

· 临床研究 ·

原发性骨质疏松症 PRO 量表的科学性考核

孙友强¹ 邵敏^{2*} 何挺³ 曾振明⁴ 姜涛¹ 汪钦生² 王茂澎⁵ 刘庆思² 陈雷雷⁶ 何伟⁶

1. 广州中医药大学, 广州 510405
2. 广州中医药大学第三附属医院, 广州 501240
3. 广州市荔湾区骨伤科医院, 广州 510140
4. 深圳市宝安区中医院, 深圳 518133
5. 佛山市顺德中医院, 佛山 528333
6. 广州中医药大学第一附属医院, 广州 510405

中图分类号: R681 文献标识码: A 文章编号: 1006-7108(2016) 08-1008-04

摘要: **目的** 对原发性骨质疏松症患者报告的临床结局(patient-reported outcomes, PRO)量表的科学性进行考核评价。方法 重测 52 位原发性骨质疏松症患者及 92 名健康者,运用经典测试理论和现代项目反应理论等方法对研制出的量表的难度、区分度、信度、效度和可行性——分析评价。**结果** 项目反应理论(item response theory, IRT)分析显示量表的难度和区分度均在合格的参数范围之内,表明量表具有较好的难度和区分度,信度方面:总量表的克朗巴哈系数(Cronbach's α)为 0.801,疾病主要症状、全身表现、心理维度、社会关系维度、日常活动能力方面、满意度方面的 Cronbach's α 分别为 0.907、0.780、0.849、0.807、0.799、0.877,具有较好的内部一致性。重测信度 Pearson 相关系数为 0.925, Kendall 相关系数为 0.759,提示良好的重测信度。效度分析:结构效度结果说明量表中多维度的测量满足专业上的预想结构。可行性分析:抽样调查原发性骨质疏松症患者,量表总完成率为 90.34%,作答时间约 15 分钟。**结论** 经过经典统计检验和项目反应理论的科学性考核,OPPRO 量表具有较好的难度、区分度、信度、效度和可行性。本研究步骤标准严谨,具有充分的中华文化特色,研究范围来源广泛,方法学和报道质量较高,具有良好的创新性和科研及临床意义,可用于原发性骨质疏松症患者的临床疗效评价。

关键词: 原发性骨质疏松症;患者报告的临床结局;量表;科学性考评;项目反应理论

The scientific quality of patient-reported outcome instrument for primary osteoporosis

SUN Youqiang¹, SHAO Min^{2*}, HE Ting³, ZENG Zhenming⁴, JIANG Tao¹, WANG Qinsheng², WANG Maopeng⁵, LIU Qingsi², CHEN Leilei⁶, HE Wei⁶

1. Guangzhou University of Chinese Medicine, Guangzhou 510405
2. The Third Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 501240
3. Guangzhou Liwan Orthopaedic Hospital of Chinese Medicine, Guangzhou 510140
4. Shenzhen Bao'an Traditional Chinese Medicine Hospital, Shenzhen 518133
5. Foshan Shunde Hospital of Chinese Medicine, Foshan 528333
6. The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510405, China

Corresponding author: SHAO Min, Email: shaomin98@aliyun.com

Abstract: **Objective** To evaluate the scientific quality of the patient-reported outcome instrument for primary osteoporosis. **Methods** We retested 52 primary osteoporosis patients and 92 healthy volunteers to evaluate the difficulty, discrimination, reliability, validity and feasibility of the OPPRO instrument by using the classical test theory and modern item response theory methods. **Results** The IRT analysis showed that the difficulty and discrimination of the OPPRO instrument are within acceptable parameters. For reliability, the Cronbach's coefficient of the whole instrument is 0.801, the coefficient of the main symptoms aspect of the instrument is 0.907, the systemic performance aspect 0.780, the psychological aspect 0.849, the social aspect 0.807, the performance in daily activities aspect 0.799, and the performance satisfaction aspect 0.877. All of these results represented a good

基金项目:2014 广东省科学技术厅-广东省中医药科学院联合科研专项(2014A032500005)

* 通讯作者:邵敏,Email:shaomin98@aliyun.com

internal consistency of the instrument. The Pearson correlation coefficient of the retest reliability of the instrument is 0.925, the Kendall correlation coefficient is 0.759, both of them suggests a good retest reliability. For validity, the structural validity of the results indicates that the multidimensional structure of the instrument meets to our professional expectations. For the feasibility analysis, the sampling survey of the primary osteoporosis patients showed that the total completion rate of the instrument is 90.34% and the time required to complete is about 15 minutes. **Conclusion** By using the classical statistical testing and the item response theory to analyze the items of the patient-reported outcomes of primary osteoporosis, we found that the OPPO instrument has good difficulty, discrimination, reliability, validity and feasibility. The study was conducted strictly in standard steps and took consideration of Chinese culture. The scope of the research, the methodology and the reporting methods are of a high quality, which represents good innovation, scientific quality and as well as clinical significance. The OPPO is a useful tool for evaluating the clinical outcomes in patients with primary osteoporosis.

Key words: Primary osteoporosis; Patient-reported outcomes; Instrument; Evaluating the Scientific quality; Item response theory

1 研究背景

目前,一种疾病的疗效评价往往会用到综合性的评价方法,量表作为一个可以比较全面反映患者的整体临床疗效的评价工具而被广大临床医生所青睐^[1]。一个好的量表,不仅限于其实用性,其内在结构的良好信度和效度是其长期存在的根本保证。本研究在课题组之前的原发性骨质疏松症患者报告的临床结局(patient-reported outcomes, PRO)量表的初步研制的基础上,运用经典统计方法对原发性骨质疏松症 PRO 量表的科学性进行考核,明确其信度和效度,以便确立其应用于原发性骨质疏松症患者的临床疗效评价的作用。

2 研究方法

重新收集 52 例患者及 92 例健康者,让患者和健康者分别填写原发性骨质疏松症 PRO 量表,回收量表,提取数据,建立数据库并进行信度和效度分析。采用 SPSS20.0 软件,运用经典的 Cronbach's 系数法、Pearson 相关系数、Kendall 相关性系数、两样本 *T* 检验对 OPPO 量表进行总体和各维度信度、区分度、效度和可行性进行分析;用 Multilog7.03 软件的 IRT 参数分析方法对量表的难度和区分度进行考核。

2.1 纳入标准

大于等于 45 周岁;符合原发性骨质疏松症诊断标准且未合并其他重大内科疾病;自我报告时能够用汉语口头交流和阅读;同意签署知情同意书。

骨质疏松症诊断标准:参照中华医学会、中国老年学学会骨质疏松委员会编著的《中国人骨质疏松症诊断标准专家共识(第三稿·2014 版)》诊断标准^[2];骨密度(bone mineral density, BMD)检测: -2.0SD 以下或者骨量下降 25% 以上。

2.2 排除标准

不符合上述纳入标准,且存在听力、视力、智力、精神障碍等。

2.3 统计方法

采用 SPSS20.0 软件对 OPPO 量表所得数据进行经典测试统计学分析, $P < 0.05$ 表示有统计学意义,相关系数及克朗巴赫系数取值应大于 0.7 表示相关性和内部一致性良好。用 Multilog7.03 软件对量表的难度进行分析,取值范围在 -3.00 ~ +3.00 表示条目的难度及区分度良好。

3 结果

3.1 信度分析

本课题的信度从同质性分析和重测信度分析两方面展开。同质性信度分析可考察测量项目的内部结构^[3]。对于采用多级计分的 PRO 量表,适合采用目前最常用的评价信度系数的方法:克朗巴赫阿尔法系数来评价,计算公式为:

$$r_{\alpha} = \left(\frac{K}{K-1} \right) \left(1 - \frac{\sum_{i=1}^K S_i^2}{S_t^2} \right)$$

式中, k 为量表条目数, S_i^2 为第 i 个条目得分的方差, S_t^2 为量表总得分的方差。一般用克朗巴赫系数表示,且系数应达 0.7 以上才算比较好。分别统计每个维度的克朗巴赫内部一致性以及总的克朗巴赫内部一致性, OPPO 量表总的克朗巴赫系数为 0.856,提示有良好的内部一致性。疾病临床表现、全身相关表现、心理、社会关系、日常活动能力、满意度等各维度的克朗巴赫系数分别为 0.907、0.780、0.849、0.807、0.799、0.877,均超过 0.7,显示量表的各维度也有良好的内部一致性。

重测信度分析法是从稳定性的角度进行条目筛选^[3]。对每个对象先后测量两次,时间相隔两周,

计算每一条目两次得分的相关系数,保留相关系数高的条目。本次研究笔者对之前进行过调查的 380 名患者两周后随机抽查了 52 名患者进行回顾调查,对 OPPro 量表进行重测,52 名患者男女比例为 1:3,年龄为(67±27.6)岁,与其之前调查的 380 例患者资料基本一致,二者无资料统计学差异($P>0.05$),重测结果见表 1、2。

表 1 重测患者的一般资料
Table 1 General characteristics of the retested patients

性别	病例数(n)	年龄(岁)	总平均分
男	13	71.5±26.5	2.47±0.59
女	39	63.2±25.4	2.69±0.66

表 2 重测患者治疗前后的相关性
Table 2 Pre and post treatment correlation in the retested patients

时间	均值	标准差	Pearson 相关性	Kendall 相关性	P 值
治疗前(n=380)	2.8561	0.3129	0.925**	0.759**	0.0428
治疗后(n=52)	2.7545	0.3427			

注:**在 0.01 水平(双侧)上显著相关。

经表 2 的数据分析结果得出结论:OPPro 量表重测前后的均数和标准差存在显著差异,分析得出 pearson 和 kendall 系数均超过 0.75,说明 OPPro 量表具有良好的重测信度。

3.2 效度分析

本课题从内容效度、结构效度两个方面进行考核。内容效度是指一个特定的条目对所属内容范畴的反映程度,即量表所选的条目能在多大程度上代表所要测的内容^[4]。本课题在量表编制前期查阅大量国内外相关文献,并参考国内外骨质疏松症相关量表,咨询专家组意见,同期对患者进行了面对面的访谈,综合以上所收集的信息对量表条目的提出、修改进行了反复深入的论证以保证量表条目良好的内容效度。

结构效度即从临床和常识的角度考察量表的维度是否符合其预想的结构^[4]。本量表在前期的研制过程中对量表的结果进行了严格的设计,并经过专家的严格把关层层筛选考核,最终的量表在结构上布局合理,具有良好的结构效度。

区分度是指调查量表能够测出的调查对象对某测定指标微小改变的能力。本课题分析调查量表对患者(病例组 52 例)和健康人(对照组 92 例)的区分能力。由于本次调查的对照组均选自医院内外的原发性骨质疏松症患者,不存在认知、听力和沟通障

碍,故调查量表应区分病例组及对照组两组调查对象在 6 个维度的平均得分的差别,采用完全随机设计两样本 t 检验, $P<0.05$ 认为差异有统计学意义,统计结果见表 3。

表 3 健康者与患者测试的区分度 t 检验
Table 3 The t -test of the discrimination of the instrument between healthy volunteers and osteoporosis patients

条目	病例组均值 (n=52)	对照组均值 (n=92)	t/t'	P
疾病临床表现	3.13±0.52	1.34±0.26	27.5298	0.00
全身相关表现	3.33±0.46	1.79±0.33	23.2461	0.00
心理	2.50±0.51	1.28±0.25	19.2463	0.00
社会关系	3.11±0.56	1.89±0.22	18.5534	0.00
日常活动能力	3.13±0.66	1.43±0.16	23.5681	0.00
满意度	2.72±0.58	2.06±0.45	7.5993	0.00

表 3 的分析数据显示,该组患者与健康者在各个维度的均值和标准差均存在显著的统计学差异,即该调查量表具有较好的区分度。

3.3 可行性分析

可行性是用来反映调查量表的可接受程度。通常用调查量表的回收率和作答时间来衡量。本课题大规模调查在 8 所医疗机构共发放量表 580 份(患者 410 份,健康人 110 份,重测患者 60 份),共回收量表 524 份(患者 380 份,健康人 92 份,重测患者 52 份),回收率为 90.34%(患者 92.68%,健康人 83.63%,重测患者 86.67%),无论是调查患者、健康人还是重测患者,调查量表的回收率均高于 83%,提示调查量表具有高的接受率。本课题对 52 名调查对象的作答时间进行了记录,均可 15 分钟左右完成调查量表的作答。

3.4 IRT 参数分析

IRT 主要应用于心理和教育测量中,现在已广泛应用于量表研究的多个方面^[7]。在本研究中,难度及区分度参数值用来鉴别条目的好坏,好的条目取值范围通常为-3.00~+3.00^[8]。本研究的 52 例患者的参数在-2.97~3.6 之间,也提示量表整体有良好的难度评测能力。OPPro 量表 IRT 参数分析结果如表 4。

4 结论与展望

经过经典统计检验和项目反应理论的科学性考核,OPPro 量表具有较好的难度、区分度、信度、效度和可行性,可用于原发性骨质疏松症患者的临床疗效评价。

当然,本研究也不可避免地存在一些局限性,如

表 4 OPPRO 量表 IRT 参数分析
Table 4 The IRT parameter analysis of the
OPPRO Instrument

条目	区分度	难度				
	A	B1	B2	B3	B4	
腰背疼痛	3.60	-1.95	-1.00	-0.16	0.61	
全身骨痛	1.00	-1.39	-0.41	0.41	1.39	
驼背畸形	1.82	-2.25	-0.62	0.57	1.86	
胸背疼痛	1.00	-1.49	-0.41	0.41	1.39	
胁肋部疼痛	1.00	-1.39	-0.41	0.41	1.39	
肌肉抽搐	0.86	-2.14	0.41	1.08	2.40	
身高变矮	1.00	-1.39	-0.41	0.41	1.39	
肌肉萎缩	1.19	-0.34	0.73	1.68	2.911	
近 5 年发生过骨折	1.00	-1.39	-0.41	0.41	1.39	
关节屈伸不利	2.23	-1.88	-0.29	0.39	1.95	
肢体无力	0.95	-0.7	1.07	1.51	2.62	
行动迟缓	1.00	-1.39	-0.41	0.41	1.39	
弯腰困难	2.80	-1.86	-0.81	0.15	1.65	
腰膝酸软	2.42	-1.54	-0.53	0.56	1.87	
脊柱侧弯畸形	1.00	-1.39	-0.41	0.41	1.39	
精神差	2.95	-1.84	-0.78	0.34	1.45	
疲劳	1.00	-1.39	-0.41	0.41	1.39	
形体瘦弱	1.83	-1.77	-0.59	0.69	2.03	
性功能差障碍	1.00	-1.39	-0.41	0.41	1.39	
更年期综合征明显	0.89	-2.52	-0.43	1.85	2.19	
怕冷	1.00	-1.29	-0.31	0.40	1.36	
小便多	1.97	-1.88	-0.73	0.33	1.76	
健忘	1.45	-2.08	-0.55	0.92	2.88	
失眠多梦	1.00	-1.39	-0.41	0.41	1.39	
孤独	0.92	-1.79	0.34	1.44	2.49	
对生活失去信心	1.00	-1.39	-0.41	0.41	1.39	
易受小事情干扰	2.76	-1.82	-0.67	0.39	1.75	
沮丧	0.81	-1.29	-0.51	0.52	1.58	
焦虑	1.94	-1.64	-0.57	0.61	1.87	
不能独自出行	1.31	-2.60	-0.8	1.01	2.45	
不能做自己喜欢的运动	1.21	-1.69	-0.71	0.65	1.78	
需家人全程照顾	0.99	-2.21	-0.88	1.37	2.49	
很依赖别人	1.00	-1.39	-0.41	0.41	1.39	
常需要住院治疗	1.28	-1.83	-1.20	1.08	2.95	
消耗很多医疗费用	1.32	-2.39	-0.21	0.86	3.00	
上卫生间困难	1.00	-1.39	-0.41	0.41	1.39	
不能做家务	0.94	-1.99	0.28	1.39	2.50	
不能和小孩玩耍	1.00	-1.39	-0.41	0.41	1.39	
不能爬 1~3 层楼梯	0.97	-2.70	-1.26	1.71	2.98	
不能提 10 kg 重物	1.00	-1.39	-0.41	0.41	1.39	
不能步行 500 米以上	0.87	-2.97	0.51	1.83	2.16	
对医疗水平不满意	0.78	-2.79	2.21	4.67	2.39	
对治疗疗效不满意	1.00	-1.39	-0.41	0.41	1.39	
不满自己的健康状况	0.79	-2.88	0.48	0.71	1.965	

验证的量表样本量相对较小,仍需要进一步扩大数据量进行定量和定性分析,还可以加入更多的指标来验证其信度和效度。期望以后可以继续大样本多中心的临床研究以进一步完善该量表,以便更好地运用于原发性骨质疏松症患者的临床疗效及生存质

量评价。下一步研究可在本研究的基础上,直接利用本次研究结果,可与其它生活量表进行对比研究,确立其有效性。下一步还可以建立“基于计算机自适应测试的原发性骨质疏松症 PRO 量表”工作场,然后将各个条目逐条纳入并进行参数界定,最终建立基于计算机自适应测试的原发性骨质疏松症 PRO 量表,以期进一步完善 OPPRO 量表,更好地用于原发性骨质疏松症患者的临床疗效评价。

【参 考 文 献】

[1] Greenhalgh J, Long AF, Flynn R. The use of patient-reported outcome measures in routine clinical practice: lack of impact or lack of theory?. Soc Sci Med, 2005, 60(4): 833-843.

[2] 张智海, 刘忠厚, 李娜, 等. 中国人骨质疏松症诊断标准专家共识(第三稿, 2014 版). 中国骨质疏松杂志, 2014, 20(9): 1008.

Zhang ZH, Liu ZH, Li N, et al. Expert consensus on the diagnosis of osteoporosis in Chinese Population (the third edition, 2014). Chinese Journal of Osteoporosis, 2014, 20(9): 1008. (in Chinese)

[3] 郝元涛, 孙希凤, 方积乾, 等. 量表条目筛选的统计学方法研究. 中国卫生统计, 2004, 21(4): 18-20.

Hao YT, Sun XF, Fang JQ, et al. The statistical methods research of item screening. China Health Statistics, 2004, 21(4): 18-20. (in Chinese)

[4] 王鹏. 计算机模拟心理咨询病例系统的编制及现代测量学研究. 南昌: 江西师范大学, 2011.

Wang P. The establishment of the psychological consulting case system by computer simulation and the modern study of measurement. Nanchang: Jiangxi normal university, 2011. (in Chinese)

[5] 张岩波. 潜变量分析. 北京: 高等教育出版社, 2009: 16.

Zhang YB. Latent variables analysising. Beijing: Higher Education Press, 2009: 16. (in Chinese)

[6] 孙振球. 医学统计学. 北京: 人民卫生出版社, 2010: 660.

Sun ZQ. Medical statistics. Beijing: People's Medical Publishing Press, 2010: 660. (in Chinese)

[7] 刘仁刚. 条目反应理论简述. 中国临床心理学杂志, 2009, 17(1): 37-41.

Liu RG. Briefly introduce of Item response theory. China Clinical Psychology, 2009, 17(1): 37-41. (in Chinese)

[8] 林岳卿, 方积乾. 多维 IRT 与单维 IRT 在多维量表中的应用的差异. 中国卫生统计, 2011, 28(3): 226-228.

Lin YQ, Fang JQ. The difference of the applications of the multidimensional IRT and unidimensional IRT in the multidimensional instrument. China Health Statistics, 2011, 28(3): 226-228. (in Chinese)

(收稿日期: 2015-12-18)