

附表 3 医学领域其他的一次研究评估工具及其主要组成部分

Supplementary table 3. Major components of the tools for assessing other primary medical studies

A. 苏格兰国家指南小组 (SIGN) 方法学清单: 经济学研究 (最后修订于 2014 年)				
网址: <a href="https://www.sign.ac.uk/what-we-do/methodology/checklists/">https://www.sign.ac.uk/what-we-do/methodology/checklists/</a>				
主要条目		选项		
<b>第 1 部分: 内部真实性</b>				
在一项良好的经济学研究中...		这项研究做到了吗?		
1.1 该研究解决了一个适当且明确针对的问题。	是	否	不清楚	/
1.2 这个问题的经济重要性是显而易见的。	是	否	不清楚	/
1.3 研究设计的选择是合理的。	是	否	不清楚	/
1.4 从研究的角度来看, 所有与之相关的成本都应包括在内, 并进行适当的测量和评估。	是	否	/	/
1.5 用来回答研究问题的结局测量与该目的相关, 并被适当的测量和评估。	是	否	不清楚	不适用
1.6 如果需要折现未来成本和结果, 被正确执行。	是	否	不清楚	不适用
1.7 明确假设并进行敏感性分析。	是	否	不清楚	/
1.8 明确制定决策规则, 并根据增加的成本和结果进行比较。	是	否	/	/
1.9 结果为决策者提供了相关信息。	是	否	/	/
<b>第 2 部分: 研究的总体评估</b>				
2.1 研究进行得如何?	高质量(++)	可接受(+)	不可接受-排除 0	
2.2 这项研究的结果是否直接适用于本指南针对的患者组?	是	否		
2.3 注释: 总结作者的结论。您在评估本研究的过程中添加任何评论, 以及多大程度回答您的问题和前面提及的任何不确定性领域。				
B. 评价经济学评价研究的文献严格评价项目 (CASP) 清单 (最后修订于 2018 年)				
网址: <a href="https://casp-uk.net/casp-tools-checklists/">https://casp-uk.net/casp-tools-checklists/</a>				
主要条目		选项		
<b>第 1 部分: 经济学评价是否可靠?</b>				
1. 是否提出了明确的问题?	是	否	不知道	
2. 是否提供了对竞争替代品的全面描述?	是	否	不知道	
是否值得继续?				
3. 本文是否提供证据表明该项目有效? (即, 该项目是否弊大于利?)	是	否	不知道	
4. 是否正确识别、测量和评估了干预措施的效果?	是	否	不知道	
<b>第 2 部分: 如何评估和比较后果和成本?</b>				
5. 是否确定了所有重要且相关的资源, 以及每种替代品的健康结果成本以适当的单位进行确定、衡量并得到可靠的估值?	是	否	不知道	
6. 成本和结果是否针对发生的不同时间进行了调整 (折扣)?	是	否	不知道	
7. 评估结果是什么?	是	否	不知道	
8. 是否对替代品的后果和成本进行了增量分析?	是	否	不知道	
9. 是否进行了足够的敏感性分析?	是	否	不知道	
<b>第 3 部分: 结果是否有助于当地居民的购买?</b>				
10. 该项目是否可能在您的上下文或环境中同样有效?	是	否	不知道	
11. 费用是否可以转换为您的环境?	是	否	不知道	
12. 在您的环境中值得做吗?	是	否	不知道	

C. 评价经济学研究的乔安娜·布里格斯学院 (JBI) 清单 (最后修订于 2014 年)				
网址: <a href="https://joannabriggs.org/critical-appraisal-tools">https://joannabriggs.org/critical-appraisal-tools</a>				
主要条目	选项			
1. 问题 / 目标的定义是否明确?	是	否	不清楚	不适用
2. 是否对有替代品进行全面描述?	是	否	不清楚	不适用
3. 是否确定了每个替代方案的所有重要且相关的成本和结果?	是	否	不清楚	不适用
4. 临床疗效是否公认?	是	否	不清楚	不适用
5. 成本和成果是否得到了准确计量?	是	否	不清楚	不适用
6. 成本和结果是否得到可信的评估?	是	否	不清楚	不适用
7. 是否针对不同的时间调整了成本和结果?	是	否	不清楚	不适用
8. 是否对成本和后果进行了增量分析?	是	否	不清楚	不适用
9. 是否进行了敏感性分析以调查成本或结果估计的不确定性?	是	否	不清楚	不适用
10. 研究结果是否包括用户关注的所有问题?	是	否	不清楚	不适用
11. 结果是否可以推广到所关注的领域?	是	否	不清楚	不适用
总体评价: 纳入 <input type="checkbox"/> 排除 <input type="checkbox"/> 寻找更多信息 <input type="checkbox"/>				
D. 英国国家健康和临床优化研究所 (NICE) 经济学评价研究方法学清单 (最后修订于 2019 年 2 月)				
网址: <a href="https://www.nice.org.uk/process/pmg20/resources/appendix-h-appraisal-checklists-evidence-tables-grade-and-economic-profiles-pdf-8779777885">https://www.nice.org.uk/process/pmg20/resources/appendix-h-appraisal-checklists-evidence-tables-grade-and-economic-profiles-pdf-8779777885</a>				
主要条目	选项			
<b>第 1 部分: 适用性</b> (与 NICE 指南制订手册第 7.5 节中所述的特定综述问题和 NICE 参考案例的相关性) 应该首先使用此清单来过滤掉不相关的研究。				
1.1 研究人群是否合适?	是 / 否	部分	不清楚	不适用
1.2 干预措施是否合适?	是 / 否	部分	不清楚	不适用
1.3 研究设计与英国国情相适应?	是 / 否	部分	不清楚	不适用
1.4 成本角度是否适合?	是 / 否	部分	不清楚	不适用
1.5 结果角度是否适合?	是 / 否	部分	不清楚	不适用
1.6 所有未来成本和结果是否都适当折现?	是 / 否	部分	不清楚	不适用
1.7 是使用 NICE 的首选方法得出的 QALY, 还是将适当的与社会护理相关的等效结果用作结果? 如果没有, 请采取分析观点描述所依据的理由和结果 (上文项目 1.5)。	是 / 否	部分	不清楚	不适用
1.8 总体判断: 直接适用 / 部分适用 / 不适用 如果研究被视为“不适用”, 则无需使用清单的第 2 部分。				
其他评论:				
<b>第 2 部分: 研究局限性 (方法学质量水平)</b> 一旦确定研究足以适用于指南背景, 则应使用此清单。				
2.1 模型结构是否充分反映了所评估主题的性质?	是 / 否	部分	不清楚	不适用
2.2 时间跨度是否足够长以反映成本和结果的所有重要差异?	是 / 否	部分	不清楚	不适用
2.3 是否包括所有重要且相关的结果?	是 / 否	部分	不清楚	不适用
2.4 基线结果的估计是否来自最佳的可用来源?	是 / 否	部分	不清楚	不适用
2.5 相对干预效果的估计是否来自最佳可用来源?	是 / 否	部分	不清楚	不适用
2.6 是否包括所有重要和相关成本?	是 / 否	部分	不清楚	不适用
2.7 资源使用的估计是否来自最佳的可用来源?	是 / 否	部分	不清楚	不适用
2.8 资源的单位成本是否来自最佳来源?	是 / 否	部分	不清楚	不适用
2.9 是否提供了适当的增量分析或可以从数据中计算得出?	是 / 否	部分	不清楚	不适用
2.10 是否对所有不确定值的重要参数都进行了适当的灵敏度分析?	是 / 否	部分	不清楚	不适用
2.11 是否声明无潜在的利益冲突?	是 / 否	部分	不清楚	不适用
2.12 总体评估: 轻微限制 / 潜在严重限制 / 非常严重的限制				

其他评论:			
<b>E. 评价质性研究的文献严格评价项目 (CASP) 清单 (最后修订于 2018 年)</b>			
网址: <a href="https://casp-uk.net/casp-tools-checklists/">https://casp-uk.net/casp-tools-checklists/</a>			
主要条目	选项		
<b>第 1 部分: 研究结果是否可靠?</b>			
1. 研究是否提出了清晰明确的问题?	是	否	不知道
2. 定性方法是否合适?	是	否	不知道
是否值得继续?			
3. 研究设计是否适合解决研究的目的?	是	否	不知道
4. 招募策略是否适合研究目的?	是	否	不知道
5. 是否以解决研究问题的方式收集了数据?	是	否	不知道
6. 是否充分考虑了研究者与参与者之间的关系?	是	否	不知道
<b>第 2 部分: 研究结果是什么?</b>			
7. 是否考虑了道德问题?	是	否	不知道
8. 数据分析是否足够严格?	是	否	不知道
9. 对于发现是否有明确的陈述?	是	否	不知道
<b>第 3 部分: 结果对当地有帮助吗?</b>			
10. 研究有多少价值?	是	否	不知道
<b>F. 评价质性研究的乔安娜·布里格斯学院 (JBI) 清单 (最后修订于 2017 年)</b>			
网址: <a href="https://joannabriggs.org/critical-appraisal-tools">https://joannabriggs.org/critical-appraisal-tools</a>			
主要条目	选项		
1. 所陈述的哲学观点与研究方法之间是否具有有一致性?	是	否	不清楚 不适用
2. 研究方法与研究问题或目标之间是否一致?	是	否	不清楚 不适用
3. 研究方法和收集数据的方法之间是否一致?	是	否	不清楚 不适用
4. 研究方法 with 数据表示和分析之间是否一致?	是	否	不清楚 不适用
5. 研究方法和结果解释之间是否一致?	是	否	不清楚 不适用
6. 是否有从文化或理论上定位研究人员的声明?	是	否	不清楚 不适用
7. 是否解决了研究人员对研究的影响, 反之亦然?	是	否	不清楚 不适用
8. 参与者及其声音是否得到充分体现?	是	否	不清楚 不适用
9. 根据当前标准进行的研究是否符合道德规范, 或者对于最近的研究, 是否符合相关标准的证据?	是	否	不清楚 不适用
10. 研究报告中得出的结论是否来自对数据的分析或解释?	是	否	不清楚 不适用
总体评价: 纳入 <input type="checkbox"/> 排除 <input type="checkbox"/> 寻找更多信息 <input type="checkbox"/>			
<b>G. 英国国家社会研究中心的质性研究评估框架 (最新修订于 2003 年)</b>			
网址: <a href="https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/498322/a_quality_framework_tcm6-38740.pdf">https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/498322/a_quality_framework_tcm6-38740.pdf</a> <a href="https://core.ac.uk/download/pdf/74376246.pdf">https://core.ac.uk/download/pdf/74376246.pdf</a>			
主要条目	选项		
<b>结果</b>			
1. 结果有多可信?			
2. 研究如何扩展知识 / 理解?			
3. 评估如何很好地实现其最初的目的和宗旨?			
4. 进行更广泛推断的范围 - 解释得如何?			
5. 评价的依据有多明确?			

设计					
6. 研究设计的合理性如何?					
样本					
7. 案例 / 文件的样本设计 / 目标选择如何维护?					
8. 样本组成 / 案例纳入 - 描述的最终覆盖范围如何?					
数据收集					
9. 数据收集进行得如何?					
分析					
10. 分析的方法和公式表达得如何?					
11. 数据源的上下文 - 保留和描绘的程度如何?					
12. 探索的角度和内容的多样性如何?					
13. 数据的细节, 深度和复杂性 (即丰富性) 的表达程度如何?					
报告					
14. 数据, 解释和结论之间的联系有多清楚? 即, 可以如何很好地得出结论的途径?					
15. 报告有多清晰和连贯?					
灵活性和中立性					
16. 决定评估形式和结果的假设 / 理论观点 / 价值观是否清楚?					
伦理					
17. 对于道德问题, 有哪些证据值得关注?					
可听性					
18. 研究过程被充分记录了吗?					
H. 预后研究质量 (QIPS) 工具 (最新发布于 2013 年)					
主要条目					选项
<b>1. 研究参与者</b>					
目标: 判断选择偏倚的风险 (参与者和合格的非参与者有可能预后因素与结果之间的关系不同)					
目标人群的来源	针对关键特征充分描述了来源人群或感兴趣的人群 (列表)	高	中	低	
识别人群的方法	充分描述了抽样框架和招募, 包括确定足以限制潜在偏见的样本的方法 (使用的数量和类型, 例如卫生保健中的转诊模式)	高	中	低	
招募时期	已充分描述招聘时期	高	中	低	
招聘地点	充分描述了招聘地点 (设置和地理位置)	高	中	低	
纳入和排除标准	充分描述了纳入和排除标准 (例如, 包括明确的诊断标准或“零时间”描述)	高	中	低	
足够的研究参与者	合格的个体充分参与了研究	高	中	低	
基线特征	已充分描述了基线研究样本 (即进入研究的个人) 的关键特征 (清单)	高	中	低	
研究参与者总结	该研究样本代表了关键特征上的关注人群, 足以限制所观察到的预后因素与预后之间关系的潜在偏倚	高	中	低	
<b>2. 研究损耗</b>					
目标: 判断损耗的风险 (对于完成和未完成的参与者, 预后因素与结果之间的关系可能不同)					
可用于分析的基线样本比例	应答率 (即完成研究并提供结果数据的研究样本比例) 足够	高	中	低	
尝试收集有关脱落参与者的信息	描述了尝试收集有关脱落参与者的信息	高	中	低	
受试者失去随访的原因和潜在影响	提供了失访的原因	高	中	低	
有关失访者的结果和预后因素信息	脱落受试者的关键特征被描述 (清单)	高	中	低	
	完成研究的参与者和未完成研究的参与者的关键特征 (LIST) 和结果之间没有重要差异	高	中	低	
研究损耗摘要	失访 (从基线样本到所分析的研究人群) 与关键特征 (即研究数据足以代表样本) 无关, 足以将潜在偏倚限制在观察到的预后因素与预后关系之间	高	中	低	

<b>3. 预后因素测量</b>				
目的：判断与预后因素测量相关的测量偏倚风险（与结局水平相关预后因素的差异测量）				
预后因素的定义	提供了对“预后因素”的明确定义或描述（例如，包括剂量、水平、暴露持续时间以及测量方法的明确说明）	高	中	低
有效和可靠地评估预后因素	预后因素测量方法足够有效且可靠，可以限制误分类偏倚（例如，可能包括相关外部信息源的有关测量属性及特征，如盲法测量和有限依赖召回）	高	中	低
	报告连续变量或使用适当的切入点（即不依赖于数据）	高	中	低
预后因素测量的方法和设置	所有研究参与者的预后因素测量方法和设置均相同	高	中	低
可用的预后因素数据比例	足够比例的研究样本具有完整的预后因素变量数据	高	中	低
缺失数据的方法	缺失的“预后因素”数据使用适当的插补方法	高	中	低
预后因素测量总结	在研究参与者中充分评估了预后因素，以充分限制潜在的偏倚	高	中	低
<b>4. 结果测量</b>				
目标：判断与结果测量相关的偏倚风险（与基线预后因素水平相关的结果差异测量）				
结果的定义	提供了明确的结果定义，包括随访时间，结果构建的水平和程度	高	中	低
有效可靠地测量结果	所使用的结果测量方法足够有效且可靠，可以限制误分类偏差（例如，可能包括相关外部信息源的有关测量属性以及特征，如盲法测量以及通过有效且可靠的测试确认结果）	高	中	低
结果测量的方法和设置	所有研究参与者的结果测量方法和设置均相同	高	中	低
结果测量总结	在研究参与者中适当测量了感兴趣的结果，以充分限制潜在的偏倚	高	中	低
<b>5. 研究混杂因素</b>				
目标：判断混杂因素导致的偏倚风险（即，预后因素的影响因与预后因素和结果相关的另一因素而失真）				
测量重要混杂因素	测量所有重要的混杂因素，包括治疗方法（概念模型中的关键变量：列表）	高	中	低
混杂因素的定义	提供了对所测量的重要混杂因素的明确定义（例如，包括剂量，水平和暴露时间）	高	中	低
有效和可靠的混杂因素测量	对所有重要混杂因素的测量是足够有效和可靠的（例如，可能包括相关外部信息源的有关测量属性及特征，如盲法测量和有限依赖召回）	高	中	低
混杂测量的方法和设置	所有研究参与者的混杂测量方法和设置都相同	高	中	低
缺失数据的方法	使用适当的方法将插补用于丢失混杂因素数据	高	中	低
适当的混杂比例	研究设计中考虑了重要的潜在混杂因素（例如，关键变量匹配，分层或组间的初始集合）	高	中	低
	分析中考虑了重要的潜在混杂因素（即适当的调整）	高	中	低
研究混杂因素总结	对重要的潜在混杂因素进行了适当的说明，从而限制了与预后因素和结果之间的关系有关的潜在偏见	高	中	低
<b>6. 统计分析和报告</b>				
目标：判断与统计分析和结果显示有关的偏倚风险				
提出分析策略	有足够的数据来评估分析的充分性	高	中	低
模型开发策略	建立模型的策略（即在统计模型中包含变量）是适当的，并且基于概念框架或模型	高	中	低
	选择的统计模型足以进行研究设计	高	中	低
结果报告	没有选择性的结果报告	高	中	低
统计分析和报告总结	统计分析适用于研究设计，限制了呈现无效或虚假结果的可能性	高	中	低
<b>I. 预测模型偏倚风险评估工具（PROBAST）（最新发布于2019年）</b>				
网址： <a href="http://www.probast.org">http://www.probast.org</a>				
主要条目	选项			
<b>1. 受试者</b>				

信号问题			
1.1. 是否使用了适当的数据源, 例如队列, 随机对照试验或巢式病例-对照研究数据?	是 / 可能是	否 / 可能否	无信息
1.2. 参与者的纳入和排除标准都合适吗?	是 / 可能是	否 / 可能否	无信息
偏倚风险			
选择受试者	高	低	不清楚
适用性			
纳入的受试者或设置与研究问题不匹配	高	低	不清楚
2. 预测变量			
2.1. 对于所有受试者, 是否以相似的方式定义和评估了预测变量?	是 / 可能是	否 / 可能否	无信息
2.2. 是否在了解结果数据的情况下评估了预测变量?	是 / 可能是	否 / 可能否	无信息
2.3. 打算使用该模型时, 所有预测变量都可用吗?	是 / 可能是	否 / 可能否	无信息
偏倚风险			
预测变量及其评估	高	低	不清楚
适用性			
预测变量的定义, 评估或时间与研究问题不符	高	低	不清楚
3. 结局			
3.1. 结局确定是否合适?	是 / 可能是	否 / 可能否	无信息
3.2. 是否使用预先指定或标准的结果定义?	是 / 可能是	否 / 可能否	无信息
3.3. 结果定义中是否排除了预测变量?	是 / 可能是	否 / 可能否	无信息
3.4. 是否以相似的方式为所有参与者定义和确定了结果?	是 / 可能是	否 / 可能否	无信息
3.5. 是否在了解预测信息的情况下确定结果?	是 / 可能是	否 / 可能否	无信息
3.6. 预测因子评估和结果确定之间的时间间隔是否合适?	是 / 可能是	否 / 可能否	无信息
偏倚风险			
结局及其确定	高	低	不清楚
适用性			
其定义、时间或确定与研究问题不符	高	低	不清楚
4. 分析			
4.1. 有结局的参加人数是否合理?	是 / 可能是	否 / 可能否	无信息
4.2. 连续预测和分类预测是否得到适当处理?	是 / 可能是	否 / 可能否	无信息
4.3. 分析中是否包括所有被招募的参与者?	是 / 可能是	否 / 可能否	无信息
4.4. 数据缺失的参与者是否得到了适当的处理?	是 / 可能是	否 / 可能否	无信息
4.5. 是否避免了基于单变量分析的预测变量选择? <sup>1</sup>	是 / 可能是	否 / 可能否	无信息
4.6. 数据的复杂性(例如检查, 竞争风险, 对照组的抽样)是否得到适当考虑?	是 / 可能是	否 / 可能否	无信息
4.7. 相关模型性能指标是否得到适当评估?	是 / 可能是	否 / 可能否	无信息
4.8. 模型过拟合, 拟合不足和乐观对模型的性能造成了影响吗? <sup>1</sup>	是 / 可能是	否 / 可能否	无信息
4.9. 最终模型中的预测变量及其分配的权重是否与报告的多变量分析的结果相对应? <sup>1</sup>	是 / 可能是	否 / 可能否	无信息
偏倚风险			
	高	低	不清楚
<sup>1</sup> 仅发展研究			
J. 评价临床预测规则的文献严格评价项目 (CASP) 清单 (最后修订于 2018 年)			
网址: <a href="https://casp-uk.net/casp-tools-checklists/">https://casp-uk.net/casp-tools-checklists/</a>			
主要条目			选项
第 1 部分: 试验结果是否可靠?			
1. 是否明确定义了临床预测规则?	是	否	不知道
2. 得出规则的人群是否包括适当范围的患者?	是	否	不知道

3. 该规则是否在另一组患者中得到验证?	是	否	不知道
是否值得继续?			
4. 预测变量和结果是否采用盲法?	是	否	不知道
5. 在最初选择的整个样本中, 预测变量和结果是否得到评估?	是	否	不知道
6. 是否清楚描述了用于构建和验证规则的统计方法?	是	否	不知道
<b>第 2 部分: 研究结果是什么?</b>			
7. 可以计算规则的执行情况吗?			
8. 对治疗效果的估计有多精确? (他们是否尝试使用其他变量完善规则, 以提高精度或简化规则?)			
<b>第 3 部分: 结果对当地有帮助吗? 这些发现是否适用于该环境?</b>			
9. 如果用于您的患者, 预测规则是否可靠并且结果可以解释?	是	否	不知道
10. 该规则在您的病例中是否可以接受?	是	否	不知道
11. 规则的结果会修改您对患者管理的决定, 还是可以提供给他/她的信息?	是	否	不知道
<b>K. 评价文献和专家观点的乔安娜·布里格斯学院 (JBI) 清单 (最后修订于 2017 年)</b>			
网址: <a href="https://joannabriggs.org/critical-appraisal-tools">https://joannabriggs.org/critical-appraisal-tools</a> <a href="https://wiki.joannabriggs.org/display/MANUAL/Appendix+4.1%3A+JBI+Critical+Appraisal+Checklist+for+Text+and+Opinion+Papers">https://wiki.joannabriggs.org/display/MANUAL/Appendix+4.1%3A+JBI+Critical+Appraisal+Checklist+for+Text+and+Opinion+Papers</a>			
主要条目		选项	
1. 是否明确指出了观点来源?	是	否	不清楚 不适用
2. 观点来源是否具有专业知识?	是	否	不清楚 不适用
3. 相关人群的利益是否是观点的中心焦点?	是	否	不清楚 不适用
4. 陈述的立场是分析过程的结果吗, 观点表达是否有逻辑?	是	否	不清楚 不适用
5. 是否参考了现有文献?	是	否	不清楚 不适用
6. 在逻辑上与文献/资料来源有任何不一致之处吗?	是	否	不清楚 不适用
总体评价: 纳入 <input type="checkbox"/> 排除 <input type="checkbox"/> 寻找更多信息 <input type="checkbox"/>			
<b>L. 基于共识的健康测量工具选择标准倡议 (COSMIN) 偏倚风险清单 (最后修订于 2018 年 7 月)</b>			
网址: <a href="https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/">https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/</a>			
<b>LA. 患者 - 报告结果测量 (PROM) 的开发</b>			
主要条目		选项	
<b>1a. PROM 的设计</b>			
一般设计要求			
1. 是否对要测量的结构提供了清晰的描述?	非常好	充分	有疑问 不充分 不适用
2. 结构的起源是否明确: 是否使用了理论, 概念框架或疾病模型, 或者提供了明确的理由来定义要测量的结构?	非常好	充分	有疑问 不充分 不适用
3. 是否明确描述了针对 PROM 开发的目标人群?	非常好	充分	有疑问 不充分 不适用
4. 是否明确说明了使用环境	非常好	充分	有疑问 不充分 不适用
5. 是否在代表开发 PROM 的目标人群的样本中进行了 PROM 开发研究?	非常好	充分	有疑问 不充分 不适用
概念启发 (相关性和全面性)			
6. 是否使用适当的定性数据收集方法来识别新 PROM 的相关项目?	非常好	充分	有疑问 不充分 不适用
7. 是否使用了熟练的小组主持人/采访员?	非常好	充分	有疑问 不充分 不适用
8. 小组会议或访谈是否基于适当的主题或访谈指南?	非常好	充分	有疑问 不充分 不适用
9. 小组会议或访谈是否被记录并逐字记录?	非常好	充分	有疑问 不充分 不适用
10. 是否使用适当的方法来分析数据?	非常好	充分	有疑问 不充分 不适用
11. 是否至少部分数据是独立编码的?	非常好	充分	有疑问 不充分 不适用
12. 数据收集是否一直持续到达到饱和?	非常好	充分	有疑问 不充分 不适用

13. 对于定量研究（调查）：样本量合适吗？	非常好	充分	有疑问	不充分	不适用
<b>1b. 认知访谈研究或其他先导测试</b>					
14. 是否进行了认知访谈研究或其他先导测试？	非常好	充分	有疑问	不充分	不适用
一般设计要求					
15. 是否在代表目标人群的样本中进行了认知访谈研究或其他先导测试？	非常好	充分	有疑问	不充分	不适用
可理解性					
16. 是否询问患者有关 PROM 的可理解性？	非常好	充分	有疑问	不充分	不适用
17. 所有项目都以最终形式进行了测试吗？	非常好	充分	有疑问	不充分	不适用
18. 是否使用适当的定性方法来评估 PROM 指令、项目、选项和召回期限的可理解性？	非常好	充分	有疑问	不充分	不适用
19. 是否在适当数量的患者中测试了每个条目？ 对于定性研究 对于定量（调查）研究	非常好	充分	有疑问	不充分	不适用
20. 是否使用了熟练的采访者？	非常好	充分	有疑问	不充分	不适用
21. 访谈是否基于适当的访谈指南？	非常好	充分	有疑问	不充分	不适用
22. 访谈是否记录下来并逐字记录？	非常好	充分	有疑问	不充分	不适用
23. 是否使用适当的方法来分析数据？	非常好	充分	有疑问	不充分	不适用
24. 是否至少两名研究人员参与了分析？	非常好	充分	有疑问	不充分	不适用
25. 是否通过改编 PROM，解决有关 PROM 指令、条目、选项和召回期限的可理解性存在问题？	非常好	充分	有疑问	不充分	不适用
综合性					
26. 是否询问患者有关 PROM 的全面性？	非常好	充分	有疑问	不充分	不适用
27. 是否测试了最终的条目？	非常好	充分	有疑问	不充分	不适用
28. 是否使用适当的方法来评估 PROM 的全面性？	非常好	充分	有疑问	不充分	不适用
29. 是否在适当数量的患者中测试了每个条目？ 对于定性研究 对于定量（调查）研究	非常好	充分	有疑问	不充分	不适用
30. 是否使用熟练的采访者？	非常好	充分	有疑问	不充分	不适用
31. 访谈是否基于适当的访谈指南？	非常好	充分	有疑问	不充分	不适用
32. 访谈是否记录并逐字记录？	非常好	充分	有疑问	不充分	不适用
33. 是否使用适当的方法来分析数据？	非常好	充分	有疑问	不充分	不适用
34. 是否至少两名研究人员参与了分析？	非常好	充分	有疑问	不充分	不适用
35. 通过调整 PROM 是否适当地解决了 PROM 的全面性问题？	非常好	充分	有疑问	不充分	不适用
<b>LB. 内容真实性</b>					
<b>2a. 向患者询问相关性</b>					
设计要求					
1. 是否使用一种适当的方法来询问患者每个条目与他们的病情经验相关性？	非常好	充分	有疑问	不充分	不适用
2. 是否在适当数量的患者中测试了每个条目？ 对于定性研究 对于定量（调查）研究	非常好	充分	有疑问	不充分	不适用
3. 是否使用熟练的采访者？	非常好	充分	有疑问	不充分	不适用
4. 小组会议或访谈是否基于适当的主题或访谈指南？	非常好	充分	有疑问	不充分	不适用
5. 小组会议或访谈是否被记录并逐字记录？	非常好	充分	有疑问	不充分	不适用
分析					
6. 是否使用适当的方法来分析数据？	非常好	充分	有疑问	不充分	不适用

7. 是否至少两名研究人员参与了分析?	非常好	充分	有疑问	不充分	不适用
<b>2b 向患者询问全面性</b>					
设计要求					
8. 是否使用适当的方法来评估 PROM 的全面性?	非常好	充分	有疑问	不充分	不适用
9. 是否在适当数量的患者中测试了每个条目? 对于定性研究 对于定量(调查)研究	非常好	充分	有疑问	不充分	不适用
10. 是否使用熟练的采访者?	非常好	充分	有疑问	不充分	不适用
11. 小组会议或访谈是否基于适当的主题或访谈指南?	非常好	充分	有疑问	不充分	不适用
12. 小组会议或访谈是否被记录并逐字记录?	非常好	充分	有疑问	不充分	不适用
分析					
13. 是否使用适当的方法来分析数据?	非常好	充分	有疑问	不充分	不适用
14. 是否至少两名研究人员参与了分析?	非常好	充分	有疑问	不充分	不适用
<b>2c 向患者询问可理解性</b>					
设计要求					
15. 是否使用适当的定性方法来评估 PROM 指令、项目、选项和召回期限的可理解性?	非常好	充分	有疑问	不充分	不适用
16. 是否在适当数量的患者中测试了每个条目? 对于定性研究 对于定量(调查)研究	非常好	充分	有疑问	不充分	不适用
17. 是否使用了熟练的小组主持人/采访员?	非常好	充分	有疑问	不充分	不适用
18. 小组会议或访谈是否基于适当的主题或访谈指南?	非常好	充分	有疑问	不充分	不适用
19. 小组会议或访谈是否被记录并逐字记录?	非常好	充分	有疑问	不充分	不适用
分析					
20. 是否使用适当的方法来分析数据?	非常好	充分	有疑问	不充分	不适用
21. 是否至少两名研究人员参与了分析?	非常好	充分	有疑问	不充分	不适用
<b>2d. 向专业人士询问相关性</b>					
设计要求					
22. 是否使用一种适当的方法来询问专业人员, 每个项目是否与感兴趣的结构相关?	非常好	充分	有疑问	不充分	不适用
23. 是否包括所有相关学科的专业人员?	非常好	充分	有疑问	不充分	不适用
24. 是否在适当数量的专业人员中对每个项目进行了测试? 对于定性研究 对于定量(调查)研究	非常好	充分	有疑问	不充分	不适用
分析					
25. 是否使用适当的方法来分析数据?	非常好	充分	有疑问	不充分	不适用
26. 是否至少两名研究人员参与了分析?	非常好	充分	有疑问	不充分	不适用
<b>2e. 向专业人员询问全面性</b>					
设计要求					
27. 是否使用适当的方法来评估 PROM 的全面性?	非常好	充分	有疑问	不充分	不适用
28. 是否包括所有相关学科的专业人员?	非常好	充分	有疑问	不充分	不适用
29. 是否在适当数量的专业人员中对每个项目进行了测试? 对于定性研究 对于定量(调查)研究	非常好	充分	有疑问	不充分	不适用
分析					
30. 是否使用适当的方法来分析数据?	非常好	充分	有疑问	不充分	不适用
31. 是否至少两名研究人员参与了分析?	非常好	充分	有疑问	不充分	不适用

LC. 结构真实性					
量表是否由效果指标组成, 即是否基于反射模型? <sup>1</sup>	是	否	/	/	/
研究是否涉及一维性或结构真实性? <sup>2</sup>	一维性		结构真实性		/
统计方法					
1. 对于经典测试理论(CTT): 是否进行了探索性或验证性因素分析?	非常好	充分	有疑问	不充分	不适用
2. 对于项目应答理论(IRT)/Rasch 模型: 所选模型是否适合研究问题?	非常好	充分	有疑问	不充分	不适用
3. 分析中包括的样本量是否足够?	非常好	充分	有疑问	不充分	不适用
其他					
4. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<sup>1</sup> 如果量表不是基于反思模型, 则一维性或结构真实性不相关。					
<sup>2</sup> 在系统评价中, 对于区分对每个(子)量表进行因子分析以评估(子)量表是否为一维(一维研究)的研究, 与对所有(子)量表进行因子分析以评估(预期)数量的子量表和子量表中聚类(结构真实性研究)的研究, 是有帮助的。					
LD. 内部一致性					
量表是否由效果指标组成, 即是否基于反射模型? <sup>1</sup>	是	否	/	/	/
设计要求					
1. 是否分别针对每个一维尺度或子尺度计算内部一致性?	非常好	充分	有疑问	不充分	不适用
统计方法					
2. 对于连续变量: 是否计算 Cronbach 的 alpha 值或 omega 值?	非常好	充分	有疑问	不充分	不适用
3. 对于二分类变量: 是否计算了 Cronbach 的 alpha 值或 KR-20?	非常好	充分	有疑问	不充分	不适用
4. 对于基于项目响应理论(IRT)的变量: 是否计算了 theta 的标准误差(SE( $\theta$ ))或估计的潜在特征值的可靠性系数(目的或条目)分离指数)?	非常好	充分	有疑问	不充分	不适用
其他					
5. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<sup>1</sup> 如果量表不是基于反射模型, 则内部一致性无关紧要。					
LE. 跨文化真实性 / 测量不变性					
设计要求					
1. 除了组变量外, 样本的相关特征是否相似?	非常好	充分	有疑问	不充分	不适用
统计方法					
2. 是否使用适当的方法来分析数据?	非常好	充分	有疑问	不充分	不适用
3. 分析中纳入的样本量是否足够?	非常好	充分	有疑问	不充分	不适用
其他					
4. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
LF. 信度					
设计要求					
1. 患者在中期要测量的结构是否稳定?	非常好	充分	有疑问	不充分	不适用
2. 时间间隔是否合适?	非常好	充分	有疑问	不充分	不适用
3. 测试条件的测量是否相似?(例如管理类型, 环境, 说明)	非常好	充分	有疑问	不充分	不适用
统计方法					
4. 对于连续型变量: 是否计算了组内相关系数(ICC)?	非常好	充分	有疑问	不充分	不适用
5. 对于二分/标称/顺序变量: 是否计算了 kappa 值?	非常好	充分	有疑问	不充分	不适用
6. 对于序数变量: 是否计算了加权 kappa 值?	非常好	充分	有疑问	不充分	不适用
7. 对于序数变量: 是否描述了加权方案? 例如线性、二次	非常好	充分	有疑问	不充分	不适用

其他					
8. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>LG. 测量误差</b>					
设计要求					
1. 患者在中期要测量的结构是否稳定?	非常好	充分	有疑问	不充分	不适用
2. 时间间隔是否合适?	非常好	充分	有疑问	不充分	不适用
3. 测试条件的测量是否相似? (例如管理类型、环境、说明)	非常好	充分	有疑问	不充分	不适用
统计方法					
4. 对于连续变量: 是否计算了测量标准误差 (SEM), 最小可检测变化 (SDC) 或协议限制 (LoA) ?	非常好	充分	有疑问	不充分	不适用
5. 对于二分 / 标称 / 顺序变量: 是否计算了百分比 (正负) 一致性?	非常好	充分	有疑问	不充分	不适用
其他					
6. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>LH. 标准真实性</b>					
统计方法					
1. 对于连续变量: 是否计算了相关性, 还是计算出了受试者操作曲线下面积?	非常好	充分	有疑问	不充分	不适用
2. 对于二分类变量: 是否确定了敏感性和特异性?	非常好	充分	有疑问	不充分	不适用
其他					
3. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>LI. 结构真实性的假设检验</b>					
<b>9a. 与其他结果测量工具的比较 (聚合真实性)</b>					
设计要求					
1. 比较仪测量的内容是否清楚?	非常好	充分	有疑问	不充分	不适用
2. 比较仪测量的特性是否足够?	非常好	充分	有疑问	不充分	不适用
统计方法					
3. 统计方法是否适合检验的假设?	非常好	充分	有疑问	不充分	不适用
其他					
4. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>9b. 亚组之间的比较 (区分或已知组的真实性)</b>					
设计要求					
5. 亚组的重要特征是否被适当描述?	非常好	充分	有疑问	不充分	不适用
统计方法					
6. 统计方法是否适合检验的假设?	非常好	充分	有疑问	不充分	不适用
其他					
7. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>LJ. 应答性</b>					
<b>10a. 标准方法 (即与金标准进行比较)</b>					
统计方法					
1. 对于连续变量: 是否计算了相关性, 还是计算出了受试者操作曲线 (ROC) 下面积?	非常好	充分	有疑问	不充分	不适用
2. 对于二分类变量: 是否确定了敏感性和特异性 (变化还是不变)?	非常好	充分	有疑问	不充分	不适用
其他					

3. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>10b. 构建方法 (即假设检验; 与其他结果度量工具的比较)</b>					
设计要求					
4. 比较仪测量的内容是否清楚?	非常好	充分	有疑问	不充分	不适用
5. 比较仪测量的特性是否足够?	非常好	充分	有疑问	不充分	不适用
统计方法					
6. 统计方法是否适合检验的假设?	非常好	充分	有疑问	不充分	不适用
其他					
7. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>10c. 构建方法: (即假设检验: 亚组之间的比较)</b>					
设计要求					
8. 是否充分描述了有关亚组重要特征?	非常好	充分	有疑问	不充分	不适用
统计方法					
9. 统计方法是否适合检验的假设?	非常好	充分	有疑问	不充分	不适用
其他					
10. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用
<b>10d. 构建方法: (即假设检验: 干预前和干预后)</b>					
设计要求					
11. 是否充分描述了有关干预措施?	非常好	充分	有疑问	不充分	不适用
统计方法					
12. 统计方法是否适合检验的假设?	非常好	充分	有疑问	不充分	不适用
其他					
13. 研究的设计或统计方法是否还有其他重要缺陷?	非常好	充分	有疑问	不充分	不适用