

基于中国腰痛患者 Oswestry 功能障碍指数测量性能证据的 COSMIN 系统评价*

陈千吉¹ 陈红² 张英¹ 万颖¹ 刘文可¹ 王悦¹ 于长禾^{2,3}

腰痛患病率稳居全球第一,是人类致残的首要原因^[1]。我国慢性腰痛患病率为20.60%—41.54%^[2],给个人及家庭生活、社会医疗系统均带来极大负担。功能障碍为腰痛临床试验的核心结局指标^[3],Oswestry 功能障碍指数(Oswestry disability index, ODI)是功能障碍的测量工具^[4]。ODI量表基于中国腰痛人群的测量性能评价情况还未知。

基于共识健康测量工具的选择标准(consensus-based standards for the selection of health measurement instruments, COSMIN),通过评价患者报告结局指标测量工具的研究方法学质量及测量性能优劣,帮助研究人员选择合适的测量工具,以提高临床研究质量^[5]。COSMIN 系统评价指南(COSMIN methodology for systematic reviews of patient-reported outcome measures)^[6]详细说明研究步骤以指导测量工具的系统评价。

本研究按照该指南系统评价 ODI 量表,旨在判断其是否适合中国腰痛人群使用。

1 资料与方法

1.1 测量工具

ODI 由 John O'Brien 及 Fairbank 为评价慢性腰痛患者的功能和治疗于1976年所研制^[7-8]。1989年英国医学研究委员会在 ODI 1.0 版本基础上提出 ODI 2.0 版本^[9],2001年改进为 ODI 2.1 版本^[10]。ODI 量表由患者自行填写,5min 内可完成评分,内容包含疼痛强度、个人护理、举重、行走、坐、立、睡眠、性生活、社交生活和旅行等10个问题,每个问题包括5个答案,总分越高表明症状越重^[11]。

1.2 检索策略与数据库建立

课题组构建腰痛结局测量工具性能研究的文献库。计算机检索 PubMed、Embase、SinoMed、CNKI、Web of Science、万方数据库,时间为建库至2020年7月2日。检索词由以下三部分搜索项和布尔运算符“AND”组成,(1)腰痛(back pain, low back pain);(2)中国;(3)测量性能相关词

例如:效度、信度、反应度、克隆巴赫系数、组内相关系数(validity、reliability、responsiveness、Cronbach、ICC)等。纳入文献或综述的参考文献以补充潜在研究。

1.3 文献纳入与排除标准

纳入标准:①基于中国腰痛人群临床测量;②针对 ODI 量表至少一种测量性能;③可获得全文。排除标准:①仅使用 ODI 量表作为结局测量工具;②使用 ODI 量表为对比工具研究其余测量工具测量性能。

1.4 文献筛选

2名研究者按照纳排标准阅读文献题目和摘要,独立筛选、评价,分歧通过讨论或与第3位研究员协商解决。

1.5 评价流程

1.5.1 方法学质量评价:通过“COSMIN 偏倚风险检查表”^[12]对应框中的条目进行评价。每个条目都可以被评为“非常好”、“足够”、“不确定”或“不足”。每项研究的质量由框中条目的最低评分确定。如果一篇文献同时评价了几种测量性能,则应把每种测量性能的研究看作一项单独的研究,并进行偏倚风险检查表的多次填写。

1.5.2 测量性能评价:根据“良好内容效度的10条评价标准”^[13]及“单个研究内容效度评定指南”评价内容效度为充分(+)、不足(-)、不一致(±)或不确定(?)。根据“更新测量性能优劣评价标准”^[13-14]评价其余测量性能为充分(+)、不足(-)或不确定(?)。

1.5.3 证据综合与证据质量分级:证据综合:首先判断每个测量性能的研究结果是否一致。若结果一致,可以定量汇总或定性总结,并与良好测量性能的标准进行比较,以确定测量性能在总体上是充分的(+)、不足的(-)、不一致的(±)或不确定的(?);若结果不一致,可以进行亚组分析或根据大多数一致的结果得出结论,并因不一致性而使证据质量降级。

证据质量分级:根据“改进 GRADE 证据质量评级”标准^[15],假设证据为“高质量”,考虑偏倚风险^[16]、不一致性、不精确性及间接性进行降级。

DOI:10.3969/j.issn.1001-1242.2022.01.015

*基金项目:国家自然科学基金青年科学基金项目(81803956)

1 北京中医药大学第一临床医学院,北京市,100700; 2 北京中医药大学东直门医院推拿科; 3 通讯作者

第一作者简介:陈千吉,女,硕士研究生; 收稿日期:2020-09-28

2 结果

2.1 纳入研究一般情况

共检索3559篇文献,剔除重复文献1257篇,阅读题目摘要剔除不相关文献2179篇,纳入123篇文献建立数据库。全文阅读后纳入19篇,剔除无法获得全文文献2篇与重复发表文献3篇(会议论文2篇,中英互译1篇),最终纳入文献14篇^[17-30],见图1,基本情况见表1。

2.2 方法学质量评价

2项研究^[19,25]涉及内容效度,由于研究设计及统计方法

不合适方法学质量均被评为不足。5项研究涉及结构效度,其中4项^[24-25,28,30]均因为只计算了公因子方差贡献率,方法学质量被评为不确定;另1项^[19]因为样本量较小被评为不足。7项研究涉及内部一致性,其中5项^[17,23-24,26,30]方法学质量由于只计算了总量表的内部一致性系数被评为不确定,另2项^[19,28]为非常好。6项研究涉及重测信度,其中4项^[17,23,26,30]由于无法判断两次填写条件是否相似方法学质量被评为不确定,另2项由于填写间隔时间不合适^[24]及两次填写条件不相似^[28]被评为不足。1项研究^[17]涉及测量误差,由于无法评价两次填写条件是否相似方法学质量被评为不确定。8项研究涉及校标效度,其中3项研究^[23,26,28]方法学质量非常好,另5项涉及2篇文献^[24-25]由于设计缺陷被评为不确定。4篇文献的14项研究涉及假设检验,其中12项方法学质量被评为不确定,另2项为充分。4篇文献^[17,20,22,27]的5项研究涉及反应度,其中4项^[17,20,22,27]方法学质量评为非常好,另1项^[27]由于对于所比较的测量工具测量性能是否充分无相关描述而被评为不确定。结果见表2。

2.3 测量性能评价

2项内容效度研究设计及分析方法与COSMIN中的标准不相符,本研究未进行评价。研究人员将因子贡献率大于60%的结构效度视为充分,故5项研究结构效度均充分。7项研究内部一致性均充分。6项研究重测信度均充分。1项研究测量误差充分。14项校标效度研究中,3项充分、3项不确定、2项不足。14项假设检验研究依据文中假设或标准,11项不确定,2项充分。5项研究反应度均充分。结果见表2。

2.4 证据综合与证据质量分级

只有同一个版本的结局测量工具可以进行证据综合。刘晖等^[26]自行改良2.1英文版ODI,改良后的量表被纳入ODI官网,因此本研究将其与5篇文献中2.1中文版ODI研究证据进行综合。综合结果见表3。

图1 文献检索及筛选流程

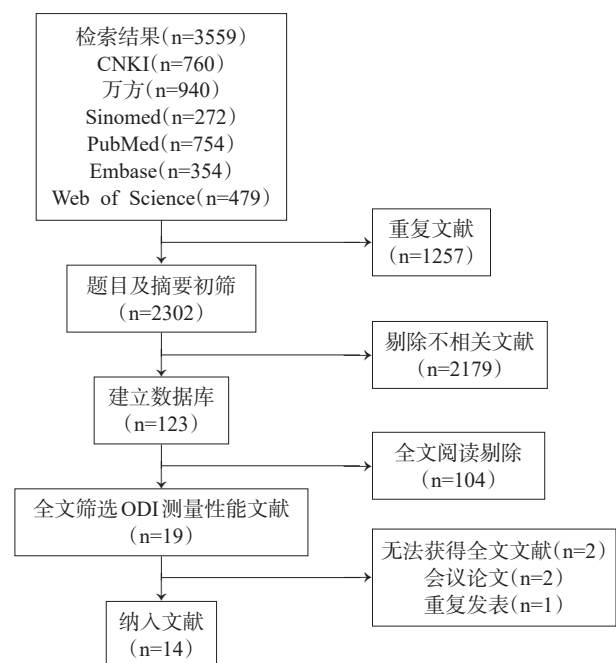


表1 纳入文献基本情况

作者/出版时间	ODI版本	总样本量(人)	平均年龄(岁)	性别(女,%)	疾病特点	条目8分析与否
Yao M/2020 ^[17]	2.1中文版	353	50.53	55.81	LDH	是
Lee C/2017 ^[18]	2.1中文版	225	40.7	45.80	CBP	是
程继伟/2017 ^[19]	自行改良版	41	51.4	36.59	DLS	无条目8
方贤聪/2017 ^[20]	汉化版	140	45.20	55.71	LS	是
Lu YM/2013 ^[21]	2.1中文版	408	59.0	64.50	BP	是
Ma C/2011 ^[22]	2.1中文版	87	42.8	34.20	NSCBP	是
Volinn E/2010 ^[23]	2.1中文版	89	39	59.00	NSCBP	是
俞红/2010 ^[24]	简氏中文版	140	36.37	45.71	LBP(LCCS)	是
刘绮/2010 ^[25]	简体中文版	107	41.5	32.71	CBP	否
刘晖/2010 ^[26]	自行改良版	179	44.46	58.10	BP	是
刘绮/2010 ^[27]	简体中文版	87	48.3	35.63	CBP	否
谭柯/2009 ^[28]	自行翻译版	86	39.5	59.30	BP(ex Sciatica)	否
郑光新/2002 ^[29]	汉化版	22	A组47.7;B组36.6	40.90	BP(LDH、LSS、LS)	否
Chow JH/2005 ^[30]	2.0中文版	79	42	31.64	BP	是

注:LDH:腰椎间盘突出症;CBP:慢性腰痛;DLS:腰椎退变性疾病;LS:腰椎滑脱;BP:腰痛;NSCBP:是非特异性腰痛;LCCS:腰骶部慢性骨筋膜间隔综合征;Sciatica:坐骨神经痛;LSS:腰椎管狭窄症;条目8是ODI量表关于腰痛对性生活影响的条目。

表2 纳入研究方法学质量评价及测量性能评价结果

测量性能	作者/出版时间	方法学质量评价	测量性能评价
内容效度	程继伟/2017 ^[19]	不足	无法评价
	刘绮/2010 ^[25]	不足	无法评价
结构效度	程继伟/2017 ^[19]	不足	+
	俞红/2010 ^[24]	不确定	+
	刘绮/2010 ^[25]	不确定	+
	谭柯/2009 ^[28]	不确定	+
	Chow JH/2005 ^[30]	不确定	+
内部一致性	Yao M/2020 ^[17]	不确定	+
	程继伟/2017 ^[19]	非常好	+
	Volinn, E/2010 ^[23]	不确定	+
	俞红/2010 ^[24]	不确定	+
	刘晖/2010 ^[26]	不确定	+
	谭柯/2009 ^[28]	非常好	+
	Chow JH/2005 ^[30]	不确定	+
重测信度	Yao M/2020 ^[17]	不确定	+
	Volinn E/2010 ^[23]	不确定	+
	俞红/2010 ^[24]	不足	+
	刘晖/2010 ^[26]	不确定	+
	谭柯/2009 ^[28]	不足	+
	Chow JH/2005 ^[30]	不确定	+
测量误差	Yao M/2020 ^[19]	不确定	+
校标效度	Volinn E/2010 ^[23]	非常好(以SF-36 PF为校标)	-
	俞红/2010 ^[24]	不确定(以VAS为校标)	+
	俞红/2010 ^[24]	不确定(以ICF核心要目为校标)	?
	俞红/2010 ^[24]	不确定(以RMDQ为校标)	+
	刘绮/2010 ^[25]	不确定(以VAS为校标)	+
	刘绮/2010 ^[25]	不确定(以SF-36为校标)	?
	刘晖/2010 ^[26]	非常好(以SF-36为校标)	?
	谭柯/2009 ^[28]	非常好(以SF-36为校标)	-
假设检验	Yao M/2020 ^[17]	不确定(与JOA BPEQ问题1比较)	?
	Yao M/2020 ^[17]	不确定(与JOA BPEQ问题2比较)	?
	Yao M/2020 ^[17]	不确定(与JOA BPEQ问题3比较)	?
	Yao M/2020 ^[17]	不确定(与JOA BPEQ问题4比较)	?
	Yao M/2020 ^[17]	不确定(与JOA BPEQ问题5比较)	?
	Yao M/2020 ^[17]	不确定(与NPRS比较)	?
	Yao M/2020 ^[17]	不确定(与RMDQ比较)	?
	Yao M/2020 ^[17]	不确定(与SF-PCS比较)	?
	Yao M/2020 ^[17]	不确定(与SF-MCS比较)	?
	Volinn, E/2010 ^[23]	充分(与SF-36 energy scale比较)	+
	Volinn E/2010 ^[23]	不确定(与过去两周疼痛量表比较)	+
	刘晖/2010 ^[26]	不确定(与VAS比较)	+
	刘绮/2010 ^[27]	不确定(与VAS比较)	?
	刘绮/2010 ^[27]	充分(与JOA比较)	?
反应度	Yao M/2020 ^[17]	非常好	+
	方贤聪/2017 ^[20]	非常好	+
	Ma C/2011 ^[22]	非常好	+
	刘绮/2010 ^[27]	非常好	+
	刘绮/2010 ^[27]	不确定	+

注: +表示测量性能充分; -表示测量性能不足; ?表示测量性能不确定

表3 2.1中文版ODI证据综合与证据质量评级结果

测量属性	证据综合	证据质量评级	证据质量降级原因
内部一致性	充分	中等质量	偏倚风险严重
重测信度	充分	中等质量	偏倚风险严重
测量误差	充分	低质量	偏倚风险非常严重
校标效度	不确定	中等质量	不一致性严重
假设检验	不确定	低质量	偏倚风险严重
反应度	充分	高质量	结果不一致性严重

其余8篇文献对于ODI版本的描述不统一,未能全面获得各自使用的ODI量表,无法判断其差异及差异大小,且每个版本相关研究极少,因此未整合其研究结果。

3 讨论

研究结果显示2.1中文版ODI内部一致性及重测信度充分,反映其各条目之间的相互关联性较好且在不同时间对于同一受试者的测量稳定性充分。以SF-36为校标时,中等质量证据证明其校标效度不确定,反映其与SF-36量表的相关性不确定。低质量证据证明其结构效度假设检验不确定。有低质量的证据证明其测量误差充分,反映其分数变化不是由测量误差所致,代表其可以真实地反映功能障碍的变化,但证据质量低。有高质量的证据证明其反应度充分,代表该量表反映微小特性变化的能力充分。但尚不能推荐其使用于中国人群,因为缺少内容效度研究。Chiarotto A等^[31]对2.1英文版ODI的系统评价同样显示对于2.1英文版ODI的使用持疑问态度,需要通过对患者的定性研究来全面评价内容效度。

研究涉及8个ODI版本:2.1中文版^[17-18, 21-23]、2.0中文版(繁体)^[30]、简体中文版ODI(CODI)^[25, 27]、简式中文版ODI(SCODI)^[24]、汉化版ODI^[20]、及4个研究^[19, 26, 28-29]自行改良的ODI(其中刘晖等^[26]自行改良版本视为2.1中文版)。这提示当前中文版ODI版本繁多,且名称不统一。

ODI测量性能研究的方法学质量整体不高。建议量表测量性能研究应注意研究设计与报告的规范与详尽,以提高证据质量。根据COSMIN标准,内容效度的研究应包括相关人群(专家与受试者)的定性或定量研究、对量表相关性、全面性及可理解性全方位评价(相关性包括量表条目与待测结构、目标人群及使用环境是否相关、条目选项及回忆的时间间隔是否合适5项内容;全面性指量表条目所涵盖的概念是否全面,是否存在缺失;可理解性包括量表的说明、条目的选项、条目的措辞是否能被受试对象理解、量表选项是否与问题匹配4项内容)、严格的数据管理与分析(被广泛认可的数据处理方法与多名研究人员的参与)等。结构效度的研究应致力于采用大样本(大于条目数的7倍且≥100例),并且采用

验证性因子分析的分析方法,计算比较拟合指数、塔克刘易斯指数、近似均方根误差或标准化的均方根残差,详细报告相关数据等。内部一致性的研究应在明确结构效度的前提下,除了量表总的Cronbach α 系数,还应有每个条目或亚量表的计算。重测信度的研究应对测量条件和/或两次填写间隔期间内待测结构是否稳定进行说明。跨文化效度的研究存在空缺,其需要在文化调适的基础上再进行结构效度的评价,评价时应选用两组样本,除了语言的差异外,应控制其余人群特征(如年龄、性别、疾病特点等)的一致。反应度的研究可以通过其与“金标准”比较、3种假设检验(与其余测量工具比较、亚组间比较或干预前后比较)4种方法进行评价。应用假设检验的方法应明确给出假设,此外,选取与其余测量工具比较的方法时,应选择测量性能较为充分的测量工具,并提供其充分的相关证据;选取亚组间比较的方法时,应对亚组间的特征进行详细的描述;选取干预前后的方法时,应对干预进行详细的说明。

研究发现纳入文献与COSMIN测量性能的定义存在不一致性。COSMIN将PROMs的测量性能分为三大类:信度(reliability)、效度(validity)和反应度(responsiveness)。其中信度又包括内部一致性(internal consistency)、测量误差(measurement error)和信度(reliability),信度又再划分为重测信度(test-retest reliability)、评分者间信度(inter-rater reliability)及评分者内信度(intra-rater reliability)。效度包括内容效度(content validity)主要指表面效度(face validity)、建构效度(construct validity),建构效度又再划分为结构效度(structural validity)、假设检验(hypotheses testing)、跨文化效度(cross-cultural validity)及效标效度(criterion validity)。Yao等^[17]通过分析ODI与其他结局测量工具的相关性评价“结构效度”,本研究根据COSMIN系统评价指南将其归类为假设检验中聚合效度的模块以进行评价。

5篇文献^[19,25,27-29]由于条目8(涉及腰痛对性生活的影响)填写缺失率较大而未将其纳入分析。Chow等^[30]于2005年在中国香港进行的样本量为79例的研究,在评价2.0中文版ODI内部一致性时,剔除条目8前后计算克朗巴赫系数结果分别为0.81及0.89。此外,Liu等^[32]于2009年发表的文献通过对受试者年龄分布进行分析,发现50岁以前条目8的回答率呈上升趋势,30—39岁年龄组回答率最高(89.74%),年龄超过法定婚龄(22岁以上)的受试者回答率为77.70%。这些信息提示ODI涉及性生活的条目不适合中国避讳性生活话题的文化环境。但这5项研究中一项^[19]于2017年发表,其余均发表在2010年以前。随着时间推移,腰背痛发病的平均年龄降低,中国人群对于婚前性行为的回避态度与对于性话题更为开放的态度相矛盾,使得如何优化条目8成为新的问题。

本研究存在一定局限。COSMIN发起时间为2005年,COSMIN偏倚风险清单^[30]发表时间为2010年,纳入14篇文献中有8篇是在2010年及之前所发表,根据较新的标准评价其研究方法学质量存在争议。此外,COSMIN系统评价指南对于信度的评价主要针对重测信度,对于分半信度没有给出相关标准,因此对于涉及此测量性能研究^[19]未进行分析。最后,本研究未能获取所有纳入文献所使用的ODI量表,未能详细地对比分析各版本之间差异,仅对2.1中文版ODI的研究证据进行综合,可能遗漏与2.1中文版ODI仅有较小差异ODI的研究证据。

4 结论

本系统评价纳入文献涉及ODI版本繁杂,仅整合2.1中文版的研究证据。结果显示不能推荐其用于中国腰痛临床实践或试验,即使有中等质量的证据证明其内部一致性及重测信度充分、有高质量的证据证明其反应度充分,因为暂无内容效度研究,且结构效度研究不规范。未来的研究应按照规定,填补内容效度及跨文化效度研究的空白,并完善结构效度的研究,以全面准确地判断其是否适合中国人群使用。

参考文献

- [1] GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990—2016: a systematic analysis for the global burden of disease study 2016[J]. Lancet, 2017, 390(10100):1211—1259.
- [2] 刘延青,郑拥军. 中国慢性腰背痛流行病学调查[C]//中国医师协会. 2018中国医师协会疼痛科医师分会年会资料汇编. 2018;22—27.
- [3] Chiarotto A, Deyo RA, Terwee CB, et al. Core outcome domains for clinical trials in non-specific low back pain[J]. Eur Spine J, 2015, 24(6):1127—1142.
- [4] Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain[J]. Pain, 2018, 159(3):481—495.
- [5] 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.
- [6] Prinsen C, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures[J]. Qual Life Res, 2018, 27(5):1147—1157.
- [7] Fairbank J, Couper J, Davies J, et al. The Oswestry low back pain questionnaire[J]. Physiotherapy, 1980, 66(8):

- 271—273.
- [8] Baker D, Pynsent P, Fairbank J. The Oswestry disability index revisited.in: Roland M, Jenner J, eds.Back Pain: New Approaches to Rehabilitation and Education [M]. Manchester: Manchester University Press, 1989: 174—186.
 - [9] Meade TW, Dyer S, Browne W, et al. Randomized comparison of chiropractic and hospital outpatient management for low-back-pain: results from extended follow-up[J]. BMJ, 1995, 311(7001):349—351.
 - [10] Roland M, Fairbank J. The Roland-Morris disability questionnaire and the Oswestry disability questionnaire[J]. Spine, 2000, 25(24):3115—3124.
 - [11] Grotle M, Brox JI, Vøllestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires[J]. Spine, 2005, 30(1):130—140.
 - [12] Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures[J]. Qual Life Res, 2018, 27(5):1171—1179.
 - [13] Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set": a practical guideline[J]. Trials, 2016, 17(1):449.
 - [14] 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.
 - [15] GRADE. GRADE Handbook: Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. 2013. <https://gdt.gradeapro.org/app/handbook/handbook.html>
 - [16] 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.
 - [17] Yao M, Xu BP, Li ZJ, et al. A comparison between the low back pain scales for patients with lumbar disc herniation: validity, reliability, and responsiveness[J]. Health Qual Life Outcomes, 2020, 18(1):175.
 - [18] Lee CP, Fu TS, Liu CY, et al. Psychometric evaluation of the Oswestry Disability Index in patients with chronic low back pain: factor and Mokken analyses[J]. Health Qual Life Outcomes, 2017, 15(1):192.
 - [19] 程继伟,王振林,刘伟,等.Oswestry功能障碍指数的改良及信度和效度检验[J].中国脊柱脊髓杂志,2017,27(3):235—241.
 - [20] 方贤聪,程继伟,盛锡华.Oswestry功能障碍指数评定腰椎滑脱症患者的反应度和最小有意义变化研究[J].中国现代医生, 2017,55(28):16—20.
 - [21] Lu YM, Wu YY, Hsieh CL, et al. Measurement precision of the disability for back pain scale- by applying Rasch analysis[J]. Health Qual Life Outcomes, 2013, 11: 119.
 - [22] Ma C, Wu S, Xiao L, et al. Responsiveness of the Chinese version of the Oswestry disability index in patients with chronic low back pain[J]. Eur Spine J, 2011, 20(3): 475—481.
 - [23] Ernest Volinn 1, Bangxiang Yang, Jing He, et al. West china hospital set of measures in Chinese to evaluate back pain treatment[J]. Pain Medicine, 2010,11(5): 637—647.
 - [24] 俞红,白跃宏.简式中文版Oswestry功能障碍指数评定下背痛患者的信度及效度分析[J].中华物理医学与康复杂志,2010, (2):125—127.
 - [25] 刘绮,马超,伍少玲,等.Oswestry功能障碍指数评定慢性腰痛患者的效度分析[J].中国康复医学杂志,2010,25(3):228—231.
 - [26] 刘晖.腰痛患者的临床评估研究[D].第四军医大学,2010.
 - [27] 刘绮,麦明泉,肖灵君,等.中文版Oswestry功能障碍指数评定慢性腰痛患者的反应度研究[J].中国康复医学杂志,2010,25 (7):621—624.
 - [28] 谭柯,郑敏,杨邦祥,等.欧氏失能指数用于中国四川腰背痛患者功能测量的评价研究[J].四川大学学报(医学版),2009,40 (3):559—561.
 - [29] 郑光新,赵晓鸥,刘广林,等.Oswestry功能障碍指数评定腰痛患者的可信性[J].中国脊柱脊髓杂志,2002(1)13—15.
 - [30] Chow JH, Chan CC. Validation of the Chinese version of the Oswestry Disability Index[J]. Work, 2005, 25(4): 307—314.
 - [31] Chiarotto A, Ostelo RW, Boers M, et al. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain[J]. J Clin Epidemiol, 2018, 95:73—93.
 - [32] Liu H, Tao H, Luo Z. Validation of the simplified Chinese version of the Oswestry Disability Index[J]. Spine, 2009, 34(11):1211—1216.
 - [33] 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.