

PROBAST

(预测模型风险偏倚评估工具)

请参阅已发表的 PROBAST 系列论文...

PROBAST 用于评价什么？

PROBAST 可用于评价旨在评估 (包括建立、验证或更新) 某多因素诊断或预后预测模型的研究之 *偏倚风险* 和 *适用性*。该工具的设计是为了评价一项系统评价中所纳入的原始研究。

当一项原始研究的设计、实施或分析出现系统性缺陷和局限，以至于研究结果被扭曲时，**偏倚** 即会出现。针对预测模型研究，我们已经明确了会发生偏倚风险的情况是：研究设计、实施或数据分析上的不足，可导致一个模型的预测性能估值出现系统性失实，或造成一个预测模型不能够解决研究问题。通常情况下，模型预测性能可以经由校准度 (calibration)、区分度 (discrimination) 以及常用的分类指标 (classification measures) 来评估。在一项高偏倚风险的研究，这些指标很可能无法被准确评估出来。**模型适用性**指的是一项原始研究中的某个预测模型在多大程度上与你系统评价研究问题 (在研究对象、预测因子或感兴趣的临床结局等方面) 相匹配。

一项原始研究中可以涵盖多个预测模型的建立和 (或) 验证，或者更新。针对每个不同的、为个体化预测而建立、验证或者更新的模型，均应该完成一次 PROBAST 的评价。当一个文献同时评估多个预测模型时，仅需对符合你系统评价纳入标准的那些模型做一次性的 PROBAST 评价。请注意，下文用到的“模型”一词即包括诸如简化风险评分、列线图 (又称诺莫图) 或模型再校准等模型衍生词汇。

PROBAST 并不适用于所有类型的多因素诊断或预后研究。例如，PROBAST 不可用于那些尽管用了多变量模型作为数据统计分析方法，但研究目的是为了识别与临床结局相关的预测因子，而不是为了建立一个个体化预测模型的原始研究。

PROBAST 评价包括四个步骤：

步骤	任务	何时完成
1	明确系统评价的研究问题	对每个系统评价各评价一次[，均进行一次步骤 1 相应的任务]
2	区分预测模型评估的类型	对每个待评价研究中的每个预测模型以及每个相关临床结局各评价一次[，均进行一次步骤 2 相应的区分]
3	评价偏倚风险和模型适用性	对研究中的每个预测模型的每次建立与验证各评价一次[，均进行一次步骤 3 相应的评价]
4	总体评价	对研究中每个预测模型的每次建立与验证各评价一次[，均进行一次步骤 4 相应的评价]

若这是您第一次使用 PROBAST，我们强烈建议您阅读更详细的说明和详述文本，并可查阅我们网站上提供的案例示范[Ann Intern Med M18-1377, www.probast.org].

第一步：明确系统评价的研究问题

为帮助评价所评估模型之于您研究问题的适用性，请陈述您的系统评价研究问题。每个系统评价都应该完成一次以下表格。

标准	明确系统评价的研究问题
模型的使用意图：	
研究对象，包括选择标准和临床背景	
预测因子（预测模型中使用到的），包括预测因子类型（如，病史、临床检查、生化指标、影像检查），测量时间，具体的测量问题（如，对特定设备的需求和禁忌）：	
待预测的临床结局：	

第二步：区分预测模型评估的类型

用下表将预测模型评估的类型区分为模型建立、模型验证或模型更新、或任意组合。不同的信号问题适用于不同类型的预测模型评估。如果该模型评估无法和上述分类中的任何一个相匹配，即不该使用 PROBAST。

依据其目的区分预测模型评估			
预测研究类型	PROBAST 相应的方框	请打勾	预测模型研究类型的定义
仅以模型建立为目的	建立		预测模型建模研究（无外部验证）。 这类研究可以包含模型的内部验证，

			例如使用自助抽样法 (bootstrapping) 和交叉验证技术 (cross-validation techniques)
模型建立和验证	建立和验证		在同一篇文献中，既有预测模型建立，又包含在其他研究对象对此模型所做的外部验证
仅以模型验证为目的	验证		基于其他研究对象对已知 (先前建立的) 预测模型进行的外部验证

针对您系统评价中待评价的每项研究和每个相关临床结局，均需要填写一次该表。

纳入研究的参考文献	
待评价的预测模型	
感兴趣的临床结局	

第三步：评价偏倚风险和模型适用性

PROBAST 包括四个关键领域。每个领域均是用于判断 (低、高或不清楚) 偏倚风险，且包括有一些信号问题以助于做出判断。信号问题都可用“是 (Y)”、“可能是 (PY)”、“可能否 (PN)”、“否 (N)”或“缺少信息 (NI)”来回答。所有信号问题都措辞明确，因此“是”即表示不存在任何偏倚。任何定级为“否”或“可能否”的信号问题，均提示有出现偏倚的可能。你需要基于自己的判断，来决定是否一个领域应该被判定为偏倚风险“高”、“低”或“不清楚”。该指导文本包含更多指导和示例，以帮助判定每一领域的信号问题和偏倚风险。

前三个领域同时还可用于判定[预测模型]之于上文已定义系统评价研究问题的适用性关切

(“高”、“低”或“不清楚”)。

针对各不相同的每个预测模型的每次模型评估，均需要分别评价所有领域。有阴影的方框表示该信号问题在相应的地方并不适用，因此无需回答。

领域 1：研究对象

A. 偏倚风险

描述数据来源和研究对象的筛选标准：

	建立 (Dev)	验证 (Val)
1.1 是否使用了恰当的数据来源，如队列、RCT 或巢式病例对照研究数据？		
1.2 研究对象的纳入和排除标准是否恰当？		
由研究对象筛选所致 的偏倚风险	风险： (低/高/不清楚)	

偏倚评判的科学依据：

B. 适用性

描述纳入的研究对象、临床场所和日期：

研究对象和场所与系 统评价研究问题不匹 配而导致的关切	关切： (低/高/不清楚)		
-----------------------------------	--------------------	--	--

适用性评判的科学依据：

领域 2：预测因子			
A . 偏倚风险			
列出并描述最终预测模型中纳入的预测因子，例如因子评测的定义和时间点：			
		建立 (Dev)	验证 (Val)
2.1 是否对所有研究对象均通过相似的方法来定义和评测预测因子？			
2.2 预测因子的评测是否是在不知晓临床结局数据的前提下做出的？			
2.3 在想要使用预测模型的时间节点上，模型中所有预测因子的信息是否都可以获取？			
由预测因子或其评估所导致的偏倚风险	风险： (低/高/不清楚)		
偏倚评判的科学依据：			
B . 适用性			
预测模型中最终纳入的预测因子定义、测量评价或时间点与系统评价研究问题不匹配而导致的关切	关切： (低/高/不清楚)		
适用性评判的科学依据：			

领域 3：临床结局

A. 偏倚风险

描述临床结局，它是怎样被定义和判定的，以及预测因子测量和临床结局确定之间的时间间隔：

	建立 (Dev)	验证 (Val)
3.1 临床结局是否被恰当地判定？		
3.2 是否使用了预先设定的或标准的临床结局定义？		
3.3 预测因子是否从临床结局定义中被排除掉？		
3.4 对所有研究对象来说，临床结局是否均经相似的方法而被定义和判定？		
3.5 结局指标是否是在不知晓预测因子信息的情况下而被判定？		
3.6 预测因子评测和临床结局判定之间的时间间隔是否恰当？		
由临床结局或其确定所致的偏倚风险	风险： (低/高/不清楚)	

偏倚评判的科学依据：

B. 适用性

临床结局在哪个时间点上被判定：

如果使用一个复合型临床结局的话，请描述其所包含的每单个临床结局的相对频率或分布：

<p>临床结局、其定义、 测量时机或确定方法 与系统评价研究问题 不匹配而导致的关切</p>	<p>关切： (低/高/不清楚)</p>		
<p>适用性评判的科学依据：</p>			

<p>领域 4：数据分析</p>
<p>偏倚风险</p>
<p>描述研究对象例数、候选预测因子数目、临床事件发生例数以及候选预测因子平均事件发生数：</p>
<p>描述模型是如何被建立起来的（例如，建模方法（如，生存模型或逻辑模型）、预测因子筛选和风险区组定义等方面）：</p>
<p>描述模型是否以及如何被验证的，不管内部验证（如，自助抽样法、交叉验证技术或随机分割样本）或是外部验证（如，时序验证、地理验证、不同场所、不同研究对象类型）：</p>
<p>描述模型性能测量指标，如校准或再校准、区分度、分类或再分类、净获益、以及是否这些测量指标曾调整过乐观偏差：</p>
<p>描述从数据分析中被排除的研究对象：</p>
<p>描述预测因子和临床结局有关的删失数据和删失数据处理方法：</p>

		建立 (Dev)	验证 (Val)
4.1 有特定临床结局事件的研究对象例数是否足够合理？			
4.2 连续型预测因子和分类型预测因子是否被恰当地处理？			
4.3 是否所有参与研究的研究对象均被纳入到数据分析？			
4.4 出现缺失数据的研究对象是否被恰当地处理？			
4.5 是否避免依据单因素分析来筛选预测因子？			
4.6 是否恰当地考量处理了数据中出现的复杂问题（如，截尾数据、竞争风险数据、对照组研究对象的抽样）？			
4.7 是否恰当地评估了相应的模型性能测量指标？			
4.8 模型过度拟合和模型性能上的乐观偏差是否被恰当地考量处理？			
4.9 最终模型中的预测因子及其权重分配是否和多因素分析中的结果相符？			
由数据分析所致的偏倚 风险	风险： (低/高/不清楚)		
偏倚评判的科学依据：			

--

第四步：总体评价

可以采用下面的表格对该预测模型评价（建立和/或验证）的偏倚风险和适用性有关的关切做总体评价；该总体评价基于全部已评价的领域完成。

注意需要对每个模型的模型评价做总体评价

预测模型评估的偏倚风险总体评价

低偏倚风险	如果所有领域均被判定为低风险； 如果一个预测模型被建立起来却没有对其做外部验证，尽管所有领域均判定为低风险，也建议考虑将总体评价结果降级为高偏倚风险。除非，该模型的建立是基于极大样本数据集而实现的，且已经进行过某种内部验证，才可认为该模型存在低风险
高偏倚风险	如果至少一个领域被判定为高风险
不清楚	如果至少一个领域被判定为不清楚，且其他领域均为低风险

预测模型评估的适用性总体评价

低度关切	如果所有领域均被判定为低度关切
高度关切	如果至少一个领域被判定为高度关切
不清楚	如果至少一个领域被判定为不清楚，且不存在高度关切

预测模型评估偏倚风险和适用性的总体评价		
偏倚风险的总体评价	风险： (低/高/不清楚)	
潜在偏倚来源小结：		
适用性的总体评价	关切： (低/高/不清楚)	
适用性关切小结：		

(时春虎翻译 , Zhujie Gu 审校)