



COSMIN 报告规范：用于患者报告结局测量工具的测量属性研究

版本：2021 年 8 月

作者：

Joel J Gagnier¹⁻²

Jianyu Lai¹

Lidwine B.Mokkink³

Caroline B.Terwee³

1. 美国密歇根州，密歇根大学安娜堡分校，骨外科
2. 美国密歇根州，密歇根大学安娜堡分校，公共卫生学院，流行病学系
3. 荷兰阿姆斯特丹，阿姆斯特丹自由大学医学中心，公共卫生研究所，流行病学和数据科学系

译者：

胡雁¹

贾凌莹¹ 何梦雪^{1,2}

袁书琪¹ 王静¹

郑苏娜¹ 成磊¹

1. 中国上海，复旦大学护理学院，上海市循证护理中心
2. 中国上海，上海交通大学医学院附属上海儿童医学中心

联系方式：

Joel J Gagnier

电子邮箱：jgagnier@umich.edu

网站：www.cosmin.nl

胡雁

电子邮箱：huyan@fudan.edu.cn

电话：+86 21 6443 1273

目录

缩略语表	2
引用&资金支持	3
前言	4
测量属性研究报告的一般建议	5
内容效度的具体报告建议	8
结构效度的具体报告建议	8
内部一致性的具体报告建议	8
跨文化效度\测量不变性的具体报告建议	9
稳定性的具体报告建议	10
测量误差的具体报告建议	10
效标效度的具体报告建议	11
构念效度中假设检验的具体报告建议	11
反应度的具体报告建议	11
参考文献	13

缩略语表

CTT:	classical test theory, 经典测量理论
IRT/Rasch:	Item Response Theory and Rasch analyses, 项目反应理论和 Rasch 分析
NA:	not applicable, 不适用
Original CC:	original COSMIN checklist, 初版 COSMIN 清单
PROM:	patient-reported outcome measure, 患者报告结局测量工具
RoB:	Risk of Bias, 偏倚风险; 指 COSMIN 偏倚风险清单 ²

引用：

Joel J Gagnier, Jianyu Lai, Lidwine B Mokkink, Caroline B Terwee. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021 Aug; 30(8):2197-2218. doi:10.1007/s11136-021-02822-4.

资金支持：

本研究由 Patient-Centered Outcomes Research Institute (PCORI) 提供方法学基金支持。

前言

对于评估现有患者报告结局测量工具(PROM)测量属性的研究, 撰写报告时推荐参考本 COSMIN 报告规范。

对科学研究的充分报告将提高科学知识的适用性和研究价值³。效验测量属性的研究通常会缺少一些关键信息, 然而这些信息可帮助读者(例如: 临床专业人员、科研人员、基金赞助者) 确定研究方法、研究结果, 以及该研究最终在证明其 PROM 质量中的作用。本报告规范可指导 PROMs 测量属性研究报告的撰写并提高报告质量, 以增加研究透明性, 从而明确研究中存在的偏倚风险, 使引入的科学知识更加清晰和明确。同时, 本规范能帮助进行准确的研究方法学评估和可靠的研究结果应用(例如: 临床研究), 也使研究者们在未来能够构建或改进该领域的研究。

COSMIN 报告规范作为一份详细而具体的报告指南, 适用于 PROM 测量属性研究文稿中的所有部分, 以及其他 PROM 研究中的所有测量属性。关于本报告规范的更多信息详见: <https://doi.org/10.1007/s11136-021-02822-4>⁴。

本 PROMs 测量属性研究的报告规范共 71 个条目, 其中包含 35 条适用于所有测量属性研究报告的一般建议, 以及 36 条与特定测量属性报告对应的具体建议。*测量属性研究报告的一般建议*分为题目(n=3)、摘要(n=7)、前言(n=6)、方法(n=8)、结果(n=3)、讨论(n=6)、结论(n=1)和其他信息(n=1)。*具体报告建议*分为内容效度(n=7)、结构效度(n=2)、内部一致性(n=3)、跨文化效度\测量不变性(n=5)、稳定性(n=3)、测量误差(n=2)、效标效度(n=3)、构念效度的假设检验(n=5)、反应度(n=6)。

如同其他报告规范(如 CONSORT)一样, 我们希望所有相关的国际科研团队都能采纳这些建议, 并希望相关的同行评议期刊能够认可并推广其应用。

测量属性研究报告的一般建议		
条目编号	条目名称	内容描述
报告部分：题目 (Title)		
T1	患者报告结局测量工具 (PROM)	本研究中 PROM 工具（及相关版本）的名称。
T2	测量属性 (Measurement Property, MP)	研究了哪些 MPs。通常（若有多种具体属性需要调查时），指所研究测量属性的类型（如：信度、效度、反应度等）。
T3	研究对象	对相关研究对象特征（例如：所关注的疾病情况，语种等），以及使用的所有干预或暴露方法（例如：治疗手段）进行总体概述。
报告部分：摘要 (Abstract)		
A1	PROM	本研究中 PROM 工具（及相关版本）的名称（例如：SF-36 或 SF-12、不同语种版本）；是否设置题库（例如：PROMIS 工具）；工具的类型（例如：自我报告问卷、访谈）。
A2	测量属性	研究了哪些 MPs。通常（若有多种具体属性需要调查时），指所研究测量属性的类型（如：信度、效度、反应度等）。
A3	研究设计	用于测试 MP 的研究设计类型（例如：重测设计、纵向、队列、横断面、病例系列、随机等）；研究设计的其他相关细节（例如：干预/暴露方法、被比较工具的描述、除 PROMs 以外的结局指标）。
A4	样本	纳入/排除标准；研究样本特征相关的一般描述（例如：所关注的情况、地理位置、语种、其他相关的人口统计和基线特征）。
A5	方法	简要说明调查每种 MP 的方法，包括统计分析。
A6	结果	主要结果应包含本研究调查的所有 MPs。在此部分需报告每一 MP 的统计结果，且使用适当的统计精确度。
A7	结论/讨论	结合已有的证据，简述研究结果；主要的优势与不足；未来进一步研究该 PROM 的需要。
报告部分：前言 (Introduction)		
I1	所关注 PROM 的名称和介绍	具体说明本研究中 PROM 的名称、类型、语种、版本以及它是如何开发的。描述 PROM 旨在测量的构念及其分量表；描述 PROM 的结构（例如：因子数量、条目数量、计分方法）；描述相关的使用说明（如测量间隔期），以及答案类别的数量或类型。说明 PROM 是基于反映模型或是形成模型。注意：这些信息在方法部分有更详细地介绍。
I2	目标人群	描述 PROM 针对的特定目标人群。作者需提供恰当且必需的人群特征信息。

I3	PROM 最初开发研究的标引 (Citation for the original development of the PROM)	应提供最初开发文章的引文；同时，其他与本研究 PROM 质量高度相关的文章也应援引。
I4	研究现状&理论依据的陈述 (State of knowledge & Rationale)	描述（已知的）与本研究中 PROM 测量属性相关的研究现状。 同时，作者应提供文献综述；或参考最新的，关于特定版本 PROM（例如：特定语种或简版）既有证据的综述。且应说明进行新研究的必要性和重要性。此外，应给出进行当前研究的合理理由。
I5	定义	应对特定术语进行定义或解释。
I6	研究目的与假设	说明本研究的具体目标，以及与本研究中 PROM 相关的假设。
报告部分：一般方法（General Methods）		
GM1	研究设计	说明研究设计的关键要素。
GM2	研究对象	说明如何进行抽样，以及研究对象的纳入和排除标准（例如：若 PROM 仅用于特定情境*，应在纳排标准中体现）。
GM3	PROM 的测量	应明确描述测量 PROM 的方式和时间（例如：在什么场景设定下），同时包括所使用的的数据收集设备或系统（例如：纸质、电子管理系统/ePRO 等）。
GM4	资料收集步骤	提供有关其他资料收集、暴露方法（例如：干预的分配）和测量时间点/随访点的信息。
GM5	检验效能/样本量计算	为所有测量属性分析提供检验效能计算(power calculation)；或者，如若使用经验法，则需进行陈述，并说明出处/标引。
GM6	统计分析	应报告与所有研究假设或研究目标相对应的所有测量属性统计分析和测试。必要时，应报告统计显著性的分界值（例如：P 值小于 0.05）。 还应报告用于评估影响程度和方向的所有统计信息，以及测量的变异度(variability)或精确度(precision)。 此外，报告所使用的统计软件包 (statistical package)。
GM7	数据缺失	说明处理缺失数据的方法或计划。
GM8	事后分析 (post hoc analysis)	应明确指出在数据收集阶段结束后决定进行的数据分析（即：事后分析；二次分析），并说明进行事后分析的理由。
报告部分：一般结果（General Results）		
GR1	数据缺失	所有 PROM（或其他结局测量工具）和相关组别的分析都应解释缺失数据的数量及缺失原因。
GR2	研究对象/患者特征	应描述研究对象特征，包括 PROM 的基线分数。
GR3	样本量	如果本研究包含多个不同样本量的分析，则需报告每个分析的样本量。

报告部分：讨论 (Discussion)		
D1	MP 的证据 (MP evidence)	应就研究中的每一测量属性与良好测量学属性标准（例如：COSMIN 标准 ²⁷ ）进行比较，以确定特定的 MP 是否充分。 注意：这些信息可能在后续结果部分（例如：表格）中有更详细的介绍。
D2	实际相关性	需讨论研究结果的现实意义(practical relevance)。
D3	优势与局限性	应讨论本研究的优势和局限性。例如，讨论研究中是否存在可能影响结果的重要潜在误差。
D4	普适性	应讨论与 PROM 结果相关的普适性问题。例如，讨论本研究样本的结果是否可以推广到其他人群。
D5	量表变动 (Instrument changes)	讨论是否需要现有 PROM 进行修订，或需开发新 PROM。 如果您认为其中一个测量属性不达标，可以建议进行一些修订；或者如若测量属性过差，则可以建议（在特定人群或一般人群中）停止使用该 PROM。
D6	未来研究方向	具体说明未来所需的研究类型，以回答针对特定 MP 和 PROM 进行研究调查后所发现的新问题。
报告部分：结论 (Conclusions)		
C1	结论	给出本研究中每个 MP 以及所使用 PROM 的总体结论。
报告部分：其他信息 (Other information)		
O1	利益冲突	说明与本研究 PROM 相关的所有利益冲突（例如：作者是该 PROM 的开发人员、资助机构等）。

*使用情境 (context)：使用情境可以指 PROM 测量的目的（如用于诊断、评价或预测），也可以指开发 PROM 的特定环境（如在医院或家中使用）或特定的使用方法（如纸张或计算机）。如果 PROM 是为多个使用情境而开发，那么对此也应清楚描述。

内容效度的具体报告建议 (Content Validity)		
条目编号	条目名称	内容描述
CV1	相关性	报告患者和/或健康领域专家有无以及如何被问及：是否每个条目与其实际经历和/或经验相关
CV2	全面性	报告患者和/或健康领域专家有无以及如何被问及：是否已包含所有关键概念
CV3	可理解性	报告有无以及如何评估：PROM 指导语、条目、选项、回忆期*的可理解性
CV4	相关性结果	报告是否：每个条目均与患者和/或健康领域专家所关注的构念、人群和使用情境相关
CV5	答案选项和回忆期	报告患者和/或健康领域专家是否认为：答案的选项和回忆期是恰当的
CV6	全面性结果	报告患者和/或健康领域专家是否认为：已将所有关键概念纳入 PROM 中
CV7	可理解性结果	报告患者是否能够：按预期理解 PROM 指导语、条目和选项

*回忆期: recall period, 即患者填写 PROM 时所忆及的时间周期。常见回忆期为 24 小时、7 天或无框定的回忆期。例如：一周内的服药次数、24 小时内的母乳喂养频率等

结构效度的具体报告建议 (Structural Validity)		
条目编号	条目名称	内容描述
SV1	因子分析: PROMs 经典测量理论 (CTT)	报告探索性因子分析或验证性因子分析相关方法和结果的所有细节。 说明进行探索性因子分析的理由 (例如: 没有明确的先验假设*)。 对于验证性因子分析 (CFA), 描述并解释被测模型的因子结构。描述检验假设的方法和结果、估计方法、模型拟合优度和模型拟合良好的分界值, 包括最优拟合模型的因子载荷。
SV2	项目反应理论 (IRT) 分析	报告 IRT/Rasch 分析类型、估计方法、检验假设的方法和结果, 包括但不限于: 单维性** (通过因子分析可知)、局部依赖性*** (通过残差相关性可知)、单调性**** (可参考 Mokken 量表分析)、模型拟合优度、模型拟合良好的分界值以及所有项目参数。

*先验假设: priori hypothesis, 对一种现象的解释, 是基于理论和逻辑, 而不是基于实验和观察得出的结论。

**单维性: unidimensionality, PROM 中各维度/分量表的条目是否度量单个构念, 通常对每个维度单独进行因子分析进行评估。在单维 PROM 中, 单维性与结构效度一致; 在多维 PROM 中, 不可通过计算各个维度的单维性来替代计算 PROM 整体的结构效度。

***局部依赖性: local dependency; 与局部独立性相对, 指患者在某一条目上的作答影响到其他条目的作答, 或受其他条目作答情况的影响。当局部依赖性存在时, 去除所测构念的效应, 实际测得分数与反映构念的真分数的残差之间仍然存在显著水平的相关。

****单调性: monotonicity, 指患者答对条目/得分高的概率随潜在构念/特质的增加而增加, 可通过项目特征曲线反映。

(单维性、局部独立性、单调性为非参数项目反应理论中单调匀质模型三个基本假设, 由 Mokken 率先提出)

内部一致性的具体报告建议 (Internal Consistency)		
条目编号	条目名称	内容描述
IC1	测量单元 (Unit of measurement)	报告内部一致性的测量方法和每一个单维(分)量表的结果。报告所有与单维性相关的证据或假设。
IC2	定量数据	报告 Cronbach's alpha 系数或 Omega 系数。报告其他定量数据的内部一致性统计结果。
IC3	二分类数据	报告 Cronbach's alpha 系数或 KR-20 值。报告其他二分类数据的内部一致性统计结果。

跨文化效度\测量不变性的具体报告建议 (Cross-Cultural Validity\Measurement Invariance)		
条目编号	条目名称	内容描述
CCV1	比较组	报告被比较 (亚) 组的特征, 包括每组的样本量。
CCV2	因子分析: PROMs 经典测量理论 (CTT)	报告多组验证性因子分析、logistic 回归分析或其他分析的方法和结果。 在验证性因子分析中, 描述并解释被测模型的相关内容, 包括因子载荷限制、截距和方差。 描述检验假设的方法和结果、定义不变性的标准、估计方法、模型拟合优度以及用于标明项目测量不变性的标准。
CCV3	项目反应理论 (IRT) 分析	报告 IRT/Rasch 分析类型、估计方法、检验假设的方法和结果, 包括但不限于: 单维性 (通过因子分析可知)、局部依赖性 (通过残差相关性可知)、单调性 (可参考 Mokken 量表分析)。 描述统计软件包、估计方法、用于检测项目功能差异* (differential item functioning, DIF) 的标准, 以及所有模型比较的方法和结果。

*项目功能差异: differential item functioning, 指量表在潜在构念/特质水平相同的不同群体中的测量结果差异 (比如性别、种族等), 检测 DIF 用以保证测量的公平性和有效性

稳定性的具体报告建议 (Reliability)		
条目编号	条目名称	内容描述
R1	PROM 测量的实施 (PROM Administrations)	报告测量总次数和是否使用相同的 PROM 对同一样本进行测量。 描述对受试者实施测量的过程, 包括测量对象 (即是由受试者自己还是其代理人完成)、测量时间、测量方式以及报告所有的测量间隔期 (time intervals)。 应包括重复实施测量之间的时间间隔 (例如: 患者稳定性)、测量类型 (例如: 自我报告问卷或基于访谈的 PROM)、测量场所 (例如: 医院或家中) 以及帮助完成测量的指导语。 如必要, 需报告在 PROM 重复测量时, 伴随使用的其他测量工具; 同时, 还应报告测量的独立性 (是否在不知晓先前测量信息的情况下完成 PROM)
R2	统计分析	描述针对稳定性的所有统计分析和结果, 并说明使用理由 (例如: 组内相关系数 (ICC) 或 Kappa 值)。同时, 描述方差成分和用于分类/有序数据的加权法 (例如: 线性或二次加权法)。
R3	提高稳定性的方法	报告所有用于提高稳定性的方法, 如样本限制、研究人员培训、研究方法标准化和取重复测量的平均值。

测量误差的具体报告建议 (Measurement Error)		
条目编号	条目名称	内容描述
ME1	PROM 测量的实施 (PROM Administrations)	<p>报告测量总次数和是否使用相同的 PROM 对同一样本进行测量。</p> <p>描述对受试者实施测量的过程, 包括测量对象(即是由受试者自己还是其代理人完成)、测量时间、测量方式以及报告所有的测量间隔期。</p> <p>应包括重复实施测量之间的时间间隔(例如: 患者稳定性)、测量类型(例如: 自我报告问卷或基于访谈的 PROM)、测量场所(例如: 医院或家中)以及帮助完成测量的指导语。</p> <p>如必要, 需报告在 PROM 重复测量时, 伴随使用的其他测量工具; 同时, 还应报告测量的独立性(the independence, 即是否在不知晓先前测量信息的情况下完成 PROM)</p>
ME2	统计分析	<p>描述所有针对测量误差的统计分析和结果, 并说明使用理由。具体而言:</p> <p>对于定量数据, 应报告测量标准误(即 SEM; 具体说明用于计算 SEM 的对应模型(即 SEM 一致性))、最小可测变化值(即 SDC; 具体说明使用的公式, 包括基于 SEM 的 SEM 模型)或一致性限度(limits of agreement, LoA)。</p> <p>对于二分类/多分类/有序数据, 报告(原始数据)边界值和一致性百分比(例如: 阳性/阴性)。</p>

效标效度的具体报告建议 (Criterion Validity)		
条目编号	条目名称	内容描述
CriV1	效标	<p>报告所使用效标的具体信息以及使用原因。</p> <p>须报告所使用的效标可被视作金标准的理由;(如果适用,) 描述该效标进行二分类或多分类的方法和原因; 描述效标实施的方法和时间(例如: 效标是否独立于 PROM 进行测量)。</p>
CriV2	定量数据	<p>(当效标中有定量数据时) 报告相关性;</p> <p>或(当效标是二分类数据时) 报告受试者工作特征曲线下的面积(ROC)。</p>
CriV3	分类数据	<p>描述 PROM 二分类或多分类的方法(和原因)。</p> <p>报告其敏感性和特异性。</p>

构念效度中假设检验的具体报告建议 (Construct Validity)		
条目编号	条目名称	内容描述
ConV1	被比较工具	描述被比较工具所测量的构念。 报告被比较工具的测量属性和相关的引文或数据。
ConV2	比较组	报告比较组的特征，包括每组的样本量。
ConV3	假设	报告所有的假设，包括所关注的 PROM 和其他被比较工具之间预期相关性的方向和大小，或 PROM 组间分数差异的方向和大小。
ConV4	统计分析	报告用于检验每个假设的所有统计方法和结果。
ConV5	结果	根据研究假设，报告相应结果。

反应度的具体报告建议 (Responsiveness)		
条目编号	条目名称	内容描述
Resp1	被比较工具	描述被比较工具所测量的构念。 报告被比较工具的测量属性和相关的引文或数据。
Resp2	比较组	报告比较组的特征，包括每组的样本量。
Resp3	假设	报告所有的假设，包括所关注 PROM 的变化和其他被比较工具的变化之间预期相关性的方向和大小，或 PROM 组间分数变化差异的方向和大小。
Resp4	测量步骤	报告是否使用相同的工具对同一样本进行了测量。 描述测量步骤，包括不同测量工具之间的时间间隔。
Resp5	测量间隔期	描述各时间点之间的测量间隔期。
Resp6	干预/暴露	如必要，描述在间隔期给予的干预或暴露方法。
Resp7	患者变化	报告受试者在 PROMs 测量的构念上发生改善或退化的比例（以及发生心锚效应*的具体细节）。 报告研究目标人群的 PROM 分数相较于先前假设的变化。
Resp8	统计分析	报告用于检验每个假设的所有统计方法和结果。
Resp9	结果	根据研究假设，报告相应结果。

*心锚效应：心锚 (Anchor) 是条件反射 (conditioning) 的一种形式和技术，由内心主观情绪与感知到的外在事物或行为（某一动作或表情）的连接而产生。心锚可自然发生，也可人为设置，用于改变内心状态。常见的心锚效应包括：“触景生情”、“睹物思人”、“爱情与钻石”、“一遭被蛇咬，十年怕井绳”等。

参考文献

1. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19(4):539-49. doi:10.1007/s11136-010-9606-8 [doi]
2. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res* 2018;27(5):1171-79. doi:10.1007/s11136-017-1765-4
3. Ioannidis JP. How to make more published research true. *PLoS Med* 2014;11(10):e1001747. doi:10.1371/journal.pmed.1001747 [published Online First: 2014/10/22]
4. Gagnier JJ, Lai J, Mokkink LB, et al. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res* 2021 doi:10.1007/s11136-021-02822-4 [published Online First: 2021/04/06]