

# 化疗致恶性肿瘤患儿骨髓抑制风险预测模型的系统评价<sup>Δ</sup>

何莉\*, 林欣, 蒋小平<sup>#</sup>(重庆医科大学附属儿童医院护理部/儿童少年健康与疾病国家临床医学研究中心/儿童发育疾病研究教育部重点实验室/儿童神经发育与认知障碍重庆市重点实验室, 重庆 400014)

中图分类号 R979.1 文献标志码 A 文章编号 1001-0408(2026)07-0954-06

DOI 10.6039/j.issn.1001-0408.2026.07.22



**摘要** **目的** 系统评价儿童恶性肿瘤化疗后骨髓抑制风险预测模型的建模策略、关键预测因子与预测性能,为临床决策与研究提供循证依据。**方法** 检索2025年4月之前发表于中国知网、万方数据、PubMed等11个数据库中的相关文献。文献筛选和信息提取由2位研究人员独立完成,模型的偏倚风险和适用性依照PROBAST工具进行严格评估。**结果** 最终筛选得到7项研究,其中英文文献4篇、中文文献3篇,涉及12个风险预测模型。多数模型在判别能力方面表现良好,受试者工作特征曲线下面积(AUC)为0.748~0.981,但仅有2项研究实施了外部验证;有3项研究未充分报告模型的校准信息。PROBAST评估结果显示,所有模型在偏倚风险方面均为高水平,主要问题包括以回顾性设计为主、样本代表性不足以及缺乏盲法评估等;但在适用性方面,所有模型的评价结果均为良好。在建模方法方面,多数研究采用传统logistic回归方法构建模型,仅少数研究引入了机器学习算法并对多种算法进行了系统性比较;采用机器学习方法的模型表现明显优于传统统计方法构建的模型。**结论** 现有的儿童恶性肿瘤化疗后骨髓抑制风险预测模型在临床风险预警中显示出潜力,但普遍存在以回顾性单中心设计为主、每变量事件数较低、缺失数据处理不透明、模型系数报告不一致等设计与方法学局限。未来研究应推进前瞻性设计,融合机器学习与关键临床变量,并严格遵循TRIPOD声明等报告规范,以提升模型的科学严谨性与临床适用性。

**关键词** 风险预测模型;恶性肿瘤;化疗;骨髓抑制;儿童;预测方法

## Systematic review of risk prediction models for chemotherapy-induced myelosuppression in pediatric patients with malignant tumors

HE Li, LIN Xin, JIANG Xiaoping (Dept. of Nursing, Children's Hospital of Chongqing Medical University/National Clinical Medical Research Center for Children and Adolescents' Health and Diseases/Ministry of Education Key Laboratory of Child Development and Disorders/Chongqing Key Laboratory of Child Neurodevelopment and Cognitive Disorders, Chongqing 400014, China)

**ABSTRACT** **OBJECTIVE** To systematically evaluate risk prediction models for chemotherapy-induced myelosuppression in pediatric patients with malignant tumors, evaluate their modeling strategies, key predictors, and predictive performance, and provide evidence-based references for clinical decision-making and research. **METHODS** A literature search was conducted across 11 databases, including CNKI, Wanfang Data, and PubMed, for relevant studies published before April 2025. Two reviewers independently performed literature screening and data extraction, and the risk of bias and applicability of the models were evaluated using the PROBAST tool. **RESULTS** Ultimately, seven studies were selected, of which four were English articles and three were Chinese articles, involving 12 risk prediction models. Although model discrimination was good (AUC 0.748-0.981), only two models underwent external validation; furthermore, calibration was inadequately reported in three studies. PROBAST indicated that all models exhibited a high risk of bias, with major issues including a predominance of retrospective designs, inadequate sample representativeness, and lack of blinding. However, in terms of applicability, all models received favorable evaluations. In terms of modeling methods, most studies employed traditional logistic regression approaches to construct models, while only a minority introduced machine learning algorithms and conducted systematic comparisons among multiple algorithms. Models developed using machine learning methods significantly outperformed those constructed with traditional statistical methods. **CONCLUSIONS** The existing risk prediction models for myelosuppression after chemotherapy in children with malignant tumors demonstrate potential in clinical risk early warning. However, they generally suffer from design and methodological limitations, such as a predominance of retrospective single-center designs, few events per variable, opaque handling of missing data, and inconsistent reporting of model

coefficients. Future studies should adopt prospective designs, incorporate machine learning with key clinical predictors, and follow TRIPOD reporting guidelines to enhance scientific rigor and clinical utility.

**KEYWORDS** risk prediction model; malignant tumor; chemotherapy; myelosuppression; child; prediction method

<sup>Δ</sup> 基金项目 重庆医科大学护理学院院级科学研究项目 (No. 20230308)

\* 第一作者 护士, 硕士研究生。研究方向: 儿童肿瘤、心理护理。  
E-mail: 2550650722@qq.com

<sup>#</sup> 通信作者 教授, 主任护师, 硕士生导师, 硕士。研究方向: 儿童肿瘤、心理护理。E-mail: 1439638239@qq.com

据《国家儿童肿瘤监测年报(2022)》统计,2019—2020年度我国0~18岁人群的新发肿瘤病例达125.72/100万,其中以白血病及中枢神经系统肿瘤最为常见<sup>[1]</sup>。在临床实践中,化疗为治疗恶性肿瘤的主要方式,但其引发的骨髓抑制比例高达30%~60%,是影响治疗连续性的主要因素;而严重的骨髓抑制不仅可能迫使治疗方案中断,还会显著增加感染风险,甚至可能导致死亡,从而显著影响患儿的整体治疗效果和生活质量<sup>[2-3]</sup>。

随着精准医学的发展,恶性肿瘤患儿的化疗逐步走向个体化,这使得骨髓抑制的风险评估与干预面临诸多挑战<sup>[4-5]</sup>。虽然基线中性粒细胞数、血小板水平等临床指标被证实对骨髓抑制具有一定的预测价值<sup>[6]</sup>,但当前的风险预测模型仍存在诸多不足:一方面,传统的统计建模方法难以有效捕捉变量之间复杂且非线性的关系;另一方面,多数现有模型缺乏跨中心外部验证,限制了其在临床中的广泛应用与推广<sup>[7]</sup>。目前,尚缺乏针对儿童群体中该风险预测模型的系统综述,也未见研究对不同建模策略的效果、外部验证情况及儿童特有因素的融入进行深入比较。鉴于此,本研究旨在系统回顾与评估儿童恶性肿瘤化疗后骨髓抑制风险预测模型现状,通过整合现有证据,为临床个体化治疗提供依据,以改善治疗质量与患儿预后。

## 1 资料与方法

### 1.1 研究问题的确立

本研究依照PICOTS框架<sup>[8]</sup>明确各要素。其中P(population)为确诊恶性肿瘤并接受化疗的儿童,年龄界定为0~18岁;I(intervention model)指包括所有已建立并可用于预测化疗相关骨髓抑制风险的模型;C(comparator)不适用本研究;O(outcome)指是否发生骨髓抑制;T(timing)指在化疗前进行风险评估;S(setting)指适用于任何环境。基于上述要素,本研究的研究问题为:探讨0~18岁恶性肿瘤患儿中,现有化疗相关骨髓抑制风险预测模型的预测效能及临床实践适用性。本研究已在PROSPERO平台完成登记,注册编号为CRD420251038713。

### 1.2 文献纳入与排除标准

本研究的纳入标准包括:(1)研究对象为符合《儿童肿瘤国际分类第3版(ICCC-3)》或世界卫生组织(WHO)相关标准<sup>[9]</sup>的0~18岁恶性肿瘤患儿;(2)研究内容涉及恶性肿瘤患儿化疗后骨髓抑制风险预测模型的构建或验证;(3)研究类型包括病例对照研究、随机对照研究、队列研究以及横断面研究;(4)研究需明确报告模型构建过程及所用统计方法。

本研究的排除标准包括:(1)语言为非中、英文的资料;(2)仅为会议摘要、信件等非正式出版文献,或综述性文章;(3)重复发表的研究;(4)全文无法获取或摘要信息不全的文献;(5)仅分析影响因素或危险因素,而未建立风险预测模型的研究。

### 1.3 文献检索策略

本研究结合自由词和主题词,对11个数据库进行了系统检索,包括中国知网、维普、万方数据、中国生物医学文献库、PubMed、The Cochrane Library、Web of Science、CINAHL、Scopus、Embase以及ScienceDirect;同时,对所选文献的参考文献进行追溯分析,检索时限均为建库至2025年4月。中文检索词为“儿童”“患儿”“肿瘤”“癌症”“恶性肿瘤”“化疗”“骨髓抑制”“中性粒细胞减少”“预测”“预测因子”“风险预测”“风险评估”“风险评分”“风险因素”“预测因素”“模型”“列线图”“诊断模型”“临床预测模型”;英文检索词为“child”“pediatric”“neoplasms”“neoplasia”“tumor”“cancer”“malignant neoplasm”“chemotherapy”“forecasting”“prediction methods”“risk assessment”“nomograms”“prediction”“risk analyse”“model”“myelosuppression”“neutropenia”。

### 1.4 文献筛选与资料提取

将初步筛选的文献导入NoteExpress文献管理软件,由2位研究者分别完成文献遴选与信息提取,随后进行交叉核对。若核对结果不一致,双方将讨论解决;若无法解决,则邀请第3位研究者共同决策,直到达成一致。文献筛查过程中,先依据标题与摘要初步剔除无关文献,随后阅读全文以判断其是否符合纳入标准。完成筛选后,依据CHARMS框架<sup>[10]</sup>提取相关信息,包括第一作者、研究地区、研究类型、样本数量、模型性能等内容。

### 1.5 偏倚风险与适用性评价

2位研究者分别采用PROBAST工具<sup>[11]</sup>,对所纳入研究的偏倚风险及适用性进行独立评估,若评估结果出现分歧,则邀请第3位研究者进行讨论,并在协商后达成共识,最终确定评估结论。

#### 1.5.1 偏倚风险评估

PROBAST工具包括20个条目,围绕研究对象、预测因子、结局及统计分析4个方面设有标准化问题,选项包括“是”“可能是”“否”“可能否”“不清楚”。若某一方面内所有问题均选“是”或“可能是”,则其偏倚风险判断为低;如存在“否”或“可能否”,则该方面偏倚风险为高;若某项为“不清楚”、其余项为低风险时,则该方面偏倚风险评定为“不清楚”。

#### 1.5.2 适用性评价

适用性评价从研究对象、预测因子及结局3个方面展开。评估人员依据实际情况对每个方面的适用性进行判定,结果分为“适用性好”“适用性差”和“适用性不清楚”。若3个方面均为“适用性好”,则整体判定为适用性好;若任一方面被评为“适用性差”,则整体适用性为差;如果某些方面的适用性被评为“适用性不清楚”,但其他方面为“适用性好”,则整体判定为“适用性不清楚”。

## 2 结果

### 2.1 文献筛选结果

初步检索11个数据库,共获得相关文献767篇。经去重处理,剔除重复文献37篇,剩余文献730篇。随后,通过阅读文献标题和摘要进行初筛,共排除文献382篇,排除原因包括主题不符、研究对象不符、动物或细胞研究、Meta分析及综述类研究,初筛后获得文献348篇。进一步通过阅读全文进行复筛,共排除文献341篇,排除原因包括综述类文献、主题不符、研究对象不符、仅分析危险因素而未构建模型的研究。最终,共筛选符合纳入标准的研究文献7篇<sup>[2,5,12-16]</sup>。

### 2.2 纳入研究的基本特征

纳入的7项研究<sup>[2,5,12-16]</sup>中,3项<sup>[2,5,12]</sup>为近4年内发表。4项<sup>[2,5,12-13]</sup>研究来自中国(3篇中文文献<sup>[5,12-13]</sup>,1篇英文文献<sup>[2]</sup>),另各有1项研究在荷兰<sup>[14]</sup>、美国<sup>[15]</sup>、瑞士<sup>[16]</sup>开展,均为英文文献;4项<sup>[2,5,12,16]</sup>为单中心研究,3项<sup>[13-15]</sup>为多中心研究;4项<sup>[2,5,13,16]</sup>为回顾性队列设计,3项<sup>[12,14-15]</sup>为前瞻性队列研究;在结局判定方面,2项<sup>[12,15]</sup>依据美国国立综合癌症网络(National Comprehensive Cancer Network, NCCN)临床实践指南进行评估,2项<sup>[2,13]</sup>参考WHO抗癌药物急性及亚急性毒性反应分度标准,1项<sup>[15]</sup>同时采用美国国家癌症研究所和欧洲医学肿瘤学会的标准,1项<sup>[15]</sup>采用医院感染暴发控制指南,2项<sup>[14,16]</sup>采用由瑞士儿科肿瘤协作组(Swiss Paediatric Oncology Group, SPOG) Ammann等提出的风险评估标准(以下简称“SPOG Ammann标准”)。纳入研究的基本特征见表1。

### 2.3 模型建立情况

纳入的7项研究<sup>[2,5,12-16]</sup>共构建了12个模型,各研究中模型所涉及的候选预测因子数量为4~46个;有2项研究<sup>[5,13]</sup>将原始连续变量转化为分组后的分类变量进行处理;纳入研究的样本总量为180~3 000例,各研究中结果事件的数量为46~1 129例;仅有1项研究<sup>[2]</sup>提及了缺失值及其处理方式。在建模方法方面,4项研究<sup>[5,12-13,15]</sup>采用了线性回归(linear regression, LR)进行分析,1项研究<sup>[14]</sup>使用加权评分模型(weighted scoring model, WSM),1项研究<sup>[16]</sup>为泊松回归(Poisson regression, PR),1项研究<sup>[2]</sup>采用机器学习算法建立预测模型,包括极限梯度提升(extreme gradient boosting, XGBoost)、分类提升算法(categorical boosting, CatBoost)、

随机森林(random forest, RF)、支持向量机、最小绝对收缩和选择算子回归及LR等方法。在变量筛选方面,5项研究<sup>[5,12-13,15-16]</sup>先进行了单因素分析,随后纳入多因素建模;1项研究<sup>[2]</sup>采用信息值和变量相关性分析,1项研究<sup>[14]</sup>基于临床和实验室参数进行筛选。

### 2.4 模型性能与临床表征

在所纳入的7项研究<sup>[2,5,12-16]</sup>中,有5项<sup>[2,5,12-13,15]</sup>报告了受试者工作特征曲线下面积(AUC)值,范围为0.748~0.981;有3项研究<sup>[2,13-14]</sup>通过敏感性、特异性或预测准确率等指标对模型性能进行评估。在模型校准方面,3项研究<sup>[14-16]</sup>未报告校准方法,2项研究<sup>[2,12]</sup>采用Hosmer-Lemeshow检验(H-L检验)并绘制校准曲线,1项研究<sup>[13]</sup>只采用了H-L检验。在模型验证方面,共有6项研究<sup>[2,5,12-13,15-16]</sup>完成了内部验证,其中1项<sup>[2]</sup>采用了五折交叉验证,2项研究<sup>[14-15]</sup>实施了外部验证。在模型构建阶段,纳入的候选预测因子涵盖人口学特征、实验室指标、既往疾病史、治疗方案及生活方式相关变量;最终预测模型中出现频率较高的预测因子包括性别、年龄、白细胞计数和红细胞计数等<sup>[2,5,12,16]</sup>,这些变量较为容易获取和测量。模型最终的呈现形式各有不同,1项研究<sup>[15]</sup>为风险分层,2项研究<sup>[2,14]</sup>为风险评分,2项研究<sup>[5,12]</sup>为列线图,1项研究<sup>[13]</sup>为回归方程,1项研究<sup>[16]</sup>未报告模型最终呈现形式。

### 2.5 偏倚风险和适用性评价

在纳入的7项研究<sup>[2,5,12-16]</sup>中,整体偏倚风险评价均为高风险,但所得预测模型在不同领域及总体上均显示出较好的适用性。结果见表2。

#### 2.5.1 研究对象

在纳入的7项研究<sup>[2,5,12-16]</sup>中,4项<sup>[2,5,13,16]</sup>被判定存在高偏倚风险。因为这些研究<sup>[2,5,13,16]</sup>均采用回顾性设计,数据的完整性和准确性难以保证,同时在研究对象选择过程中可能引入选择性偏倚。

#### 2.5.2 预测因子

在预测因子相关内容中,4项研究<sup>[2,5,13,16]</sup>被判定存在高偏倚风险,这也是由于所纳入文献<sup>[2,5,13,16]</sup>均采用回顾性设计。该类研究在知晓结局后,回顾性地获取潜在预测因子的数据,往往难以实施盲法,这可能导致结局信息对预测因子的测量过程产生潜在影响,从而增加偏倚风险。

表1 纳入研究的基本特征

第一作者	发表年份	国家	研究类型	单中心/ 多中心	研究对象	结局评估方法	预测结局
刘若彤 <sup>[5]</sup>	2024	中国	回顾性队列研究	单中心	化疗后的肿瘤患儿(急性淋巴细胞白血病、急性髓系白血病和实体肿瘤)	医院感染暴发控制指南	骨髓抑制
李伟涵 <sup>[12]</sup>	2022	中国	前瞻性队列研究	单中心	化疗后的肿瘤患儿(急性白血病和实体肿瘤)	NCCN临床实践指南	骨髓抑制
徐蕊 <sup>[13]</sup>	2018	中国	回顾性队列研究	多中心	接受化疗的淋巴瘤系恶性肿瘤患儿	WHO抗癌药物急性及亚急性毒性反应分度标准	骨髓抑制
Li <sup>[2]</sup>	2023	中国	回顾性队列研究	单中心	接受化疗的肾母细胞瘤患儿	WHO抗癌药物急性及亚急性毒性反应分度标准	骨髓抑制
Miedema <sup>[14]</sup>	2011	荷兰	前瞻性队列研究	多中心	化疗后的肿瘤患儿(急性淋巴细胞白血病、急性髓系白血病、淋巴瘤等)	SPOG Ammann标准	骨髓抑制
Lyman <sup>[15]</sup>	2017	美国	前瞻性队列研究	多中心	接受化疗的实体瘤或恶性淋巴瘤患儿	NCCN临床实践指南、美国国家癌症研究所的标准和欧洲医学肿瘤学会的标准	骨髓抑制
Wicki <sup>[16]</sup>	2008	瑞士	回顾性队列研究	单中心	接受化疗的恶性肿瘤患儿和青少年(急性淋巴细胞白血病、急性髓系白血病、淋巴瘤等)	SPOG Ammann标准	骨髓抑制

表2 模型的偏倚风险和适用性评价结果

纳入研究	偏倚风险				适用性			总体	
	研究对象	预测因子	结局	统计分析	研究对象	预测因子	结局	偏倚风险	适用性
刘若彤 <sup>[8]</sup>	+	+	+	+	-	-	-	+	-
李祎涵 <sup>[12]</sup>	-	-	?	+	-	-	-	+	-
徐蕊 <sup>[13]</sup>	+	+	+	+	-	-	-	+	-
Li <sup>[2]</sup>	+	+	+	+	-	-	-	+	-
Miedema <sup>[14]</sup>	-	-	+	+	-	-	-	+	-
Lyman <sup>[15]</sup>	-	-	+	+	-	-	-	+	-
Wicki <sup>[16]</sup>	+	+	+	+	-	-	-	+	-

注：“-”表示低偏倚风险/适用性好；“+”表示高偏倚风险/适用性差；“?”表示偏倚风险/适用性不清楚。

### 2.5.3 结局

在结局方面,有6项研究<sup>[2,5,13-16]</sup>被判定为高偏倚风险,其主要问题在于:这些研究所构建的模型中包含的部分预测因子,实质上也是结局定义中相关因素的一部分,从而可能导致预测因子与结局之间的相关性被人为高估。此外,有1项研究<sup>[12]</sup>被评为偏倚风险不清楚,原因在于其未明确说明预测因子的测量时间与结局评估时间之间的间隔,同时在预测信息不清的情况下,也未说明是否采取盲法进行结局评估,从而影响了评估的可信度。

### 2.5.4 统计分析

纳入的7项研究<sup>[2,5,12-16]</sup>均被判定为高偏倚风险。6项研究<sup>[2,5,12,14-16]</sup>的主要问题是样本量过小,具体表现为每变量事件数(events per variable, EPV) < 20,其中4项研究<sup>[12-14,16]</sup>的EPV < 10;另外,有6项研究<sup>[5,12-16]</sup>未说明缺失值情况及其处理策略;2项研究<sup>[2,14]</sup>在选择预测因子时未进行单因素分析;3项研究<sup>[14-16]</sup>未提供模型的校准信息;有1项研究<sup>[13]</sup>仅通过H-L检验评估模型拟合情况;另有1项研究<sup>[14]</sup>未实施内部验证。所有7项研究均未说明数据复杂性的处理情况,且未对开发过程中使用的最终模型的预测因子系数或截距与多变量分析报告结果之间的一致性进行评估。

## 3 讨论

### 3.1 模型的主要预测因子

本研究通过对7项儿童恶性肿瘤化疗后骨髓抑制风险预测模型研究进行系统整合与分析,发现与初步筛选的候选预测因子相比,最终纳入模型的预测因子(如基线中性粒细胞计数、血小板水平、化疗方案强度和癌症类型)在预测准确性和临床适用性方面具有明显优势。这些关键变量的选取得到了循证医学的支持——研究已证实,化疗方案的强度、血小板以及血红蛋白水平与骨髓抑制的发生密切相关<sup>[17-18]</sup>。此外,血液系统恶性肿瘤的儿童患者由于骨髓浸润及接受高强度化疗,其发热性中性粒细胞减少的发生率显著高于实体瘤患者<sup>[2,19]</sup>,且不同类型肿瘤中,中性粒细胞计数的变化存在明显差异,这使其成为预测骨髓抑制风险的关键指标。

值得注意的是,尽管某些候选预测因子(如基因多

态性)在理论上可能具备潜在的预测价值,但由于其检测成本较高且在临床中普及性较低,因此尚未广泛应用于实际的预测模型中<sup>[20]</sup>。这些结果提示,在进行骨髓抑制风险预测时,应结合儿童肿瘤患者的生理特征与治疗差异,优先选择那些易于获取、便于量化且与疾病机制紧密相关的指标,从而为临床上个体化风险评估方案的制定提供重要依据。

### 3.2 预测因子筛选与模型建立方法

在本研究纳入的文献中,多数预测模型在变量筛选阶段仍主要依赖传统统计分析方法,只有少数研究尝试引入基于机器学习的特征筛选技术,如信息值评估和变量间相关性分析<sup>[2]</sup>。尽管传统统计方法在识别独立预测因子方面有一定的优势,但其在捕捉变量间复杂的非线性关系和潜在交互效应时存在一定的局限性<sup>[21]</sup>;相比之下,采用XGBoost和RF等机器学习算法的研究不仅能更高效地处理高维数据、挖掘变量之间的深层交互模式,其模型性能(AUC=0.981)也显著优于传统逻辑回归模型<sup>[22]</sup>。

在模型构建策略方面,大多数研究<sup>[5,12-13,15]</sup>采用的是LR方法,仅少数研究<sup>[2,14,16]</sup>对WSM、PR和多种机器学习算法进行了系统性比较。这种传统建模方法占主导地位的趋势与文献<sup>[23]</sup>在乳腺癌风险预测模型研究中的发现相似,这反映了儿童恶性肿瘤化疗后骨髓抑制风险建模领域尚未充分利用现代机器学习技术。

基于此,未来的研究应考虑在模型设计中采用“混合方法”策略,即在预测因子筛选阶段结合机器学习算法与经典统计方法,并在模型建立过程中对多种算法进行系统性的性能评估<sup>[24]</sup>,以降低单一方法带来的偏倚,同时提高模型的泛化能力和临床适配性。

### 3.3 预测效能与临床适配性

本研究共纳入12个风险预测模型,AUC为0.748~0.981,多数模型在判别能力方面表现良好。然而,模型的校准与验证方面仍存在一定的局限性。在校准方面,只有少数研究<sup>[2,12]</sup>使用了H-L检验并结合校准曲线进行评估,而大多数研究未充分报告校准方法,这可能影响对模型预测精度的准确评估。有研究<sup>[2]</sup>通过机器学习算法建立模型,并采用五折交叉验证进行内部验证,最终模型的AUC达到了0.981,这表明现代算法在提升模型精度方面具有明显优势。然而,尽管机器学习模型在区分度上表现优异,但其临床转化仍面临可解释性与可操作性方面的挑战。该类模型结构复杂,预测过程缺乏透明性,并且该研究仅提供了全局的特征重要性,未使用如沙普利加性解释或局部可解释模型-不可知解释等更先进的技术来揭示个体化预测结果的决策依据,不利于获得临床医生的理解与信任;同时,多数模型未提供便捷的评分工具或与医院信息系统整合的接口,限制了其在实际医疗环境中的推广。因此,在认可其高AUC的

同时,也需清醒认识到其在临床落地过程中的现实瓶颈。

在模型验证方面,虽然大多数研究进行了内部验证,但仅有少数研究<sup>[14-15]</sup>完成了外部验证,这使得这些模型在不同人群和临床环境中的应用推广受到了限制<sup>[25]</sup>。此外,模型展示的形式存在较大差异,包括列线图、风险评分系统和回归方程等多种方式,导致各研究结果缺乏可比性,并增加了临床应用的复杂性<sup>[26]</sup>。针对这些问题,未来研究应更加关注模型校准的过程,如采用Bootstrap重抽样等方法来增强模型的稳定性,同时加强外部验证以提高模型的推广适用性。此外,还应尽量规范和统一模型展示的方式,以提高其在临床实践中的可接受性和实用性<sup>[27]</sup>。

从临床决策角度看,这些模型仍具有重要的启发意义,其预测因子多源自常规化疗参数与实验室指标<sup>[17-18]</sup>,为化疗前风险评估提供了量化参考。未来应在多中心样本中强化外部验证与动态校准,结合临床路径、药物毒性监测及人工智能辅助决策系统,实现模型从“统计预测”向“临床决策支持”的转化,以真正提升化疗后骨髓抑制风险管理的科学性 with 精确性。

### 3.4 偏倚风险与适用性

本研究对7项儿童恶性肿瘤患者化疗后骨髓抑制风险预测模型的研究进行了全面评估。结果显示,所有研究均存在较高的偏倚风险,这些偏倚主要来源于研究设计、数据收集以及统计分析方法等方面的不足。具体而言,在研究对象的选择上,大多数回顾性研究因存在数据缺失和选择性偏倚而被判定为高风险,这表明今后研究应优先采用前瞻性设计,以确保在结局发生之前能够收集到预测因子,从根本上减少回顾性设计所带来的选择偏倚<sup>[28]</sup>。在预测因子的评估上,大多数研究因未采取盲法评估,可能导致研究产生测量偏倚,因此,实施严格的数据收集盲法对于提高测量的准确性至关重要。关于统计分析,大部分纳入研究(6项研究的EPV<20,其中4项研究的EPV<10)均存在样本量不足的问题,而确保每个预测因子有足够的事件数(建议EPV≥20)是提升模型稳定性和可靠性的关键——EPV偏低时,模型回归系数容易出现不稳定,判别效能指标往往被“虚高”估计,从而产生“小样本研究看似效果更好”的假象。这种由样本量不足和过度拟合共同导致的偏倚,可能在一定程度上放大部分模型的预测性能,尤其是复杂度较高的机器学习模型。因此,未来研究应对模型的泛化能力和真实临床效用保持审慎态度,并通过详细报告缺失数据的处理方法等措施,以提高研究的透明度。在模型评估方面,未来研究应结合区分度、校准度等多项评估指标,更全面地衡量模型的预测性能<sup>[29]</sup>。

综上,当前纳入研究存在的局限性在一定程度上限制了所建预测模型在临床中的直接推广。未来研究可

引入机器学习技术,并结合交叉验证等较为稳健的统计方法,以提高模型的预测性能和泛化能力。同时,建议未来预测模型的开发与报告严格遵循TRIPOD声明和PROBAST工具的指导,以规范研究流程和提高研究质量<sup>[30]</sup>,从而为儿童恶性肿瘤化疗后骨髓抑制的风险评估提供更加可靠的决策支持工具。

### 3.5 本研究的局限性

本研究纳入研究数量较少,且在研究对象、建模方法及结局定义等方面异质性明显,因此降低了结论的普适性,也难以进行定量合并分析。此外,部分研究的方法学报告不完整,如缺失数据处理方式不明、关键评估指标报告不全,影响了模型预测性能的全面评估。未来研究应遵循TRIPOD声明及PRISMA等国际报告规范,以提高方法透明度与研究质量,增强研究结果在临床实践中的适用价值。

## 4 结语

本研究系统评估了儿童恶性肿瘤化疗后骨髓抑制风险预测模型的构建方法、预测因子及其性能。结果发现基线中性粒细胞计数、化疗方案强度等核心指标具有较高预测价值,且机器学习模型在区分度方面优于传统统计方法。然而,现有模型普遍存在校准不足、缺乏外部验证及方法学偏倚较高等问题,限制了其在临床的直接应用。本文通过系统评价识别了当前模型的共性问题,总结了方法学优化方向,对后续高质量模型的构建具有循证参考价值。建议未来在该领域积累更多高质量、同质化研究后,再对特定模型进行深入比较与验证;同时,应结合儿童肿瘤的临床特征,整合多模态变量与多中心数据,构建并系统开展外部验证的动态预测模型,推动此类工具从风险评估向智能临床决策支持转化。

## 参考文献

- [1] 倪鑫. 国家儿童肿瘤监测年报:2022[M]. 北京:人民卫生出版社,2022:4.
- [2] LI M J, WANG Q, LU P, et al. Development of a machine learning-based prediction model for chemotherapy-induced myelosuppression in children with Wilms' tumor[J]. *Cancers*, 2023, 15(4):1078.
- [3] QIU B, MATTHAY K K. Advancing therapy for neuroblastoma[J]. *Nat Rev Clin Oncol*, 2022, 19(8):515-533.
- [4] HUANG J T, LU S Y, WANG J, et al. A multicenter phase II trial of primary prophylactic PEG-rhG-CSF in pediatric patients with solid tumors and non-Hodgkin lymphoma after chemotherapy: an interim analysis[J]. *Cancer Med*, 2023, 12(13):14130-14137.
- [5] 刘若彤. 肿瘤患儿化疗后骨髓抑制合并医院感染临床特点及风险预测模型建立的研究[D]. 济南:山东大学,2024.
- [6] SALVADOR C, MERYK A, HETZER B, et al. Immature platelet fraction predicts early marrow recovery after

- severe chemotherapy associated neutropenia[J]. *Sci Rep*, 2023,13(1):3371.
- [7] GURUMURTHY G, GURUMURTHY J, GURUMURTHY S. Machine learning in paediatric haematological malignancies: a systematic review of prognosis, toxicity and treatment response models[J]. *Pediatr Res*, 2025, 97(2): 524-531.
- [8] 刘阳,李红健,吴建华,等. 乳腺癌化疗致骨髓抑制风险预测模型的系统评价[J]. *中国药房*,2025,36(5):612-618.
- [9] WHO Classification of Tumours Editorial Board. WHO classification of paediatric tumours[EB/OL].[2026-03-18]. <https://publications.iarc.who.int/Book-And-Report-Series/Who-Classification-Of-Tumours>.
- [10] MOONS K G M, DE GROOT J A H, BOUWMEESTER W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist[J]. *PLoS Med*, 2014, 11(10): e1001744.
- [11] 陈香萍,张奕,庄一渝,等. PROBAST:诊断或预后多因素预测模型研究偏倚风险的评估工具[J]. *中国循证医学杂志*, 2020, 20(6): 737-744.
- [12] 李祎涵. 基于自我报告的血液肿瘤患儿治疗相关发热性中性粒细胞减少的预测模型研究[D]. 上海:上海交通大学, 2022.
- [13] 徐蕊. 淋巴系恶性肿瘤患者大剂量甲氨蝶呤化疗后不良反应预测模型的初步建立[D]. 乌鲁木齐:新疆医科大学, 2018.
- [14] MIEDEMA K G E, DE BONT E S J M, OUDE NIJHUIS C S M, et al. Validation of a new risk assessment model for predicting adverse events in children with fever and chemotherapy-induced neutropenia[J]. *J Clin Oncol*, 2011, 29(7): e182-e184.
- [15] LYMAN G H, PONIEWIERSKI M S. A patient risk model of chemotherapy-induced febrile neutropenia: lessons learned from the ANC study group[J]. *J Natl Compr Canc Netw*, 2017, 15(12): 1543-1550.
- [16] WICKI S, KEISKER A, AEBI C, et al. Risk prediction of fever in neutropenia in children with cancer: a step towards individually tailored supportive therapy[J]. *Pediatr Blood Cancer*, 2008, 51(6): 778-783.
- [17] JIA H Y, LIANG L, CHEN X, et al. Predictive value of peri-chemotherapy hematological parameters for febrile neutropenia in patients with cancer[J]. *Front Oncol*, 2024, 14: 1380195.
- [18] VENÄLÄINEN M S, HEERVÄ E, HIRVONEN O, et al. Improved risk prediction of chemotherapy-induced neutropenia-model development and validation with real-world data[J]. *Cancer Med*, 2022, 11(3): 654-663.
- [19] CHEN W Q, PENG L, ZENG X L, et al. Predictors of myelosuppression for patients with head and neck squamous cell carcinoma after induction chemotherapy[J]. *Clin Med Insights Oncol*, 2024, 18: 11795549231219497.
- [20] MIELNIK M, PODGAJNA-MIELNIK M, SZUDY-SZCZYREK A, et al. Predicting chemotherapy toxicity in multiple myeloma: the prognostic value of pre-treatment serum cytokine levels of interleukin-6, interleukin-8, monocyte chemoattractant protein-1, and vascular endothelial growth factor[J]. *Front Immunol*, 2024, 15: 1377546.
- [21] AL TAWIL A, ALMAZAYDEH L, ALQUDAH B, et al. Predictive modeling for breast cancer based on machine learning algorithms and features selection methods[J]. *Int J Electr Comput Eng*, 2024, 14(2): 1937.
- [22] XIAO J L, MO M, WANG Z Z, et al. The application and comparison of machine learning models for the prediction of breast cancer prognosis: retrospective cohort study[J]. *JMIR Med Inform*, 2022, 10(2): e33440.
- [23] MONIRUJJAMAN KHAN M, ISLAM S, SARKAR S, et al. Machine learning based comparative analysis for breast cancer prediction[J]. *J Healthc Eng*, 2022, 2022: 4365855.
- [24] LOBATO-DELGADO B, PRIEGO-TORRES B, SANCHEZ-MORILLO D. Combining molecular, imaging, and clinical data analysis for predicting cancer prognosis[J]. *Cancers*, 2022, 14(13): 3215.
- [25] BUICK J E, AUSTIN P C, CHESKES S, et al. Prediction models in prehospital and emergency medicine research: how to derive and internally validate a clinical prediction model[J]. *Acad Emerg Med*, 2023, 30(11): 1150-1160.
- [26] TOUSIGNANT-LAFLAMME Y, HOULE C, COOK C, et al. Mastering prognostic tools: an opportunity to enhance personalized care and to optimize clinical outcomes in physical therapy[J]. *Phys Ther*, 2022, 102(5): pzac023.
- [27] MISHRA A, MCCLELLAND R L, INOUE L Y T, et al. Recalibration methods for improved clinical utility of risk scores[J]. *Med Decis Making*, 2022, 42(4): 500-512.
- [28] DHIMAN P, MA J, ANDAUR NAVARRO C L, et al. Risk of bias of prognostic models developed using machine learning: a systematic review in oncology[J]. *Diagn Progn Res*, 2022, 6(1): 13.
- [29] BRACHER-SMITH M, CRAWFORD K, ESCOTT-PRICE V. Machine learning for genetic prediction of psychiatric disorders: a systematic review[J]. *Mol Psychiatry*, 2021, 26(1): 70-79.
- [30] COLLINS G S, DHIMAN P, ANDAUR NAVARRO C L, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence[J]. *BMJ Open*, 2021, 11(7): e048008.

(收稿日期:2025-10-19 修回日期:2026-02-28)

(编辑:刘明伟)