

DOI: 10.13288/j.11-2166/r.2021.03.008

循证中医药

冠心病中医临床疗效评价量表的测量学性能及研究方法学质量评价

杨洋¹, 孙亚男^{2△}, 于长禾^{3✉}

1. 中国中医科学院西苑医院, 北京市海淀区西苑操场 1 号, 100091; 2. 首都医科大学宣武医院; 3. 北京中医药大学东直门医院

摘要 目的 系统评价冠心病中医临床疗效评价量表的测量性能及相关研究的方法学质量。方法 检索中国知网、万方、重庆维普、PubMed、Embase 及 OVID 数据库, 检索时间范围为建库至 2019 年 11 月 1 日。纳入测量冠心病中医临床疗效评价相关的工具及其测量学性能评价的文献, 分别采用量表测量学性能评价标准和 COSMIN 清单对量表测量学性能、量表研究的方法学质量进行评价。结果 纳入文献 17 篇, 涉及 14 项研究共计 8 个量表, 分别为冠心病稳定型心绞痛(痰瘀互结证)疗效评价证候计分表(TCMsys)、冠心病稳定型心绞痛(痰瘀互结证)自评量表(CHD-SF)、冠心病心绞痛血瘀证疗效评价量表(CHD-XY)、冠心病心绞痛中医证候疗效评价量表(CHD-TCM)、冠心病心绞痛患者报告临床结局疗效评价量表(CHD-PRO)、中医特色冠心病生存质量量表(CAD-QOL)、冠心病心绞痛(气虚血瘀证)症状疗效评分量表(CHD-QX)、冠心病中西医结合生存质量量表(CHD-QOL)。测量学性能方面, 8 个量表的理论框架和条目研发满足内容效度要求, 且反应度评价均在一般及以上; 除 CHD-QOL 外其余 7 个量表的内部一致性、效标效度、结构效度评价多为差或未提及; 解释性评价 5 个量表未提及, 可重复性中的一致性、天花板/地板效应评价 8 个量表均未提及。相关研究的方法学质量方面, 量表总体研究方法学质量一般偏低, 其中 CHD-SF 研究质量相对最好, 其次是 CHD-QOL。结论 现有证据提示冠心病中医临床疗效评价量表的测量性能还不够全面、总体研究方法学质量偏低, 其中 CHD-QOL 的性能评价及研究质量相对尚可, 暂时推荐其作为以冠心病为目标人群的临床对照和纵向研究的测量工具。

关键词 冠心病; 中医临床疗效评价量表; 测量学性能; 方法学; 质量评价

随着中医药临床试验的逐年增多, 已有研究者认识到核心结局指标集(core outcome sets, COS)能规范化、标准化结局指标及其测量工具, 减少发表偏倚或结局指标选择性报告, 提高不同干预横向比较及不同研究纵向 Meta 分析效果^[1]。中医药治疗稳定型心绞痛的 COS 研究^[2]显示, 中医药治疗稳定型心绞痛常用结局指标包括心绞痛相关指标(如心绞痛发作次数、持续时间、硝酸甘油停减率/用量等)、生活质量(西雅图心绞痛量表)、安全性指标以及中医证候疗效、中医证候积分、中医症状积分等中医特色指标。中医药临床疗效评价工具指在中医药理论指导下或中医学者参加设计的概念

框架、实施研制和评价性能, 具有收集临床数据并进行中医临床疗效评价的工具^[3]。研究^[4-5]证明, 基于中医理论或者中国人群的普适性健康相关生活质量量表具有中国跨文化特点, 与单纯翻译引进的国外量表存在差异。中医理论指导下建立的冠心病相关量表较多, 但是缺少对这些量表系统的评价和推荐, 因此, 本系统评价在全面检索冠心病中医临床疗效的测量工具的基础上评价其测量学证据及研究方法学质量, 以期对相关量表的临床正确使用提供理论依据。

1 资料与方法

1.1 检索策略与文献纳入、排除标准

计算机检索中国知网、万方数据库、重庆维普数据库、PubMed、Embase 及 OVID 数据库, 检索

基金项目: 国家自然科学基金(81803956)

△与第一作者同等贡献

✉ 通讯作者: yakno2@163.com

时间为建库至 2019 年 11 月 1 日。英文检索词: traditional Chinese medicine, scale, questionnaire; 中文检索词以“中医”作为主题词、题目或关键词,“量表”“问卷”作为题目或关键词,并以纳入文献或综述的参考文献补充潜在相关研究。纳入中医药理论指导下应用于冠心病中医临床疗效评价的工具及对工具进行性能评价的可获得全文的研究。排除单纯译制引进或仅作为收集中医信息等非临床疗效评价使用的工具如中医认知度调查等。

1.2 文献筛选与数据提取

由两名研究者独立筛选,根据纳入、排除标准筛选文献,并对全文排除的文献进行复检。不一致的地方由双方共同协商达成共识,无法取得共识时由第三方仲裁。提取数据包括:量表名称、发表年份与类型、目标人群、理论框架、条目池来源、条目筛选、量尺范围、计分方法、可行性分析以及量表信度、效度、反应度、重复性、解释性等。

1.3 评价标准

量表测量学性能评价标准^[6]包括内容效度、内部一致性、效标效度、结构效度、可重复性、反应度、天花板/地板效应、解释性。每项对应好、一般、差、未提及 4 个判断量表质量的选项,以一个描述量表相关测量性能证据的表格展示量表的质量。

量表性能相关研究的方法学质量评价采用经过德尔菲法共识形成的 COSMIN 清单^[7-9]。清单评价分 4 步、12 个框架,共计 119 项。框架涵盖内部一致性、重测信度、内容效度、结构效度、假设检验、效标效度、反应度、解释性和条目反应理论等,框架针对不同测量人群进行量表测量性能评价的结果进行归纳^[6]。评价研究的方法学质量有 4 个原则:如评价所有相关项的证据充足,为优秀;如缺少某些证据,但可假设证据充足,为好(如假定患者重复测量时间内的病情可能未发生改变);如性能被低估(如患者异质性大或重测时间设计长),或样本量不足,或设计/统计分析的小问题,为一般;如因设计/统计分析(如样本量小或应用了不恰当的统计方法)导致的重大缺陷而结果不被信任,为差。最后按照“最差评价原则”,以框架内各项评价中最低评价代表此框架性能的综合评价。为了避免偏倚,由两名研究者独立评价纳入文献的相关数据,而后评价量表测量学性能质量及其研究的方法学质量。

2 结果

2.1 纳入文献的基本信息

共检索 13 462 篇文献,纳入全文文献 528 篇,综述和纳入文献的参考文献补充 9 篇,最后纳入冠心病中医临床疗效评价量表相关研究文献 17 篇,涉及 14 项研究、8 个量表,分别为冠心病稳定型心绞痛(痰瘀互结证)疗效评价证候计分表(TCMsys)、冠心病稳定型心绞痛(痰瘀互结证)自评量表(CHD-SF)、冠心病心绞痛血瘀证疗效评价量表(CHD-XY)、冠心病心绞痛中医证候疗效评价量表(CHD-TCM)、冠心病心绞痛患者报告临床结局疗效评价量表(CHD-PRO)、中医特色冠心病生存质量量表(CAD-QOL)、冠心病心绞痛(气虚血瘀证)症状疗效评分量表(CHD-QX)、冠心病中西医结合生存质量量表(CHD-QOL)。

文献发表于 2008—2013 年,以冠心病患者为研究对象的文献 3 篇,明确冠心病心绞痛患者的 5 篇;研究人群年龄 18~89 岁,平均年龄大于 60 岁;研究均为多中心,样本量 35~350 例,量表完成率 85% 以上,纳入文献基本特征详见表 1。

2.2 纳入量表的基本信息

纳入量表的基本信息见表 2。研究目的和目标人群决定量表理论框架的构建^[10]。TCMsys、CHD-SF、CHD-PRO 以中医基础理论为指导构建以疾病、证候为维度的中医证候疗效评价量表,其中 TCMsys 根据病证模型构建了 2 个维度,CHD-SF 以“以心为主,五脏相关”“内伤七情”“天人相应”理论为指导形成病、证维度,CHD-PRO 根据五脏学说主病的情况构建了 5 个维度。CAD-QOL 参考西雅图心绞痛问卷(SAQ)、生活质量量表(SF-36),并结合冠心病中西医结合诊疗特点形成中医特色框架。其余 4 个量表通过问卷调查、文献研究、病例调研或多种方法相结合构建理论框架,其中 CHD-XY 基于前期文献研究和专家调查形成 4 维度 12 条目;CHD-TCM 通过查阅文献,采取专家调查法构建 20 条目量表;CHD-QX 对主观症状进行病情和疗效评价并形成 6 个症状条目量表;CHD-QOL 包含中医症状、生理、心理、社会维度等维度共 21 条目。

2.3 纳入量表测量学性能的质量评价

冠心病中医临床疗效评价量表测量学性能的质量评价详见表 3。现有证据表明,TCMsys 在冠心病稳定型心绞痛痰瘀互结证人群中具有很好的可重复性,人群与研制量表的人群相同评价内容效度为

表 1 冠心病中医临床疗效评价量表相关研究纳入文献基本特征

量表名称	研究 ID	研究人群	年龄/岁 范围或($\bar{x} \pm s$)	人数 (男性例数/%)	证候类型	研究所在地	中心数	完成率/%
TCMsys	王鸿琳 2012 ^[1]	SA	68 ± 22.5	184(41.21)	痰瘀互结证	沈阳	6	99.46
	于长禾 2013a ^[2]	SA	68 ± 22.5	182(41.21)	痰瘀互结证	沈阳	6	—
CHD-SF	张会永 2011b ^[3]	SA/H	—	—	—	—	—	—
	刘丽星 2011 ^[4]	SA/H	(59.20 ± 12.38) / (63.29 ± 12.29)	70(45.71) / 35(48.57)	—	沈阳	6	—
	杜蕊 2011a ^[5-14]	—	—	—	—	—	—	—
	刘丽星 2011a ^[15]	—	—	—	—	—	—	—
	陈智慧 2011a ^[16]	—	—	—	—	—	—	—
CHD-XY	何庆勇 2011 ^[17]	AP	62.48 ± 10.28	350(71.1)	血瘀证	北京、河南、湖北	5	—
CHD-TCM	何庆勇 2011 ^[18]	AP	60.98 ± 10.04	219(72.1)	—	北京	—	—
CHD-PRO	何庆勇 2010a ^[19]	—	—	—	—	—	—	—
	何庆勇 2010b ^[20]	—	—	—	—	—	—	—
	何庆勇 2011 ^[19,21]	CAD/SA/UA	60.57 ± 9.23	237(59.49)	—	北京、河南、湖北	4	98.80
	何庆勇 2011b ^[22]	—	—	—	—	—	—	—
CAD-QOL	朱婷 2008a ^[23]	CHD	—	—	—	—	—	—
	朱婷 2008 ^[24]	CHD	—	120	—	—	—	85.00
CHD-QX	吕映华 2008 ^[25]	AP	18 ~ 65	—	气虚血瘀证	—	8	96.29
CHD-QOL	王伟 2009 ^[26]	CHD/H	—	101(77.2) / 50(64.0)	心气不足、 痰瘀痹阻证	广东	3	—

注: TCMsys, 冠心病稳定型心绞痛(痰瘀互结证)疗效评价证候计分表; CHD-SF, 冠心病稳定型心绞痛(痰瘀互结证)自评量表; CHD-XY, 冠心病心绞痛血瘀证疗效评价量表; CHD-TCM, 冠心病心绞痛中医证候疗效评价量表; CHD-PRO, 冠心病心绞痛患者报告临床结局疗效评价量表; CAD-QOL, 中医特色冠心病生存质量量表; CHD-QX, 冠心病心绞痛(气虚血瘀证)症状疗效评分量表; CHD-QOL, 冠心病中西医结合生存质量量表; a, 条目筛选; b, 理论框架; SA, 冠心病稳定型心绞痛; UA, 冠心病不稳定型心绞痛; H, 健康人; AP, 冠心病心绞痛; CHD, 冠心病; CAD, 冠状动脉粥样硬化性心脏病; —, 原文未说明。

表 2 冠心病中医临床疗效评价量表基本信息

量表名称	理论框架	量尺 (范围)	计分方向	计算得分	测量方式	测量时间 /min	重测间隔 /h
TCMsys	2 维度 8 条目: 疾病维度 2 条目、证候维度 6 条目	4(0 ~ 3)	B-W	总分/1.28 (百分制转化)	他评	8	24
CHD-SF	2 维度 7 条目: 疾病维度 5 条目、证候维度 2 条目	5(1 ~ 5)	W-B	—	自评	3 ~ 5	24 ~ 48
CHD-XY	4 维度 12 条目: 疾病、血瘀证证候、舌脉象、理化病理检查各维度均 3 条目	2,3,4(0 ~)	B-W	—	他评	—	—
CHD-TCM	20 条目: 胸痛、胸闷、倦怠乏力、短气、心悸、自汗、盗汗、五心烦热、两目干涩、不寐、太息、多痰、畏寒肢冷、舌红、苔厚腻、舌紫暗、舌有瘀斑瘀点、脉细弱、脉滑数、脉沉细	2,3,4(0 ~)	B-W	—	他评	—	—
CHD-PRO	5 维度 30 + 1 条目: 生理领域 13 条目、心理领域 7 条目、独立性领域 3 条目、社会关系领域 5 条目、社会环境领域 2 条目、总体评价 1 条目	5(1 ~ 5)	W-B	维度、总分	自评	15.77 ± 6.82	—
CAD-QOL	7 维度 36 + 1 条目: 临床主要症状 6 条目、躯体活动受限程度 9 条目、疾病发作情况 1 条目、疾病诱发因素 4 条目、疾病伴随情况 9 条目、治疗满意程度 4 条目、疾病认识程度 3 条目、总满意度评估 1 条目	5	W-B	—	自评 (帮助下完成)	15	9 ~ 24
CHD-QX	6 条目: 胸痛、胸闷、心悸、气短、神疲乏力、唇色紫暗	2,4(0 ~)	B-W	—	他评	—	—
CHD-QOL	4 维度 21 条目: 中西医症状维度 8 条目、生理维度 5 条目、心理维度 5 条目、社会维度 3 条目	5	(W-B) R	—	自评	4 ~ 16 (平均 11.36)	72

注: TCMsys, 冠心病稳定型心绞痛(痰瘀互结证)疗效评价证候计分表; CHD-SF, 冠心病稳定型心绞痛(痰瘀互结证)自评量表; CHD-XY, 冠心病心绞痛血瘀证疗效评价量表; CHD-TCM, 冠心病心绞痛中医证候疗效评价量表; CHD-PRO, 冠心病心绞痛患者报告临床结局疗效评价量表; CAD-QOL, 中医特色冠心病生存质量量表; CHD-QX, 冠心病心绞痛(气虚血瘀证)症状疗效评分量表; CHD-QOL, 冠心病中西医结合生存质量量表; B-W, 随量表评分增加健康程度越差; W-B, 随量表评分增加健康程度越好; R, 条目中含得分逆转条目; —, 原文未说明。

表 3 冠心病中医临床疗效评价相关量表测量学性能的质量评价

量表名称	研究对象	内容效度	内部一致性	校标效度	结构效度	可重复性		反应度	天花板/地板效应	解释性
						可靠性	一致性			
TCMsys	SA	一般	差	差	差	好	未提及	未提及	未提及	未提及
CHD-SF	SA + H	好	差	差	差	好	未提及	未提及	未提及	未提及
CHD-XY	AP	差	差	未提及	差	好	未提及	一般	未提及	一般
CHD-TCM	AP	差	差	未提及	差	好	未提及	一般	未提及	一般
CHD-PRO	CAD + SA + UA	好	差	未提及	差	好	未提及	一般	未提及	一般
CAD-QOL	CHD	好	差	未提及	差	好	未提及	一般	未提及	未提及
CHD-QX	AP	一般	一般	未提及	未提及	未提及	未提及	好	未提及	未提及
CHD-QOL	CHD + H	差	好	好	好	好	未提及	一般	未提及	未提及

注, TCMsys, 冠心病稳定型心绞痛 (痰瘀互结证) 疗效评价证候计分表; CHD-SF, 冠心病稳定型心绞痛 (痰瘀互结证) 自评量表; CHD-XY, 冠心病心绞痛血瘀证疗效评价量表; CHD-TCM, 冠心病心绞痛中医证候疗效评价量表; CHD-PRO, 冠心病心绞痛 PRO 疗效评价量表; CAD-QOL, 中医特色冠心病生存质量量表; CHD-QX, 冠心病心绞痛 (气虚血瘀证) 症状疗效评分量表; CHD-QOL, 冠心病中西医结合生存质量量表; SA, 冠心病稳定型心绞痛; H, 健康人; AP, 冠心病心绞痛; CAD, 冠心病; UA, 冠心病不稳定型心绞痛。

一般; 量表证候维度 Cronbach α 系数仅为 0.409, 内部一致性较差; TCMsys 总分与中医症状积分相关系数为 0.658, 与 SAQ 的相关系数为 -0.512, 效标效度较差; 结构效度显示 4 个公因子累积贡献率仅为 69.50%, 结构效度较差。CHD-SF 在冠心病稳定型心绞痛痰瘀互结证人群中具有很好的内容效度和可靠性, 证候维度的 Cronbach α 系数为 0.517, 量表总分相关系数为 0.611, 因子分析的累积贡献率为 62.10%, CHD-SF 的内部一致性、效标效度、结构效度均较差。CHD-XY、CHD-TCM 分别在冠心病心绞痛血瘀证人群、冠心病心绞痛人群中具有很好的可靠性, 但由于研究缺少与最小重要变化差异 (MIC) 比较而反应度评价为一般; 由于量表均值和标准差无法解释亚组或未定义 MIC 而评价解释性为一般; 量表前期条目池的形成无目标人群的参与而内容效度差, 两个量表各个维度的 Cronbach α 系数分别为 0.63 ~ 0.91、0.22 ~ 0.85, 内部一致性均较差; 因子分析结果显示结构效度较差。CHD-PRO 在冠心病心绞痛人群中具有较好的内容效度、可靠性; 研究缺少 MIC 比较而评价反应度为一般; 量表均值和标准差无法解释亚组或未定义 MIC 而解释性评价为一般; 量表各维度 Cronbach α 系数为 0.58 ~ 0.85, 内部一致性差; 因子分析结果显示公因子能解释 50.08% 变量, 结构效度差。CAD-QOL 在冠心病人群中内容效度、可靠性好, 但因缺少 MIC 比较而评价反应度为一般; 样本量小于 7 倍条目数, 7 个维度中有 4 个维度 Cronbach α 系数小于 0.7, 内部一致性差; 因子分析显示公因子能解释 57.76% 变量, 评价结构效度为差。CHD-QX 在冠心病心绞痛气虚血瘀证人群中反应度好, 但由于缺少对量表目标人群、测量概

念、条目筛选等过程的明确描述, 内容效度和内部一致性评价为一般。CHD-QOL 在冠心病心气不足、痰瘀痹阻证人群中内部一致性、效标效度、结构效度、可靠性好; 因缺少 MIC 比较而评价反应度为一般; 无目标人群参加条目池形成及条目筛选, 内容效度评价为差。

2.4 纳入量表性能相关研究的方法学质量评价

冠心病中医临床疗效评价量表性能相关研究的方法学质量的 COSMIN 清单评价结果详见表 4。内部一致性框架中, CHD-SF 研究的方法学质量优秀; CHD-XY、CHD-TCM、CHD-PRO、CHD-QOL 由于未描述缺失数据及其处理方法而可能导致偏倚, 方法学质量评价为好; TCMsys、CAD-QOL 由于评价量表性能的研究人群与条目筛选的人群相同, 增加了测量系统偏倚的风险, 试验设计存在缺陷, 因此研究的方法学质量为一般; CHD-QX 未做因子分析评价内部一致性, 评价其方法学质量为差。重测信度框架中, CHD-XY、CHD-PRO 由于患者重测期间病情稳定, 可推断重测条件相似, 虽然未描述模型, 但组内相关系数可推断实施的独立性, 因此评价重测信度研究方法质量为好; TCMsys、CAD-QOL 由于量表评价和量表研制用相同样本, 评价其方法学质量一般; CHD-SF、CHD-TCM、CHD-QOL 重测样本量分别为 30、30、40 例, 均小于 50 例, 因此评价其方法学质量为一般。内容效度框架中, CHD-SF、CAD-QOL、CHD-QX 内容效度研究的方法学质量达到了优秀; TCMsys、CHD-PRO 条目定性研究中目标人群小于 5 例或例数不能确定, 且 TCMsys 条目生成与性能评价使用相同人群, 因此其方法学质量为一般; CHD-XY、CHD-TCM、CHD-QOL 条目定性研究中缺少目标人群参与, 方

表 4 冠心病中医临床疗效评价量表性能相关研究的方法学质量评价

量表名称	条目反应理论	内部一致性	重测信度	内容效度	结构效度	假设检验	效标效度	反应度	解释性
TCMsys	SA	一般	一般	一般	一般	一般	一般	—	差
CHD-SF	SA + H	优秀	一般	优秀	优秀	一般	好	—	一般
CHD-XY	AP	好	好	差	好	差	—	差	一般
CHD-TCM	AP	好	一般	差	好	差	—	差	好
CHD-PRO	CAD + SA + UA	好	好	一般	好	差	—	差	一般
CAD-QOL	CHD	一般	一般	优秀	一般	—	—	差	差
CHD-QX	AP	差	—	优秀	—	—	—	差	差
CHD-QOL	CHD + H	好	一般	差	优秀	一般	一般	一般	好

注: TCMsys, 冠心病稳定型心绞痛 (痰瘀互结证) 疗效评价证候计分表; CHD-SF, 冠心病稳定型心绞痛 (痰瘀互结证) 自评量表; CHD-XY, 冠心病心绞痛血瘀证疗效评价量表; CHD-TCM, 冠心病心绞痛中医证候疗效评价量表; CHD-PRO, 冠心病心绞痛 PRO 疗效评价量表; CAD-QOL, 中医特色冠心病生存质量量表; CHD-QX, 冠心病心绞痛 (气虚血瘀证) 症状疗效评分量表; CHD-QOL, 冠心病中西医结合生存质量量表; —, 原文未说明。

法质量为差。结构效度框架中, CHD-SF、CHD-QOL 结构效度的方法学质量达到了优秀; CHD-XY、CHD-TCM、CHD-PRO 由于未描述缺失数据或其处理方法可能导致偏倚, 方法学质量评为好; TCMsys、CAD-QOL 条目生成与性能评价用相同人群导致方法学质量为一般。假设检验框架中, TCMsys、CHD-SF、CHD-QOL 假设模糊或者缺少可以推导的预期结果, 因此评价假设检验的方法学质量为一般; CHD-XY、CHD-TCM、CHD-PRO 未说明对照工具测量的结构或其与测量人群的关系, 评价方法学质量为差。效标效度框架中, CHD-SF 虽然未说明选择金标准依据, 但可推断其为金标准, 因此效标效度的方法学质量为好; TCMsys 由于条目生成与性能评价使用相同人群, CHD-QOL 由于校标检验的样本量为 40 例, 因此两者方法学质量均为一般。反应度框架中, CHD-XY、CHD-TCM、CHD-PRO、CHD-QX 没有分值变化的先验假设, 未说明测量工具或其与测量人群的关系, 因此反应度的方法学质量为差。解释性框架中, CHD-TCM、CHD-QOL 描述量表得分的分布, 但未说明最高得分或最低得分的人群比例, 研究描述了天花板/地板效应、分布的均值和标准差来解释相关亚组、MIC 值等, 有 2 项以上相关描述, 解释性的方法学质量为好; CHD-SF、CHD-XY、CHD-PRO 研究中描述了均值和标准差解释相关亚组或说明 MIC 值, 方法学质量为一般; TCMsys、CAD-QOL、CHD-QX 没有描述量表得分分布, 或最高得分或最低得分的人群比例, 或天花板/地板效应, 或均值和标准差, 或 MIC 值, 因此方法学质量为差。

3 讨论

本研究纳入 14 项研究的 8 个基于中医理论研

制的冠心病中医临床疗效评价量表, 大部分量表都有理论框架构建、条目形成以及条目筛选方法的详细描述, 充分反映了冠心病中医临床疗效评价的特点。

对于疗效评价量表, 内容效度和反应度极为重要。本研究中 8 个量表理论框架和条目研发满足了内容效度要求, 量表的反应度评价在一般或以上; 内部一致性、效标效度、结构效度评价多数量表评价为差; 天花板/地板效应、解释性多数量表未提及。导致测量学性能评价结果良莠不齐的原因大致为: 1) 样本量不足导致内部一致性和结构效度的评价方法得到的结果存在潜在偏倚; 2) 中医理论有别于现代医学理论, 以中医药理论为指导研制的量表与现代医学研制量表作“金标准”的效标效度的相关系数无法达到标准所预期的 0.7, 这也与缺少必要的研究假设有关; 3) 中医研制考评量表缺少对国际相关标准的关注和解读, 导致研究缺少标准中涉及的一些设计, 如缺少对量表在人群中得分分布和条目天花板/地板效应的描述, 缺少至少 4 个亚组的差异性分析等; 4) 量表条目池形成无目标人群参与、用于量表测量性能评价的人群与量表研制的人群相同等设计失误。

现有证据提示本研究所纳入的所有量表的性能评价均不够全面, 以冠心病为目标人群的 8 个量表中 CHD-QOL 的性能评价相对全面, 测量学性能在一般以上 (除内容效度性能为差), 目前可推荐应用于以冠心病为目标人群的临床对照和纵向研究。但需要说明的是, 此推荐仅限于现有证据, 大量的临床研究和量表的应用是证据不断累积的过程, 故不排除随着证据的增加而发生结论的变化。量表性能评价研究中多采用经典测量理论, 缺少现代测量

方法如条目反应理论 (IRT), 而 IRT 可采用非线性模型建立被试者对项目的反应 (观察变量) 与潜在特质 (潜变量) 之间的关系, 故今后可结合现代测量方法综合评价量表测量学性能。

高质量的研究是量表测量性能的保障。通过量表性能评价研究的方法学质量结果可见, 本研究所纳入量表总体研究方法学质量一般偏下, 主要由于研究无明确假设、目标人群没有参与量表条目评价、样本量偏小、研究假设和研究方法描述不清或不详所导致的研究质量降级。同时反应度亦存在问题, 如数据处理方法未描述、干预描述不清、先验假设未说明测量工具间关系是相关或差异及其效应量的大小。依据现有证据, 8 个冠心病中医临床疗效评价量表中 CHD-SF 研究质量相对最好, 其次是 CHD-QOL。

有效、可信的量表可以真实反映患者的自我感受和生活质量, 提高临床研究和新药临床试验的质量、减少偏倚^[28]。中医特色量表研究方兴未艾, 但是需要重视量表及其研究质量, 研究者可通过提出假设、制定可操作化概念、明确目标人群来提高量表内容效度的质量, 通过目标人群干预前后纵向数据或不同人群比较来提高量表反应度的质量。本研究系统评价已有量表的测量学性能, 研究者可以根据以上证据, 结合其研究目的、测量模式等因素选择合适的量表进行临床试验, 或对已有量表进行再修订。此外, 本研究纳入文献时缺少对日语、韩语相关文献的检索, 且部分量表未获得研究者提供的量表全貌; COSMIN 清单和量表性能质量评价标准虽然较为成熟, 但是在评价过程中主要依靠评价者的主观认识, 因此对评价者的培训和一致性检验也是未来研究值得注意的事项。

利益冲突声明: 本研究作者与纳入文献作者没有利益冲突; 研究者间没有利益冲突。

参考文献

- [1] 孙亚男, 于长禾, 陈文强, 等. 中医临床疗效评价中核心结局指标集理论初探 [J]. 国际中医中药杂志, 2017, 39 (8): 676-680.
- [2] 张明妍, 张俊华, 张伯礼. 2015 年中药治疗稳定性心绞痛临床试验结局指标文献研究 [J]. 中国中西医结合杂志, 2018, 38 (2): 191-197.
- [3] 于长禾, 孙亚男, 何丽云, 等. 中医普适性生活质量量表的系统评价 [J]. 中华中医药杂志, 2016, 31 (2): 432-437.
- [4] 于长禾, 孙亚男, 何丽云, 等. 健康相关生活质量在中西医测量量表中概念异同的比较研究 [J]. 中华中医药杂志, 2016, 31 (11): 4773-4778.
- [5] 于长禾, 孙亚男, 何丽云, 等. 中医生活质量量表心理测量性能证据的系统评价 [J]. 中华中医药杂志, 2015, 30 (5): 1631-1635.
- [6] TERWEE CB, BOT SD, DE BOER MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires [J]. J Clin Epidemiol, 2007, 60 (1): 34-42.
- [7] MOKKINK LB, PRINSEN CA, BOUTER LM, et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument [J]. Braz J Phys Ther, 2016, 20 (2): 105-113.
- [8] MOKKINK LB, TERWEE CB, KNOL DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content [J]. BMC Med Res Methodol, 2010, 10: 22. doi: 10.1186/1471-2288-10-22.
- [9] TERWEE CB, MOKKINK LB, KNOL DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist [J]. Qual Life Res, 2012, 21 (4): 651-657.
- [10] 于长禾, 孙亚男, 何丽云, 等. 量表研制及测量属性评价标准的比较研究 [J]. 世界科学技术-中医药现代化, 2015, 17 (6): 1127-1135.
- [11] 王鸿琳. 冠心病稳定型心绞痛 (痰瘀互结证) 疗效评价证候计分表的研制与初步考评 [D]. 沈阳: 辽宁中医药大学, 2012.
- [12] 于长禾, 王鸿琳, 张哲, 等. 中医证候疗效评价计分表条目筛选分析 [J]. 中华中医药学刊, 2013, 31 (6): 1278-1282.
- [13] 张会永, 张哲, 杜蕊, 等. 冠心病稳定型心绞痛 (痰瘀互结证) 自评量表理论模型构建方法与思路 [J]. 中华中医药学刊, 2011, 29 (8): 1738-1740.
- [14] 刘丽星. 冠心病稳定型心绞痛 (痰瘀互结证) 自评量表研制与考评 [D]. 沈阳: 辽宁中医药大学, 2011.
- [15] 杜蕊, 张哲, 刘丽星, 等. 基于临床调查法的冠心病稳定型心绞痛·痰瘀互结证自评量表条目筛选 [J]. 辽宁中医杂志, 2011, 38 (8): 1494-1498.
- [16] 杜蕊, 张哲, 陈民, 等. 冠心病稳定型心绞痛·痰瘀互结证自评量表条目筛选方法学研究 [J]. 中华中医药学刊, 2011, 29 (6): 1265-1270.
- [17] 刘丽星, 张哲, 杜蕊, 等. 基于德尔菲法的《冠心病稳定型心绞痛 (痰瘀互结证) 自评量表》条目筛选 [J]. 辽宁中医杂志, 2011, 38 (5): 827-829.
- [18] 陈智慧, 张哲, 杜蕊, 等. 基于深度访谈法的冠心病疗效自评量表条目筛选 [J]. 中华中医药学刊, 2011, 29

- (5): 996 - 999.
- [9] 何庆勇. 冠心病心绞痛病证结合疗效评价体系研究 [J]. 北京: 北京中医药大学, 2011.
- [10] 何庆勇, 王阶, 朱明军, 等. 基于冠心病心绞痛患者报告临床结局评价量表的条目筛选分析 [J]. 中华中医药杂志, 2010, 25(12): 2216 - 2221.
- [11] 何庆勇, 王阶, 施展, 等. 冠心病心绞痛中医 PRO 疗效评价量表理论结构模型构想 [J]. 中华中医药杂志, 2010, 25(1): 42 - 45.
- [12] 何庆勇, 王阶, 朱明军, 等. 基于冠心病心绞痛患者报告的临床结局评价量表的科学性考评 [J]. 中华中医药杂志, 2011, 26(5): 1138 - 1142.
- [13] 何庆勇, 王阶, 姚魁武, 等. 冠心病心绞痛患者报告临床结局评价量表条目池的建立和初步筛选 [J]. 中国中西医结合杂志, 2011, 31(1): 15 - 18.
- [14] 朱婷, 毛静远. 中医特色冠心病生存质量量表条目的设计思路 [J]. 辽宁中医药大学学报, 2008, 10(7): 11 - 13.
- [15] 朱婷, 毛静远. 中医特色冠心病生存质量量表的制定及考评 [J]. 辽宁中医杂志, 2008, 35(6): 854 - 855.
- [16] 吕映华, 何迎春, 杨娟, 等. 冠心病心绞痛(气虚血瘀证)症状疗效评分量表的研究 [J]. 中国临床药理学与治疗学, 2008, 13(7): 786 - 791.
- [17] 王伟. 冠心病中西医结合生存质量量表的研制及考评 [J]. 广州: 广州中医药大学, 2009.
- [18] MARSHALL M, LOCKWOOD A, BRADLEY C, et al. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia [J]. Br J Psychiatry, 2000, 176: 249 - 252. doi: 10.1192/bjp.176.3.249.

The Measurement Property and Methodological Quality of Traditional Chinese Medicine Clinical Effectiveness Evaluation Scales for Coronary Heart Disease

YANG Yang¹, SUN Yanan², YU Changhe³

1. Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, 100091; 2. Xuanwu Hospital, Capital Medical University; 3. Dongzhimen Hospital, Beijing University of Chinese Medicine

ABSTRACT Objective To systematically assess the measurement property and methodological quality of the traditional Chinese medicine (TCM) clinical effectiveness evaluation scales for coronary heart disease (CHD). **Methods** Databases including CNKI, Wanfang, VIP, PubMed, Embase and OVID were searched from the inception date till November 1st, 2019. Studies on TCM clinical effectiveness evaluation scales for CHD were included; the measurement property standards and the consensus-based standards for the selection of health measurement instruments (COSMIN) were used to evaluate the measurement performance and the methodological quality respectively.

Results Eight scales involving 14 studies of 17 papers were included, which were TCM syndrome evaluation scale of phlegm and blood stasis syndrome in CHD with stable angina pectoris (TCMsyst), self-evaluation scale of CHD stable angina with the syndrome of phlegm and blood stasis (CHD-SF), efficacy evaluation scale for CHD angina pectoris with syndrome of blood stasis (CHD-XY), effectiveness evaluation scale of TCM patterns for CHD angina pectoris (CHD-TCM), rating scale for patients reported outcomes of CHD angina (CHD-PRO), quality of life scale with TCM characteristics for coronary artery disease (CAD) (CAD-QOL), symptoms scale for evaluating efficacy of coronary angina pectoris with Qi deficiency and blood stagnation (CHD-QX), quality of life scale of integration of TCM and western medicine for CHD (CHD-QOL). In terms of measurement property, the theoretical framework and the items development of eight scales meet the content validity requirements, and the responsiveness was at average level or above. The internal consistency, criterion validity, and construct validity were assessed as poor or not mentioned in all the scales except for CHD-QOL. Explanatory evaluation was not mentioned in five scales. Consistency and ceiling and floor effect of reproducibility evaluation were not mentioned in all scales. The methodological quality of the scale-related studies was generally poor, of which the quality of CHD-SF was the best, followed by CHD-QOL. **Conclusion** The measurement property of the current TCM clinical effectiveness evaluation scales for CHD is not comprehensive, and the overall methodological quality was poor. CHD-QOL is temporarily recommended as the measurement tool for clinical control studies and longitudinal researches on CHD, as its measurement property and methodological quality are relatively acceptable.

Keywords coronary heart disease; traditional Chinese medicine clinical effectiveness evaluation scales; measurement property; methodology; quality evaluation

(收稿日期: 2019 - 11 - 30; 修回日期: 2020 - 08 - 03)

编辑: 柴倩云