

试论效度在病人报告的临床结局 评价量表中的应用*

□张艳宏 (中国中医科学院中医临床基础医学研究所 北京 100700)

刘保延** (中国中医科学院)

何丽云 (中国中医科学院中医临床基础医学研究所 北京 100700)

訾明杰 (中国中医科学院针灸研究所)

摘要:引入了病人报告的临床结局(PRO)的概念,介绍了通过 PRO 量表测得患者自觉感受的病情变化来评定疗效的方法,论述了评价量表性能的重要指标—效度的作用,并以在研的“基于中风痉挛性瘫痪患者报告的临床结局评价量表”为例,探讨了内容效度、结构效度、标准效度和实证效度的应用。

关键词:临床评价 中医效度 中医量表 病人报告的临床结局(PRO)

随着病人报告的临床资料在临床评价中的作用日益受到关注,病人报告的临床结局(Patient Reported Outcomes,PRO)的理论框架、测量方式等方面的研究逐步展开,量表(Instrument)这种形式作为临床评价 PRO 的工具已被国际医学专家接受并得到广泛应用^[1]。由于传统中医药的诊疗方式以医者经验为主,注重患者的自觉症状,通常以患者自觉症状的改善为导向判断疾病的向愈。因此,PRO 量表的研制已成为目前中医药临床评价领域中的研究热点。那么,如何研制结构合理、性能较高的 PRO 量表呢?量表研制过程是一个有目的的、合乎逻辑的过程,研究者应严格遵循量表研制的各个步骤,构建可信、有效的测量,确保得到理想的测量结构。

量表的效度(Validity)是评价量表性能的重要指

标之一,它反映了量表设计者的意图能否让被测者理解,量表能否有效地测量设计者预先打算测定的内容,是否达到了鉴别、评价或预测的目的,也反映了测定结果与预想结果的符合程度^[2]。评价量表效度的种类很多,如内容效度、结构效度、标准效度、实证效度等。下面以在研的“基于中风痉挛性瘫痪患者报告的临床结局评价量表”(以下简称“中风 PRO 量表”)为模型,探讨各种类型的效度应用情况。

一、量表简介

中风痉挛性瘫痪患者肢体功能恢复是一个漫长的过程,神经生理及生物力学等客观指标变化可能与痉挛患者自觉症状及其日常生活能力不一定相关^[3]。因此,我们引入 PRO 的概念,以量表的形式测量患者自觉感受到的病情变化,作为疗效评定的一种方式。研究小组参照 2006 年 2 月美国食品药物

收稿日期:2008-12-24

修回日期:2009-01-06

* 国家科技部“973”计划项目(2006CB504601):辨证论治临床评价基本原理、方法和技术平台研究,负责人:刘保延;国家科技部基础平台建设项目(2004DEA71040):中医临床疗效评价标准,负责人:刘保延。

** 联系人:刘保延,教授,主任医师,主要研究方向:中医临床评价方法学以及针灸的基础与临床研究,E-mail:liuby@cintcm.ac.cn。

监督管理局发布的关于 PRO 研究应用于临床药物研制和疗效评价的指南草案^[4]中对 PRO 量表研制、评价的具体说明,从概念操作化和条目筛选,条目池到建立、临床预调查,严格遵照量表学的原理和方法进行量表的设计,最后形成包含 19 个条目,3 个领域(躯体症状、心理状态、社会交往)的自评量表。

对象:北京市 5 家医院中风痉挛性瘫痪患者 106(男 64,女 42)名,平均年龄 62.5 岁,其中,首次发病 83 例,多次发病 23 例。方法:采用一对一现场调查的方式。所有调查员均在施测前接受培训,由研究者向调查员说明研究的目的、纳入测试患者的标准和施测方法。本量表为自评量表,施测时将量表发给被试者,调查员宣读指导语,由被试者独立作答,有不理解的地方可向主试询问;如果被试者存在失语、痴呆或利手侧瘫痪,由调查员将量表内容逐条念给被试者,根据被试者提供的答案信息,调查员协助被试者作答。填写量表同时要测量简易精神状态检查量表(MMSE)、抑郁自评量表(SDS)、Ashworth 痉挛量表和脑卒中专用生活质量量表(SS-QOL)。另用该量表调查 24 例中风后弛缓性瘫痪患者,以研究量表的区分效度。统计学处理:将原始数据输入计算机并编制程序进行自动评分,用 SPSS 11.5 进行相关统计学分析。

二、效度分析

1. 内容效度

内容效度指要判断的量表特征与量表内容相符合的程度。典型的问题是量表在测量内容上是否与预先测量内容相一致。内容效度的建立依赖于两个条件;一个条件是内容范围明确,另一个条件是取样有代表性^[5]。提高内容效度有以下几方面:①量表条目协调;②由专业人员组成的专家各自对量表测量内容进行重要性评估。内容效度依赖于评估者的主观判断,因此,与其他效度评估方法相比存在的主观性偏差更大。

中风 PRO 量表研究的维度设计来自于严密的理论构思和实践调查,项目的内容取自开放式访谈调查所获得的素材,然后对访谈内容整理分析,通过患者评价和专家评议,初步获得了量表评价的指标体系,该指标体系基本上反映了中风后痉挛性瘫痪患者的实际情况。在此基础上,制定了原始项目,经神经科专家审定各个项目的内容以及词语表达的准确

性,通过反复患者试测,校正各项的表达方式是否恰当。在正式施测前,研究小组对量表的结构和内容进行评估。从编制的过程来看,中风 PRO 量表应当具有较高的内容效度。

2. 结构效度

结构效度是最强有力的效度评价指标,也是我们研制量表时最为关注的内容。结构效度指研究中所采取的测量能够合理表达潜在理论结构的程度,是最基本的效度概念。“测试最重要的特性在于结构效度”^[6],它决定了量表的性质,关系到我们依据量表测评结果的解释是否恰当而有意义。既然结构效度是反映量表能够测量到理论上的结构或特质的程度,那么只要能够阐明所考虑的特质的性质,以及影响该特质发展和表现的条件,都是结构效度的合适证据。分析结构效度的具体方法有多种,下面以因子分析、聚类分析及鉴别度分析为例说明。

(1)因子分析。

因素分析的基本思想是,根据测验的相关系数表把变量分组,使得同一组的变量间的相关较高,不同组的变量间的相关较低,每组变量就代表一个因子即基本结构;通过考察测量结果的结构与量表预想结构设计是否相符,来验证假设是否成立。

①进行变量的线性检验。因子分析要求观测变量之间呈线性关系,因此(Kaiser - Meyer - Olkin)在进行因子分析前应进行变量的线性检验。KMO 和巴特利特(Bartlett)球体检验结果见表 1。巴特利特球度检测观测值为 1382.011,相应的概率接近 0,认为相关系数矩阵与单位阵有统计学差异。KMO 值为 0.827,表明数据呈较好的线性,可以进行因子分析。

表 1 KMO 测度和 Bartlett 球体检验

Kaiser - Meyer - Olkin 取样充分度 Adequacy.	0.827
Bartlett's Test of Sphericity	Approx. Chi-Square 1382.011
	df 171
	Sig. 0.000

②选用主成分分析抽取公共因素,求得初始载荷矩阵,特征根值大于 1 的因素共有 5 个,解释总变异的 75.155%(见表 2),得到碎石图(见图 1)。然后用方差最大法正交旋转,得到旋转因子载荷矩阵(见表 3)。

本研究根据以下标准确定因素数目:①因子的

表 2 各因素的特征值及解释总变异百分比

因素	特征值	解释总变异百分比(%)	累计百分比(%)
1	7.119	37.468	37.468
2	3.095	16.291	53.759
3	1.773	9.331	63.090
4	1.175	6.185	69.274
5	1.117	5.881	75.155

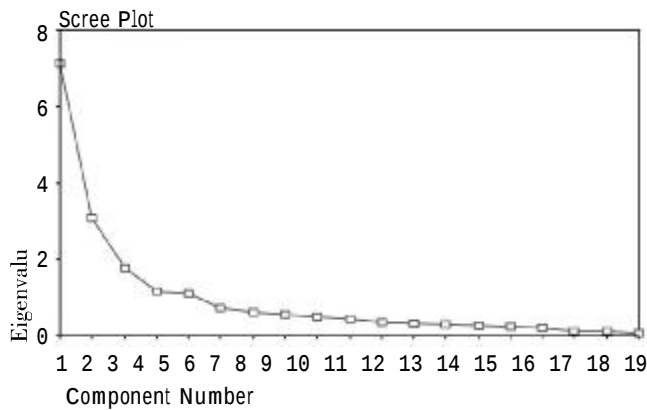


图 1 因子碎石图

表 3 旋转因子载荷矩阵

项目	因子				
	1	2	3	4	5
疼痛	0.745				
麻木	0.826				
僵硬	0.888				
沉重	0.832				
颤抖	0.513				0.691
无力	0.589				
肿胀	0.549				0.631
抽筋	0.458				0.671
出汗					0.711
想哭		0.671			0.463
着急		0.844			
发火		0.869			
低落		0.756			
信心	0.595				-0.540
担心					0.828
日常活动				0.722	
兴趣爱好				0.921	
人际交往				0.887	
经济影响				0.681	

特征值 ≥ 1 ,即因子贡献率 ≥ 1 ;②因子必须符合卡特尔“陡阶”检验原则;③每个因子至少包含三个项目;④因子比较容易命名^[7]。根据以上原则,第 5 因子中

的信心、担心两个变量合并到代表心理变量的因素中。这样,用 4 个因子解释所有的变量,丢失的信息最少。

由表 3 分析各因素所含条目的内容:第一个因子包括疼痛、麻木、僵硬、沉重、无力 5 个条目,涉及的主要是症状中自我报告的主观感觉的问题,可命名为“主观症状”;第二个因子包括想哭、着急、发火、低落、信心、担心 6 个条目,涉及的主要是与心理相关的问题,可命名为“心理感受”;第三个因子包括日常活动、兴趣爱好、人际交往、经济影响 4 个条目,涉及的主要是与社会相关的问题,可命名为“社会交往”。第四个因子包括颤抖、肿胀、抽筋、出汗 4 个条目,涉及的主要是症状中可知觉的相关问题,可命名为“客观症状”。这与预想的三个维度实际上是一致的,因子 1 和 4 是症状维度的两个方面。因子分析结果表明,量表结构清晰,具有良好的结构效度。

(2) 聚类分析。

聚类分析是按“物以类聚”的原则研究事物分类的一种多元统计分析方法。本研究对 106 例有效数据进行指标聚类,以验证因子分析结果。从图 2 分析可知,着急、发火、低落、想哭可聚为一类;兴趣爱好、人际交往、日常活动、经济影响可聚为一类;僵硬、沉重、疼痛、麻木可聚为一类;颤抖、肿胀、抽筋聚为一类;无力、出汗、担心、信心各聚一类。除无力、出汗、担心、信心项目外,其他项目的聚类结果与上述因子分析结果大体一致。

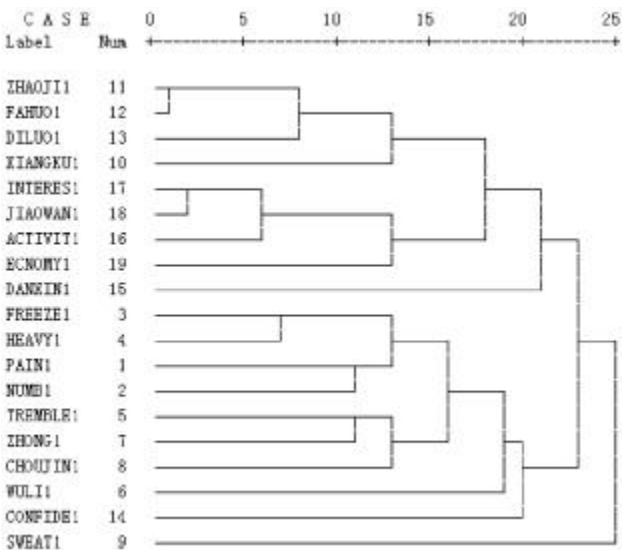


图 2 树行聚类分析图

(3) 鉴别度分析。

项目的鉴别度也叫区分度,是指测验项目对于所研究的心理特性的区分程度或鉴别能力。采用计算每个项目和量表总分之间的相关系数作为项目区分度指标^[8]。本研究采用 Spearman 相关系数法,取区分度指数在 0.3 以上的项目。从表 4 可知,量表具有很好的项目区分度,所有项目的区分度均非常显著, $P < 0.001$ 。

表 4 项目区分度分析结果

项目	系数	项目	系数
疼痛	0.564	想哭	0.453
麻木	0.594	着急	0.569
僵硬	0.664	发火	0.564
沉重	0.603	低落	0.611
颤抖	0.529	信心	0.383
无力	0.492	担心	0.346
肿胀	0.600	日常活动	0.542
抽筋	0.505	兴趣爱好	0.589
出汗	0.249	人际交往	0.660

3. 标准效度

又称标准关联效度,即以公认有效的量表作为标准,检验新量表与标准量表测定结果的相关性,用两种测定工具测得量表总分的相关系数来表示。标准量表的要求是,必须和新量表是同类量表。本量表是针对中风痉挛性瘫痪患者的特异性量表,因此采用 Ashworth 痉挛量表、脑卒中专用生活质量量表(SS-QOL)作该量表的校标。

采用 Kendal 相关系数分析前后 2 次自评量表总分与 Ashworth 量表的相关系数分别为 0.288、0.384;与 SS-QOL 量表相关系数分别为 -0.371、-0.329,表明两次量表总分与 Ashworth 量表、SS-QOL 量表均有显著相关性。自评量表与 Ashworth 量表总分越高提示患者整体健康状况越差,即二者呈正相关;而 SS-QOL 量表总分越高提示患者整体健康状况越好,与自评量表正好相反,即自评量表与 SS-QOL 量表呈负相关。可见,量表的标准效度较好。

4. 实证效度

实证效度的研究主要是对无抑郁患者与抑郁患者采用差异被试比较法。以是否抑郁为分组变量,以各维度得分和量表总分为因变量,进行独立样本的平均数差异检验。

将 106 例样本按抑郁严重度指数在 0.50 以下的

为一组,称为无抑郁组;抑郁严重度指数大于 0.50 的为一组,称为抑郁组。主观症状、社会交往两个维度得分和量表总分呈正态分布,采用单因素方差分析,结果显示抑郁组在“主观症状”、“社会交往”以及量表总分都显著高于无抑郁组;客观症状和心理感受两个维度得分不全呈正态分布,采用非参数秩和检验,结果显示抑郁组在“客观症状”、“心理感受”平均秩次均大于无抑郁组,抑郁组在“客观症状”、“心理感受”得分显著高于无抑郁组。可见,抑郁患者在症状、心理及社会方面的健康状况较无抑郁患者程度重,这同以往的研究结果^[9-10]基本相符。

三、讨论

效度反映测定工具解释预测事实和理论的有效程度,是测量和评价中最基本的要素。本文主要讨论了四种效度:①内容效度属主观指标,意在评价量表内容能在多大程度上覆盖研究目的要求达到的各个方面和领域。②结构效度的内涵就是理论原则,是证明潜在理论有效的一种效度。结构效度的评价依赖于结构理论假设,因而当结构效度被否定时,有三种可能性^[11]:一是理论假设错误;二是理论假设无误,问卷缺乏结构效度;三是两者均有误。③标准效度反映测量结果与标准测量间的接近程度。④实证效度反映所研究的结果与以往相关研究结构的一致性。其他效度评价形式,如区分效度、表面效度、聚敛效度以及结构效度中的结构方程模型等在量表性能评价中也常应用,此不赘述。

某些量表设计者往往将量表效度误认为是量表编制过程中某一特定时点的单一程序,而事实证明,量表效度在量表设计的各个阶段都要考虑,效度应该是一个过程,而不是终点事件,其目的是确保量表编制合理,减少应答偏倚。中风 PRO 量表编制过程的是:明确“临床结局”概念内涵—建立“中风痉挛性瘫痪患者报告的临床结局”结构及域组成—确定亚域—用行为术语对亚域进行操作化—编制条目形式,编制过程的每一步都是对效度的研究。同时还要注意,量表效度评价的成功与否取决于研究者既定的原则。如果某个量表的效度评价不是很理想,并不意味着量表本身的问题,也可能提示作为效度评价的原则基础存在问题。

量表研制过程中,信度和效度都是量表性能评价的主要指标。好的量表首先必须具备很好的信度,

如果信度不高,就无法获得好的效度。有时可以通过提高量表的信度来增加其效度,尽管高信度是高效度的基础,但信度高并不能保证效度就一定好。此外,不同的人群使用同一份已被证实具有较好信度及效度的量表,其结果仍可能有所不同,这提示了量表信度和效度评价在调查研究中的必要性。所有使用量表的调查研究在进行资料的统计分析前,皆应评价其信度和效度,以确保收集到的资料及其统计分析的准确性和可靠性。

病人报告的临床结局研究越来越受到国内外学术界的重视,关于 PRO 的量化测量方法已初步形成规范。近年来,中国中医科学院在科技部、国家中医药管理局的支持下,借鉴国际 PRO 研究方法,结合中医诊疗特点,先后选择中风痉挛性瘫痪、类风湿关节痛、消化性疾病、肝脏疾病、艾滋病等十类疾病,组织近百名临床医生和方法学、统计学专家,研制患者对干预效果自我感受的量化测量量表。量表测试与性能评价已经完成,目前正处于修订完善阶段。研究小组要深入理解信度、效度的内涵及评价方法,每份量表测试都应该有详细的信度、效度说明,保证研究结果真实可靠,便于推广应用,促进临床评价体系的快

速完善,提高中医辨证论治临床疗效的评价水平。

参考文献

- 1 张艳宏,刘保延,刘志顺,等. PRO 与中医临床疗效评价. 中医杂志, 2007, 48(8): 680~682.
- 2 方积乾. 生存质量测定方法及应用. 北京医科大学出版社, 2000, 73.
- 3 张艳宏,刘保延,赵宏,等. 脑卒中痉挛性瘫痪特点及其评定进展. 中国康复理论与实践, 2008, 14(2): 110~112.
- 4 FDA. Guidance for Industry-Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, draft guidance. 2006, 2.
- 5 彭凯平. 心理测验-原理与实践. 华夏出版社, 1989, 149~150.
- 6 聂建中, 汤晓媚. 试论结构效度的发展演变. 山西大学学报(哲学社会科学版). 2006, 29(3): 104-108.
- 7 谢晓庆, 王丽. 因素分析. 中国社会科学出版社, 1989, 95-98.
- 8 漆书青, 戴海崎, 丁树良. 现代教育与心理测量学原理. 高等教育出版社, 2002, 12.
- 9 区丽明, 许俭兴, 谭杰文. 中风后抑郁障碍的研究现状. 中国康复医学杂志, 2001, 16(5): 315~317.
- 10 秦绍森, 胡夏生, 文诗广. 脑卒中后并发抑郁症及其影响因素的研究. 中华医学杂志, 2000, 80(4): 292~293.
- 11 刘朝杰. 问卷的信度与效度评价. 中国慢性病预防与控制. 1997, 5(4): 174~177.

Validity Measures in Patient Reported Outcomes

Zhang Yanhong¹, Liu Baoyan², He Liyun¹, Zi Mingjie³

(1 Institute of Clinical Basic Medicine of China Academy of Chinese Medical Sciences, Beijing 100700 China;

2 China Academy of Chinese Medical Sciences, Beijing China;

3 Institute of Acupuncture and Moxibustion of China Academy of Chinese Medical Sciences)

Validity is an important measure of instrument evaluation, reflecting the validity and accuracy of the measurement. The paper discusses the use of content validity, structure validity, criteria validity, and actual case validity, based on the "self-evaluating instrument based Patient Reported Outcomes provided by apoplexy spastic-paralysis patients".

Keywords: validity; instrument; PRO

(责任编辑:王 瑀, 责任译审:邹春申)