

# 预测模型研究的偏倚风险和适用性评估 工具解读

陈茹<sup>1</sup> 王胜锋<sup>2</sup> 周家琛<sup>3</sup> 孙凤<sup>2</sup> 魏文强<sup>1</sup> 詹思延<sup>2</sup>

<sup>1</sup>国家癌症中心/国家肿瘤临床医学研究中心/中国医学科学院北京协和医学院肿瘤医院肿瘤登记办公室, 北京 100021; <sup>2</sup>北京大学公共卫生学院流行病与卫生统计学系 100191; <sup>3</sup>西安交通大学医学部公共卫生学院流行病与卫生统计学系 710061

陈茹和王胜锋对本文有同等贡献

通信作者: 詹思延, Email: siyan-zhan@bjmu.edu.cn

**【摘要】** 本文介绍了预测模型研究的偏倚风险和适用性评估工具 PROBAST (Prediction model Risk Of Bias ASsessment Tool) 的主要内容、评价步骤和相关注意事项。PROBAST 从研究对象、预测因素、结局和分析 4 个领域共 20 个信号问题对原始研究的设计、实施和分析过程中可能产生的偏倚风险和适用性进行评价。通过综合分析, 对原始研究每个领域和整体的偏倚风险和适用性做出判断, 分为高、低或不清楚。PROBAST 为个体预测模型开发、验证和更新提供了可靠的新评价工具, 它不仅可以用于预测模型的系统综述, 也可作为预测模型研究通用的方法学评价工具。

**【关键词】** 偏倚风险; 评估工具; 预测模型研究; 系统综述

**基金项目:** 国家重点研发计划 (2016YFC0901404, 2018YFC1311704); 国家自然科学基金 (81903403); 北京市自然科学基金 (7204294); 协和青年科研基金 (2017330004)

DOI: 10.3760/cma.j.cn112338-20190805-00580

## Introduction of the Prediction model Risk Of Bias ASsessment Tool: a tool to assess risk of bias and applicability of prediction model studies

Chen Ru<sup>1</sup>, Wang Shengfeng<sup>2</sup>, Zhou Jiachen<sup>3</sup>, Sun Feng<sup>2</sup>, Wei Wenqiang<sup>1</sup>, Zhan Siyan<sup>2</sup>

<sup>1</sup>Office for Cancer Registry, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; <sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; <sup>3</sup>Department of Epidemiology and Biostatistics, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an 710061, China

Chen Ru and Wang Shengfeng contributed equally to the article

Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn

**【Abstract】** This paper introduced the tool named as "Prediction model Risk Of Bias ASsessment Tool" (PROBAST) to assess the risk of bias and applicability in prediction model studies and the relevant items and steps of assessment. PROBAST is organized into four domains including participants, predictors, outcome and analysis. These domains contain a total of 20 signaling questions to facilitate structured judgment of risk of bias occurring in study design, conduct or analysis. Through comprehensive judgment, the risk of bias and applicability of original study is categorized as high, low or unclear. PROBAST enables a focused and transparent approach to assessing the risk of bias of studies that develop, validate, or update prediction models for individualized predictions. Although PROBAST was designed for systematic reviews, it can be also used more generally in critical appraisal of prediction model studies.

**【Key words】** Risk of bias; Tool for assessment; Prediction model studies; Systematic review

**Fund programs:** National Key Research and Development Program of China (2016YFC0901404, 2018YFC1311704); National Natural Science Foundation of China (81903403); Beijing Natural Science Foundation (7204294); Peking Union Medical College Youth Fund (2017330004)

DOI: 10.3760/cma.j.cn112338-20190805-00580

预测模型是指利用多个预测因素估计个体发生某种结局的概率, 又被称为风险预测模型、预测指标

或风险评分。医学研究中的预测模型一般分为诊断预测模型和预后预测模型 2 类, 前者预测的是个体

在当下是否罹患某种结局,后者预测的是个体在将来是否发生某种结局<sup>[1]</sup>。预测模型研究包括预测模型的开发、验证或更新。近些年,预测模型研究呈明显上升趋势,在肿瘤和心血管疾病领域中已有诸多应用<sup>[2-5]</sup>。

系统综述在循证医学和临床指南的制定中起着关键作用,是干预和诊断试验最可靠的证据形式。针对预测模型的系统评价是一个较新的领域,随着预测模型不断发展,越来越多的研究开始系统地评估和总结关于预测模型效果的证据<sup>[6]</sup>。由于原始研究的潜在偏倚(bias)会影响研究结果,偏倚风险(risk of bias)评估是任何系统综述的必要步骤,以便在此基础上筛选出合格研究进行证据整合,为临床循证决策提供依据。目前已有的相关质量评价工具包括针对预后研究的质量评价工具 QUIPS (quality in prognosis studies) 和针对诊断试验的偏倚风险评估工具 QUADAS-2 (quality assessment of diagnostic accuracy studies)<sup>[7-10]</sup>。此外,研究者还可以使用 Cochrane 偏倚风险评估工具 2.0 评价随机对照设计的预测模型效果研究的方法学质量<sup>[7,11]</sup>,或使用 ROBINS-I (Risk Of Bias In Nonrandomized Studies of Interventions) 工具评价非随机设计的比较研究<sup>[12-13]</sup>。然而,上述工具只针对特定的研究设计且研究目的不同,QUIPS 针对的是旨在发现危险因素预后研究,QUADAS-2 针对的是利用单一诊断方法的诊断实验,Cochrane 偏倚风险评估工具 2.0 和 ROBINS-I 针对的是旨在评价有效性或安全性的随机或非随机对照实验,由于研究目的都不是为了构建预测模型,因此无法满足日益发展的预测模型及其系统评价的需求,亟需开发针对风险预测研究的质量评价工具。

2019 年,荷兰乌得勒支大学 Moons 教授牵头的 38 位学者,在借鉴随机对照试验、诊断试验和系统综述偏倚风险评估工具的基础上,经 7 轮特尔菲法和多次预实验,提出了预测模型研究的偏倚风险评估工具 (Prediction model Risk Of Bias ASsessment Tool, PROBAST)<sup>[14]</sup>。本文将对该工具的适用范围、评估步骤、条目和注意事项进行解读,以方便读者了解、使用和推广该评价工具,更好地提升预测模型原始研究和综述的质量。

### 一、PROBAST 工具简介

PROBAST 工具适用于评估多因素诊断/预后模型的开发或验证的原始研究的偏倚风险和适用性。预测模型的开发包括在现有预测模型中增加新的预测因素,如果在对现有模型验证时进行了模型的更

新和扩展,也属于开发新模型。PROBAST 不只用于预测模型的系统综述,也可作为预测模型研究的通用评价工具。

PROBAST 工具的评价内容为预测模型研究的偏倚风险和适用性。偏倚是指导致研究结果与真实情况差异的系统误差。在预测模型开发和验证中,研究设计、实施和分析中的局限性都可能影响模型预测能力并导致偏倚的产生。模型的预测能力通常用校准度(calibration)和区分度(discrimination)进行评价,对于某些诊断模型也会用到分类(classification)能力<sup>[15]</sup>。其他医学研究领域中发现的许多偏倚来源也同样与预测模型研究相关,如结局评价者对其他研究特征是否采用盲法、预测因子或结局的定义和测量是否始终保持一致等。

当原始研究的人群、预测因素或结局与系统综述的特定研究问题存在差异时,可能产生适用性的问题。例如,预测模型的研究对象来自医院,而系统综述关注的是社区人群,这时报告的预测模型的校准度和区分度可能并不适用,因为医院来源患者的疾病通常比社区来源的更危重。而当原始研究的人群纳入标准、预测因素和结局与系统综述的问题直接匹配时,就不会出现适用性的顾虑。由于系统综述在检索原始研究时通常会设置更宽泛的检索策略,因此,需要仔细评估每个原始研究对实际问题的适用性。

### 二、评估步骤

利用 PROBAST 工具开展预测模型的系统综述时,偏倚风险的评估一般分 4 步:①明确系统综述的问题;②区分预测模型研究的类型;③评估各领域偏倚风险和适用性;④对偏倚风险和适用性的总体判断。其中,步骤 1 对每篇文献进行一次评价,步骤 2 需要对文献中每个相关结局的模型完成一次评价,而步骤 3 和 4 需要对文献中每个不同预测模型的开发和验证单独进行一次评价。

步骤 1:明确系统综述的问题:系统综述的问题明确可以参考 CHARMS 清单,从模型使用目的、目标人群、预测因素和预测结局出发进行评价<sup>[16]</sup>。

步骤 2:区分预测模型研究的类型:不同的信号问题适用于不同类型预测模型的评估,每个模型按其研究内容可以分为“仅开发”“开发加验证”和“仅验证”3 类。实际操作需要注意 3 点:①此处模型验证特指外部验证,即采用独立于构建模型所用训练数据集的新数据集;②模型更新或参数调整视为构建了新模型;③每个预测模型都要单独评价 1 次,如同

时报告模型构建及验证或模型验证及更新时,需要进行拆分。

**步骤 3: 偏倚风险和适用性评价:** PROCAST 工具将预测模型研究中所涉及的潜在偏倚分为 4 个领域(domain),即研究对象、预测因素、结局和分析,每个领域包含 4 个方面:用来支持判断的信息、2~9 个信号问题(共 20 个)、偏倚风险的判断和判断的基本原理(表 1)。评估者根据文献对每个信号问题进行判断,结果为“是”“可能是”“可能不是”“否”或“未提供信息”。“是”代表低偏倚风险,“否”代表高偏倚风险。若原始研究中确实未介绍相关信号问题的内容,判断为“未提供信息”。若原始研究中提供的信息不足以得出确定的判断,可归为“可能是”或“可能不是”。预测模型的适用性评价包含前 3 个领域,判断过程与偏倚风险相似,但没有信号问题。

**步骤 4: 总体判断:**在步骤 3 偏倚风险分类的基础上,评估者可以将预测模型整体和每个领域的偏倚风险分为低、高或不清楚。每个领域内的结果汇总借鉴了“短板理论”,只有该领域所有信号问题都回答“是”或“可能是”才视为“低风险”,只要有 1 个信号问题回答“否”或“可能不是”就有可能为“高风险”,当某个信号问题被判断为“未提供信息”而同时

其他信号都为“低风险”时,则该领域归为“不清楚”。同样,对于整体偏倚风险,只有每个领域都评为“低风险”才将整体视为“低风险”,只要 1 个领域被评为“高风险”就视为高风险,若某个领域被评为“不清楚”而同时其他领域都为“低风险”时,则整体归为“不清楚”。此外,对于构建模型的研究,即便 4 个领域都评为低风险,若没有外部验证仍降级为高风险;但验证模型的研究,只要被验证模型的构建基于庞大数据集且构建时进行了内部验证,仍可认为整体属于低风险。

同样,基于步骤 3 每个领域的适用性的评价,评估者可以将预测模型整体的适用性评为低、高或不清楚。对于研究对象、预测因素和结局 3 个领域,只有当所有领域都为低风险时,整体才判断为“低风险”,如果有 1 个或多个领域判断为高风险,则整体归为“高风险”。若某个领域判断为“不清楚”且所有其他领域都为低风险时,则整体归为“不清楚”。

对于整体汇总结果,工作组建议以表、图、文字 3 种形式呈现,以利于读者快速了解研究整体水平。其中,表格列举每个研究在每个领域及汇总的风险水平(表 2),图展示每个领域上各等级的研究所占比例(图 1),文字讨论补充各种偏倚风险构成

表 1 PROCAST 工具的偏倚风险和适用性评价的第 3 步内容汇总

1. 研究对象	2. 预测因素	3. 结局	4. 分析
判断信号问题			
1.1 所采用的数据来源是否合适? 例如队列、随机对照试验或巢式病例对照研究数据?	2.1 所有研究对象的预测因素定义和测量是否一致或类似?	3.1 结局的测量方法是否合适?	4.1 发生结局的研究对象数量是否合理?
1.2 研究对象的纳入排除标准是否合适?	2.2 预测因素的测量是否与结局无关?	3.2 结局是否采用预先确定或标准的定义?	4.2 连续变量、分类变量是否均处理恰当?
-	2.3 在模型使用的时点是否能够得到所有预测因素的信息?	3.3 结局的定义中是否排除预测因素的信息?	4.3 所有纳入对象是否都纳入了统计分析?
-	-	3.4 所有研究对象的预测结局定义和测量是否一致或类似?	4.4 研究对象的缺失数据是否处理恰当?
-	-	3.5 结局的确定是否与预测因素无关?	4.5 预测因素的选择是否基于单变量分析?
-	-	3.6 预测因素的测量与结局确定之间的时间间隔是否合适?	4.6 数据中的复杂问题(如删失、竞争风险、对照抽样等)处理是否恰当?
-	-	-	4.7 模型评价是否恰当?
-	-	-	4.8 是否考虑了模型的过度拟合、拟合不足等问题?
-	-	-	4.9 最终模型中预测因素及其权重是否与报告的多变量分析结果一致?
评价偏倚风险			
研究对象的选择	预测因素或其评估	结局或其确定	分析
评价适用性			
纳入的研究对象或环境设置与综述问题不匹配	预测因素的定义、评估或时间与综述问题不匹配	结局的定义、评估或时间与综述问题不匹配	-

注: \*仅适用于模型开发

表 2 PROBAST 评价结果示意味表

研究	偏倚风险			适用性			总体		
	研究对象	预测因素	结局	研究分析	研究对象	预测因素	结局	偏倚风险	适用性
1	+	-	?	+	+	+	+	-	+
2	+	+	+	+	+	+	+	+	+
3	+	+	+	?	-	+	+	?	-
4	-	?	?	-	+	+	-	-	-
5	+	+	+	+	+	?	+	+	?
6	+	+	+	+	?	+	?	+	?
7	?	?	+	?	+	+	+	?	+
8	+	+	+	+	+	+	+	+	+

注: + 表示低风险, - 表示高风险, ? 表示不清楚

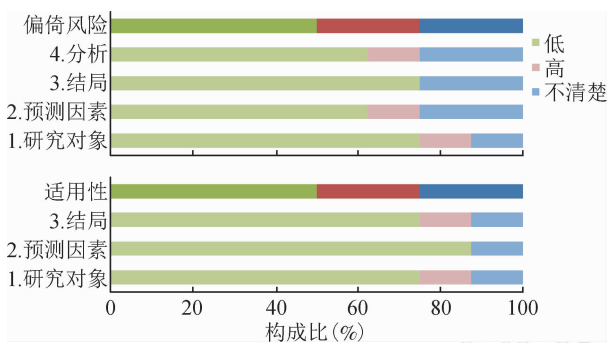


图 1 PROBAST 评价结果示意图

模式对研究问题的证据支持力度。

### 三、部分条目解读

1. 研究对象: 设有 2 个信号问题, 主要用来说明研究对象的代表性, 以便评价者判断预测模型是否正确反映了源人群中预测因素和预测结局之间的关系。实际评估时, 需要掌握两点: ①对于研究设计, 预后预测模型研究推荐前瞻性队列研究, 而诊断预测模型研究推荐横断面研究, 少部分依赖随访获取结局信息的也可采用队列研究。模型研究不推荐使用传统的病例对照研究, 巢式病例对照研究或病例队列研究经特殊分析后可视为低偏倚风险。②对于研究对象, 纳入排除标准要慎重制定, 避免诊断预测时排除合并其他疾病者、预后预测时纳入基线已患疾病尚未确诊者等不当处理。

2. 预测因素: 设有 3 个信号问题, 主要用来判断所有研究对象是否运用了相似的、真实可靠的方法来进行定义和测量预测因素。预测因素的范畴很广, 包括人口学特征、医疗史、体格检查、影像指标、电镜指标、血尿生化、组织标本、疾病分期或特征、组学及其他生物指标等。设置的 3 个问题分别对应预测因素的定义及测量方法、测量过程、测量时点。推荐所有研究对象都采用同一定义、同一测量方法, 测量时采用盲法, 所有预测因素的信息在模型应用的时点都可获取。实际评估时, 需要掌握 3 点: ①预测

因素指研究中纳入探讨范围的所有潜在预测因素, 不考虑是否纳入最终模型; ②仅针对某预测模型的系统综述, 可只关注最终纳入模型的预测因素; ③预测因素在结局发生之前测量的前瞻性研究, 默认采用了盲法, 而若预测因素在结局发生之后才测量或采用回顾性队列时, 如对既往影像资料重新读片或回顾性检测冷藏标本时, 测量过程就容易受到已发生结局的干扰, 偏倚风险较大。

3. 结局: 设有 6 个信号问题, 主要用来判断所有研究对象是否运用了相似的、真实可靠的方法来进行定义和测量预测结局。如果结局的定义或测量不当, 直接导致结局错分, 而造成模型的回归系数、截距和基线风险都发生偏倚, 影响模型效果评估。与预测因素相似, 此处的 6 个问题对应了预测结局的定义及测量方法、测量过程和测量时点。推荐所有研究对象都采用同一预先设定的定义、同一测量方法, 测量时采用盲法, 并选取恰当的时点。实际评估时, 需要掌握 5 点: ①预测结局的定义和测量需要较强的临床知识, 很多结局已经形成默认的推荐方法, 建议直接参考; ②一般结局定义根据多个条件判断时, 容易对条件拿捏不一, 偏倚风险较大; ③某些情况下测量结局时很难完全避免预测因素的干扰, 如结局本身需要专家参考尽可能多的信息才能判断时, 容易高估预测因素和结局之间的关联强度; ④诊断试验有时肯定无法对每个对象都采用完全相同的结局测量方法, 如肿瘤领域, 作为金标准的病理检测只会用在已有检查结果阳性(如影像学检查)的个体中, 阴性个体后续或者不做任何进一步检查, 或者采用其他准确性较低的替代确诊方法, 会导致偏倚; ⑤预测因素与结局测量的时间间隔是否恰当, 也需要较强的临床知识加以判断。诊断预测研究一般要求预测因素和结局指标的测量在同一时间点, 少数结局的测量和预后预测研究一样, 也需要随访。此时, 时间间隔太短会造成结局出现太少, 太长会造成结局性质可能会发生变化。

4. 分析: 设有 9 个信号问题, 主要用来说明统计方法的选择和使用是否恰当, 以便评价者判断由于统计分析和报告的问题, 哪些结果是虚假的、存在偏倚的。采用不恰当的统计方法或者忽视重要的统计原则都会增加偏倚风险。该部分评价需要较多的专业知识, 因此建议至少需要 1 位在预测模型研究领域的统计专家参与。实际评估时, 需要掌握: ①预测模型研究中, 样本量更关心的是发生预测结局的研究对象数。构建模型时, 当发生结局的研究对象

数/候选预测因素数比值 < 10 时,除要求采用各种内部验证技术定量评估外,后续还应再采用收缩回归系数等对参数进行调整;②构建模型时,尽量不要到分析阶段再将连续变量根据手头数据特征转为二分类变量,必须如此操作时,同样要补充内部验证和收缩回归系数调整,否则只要有 1 个连续变量在分析阶段临时转为二分类变量进行预测,即视为高风险。但是那些在设计阶段就决定按照某标准(如临床界值)分为 2 组的连续变量,不在该讨论范围之内;③分析过程中,对某多分类的预测因素进行亚组合并时,若合并后该因素仍未纳入最终模型,可认为偏倚风险较小,该合并是为了处理某些亚组人数相对较少,反之偏倚风险较大;④避免分析时因某些预测因素难判断、存在异常或缺失而将研究对象剔除一部分,如剔除预测因素模糊不清(如影像学检查或实验室检查)的个体,可能导致纳入分析的都是极端取值个体,高估模型区分能力;注意并非绝对的只要有数据因为上述原因剔除就一定判断为“高风险”,当只有一小部分研究对象不纳入分析且纳入部分仍具有代表性时,偏倚风险可能评价为“低风险”,但遗憾的是到底多小才认定为“可以接受的一小部

分”尚无明确标准;⑤变量筛选切忌单纯依赖单变量分析,而是要根据已有临床知识、预测因素测量的可靠性、一致性、适用性、可及性和测量成本;⑥病例-队列、巢式病例-对照研究要考虑抽样比例计算绝对结局事件概率(如采用逆概率加权的 logistic 回归);结局存在竞争风险或随访存在删失的数据,需采用整合时间-事件信息的 Cox 回归方法;结局可以复发(如癫痫)时需采用多水平模型或随机效应模型;⑦当结局事件过少、结局事件数/待筛选变量数比值过小、连续变量转为二分类变量,利用单变量分析或前进/后退自动法筛选变量时,过度拟合问题会更为突出,要补充内部验证和收缩回归系数调整。

#### 四、实例分析

乳腺癌风险评估的 Gail 模型目前已成为欧美发达国家普遍采用的乳腺癌风险预测工具<sup>[17]</sup>。本文以此为例,应用 PROBAST 工具进行偏倚风险评价,由于适用性评价需结合系统综述的具体问题,此处暂不涉及。Gail 模型的偏倚风险评价结果显示(表 3),即便已经公认的预测模型,也仅有 1 个维度可评估为低偏倚风险,但考虑到样本量不大且未进行外部验证,整体视为“高偏倚风险”。当然,该评估结果与

表 3 PROBAST 应用举例:偏倚风险判断

信号问题	问题回答	问题评价	领域评价
1 研究对象			高风险
1.1 所采用的数据来源是否合适? 例如队列、随机对照试验或巢式病例对照研究数据?	病例对照研究	否	
1.2 研究对象的纳入排除标准是否合适?	患者为新发,对照采用匹配	是	
2 预测因素			高风险
2.1 所有研究对象的预测因素定义和测量是否一致或类似?	结合引文,调查员入户询问	是	
2.2 预测因素的测量是否与结局无关?	结合引文,未提到盲法	可能不是	
2.3 在模型使用的时点是否能够得到所有预测因素的信息?	都可得到	是	
3 结局			低风险
3.1 结局的测量方法是否合适?	结合引文,标准检测技术	是	
3.2 结局是否采用预先确定或标准的定义?	原位癌和浸润性癌	是	
3.3 结局的定义中是否排除预测因素的信息?	未包含	是	
3.4 所有研究对象的预测结局定义和测量是否一致或类似?	完全一致	是	
3.5 结局的确定是否与预测因素无关?	结局早于预测因素测量	是	
3.6 预测因素的测量与结局确定之间的时间间隔是否合适?	不需要时间间隔	是	
4 分析			高风险
4.1 发生结局的研究对象数量是否合理?	发生结局的研究对象数/候选预测因素数比值 > 20, $2\ 852/7 = 407.4 > 20$	是	
4.2 连续变量、分类变量是否均处理恰当?	连续变量(年龄、初潮年龄、活产数、乳腺结节数)转为分类变量,且事先未明确分组依据	否	
4.3 所有纳入对象是否都纳入了统计分析?	剔除了存在缺失的研究对象	否	
4.4 研究对象的缺失数据是否处理恰当?	研究对象按缺失剔除	否	
4.5 预测因素的选择是否基于单变量分析?(仅适用于模型开发)	结合专业知识和实际分析	是	
4.6 数据中的复杂问题(如删失、竞争风险、对照抽样等)处理是否恰当?	推测一般人群的基线风险,但未调整抽样权重	可能不是	
4.7 模型评价是否恰当?	无相关内容	否	
4.8 是否考虑了模型的过度拟合、拟合不足等问题?(仅适用于模型开发)	无相关内容	否	
4.9 最终模型中预测因素及其权重是否与报告的多变量分析结果一致?(仅适用于模型开发)	一致	是	

当时预测模型的方法学尚未成熟、偏倚风险的意识尚未形成有关,但也侧面反映出关注和提高预测模型原始研究质量的必要性和紧迫性。

## 五、讨论

PROBAST 是严格开发的专门评价开发、验证或更新个体水平预测模型的原始研究偏倚风险的工具。它适用于诊断预测研究和预后预测研究,且不受医学领域、研究结局或预测因素的变量类型、统计方法的限制。PROBAST 的潜在使用者包括系统综述的作者、医疗保健决策者、对循证医学感兴趣或参与指南制定的研究人员和临床医生以及期刊编辑和审稿人。在实际使用时,需要评估者具备预测模型和临床两方面的专业知识,因而单纯临床医生或者方法学专家都难以单独完成。另外,预测模型的系统综述作为相对独立的研究领域,不仅偏倚风险和适用性评价有独特之处,其实从研究问题的提出、文献检索、文献纳入排除、数据分析、结果解读到论文报告都有针对性的方法学探讨,也正在逐渐形成一套独立的方法学体系<sup>[6,16,18-20]</sup>。

目前,对预测模型方法的研究仍处于起步阶段,但预测模型的系统综述却越来越多。PROBAST 的出现对今后相关原始研究和系统综述都有很好的推动和规范作用。与其他医学评价指南一样,PROBAST 也会随着预测模型研究方法的发展而不断更新。因此,PROBAST 工具的出现,至少会推动以下研究领域的兴起和发展:①针对 PROBAST 工具的细节完善和方法学探索,如对于机器学习建模的针对性条目的开发与验证、针对生存分析资料模型评价指标的优化等;②针对已有预测模型的质量评价,找出差距以便针对性完善现有预测模型的不足;③针对 PROBAST 工具的推广效果评价,通过对比工具提出前后,发表的预测模型质量变化,反映工具的推广和接受程度;④针对预测模型的系统综述,方法学的提出将会吸引不同专业领域的学者关注对预测模型进行系统综述,综合回答实际问题。而伴随上述这一系列研究的积累和探讨,预测模型的原始研究势必会愈来愈规范和科学,相关系统综述工作也势必会越来越完善和普及。

利益冲突 所有作者均声明不存在利益冲突

## 参 考 文 献

[1] Bouwmeester W, Zuihthoff NPA, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review [J]. PLoS Med, 2012, 9(5): e1001221. DOI: 10.1371/journal.pmed.1001221.

[2] Gray EP, Teare MD, Stevens J, et al. Risk prediction models for lung cancer: a systematic review [J]. Clin Lung Cancer, 2016, 17(2): 95-106. DOI: 10.1016/j.clcc.2015.11.007.

[3] Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Risk prediction models of breast cancer: a systematic review of model performances [J]. Breast Cancer Res Treat, 2012, 133(1): 1-10. DOI: 10.1007/s10549-011-1853-z.

[4] Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review [J]. BMJ, 2016, 353: i2416. DOI: 10.1136/bmj.i2416.

[5] Sahle BW, Owen AJ, Chin KL, et al. Risk prediction models for incident heart failure: a systematic review of methodology and model performance [J]. J Card Fail, 2017, 23(9): 680-687. DOI: 10.1016/j.cardfail.2017.03.005.

[6] Debray TP, Damen JAAG, Snell KIE, et al. A guide to systematic review and Meta-analysis of prediction model performance [J]. BMJ, 2017, 356: i6460. DOI: 10.1136/bmj.i6460.

[7] Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors [J]. Ann Intern Med, 2013, 158(4): 280-286. DOI: 10.7326/0003-4819-158-4-201302190-00009.

[8] 唐少文, 张渊, 陶必林, 等. 偏倚风险评估系列: (七) 预后因素研究 [J]. 中华流行病学杂志, 2018, 39(7): 1003-1008. DOI: 10.3760/cma.j.issn.0254-6450.2018.07.026.

[9] Tang SW, Zhang Y, Tao BL, et al. Risk of bias assessment: (7) assessing bias in studies of prognostic factors [J]. Chin J Epidemiol, 2018, 39(7): 1003-1008. DOI: 10.3760/cma.j.issn.0254-6450.2018.07.026.

[10] Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies [J]. Ann Intern Med, 2011, 155(8): 529-536. DOI: 10.7326/0003-4819-155-8-201110180-00009.

[11] 曲艳吉, 杨智荣, 孙凤, 等. 偏倚风险评估系列: (六) 诊断试验 [J]. 中华流行病学杂志, 2018, 39(4): 524-531. DOI: 10.3760/cma.j.issn.0254-6450.2018.04.028.

[12] Qu YJ, Yang ZR, Sun F, et al. Risk on bias assessment: (6) a revised tool for the Quality assessment on diagnostic accuracy Studies (QUADAS-2) [J]. Chin J Epidemiol, 2018, 39(4): 524-531. DOI: 10.3760/cma.j.issn.0254-6450.2018.04.028.

[13] 杨智荣, 孙凤, 詹思延. 偏倚风险评估系列: (二) 平行设计随机对照试验偏倚评估工具 2.0 介绍 [J]. 中华流行病学杂志, 2017, 38(9): 1285-1291. DOI: 10.3760/cma.j.issn.0254-6450.2017.09.028.

[14] Yang ZR, Sun F, Zhan SY. Risk on bias assessment: (2) revised cochrane risk of bias tool for individually randomized, parallel group trials (RoB2.0) [J]. Chin J Epidemiol, 2017, 38(9): 1285-1291. DOI: 10.3760/cma.j.issn.0254-6450.2017.09.028.

[15] Sterne JAC, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions [J]. BMJ, 2016, 355: i4919. DOI: 10.1136/bmj.i4919.

[16] 孙凤, 高乐, 杨智荣, 等. 偏倚风险评估系列: (五) 非随机干预性研究 [J]. 中华流行病学杂志, 2018, 39(3): 374-381. DOI: 10.3760/cma.j.issn.0254-6450.2018.03.024.

[17] Sun F, Gao L, Yang ZR, et al. Series of risk of bias assessment (5): Risk Of Bias In Non-randomized Studies-of Interventions (ROBINS-I) [J]. Chin J Epidemiol, 2018, 39(3): 374-381. DOI: 10.3760/cma.j.issn.0254-6450.2018.03.024.

[18] Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies [J]. Ann Intern Med, 2019, 170(1): 51-58. DOI: 10.7326/M18-1376.

[19] Burke HB. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [J]. Ann Intern Med, 2015, 162(10): 735-736. DOI: 10.7326/L15-5093.

[20] Moons KGM, de Groot JAH, 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. DOI: 10.1371/journal.pmed.1001744.

[17] Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually [J]. J Natl Cancer Inst, 1989, 81(24): 1879-1886. DOI: 10.1093/jnci/81.24.1879.

[18] Geersing GJ, Bouwmeester W, Zuihthoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews [J]. PLoS One, 2012, 7(2): e32844. DOI: 10.1371/journal.pone.0032844.

[19] Debray TPA, Koffijberg H, Nieboer D, et al. Meta-analysis and aggregation of multiple published prediction models [J]. Stat Med, 2014, 33(14): 2341-2362. DOI: 10.1002/sim.6080.

[20] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and Meta-analyses: the PRISMA statement [J]. Ann Intern Med, 2009, 151(4): 264-269. DOI: 10.7326/0003-4819-151-4-200908180-00135.

(收稿日期: 2019-08-05)

(本文编辑: 李银鸽)