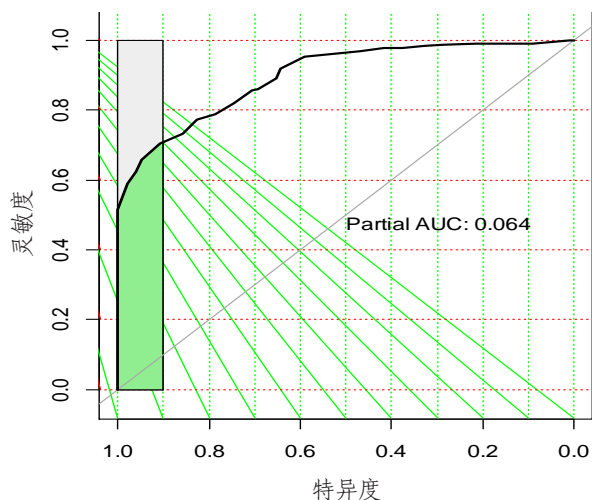


图3 溶媒种类在ROC曲线下(特异度0.9-1)偏AUC面积

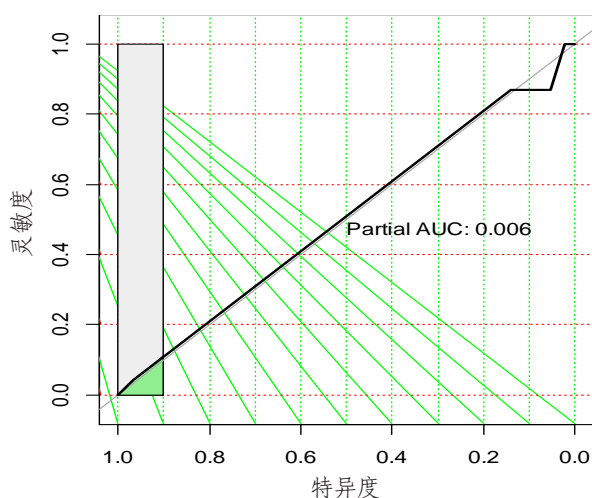


图4 每日给药剂量ROC曲线下(特异度0.9-1)偏AUC面积

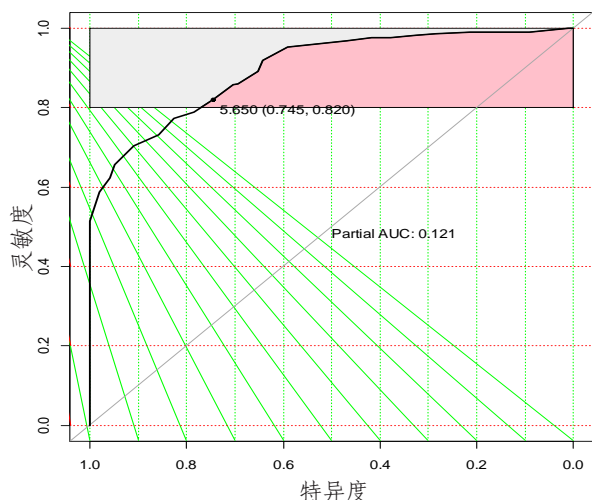


图5 溶媒种类在ROC曲线下(灵敏度0.8-1)偏AUC面积

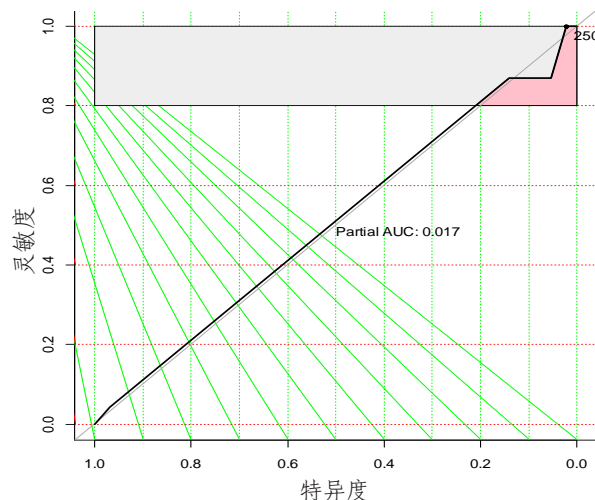


图6 每日给药剂量ROC曲线下(灵敏度0.8-1)偏AUC面积

3 讨论

合并用药指用两种以上的药物治疗一种或多种疾病,这是常见的治疗措施^[9]。多种药物配合得好可提高疗效,减少不良反应。但是合并用药过多,安全性就难以保证^[10]。从注射用血栓通标准指纹图谱可知,人参皂苷Rb1、Rd、Rg1,三七皂苷R1为注射用血栓通的主要成分^[11-12],同时又存在微量的未知成分^[12],这些复杂成分间药代动力学、药效学影响随着合并用药数增加复杂性成数量级增加,药物间相互作用尚不完全清楚。本研究表明1 229例住院患者合并用药数最少3种,最多30种;657例合并用药3~5种,303例合并用药6~10种,165例合并用药11~15种,102例合并用药15~30种,平均合并用药6.7种。模型提示当合并用药数 ≥ 7 种时,与注射用血栓通不良反应发生密切相关。

注射用血栓通说明书中用法用量为:“静脉滴注,每次250~500 mg,用5%或10%的葡萄糖或0.9%氯化钠注射液250~500 mL稀释,每日1次。”本试验数据显示,最小剂量100 mg(4例),最大用量750 mg(25例),其中500 mg(901例);年龄小于18岁者共8例,全部为500 mg,有11例70岁以上的老人给药剂量为600 mg。由此看来临床给药剂量计算的方法缺乏统一标准。本研究表明,注射用血栓通药品不良反应发生与单位体表面积单次给药剂量密切相关,当给药量 ≥ 289.2 mg/m²时,发生概率为66.7%,提示注射用血栓通在临床应用时,应按体表面积计算给药剂量,建议补充说明书中相关内容。

表5中显示,对于单位体表面积单次给药剂量(DPBSA),MDG=7.640 889;而每日给药剂量(daily dosing)MDG=1.592 243,对模型贡献度不

大。进一步说明注射用血栓通按体表面积计算给药剂量的必要性。

尽管模型中没有给出静滴速度具体警示值,变量重要性排名在第三,也是不可忽略的因素。进一步对两组静滴速度使用 Independent-Samples T-Test 检验,发现两组之间的均值有差异,但差异无统计学意义。根据本研究的结果(表6),建议静脉滴注速度 ≤ 60 滴/min。

合并用药数、用药剂量、静脉输注速度都是临床常见问题,本研究运用了数据挖掘领域常用的方法^[13-16]:PSM、Decision Tree、Random Forest,从众多影响因素中筛选出排名前十的变量,首次从理论上给出了建议值,并且给出了简便的不良反应预测模型,具有一定的临床意义。和经典统计模型相比较,数据挖掘方法中 Decision Tree、Random Forest 模型可以自动处理大量的自变量,不用担心无关变量纳入模型后干扰模型效果等问题,同时该方法为非参数方法,没有太多的使用条件限制。这一点与 Logistic 回归等方法相比有极大的优势。

作者所在单位参与了“注射用血栓通上市后安全性集中监测研究”项目,但本研究只收集到兰州地区分中心数据。因此,尽管做了 PSM 组间均衡性矫正,该数据仍无法代表其他地区的研究结论。

参考文献

- [1] 李春晓,陈玉欢,凌霄,等.基于真实世界数据的注射用血栓通(冻干)治疗脑梗死急性期疗效比较研究[J].中医杂志,2023,64(16):1661-1669.
- [2] XING W, SONG Y, LI H, et al. Fufang Xueshuantong protects retinal vascular endothelial cells from high glucose by targeting YAP[J]. Biomed Pharmacother, 2019, 120:109470.
- [3] WANG J. To use or not to use propensity score matching[J]. Pharm Stat, 2021, 20(1):15-24.
- [4] KANEL T, FANG T, GALETAM S, et al. Propensity score matching: a statistical method[J]. Clin Spine Surg, 2020, 33(3):120-122.
- [5] YUS H, WANGH L. An updated decision tree for horizontal ridge augmentation: a narrative review[J]. Int J Periodontics Restorative Dent, 2022, 42(3): 341-349.
- [6] YUAN Q, WANGF J, JIAZ Z, et al. Xueshuantong injection combined with Salvianolate lyophilized injection improves the synaptic plasticity against focal cerebral ischemia/reperfusion injury in rats through PI3K/AKT/mTOR and RhoA/ROCK pathways[J]. Brain Res, 2022, 178(7):147923.
- [7] HAN S, CHEN Y, WANG J, et al. Anti-thrombosis effects and mechanisms by Xueshuantong capsule under different flow conditions[J]. Front Pharmacol, 2019, 10:35.
- [8] 翟宇瑶,段万里,胡斌,等.活血化瘀类中药注射剂临床应用合理性分析[J].西部中医药,2021,34(5):112-115.
- [9] SHITAR G, AZULAY L, NIZRI O, et al. CDCDB: a large and continuously updated drug combination database[J]. Sci Data, 2022, 9:263.
- [10] HORN J, UENG S. The effect of patient-specific drug-drug interaction alerting on the frequency of alerts: a pilot study[J]. Ann Pharmacother, 2019, 53(11):1087-1092.
- [11] 国家药典委员会药品标准服务平台.注射用血栓通(冻干)电子标准图谱[DB/OL].(2019-12-19)[2023-06-16]. <http://fuwu.chp.org.cn/account/orders>.
- [12] PINTUSOPHON S, NIU W, DUANX N, et al. Intravenous formulation of Panax notoginseng root extract: human pharmacokinetics of ginsenosides and potential for perpetrating drug interactions[J]. Acta Pharmacol Sin, 2019, 40(10):1351-1363.
- [13] 孙雪,梁鹏,梁建庆,等.基于数据挖掘分析名老中医治疗帕金森病的临证经验[J].西部中医药,2024,37(3):32-38.
- [14] 徐小港,王钰,徐义峰,等.基于数据挖掘的痛经用药规律研究[J].西部中医药,2024,37(2):99-103.
- [15] 张丽娟,陈月莹,庞然,等.基于数据挖掘方法探讨高荣林对失眠的诊治规律[J].西部中医药,2023,36(9):72-77.
- [16] 何金丽,李淑玲,雷作汉.基于数据挖掘刘国安教授治疗老年高血压临床经验研究[J].西部中医药,2023,36(6):50-53.

收稿日期:2024-03-05

*基金项目:2022年中央转移支付医疗服务与保障能力提升(中医药事业传承与发展部分)——中医药人才培养重点学科建设项目(甘财社[2022]48号)。

作者简介:肖正国(1966—),男,主任药师。研究方向:医院药学。

△通讯作者:刘效栓(1964—),男,硕士学位,主任药师。研究方向:医院药学。Email:liuxiushuang1964@163.com。