

整合患者导向的测量工具项目编制等距量表实现 ICF 类目用于评定的可操作化

Cieza A, Hilfiker R, Boonen A, Chatterji N, Kostanjsek N, Üstün BT, Stucki G

叶祥明¹, 朱根应¹, 田亮¹译 邱卓英²审校

[摘要] 目的 通过整合不同的患者导向的测量工具的项目来编制 ICF 特定类目的等距量表结构。方法 使用有效的 122 例类风湿关节炎患者的样本资料进行心理测量研究。患者完成 6 个不同的以患者导向的测量工具。将这些测量工具项目的内容与 ICF 类目进行匹配。对项目反应等级进行 Rasch 分析用于检查这些测量工具项目是否说明了 ICF 类目 b130: 能量和驱动力功能, 从而构成一个心理测量可靠的等距量表。结果 19 个项目与 b130 能量和驱动力功能有关。根据 χ^2 检验和 Z 值统计, 19 个项目中的 16 个项目拟合 Rasch 模型(项目和个人的 $Z_{Mean} = 0.451$, $Z_{SD} = 1.085$ 和 $Z_{Mean} = -0.223$, $Z_{SD} = 1.132$)。Person 分离指数 r_{β} 是 0.93。结论 可以构建 ICF 类目等距量表使得单个 ICF 类目可操作化。每个项目的原始格式在等距量表中保持不变。本研究在 ICF 的操作和将来的实施方面迈出了重要的一步。

[关键词] 结局评定; 健康; 心理测量; 分类; 问卷

[中图分类号] N3; R49 **[文献标识码]** A **[文章编号]** 1006-9771(2011)02-0129-08

[本文著录格式] Cieza A, Hilfiker R, Boonen A, et al. 叶祥明, 朱根应, 田亮, 等译. 整合患者导向的测量工具项目编制等距量表实现 ICF 类目用于评定的可操作化[J]. 中国康复理论与实践, 2011, 17(2): 129—136.

1 引言

功能和残疾是普遍的人类体验^[1-2], 是医学^[3]和公共健康^[4]关注的核心。它们在一些领域也是基本相关的, 如劳动、教育和社会事务领域^[5]。

在服务的整个环节以及从急性期医院到社区的健康保健机构的整个过程中, 医学上功能受限的管理弥补了药物和外科治疗^[6]。提高或者保持功能或者预防残疾逐渐成为当前公共健康最迫切需要解决的问题之一^[7]。在劳动、教育和社会事务领域, 如果人们正在体验或者可能体验残疾, 预防行为的计划和实施被认为是惟一可行的^[5]。因此, 在各个专业学科和领域中功能和残疾的概念、分类和测量是非常有意义和重要的。

随着 ICF^[8]的发展, WHO 第一次提供了一个普遍和全球公认的架构和分类来描述可能受健康状况影响的人类功能和残疾的全部领域^[9]。ICF 模型确定了功能领域的 3 个成分即身体功能和结构、活动、参与。这 3 个成分出现问题或困难称为损伤、活动受限和参与局限性, 也就是说它们是残疾领域的组成成分。功能和残疾领域均受健康条件和背景因素(环境因素和个人因素)间相互作用的影响。

身体功能和结构、活动和参与及环境因素成分(有关的个人因素仍有待进一步研究和发展)基于 ICF 类目进行分类。ICF 总共包括 1424 个类目, 类目之间相互排斥并且按照等级式的结构, 结构分为 4 个不同的水平。ICF 类目以独特的字母数字编码表示, 使用这些字母数字编码在个体和人群水平对功能和残疾进行分类、测量和描述是可能的。

由于 ICF 类目总是伴随一个简短的定义、包括和不包括, 因此相应的, 基于 ICF 类目功能方面的信息可以明确地报告并

进行比较^[10-11]。表 1 列举了 ICF 类目的定义、包括和不包括。下面是一个等级式结构的例子:

“b1 精神功能”(第一级/章水平)

“b130 能量和驱动力功能”(第二级水平)

“b1301 动机”(第三级水平)

表 1 ICF 类目相应的编码、类目名称和定义的例子

b130 能量和驱动力功能

驱使个体以持久的方式为满足特殊需要和总目标而不懈追求的生理和心理机制的一般精神功能。

包括: 能量水平、动机、食欲、成瘾(包括可能导致滥用成瘾物质)以及冲动控制的功能

不包括: 意识功能(b110); 气质功能(b126); 睡眠功能(b134); 心理运动功能(b147); 情绪功能(b152)

b280 痛觉

对预示身体某处受到潜在或实际损害而感到不舒服的感觉。

包括: 在身体一处或多处的全身性或局部性疼痛、皮肤疼痛、刺疼、灼疼、钝疼、疼痛; 如肌疼、痛觉缺失和痛觉过敏的损伤

s730 上肢的结构

d450 步行

靠脚在地面一步步走动, 总是一只脚在地面, 如漫步、踱步、向前、后或两侧行走。

包括: 短距离或长距离步行; 不同地面步行; 绕障碍步行

不包括: 移动自身(d420); 到处移动(d455)

d920 娱乐和休闲

参与任何形式的游戏、娱乐或休闲活动, 如非正式或有组织的游戏和运动、体育健身、休闲、消遣娱乐、参观画展、博物馆、去电影院或剧场; 参加手工艺或业余爱好活动、休闲阅读、演奏乐器、观光、旅游和娱乐旅行。

包括: 游戏、运动、艺术和文化、手工艺、业余爱好和社会活动

不包括: 驾驭动物作为交通工具(d480); 有报酬的就业(d850)或无报酬的就业(d855); 宗教和精神性活动(d930); 政治生活和公民权

译者单位: 1. 浙江省人民医院, 浙江杭州市 310014; 2. 中国康复研究中心, 北京市 100068。译者简介: 叶祥明(1966-), 男, 浙江宁波市人, 副主任医师, 主要研究方向: 脑血管病、脊髓损伤康复。

(d950)

e1101 药品

为医疗目的而收集、加工或制造的任何天然或人造的物品,如对抗疗法和自然疗法的药物。

注:字母 b 指身体功能;s:身体结构;d:活动和参与维度;e:环境因素。

原则上讲,有两个方法可以测量某一特定的 ICF 类目,例如对变化的范围进行量化。第一种方法是使用所谓的 ICF 限定值作为专家等级评定量表,量值 0~4(表 2)。使用这种方法,损伤、活动受限、参与局限性和背景因素可以根据已经建立好的编码指南直接进行评价^[10]。然而和任何等级量表一样,专家可以使用任何可得到的信息来源^[12]。

表 2 WHO 提供的 ICF 限定值的百分等级^[10]

ICF 限定值*	问题程度的百分等级
0—没有问题(无,缺乏,微不足道,…))	0~4%
1—轻度问题(略有一点,很低,…))	5%~24%
2—中度问题(中等程度,一般,…))	25%~49%
3—重度问题(很高,非常,…))	50%~95%
4—完全问题(全部,…))	96%~100%

*:“有问题就意味着不同结构下存在的损伤、受限、局限性或障碍。”^{[10]222},即取决于我们是否对身体功能和结构(损伤)、活动和参与(受限或局限性)以及环境因素(障碍或有利)进行分类。

第二种方法是使用临床试验或以患者导向的测量工具所获得的信息,临床试验包括标准化的专家和技术检查,以及患者导向的测量工具包括患者或报告代理人自我施测或访谈形式的问卷表,然后把这些信息转换成 ICF 限定值。第一步,基于已经建立的联系规则把临床测试或者以患者导向的测量工具和 ICF 联系起来^[13]。第二步,把临床测试和以患者导向的测量工具所获得的分数转换成 ICF 限定值。

第二种方法的优点是已获得的信息可以被转换成 ICF 的标准语言,这种标准语言可以为所有有兴趣的专业人员所理解,不论他们所从事的学科或所在的部门(如健康、劳动或教育)。

当使用的临床试验或者患者导向的测量工具是等距量表时,可以直接转换为 ICF 限定值,它全面并且独一无二地涵盖了某一可轻易得到的 ICF 类目的内容。例如,评估疼痛的视觉模拟量表(VAS)首先可与 ICF 类目 b280 痛觉相联系。第二步,疼痛 VAS 的测试值可以直接转换成 ICF 的限定值,因为它代表了一个一端标记是“无痛”、另一端标记是“最痛”的 100 mm 的等距量表^[14]。考虑到表 2 中的 ICF 限定值的百分等级,某人 VAS 疼痛标记疼痛的等级从 0~4 mm 可能在 ICF 类目 b280 痛觉对应限定值 0.5~24 mm 对应限定值 1.25~49 mm 对应限定值 2.50~95 mm 对应限定值 3.96~100 mm 对应限定值 4。

如果没有可轻易得到的临床测试或者患者导向的测量工具,可以开发第 3 种方法。我们可以考虑使用临床成套测试的一部分或者患者导向的测量工具中涵盖某一特定 ICF 类目的挑选出来的项目。因此,一个 ICF 类目等距量表可以构建作为临床测试或者患者导向的测量工具和 ICF 限定值之间的联系界面。

已确定的联系规则^[13,15]可以用于确认临床测试的适合部

分和以患者导向的测量工具项目。Rasch 模型或项目反应理论模型^[16]可以用来构建等距量表。然而,如何发展 ICF 类目等距量表的完整过程以及它们怎样作为联系界面到目前为止还没有进行研究说明。

本文的目的是通过整合不同的广泛使用的以患者导向的测量工具项目来举例说明构建特定 ICF 类目的等距量表,这将通过有效的类风湿关节炎患者的样本来进行说明。具体的目的是:①从一系列强调特定 ICF 类目的以患者导向的测量工具中确定备选项目作为测量量表的备选项目;②评估挑选出说明 ICF 类目“能量和驱动力功能”的项目范围从而形成一个单维的、等级的等距量表。

2 材料和方法

2.1 研究设计 心理测量研究所使用的资料来自于荷兰马斯特里赫特大学医学院管理的类风湿关节炎患者的有效样本进行的一项横断面研究。由于患者被要求填写许多患者导向的测量工具,因此不是心理测量研究。

横断面研究的研究协议和知情同意书得到了马斯特里赫特大学医学院伦理委员会的批准。患者的纳入标准:根据修订的美国风湿病学会(ACR)诊断标准,年龄至少 18 周岁,足够的荷兰语知识,并且签署知情同意书。

2.2 测量 患者完成下列患者导向的测量工具:类风湿关节炎生活质量问卷(RAQoL^[17])、健康评估问卷(HAQ^[18])、简明医学结果研究-36(SF-36^[19])、欧洲生活质量评定量表(EQ-5D^[20])、多维疲劳量表(MFI^[21])和流行病学研究中心抑郁量表(CES-D^[22])。

2.3 确定备选项目 为整合某一 ICF 类目等距量表确定备选项目,由 2 位健康专业人员使用已确定的联系规则将患者完成的工具项目内容分别与 ICF 相联系^[13,15]。根据这些规则,某一项目包含的所有概念将被确定并且与最准确的 ICF 类目相联系。例如,在 RAQoL 项目“由于太费精力而不能外出与人见面”中,需要确定两个不同的概念,即“太费精力”和“外出与人见面”。它们分别与 ICF 类目 b1301 动机和 d9205 社会活动相联系,b1301 动机定义为“产生行动诱因的精神功能;行动的有意识或无意识的驱动力量”,d9205 社会活动定义为“与别人一起参与非正式或随意的聚会,如走访亲戚朋友或在公共场合非正式的会面”。SF-36 和 EQ-5D 的测量项目已经在另外一项研究中和 ICF 类目相联系^[23]。然而本次调查研究中,与其他以患者导向的测量工具相关的 2 位健康专业人员再一次把它们联系起来。如果在确定 ICF 类目时出现不一致情况,由第三方(AC)做出非正式的决定。

为了评定连接过程的信度,基于 2 个独立的以患者导向的测量工具的连接版本来计算观察到的一致百分率。另外,为了调查是否比随机发生有更多地一致性,需要计算 Kappa 系数^[24]和非参数引导的置信区间^[25-26]。对 ICF 所有不同水平的类目连接过程的信度均进行了研究,即一级、二级和三级水平成分。

计算测量项目的数量和相联系的二级水平的 ICF 类目数量。这意味着在连接过程中当三级或四级水平类目已经被选择时,二级水平类目然后被考虑。选择与大多数测量工具项目相关的二级水平的 ICF 类目作为典型的 ICF 类目来说明 ICF 类目等距量表的构建。

2.4 备选测量项目的有效化 内容连接中所选择出来的 ICF 类目用于测量项目的有效性研究是基于 Rasch 模型的等级反应选择模式,该模型扩展了原有的两分法模型,考虑等级量表的典型数据^[27]。根据 Rasch 模型所确定的备选项目要满足单一维度的标准,这将会在下面的章节中详细描述。

2.5 ICF 类目等距量表的构建 基于等级反应选择的 Rasch 模型,针对所选择的 ICF 类目确定等距量表的备选项目。Rasch 模型认为个人能力和测量项目难度参数都被放在一起,属于同一个维度,这表明潜在的特质将被测量。该维度的单位,正如模型所定义的那样,是组成等距量表的 logits(成功和失败的自然对数)。

以下属性将被研究:量表的单维性和信度、项目拟合度、对于某些功能的单项项目反应,及项目和个人能力之间的目标。

适合项目的单维度性通过项目-特质相互作用的 χ^2 检验^[28]和 Z 值统计^[29]来分析。显著的 χ^2 检验概率小于 0.05,说明量表与该模型不拟合。

使用 Person 分离指数 r_p 研究信度,这类类似于传统的理论测试指标 Kuder-Richardson 公式 21 或 Cronbach's α 系数范围在 0 到 1 之间,在这儿测试值 1 表示个人定位的完全可复制性^[30]。不同能力层次的数目 H_p 可以通过量表分数来可靠地确定,量表分数通过公式 $H_p = (4G_p + 1) / 3$ 计算得出^[31], G_p 是指以标准误的单位来表示的样本的标准差。

单个项目的拟合度测试也是基于 χ^2 检验来进行的。另外,标准的拟合度残差值“Z”也被考虑。取值范围应该是在 -2.5~+2.5 之间,显示模型拟合^[32]。这些值提供了观察项目数值偏离方向的有关信息。

基于对每一个 ICF 分类类目估计的阈值来研究有关功能的项目反应。阈值对应于个人潜在连续体中的位置,在该位置上某人将被分类到邻近的反应选择的位置可能相等,并且因此获得两个连续分数之一。阈值的数目等于反应选择的数目减去 1,并且它们应该有渐增的测试值,因为非等级性的阈值估计显示不能构建一种连续分数反映测量维度的渐增水平的测量^[33-34]。当阈值无序时,反应选择就混乱了,也就是说,考虑到频率分布和它们的概率曲线测量项目被重新调整。在重新调整测量项目后,阈值应该显现出预期的逐渐增加的等级。

不匹配的测量项目和没有等级的反应选择阈值就对所有深层次结论的效度提出了质疑,包括项目的单一维度性。因此数据必须反复地修改,模型需要重新校准,单一维度性和项目的拟合度要不断检查直到项目-特质相互作用的 χ^2 检验没有显著性结果。我们使用 2 个步骤来修订数据:通过分解反应选择并且逐步删除最小卡方概率项目的方式。

通过检查潜在的特质连续体中个人能力分布和项目难易度分布以研究测量项目的位置。比较平均个人和平均项目的反应点,其通过定义限定为 0 显示这个领域要达到的目标,差异越小,针对性越强。任意的平均数为 0 的 logits 可以被转换成更有意义的量表,我们称之为 ICF 类目等距量表,取值范围从 0 到 100,使用下述公式: $Y = m + (s \times location)$

$s = (\text{理想范围}) / (\text{实际范围})$, $m = (\text{理想值的最低值}) - (\text{实际最低值} \times s)$ 。每个单独项目的反应选择的阈值也被转换成 ICF 类目等距量表取值范围从 0 到 100。

为了说明患者的 ICF 类目等距量表的得分可以用来评估

ICF 限定值(从 0 到 4)的测试值,ICF 限定值的反应选择被放入 ICF 类目等距量表中。限定值 0 对应于在 ICF 类目等距量表中的 0~5 分,限定值为 1 时达到 25,限定值取值 2 时达到 50,限定值取值 3 时达到 75,限定值取值 4 时达到 100。

2.6 统计程序 使用 SPSS 14.0 进行描述性统计^[36]。使用 SAS 9.0 进行 Kappa 分析^[37]。使用 RUMM 2020 软件进行 Rasch 分析^[38]。

3 结果

3.1 研究人群 表 3 显示 122 例患者有效样本的人口学和相关的类风湿关节炎特征。

表 3 本研究中 122 例类风湿关节炎患者的人口学数据和疾病特征

患者特点	n=122
社会人口学数据	
男性	36(29.5%)
年龄(岁)	58.6±15.4
目前的工作状态	
带薪工作	12.3%
失业(由于类风湿性关节炎)	20.5%
失业(由于另外一个原因)	2.5%
待在家里/家庭主妇	23.8%
退休	33.6%
疾病特征	
病程(年)	13.6±11.5
并发症数目*;中位数[0~11]	4.3±1.98;4
类风湿关节炎疾病活动指数(RADAI)[0~10]	3.7±2.1

注:根据自我管理的并发症问卷(SCQ)的并发症数目:心脏病 10 例,高血压 32 例,肺炎 11 例,糖尿病 8 例,溃疡或胃病 12 例,肾脏疾病 3 例,肝病 2 例,贫血或其他血液疾病 4 例,肿瘤 3 例,骨关节炎 39 例,后背痛 40 例。此外,81 例患者表示他们患有问卷中没有明确提到的其他疾病。

3.2 备选项目的确定 表 4 显示连接过程的评估结果,如在 ICF 所有不同的水平观察一致性的百分率、Kappa 统计和引导的置信区间。没有一个 95%置信区间包括 0,表明一致性不是偶然的。

表 4 观察一致性的百分比,估计的 kappa 系数,以及在 ICF 的成分、章、二和三级水平的非参数引导的 95%置信区间

	观察的一致性(%)	kappa 系数	95%置信区间
成分	92	0.90	(0.83, 0.94)
第一级	98	0.98	(0.95, 0.99)
第二级	88	0.85	(0.80, 0.91)
第三级	83	0.80	(0.57, 0.90)

二级水平的 ICF 类目 b130 能量和驱力功能是 ICF 类目中联系最多的项目,共计 19 个项目。因此,选用 ICF 类目 b130 来举例说明 ICF 类目等距量表的构建。表 1 呈现的是 b130:能量和驱力功能的定义。表 5 呈现的是 19 个项目及它们相应的反应选择。

在 ICF 中,b130 能量和驱力功能所反映的是一个中性概念。然而,基于表 5 呈现的挑选出来的项目和它们相应的反应选择,强调的概念是“能量和驱力功能的损伤”,这也是 ICF 限定值所表达的概念,即损伤的程度或者幅度(参见表 2)。

3.3 备选项目的标准化和 ICF 类目等距量表的构建 根据 χ^2 检验, 19 个项目的整体拟合度显示有显著性差异, 表明正如 Rasch 模型($\chi^2_{df=38} = 69.401, P < 0.01$)所预期的那样, 不是所有的项目均测量了能量和驱力功能维度。也发现了 Z 值拟合度统计变异较高(项目和个人分别对应的是: $Z_{平均数} = 0.32, Z_{标准差} = 1.22$ 及 $Z_{平均数} = -0.15, Z_{标准差} = 1.24$)。Person 分离指数 r_{β} 为 0.94, 表明具有很高的信度。

根据单个项目和模型的拟合度, 根据 χ^2 检验结果有 3 个项目显现出了显著的不一致($\chi^2_{df=2} = 8.78, P = 0.01$; $\chi^2_{df=2} = 8.78, P = 0.01$; $\chi^2_{df=2} = 6.49, P = 0.04$)。4 个项目显现出非等级的阈值参数。因此, 这 4 个项目的反应选择被排除, 并且再次进行模型拟合度和单维度性检验。出现非等级反应选择的 4 个项目中, 每一个项目的排除策略也呈现在表 5。

在排除反应选择后, 3 个不一致项目中仅有一个项目仍然

显示出明显的不一致[“我得比我希望的时间更早上床睡觉”(RAQoL1)]。然而, 其他两个最初一致的项目显现出显著的不一致[“我不想吃东西”(CES-D 2)和“我感觉自己的身体状况很差”(MFI 14)]。到最后, 这 3 个不匹配的项目随着分析的深入而逐渐被删除。

根据 χ^2 检验, 剩余 16 个项目的整体拟合度统计并没有显现出显著差异($\chi^2_{df=32} = 38.25, P = 0.21$)。对项目和个人 Z 值拟合度统计结果分别是 $Z_{平均数} = 0.451, Z_{标准差} = 1.085$ 以及 $Z_{平均数} = -0.223, Z_{标准差} = 1.132$, 而 Person 分离指数 r_{β} 为 0.93。基于 16 个项目量表得分可以很明确地确定出 5 个不同的能力层次。

表 6 显示了计算非等级的阈值前后和不一致项目的单个项目拟合度和评估的阈值参数。

表 5 与 ICF 分类中 b130 能量和驱力功能相关联的 19 个测量项目

项目	问卷	反应选择	排除策略
MFI-20 指导: 通过下面的表述我们希望得到一个你最近如何感觉的想法。例如, “我感到放松”这个表述, 如果你认为这是完全正确的, 也就是你最近的确感觉放松, 请在左边的盒中划一个“×”号。你越是不同意这样的表述, 越应该在“不, 这是不正确的”这边写“×”。请不要漏掉任何一个表述, 并且不要忘记划“×”。			
我感觉很好。	MFI-20	从 0(是的, 那是正确的)到 4(不, 那是不正确的)	
我感到充满活力。	MFI-20	"	0 1 1 2 3
我感到疲劳。	MFI-20	"	
我休息好了。	MFI-20	"	
我感觉只能做一点点体力活动。	MFI-20	"	3 2 2 1 0
我感觉能做很多体力活动。	MFI-20	"	
我感觉身体状况很差。	MFI-20	"	2 1 1 1 0
我很容易觉得累。	MFI-20	"	
我感觉身体状况很好	MFI-20	"	
CES-D 指导: 下列是你可能感觉到或表现出的行为方式的清单。请告诉我你在最近 1 周里有这样感觉的频率。			
我不想吃东西; 我的胃口很差。	CES-D	0=无或很少的时间(小于 1 d); 1=有时或时间很少(1~2 d); 2=偶尔或适量的时间(3~4 d); 3=大部或全部时间(5~7 d)。	
我感觉做任何事情都很费力。	CES-D	"	
我不能耐受持续行走。	CES-D	"	
SF-36 指导: 下面这些问题是关于最近 4 周你如何感觉以及事情对你的影响。对每一个问题, 请给出一个最接近你感觉的答案。最近 4 周有多少时间(每行圈一个数字)			
你精力充沛吗?	SF-36	1=所有的时间; 2=大部分时间; 3=相当多的时间; 4=有些时候; 5=有点时间; 6=无时间。	
你感到过筋疲力尽吗?	SF-36	"	
你感到过疲劳吗?	SF-36	"	3 2 2 2 2 1
RAQoL 指导: 下面你会发现一些类风湿关节炎患者的表述。请仔细阅读每一条表述。我们希望你如果感觉表述适合你就勾选“对”, 否则就勾选“错”。			
我得比我希望的时间更早上床睡觉。	RAQOL	0=错; 1=对	
需要花费太多的精力以至于不能外出与人见面。	RAQOL	"	
我得停下手头的事情, 去休息。	RAQOL	"	
我无论做什么都觉得累。	RAQOL	"	

表 6 个体测量项目与模型的前后拟合度解释了非等级阈值和项目不匹配。测量项目根据他们在表 5 中所属的工具来呈现

项目	之前											之后														
	D	R	SE	FR	DF	χ^2	DF	P	T1	T2	T3	T4	T5	D	R	SE	FR	DF	χ^2	DF	P	T1	T2	T3	T4	T5
MFI 1: 我感觉很好。	-0.56	7	0.11	-0.74	111.33	3.49	2	0.17	-1.9	-0.6	0.9	1.7		-0.29	7	0.12	-0.33	110.27	1.92	2	0.38	-2.0	-0.7	0.9	1.8	
MFI 3: 我感觉很有活力。*	-0.39	10	0.10	1.19	111.33	0.37	2	0.83	-0.7	-0.9	0.6	1.0		-0.07	10	0.13	0.25	110.27	0.98	2	0.61	-1.8	0.8	1.0		
MFI 5: 我感觉累。	-0.33	11	0.10	1.33	111.33	0.67	2	0.71	-0.7	-0.6	-0.2	1.4		-0.08	9	0.10	1.52	110.27	1.97	2	0.37	-0.8	-0.6	-0.2	1.5	
MFI 12: 我休息得好。	-0.84	5	0.11	-1.85	111.33	2.14	2	0.34	-1.1	-0.8	0.4	1.5		-0.62	5	0.11	-1.47	110.27	1.33	2	0.51	-1.3	-0.7	0.5	1.5	
MFI 2: 我感觉只能做一点点体力活动。*	0.06	13	0.10	2.58	111.33	8.78	2	0.01	-1.4	0.3	0.2	1.0		0.34	12	0.13	1.35	110.27	4.16	2	0.12	-1.4	-0.3	1.8		
MFI 8: 我能进行很多体力活动。	-0.42	9	0.11	1.69	110.41	5.81	2	0.05	-1.2	-0.5	0.2	1.4		-0.15	8	0.11	1.73	109.36	4.71	2	0.09	-1.2	-0.5	0.1	1.6	
MFI 14: 我感觉身体状况不佳。*	-0.47	8	0.11	-1.35	111.33	3.60	2	0.17	-0.9	-1.0	0.9	1.0														
MFI 16: 我容易疲倦。	-0.84	6	0.10	1.04	110.41	1.32	2	0.52	-0.6	-0.5	0.1	1.0		-0.61	6	0.11	1.19	109.36	0.73	2	0.69	-0.7	-0.4	0.0	1.1	
MFI 20: 我感觉身体状况非常好。	-1.15	1	0.11	-0.04	111.33	1.87	2	0.39	-1.0	-1.0	0.6	1.4		-0.92	2	0.12	0.39	110.27	1.18	2	0.55	-1.1	-0.9	0.5	1.5	
CES-D 2: 我不想吃东西, 我的胃口很差。	3.77	19	0.19	0.81	110.41	5.91	2	0.05	-2.4	-1.4	3.8															
CES-D 07: 我感觉我所做的所有事情都很费力。	0.73	16	0.14	0.19	110.41	6.49	2	0.04	-1.9	-0.3	2.2			1.07	14	0.14	0.47	109.36	6.22	2	0.04	-2.0	-0.3	2.3		
CES-D 20: 我不能耐受持续行走。	0.82	17	0.14	1.47	109.49	3.05	2	0.22	-1.7	-0.5	2.2			1.16	15	0.14	2.16	108.45	0.92	2	0.63	-1.8	-0.5	2.3		
SF-36 9e: 你感到精力充沛吗?	-0.11	12	0.11	0.34	109.49	1.97	2	0.37	-2.7	-1.0	0.2	1.1	2.4	0.19	11	0.11	0.91	108.45	1.94	2	0.38	-2.9	-1.0	0.1	1.1	2.6
SF-36 9g: 你感到筋疲力尽吗?	0.56	14	0.11	0.73	108.57	0.61	2	0.74	-2.7	-1.3	0.0	1.3	2.7	0.89	13	0.11	1.52	107.54	2.15	2	0.34	-2.8	-1.4	-0.1	1.4	2.9
SF-36 9i: 你感到累吗? *	-0.92	3	0.12	0.16	110.41	5.60	2	0.06	-3.1	-1.9	0.5	2.4	2.2	-1.18	1	0.31	-0.27	109.36	2.17	2	0.34	-4.2	4.2			
RAQoL 1: 我必须比我希望的时间更早睡觉。	0.67	15	0.22	1.56	109.49	8.78	2	0.01	0.0																	
RAQoL 25: 需要花费太多的精力以至于不能外出与人见面。	1.39	18	0.25	-0.63	107.65	3.10	2	0.21	0.0					1.73	16	0.25	-0.33	106.62	2.75	2	0.25	0.0				
RAQoL 10: 我得停下手头的事情, 去休息。	-0.90	4	0.22	-1.19	108.57	4.11	2	0.13	0.0					-0.64	4	0.23	-0.92	107.54	3.13	2	0.21	0.0				
RAQoL 21: 我无论做什么都觉得累。	-1.06	2	0.22	-1.19	110.41	1.74	2	0.42	0.0					-0.81	3	0.23	-0.95	109.36	2.00	2	0.37	0.0				

注: 加“*”代表该项目为非等级反应选择且已被重新测量。D: 项目难度的估计; R: 根据项目难度估计的排序; SE: 项目难度估计的标准误; FR: 残差 Z 的标准化; χ^2 = 卡方; DF: 自由度; P: χ^2 检验的概率; T1: 第 1 阈值估计; T2: 第 2 阈值估计; T3: 第 3 阈值估计; T4: 第 4 阈值估计。当项目为二分法时只有一个阈值。

图 1 显示了在测量连续体中个人和项目的分布。平均个人位置是 0.15, 和平均项目位置 0.0 比较起来, 表明针对性合适。项目的反应选择范围超过 8.9 个 logits, 涵盖了很大一部分类风湿关节炎患者, 他们在能量和驱力功能上呈现出中度至严重程度的障碍。

图 2 代表了能量和驱力损伤整个分值范围的 ICF 类目等距量表, 取值范围从 0 到 100。这些测量项目以及他们相应的描述在图中以横排的形式显现。在描述值旁边的值代表所有项目阈值并且涉及项目位置或者项目难度水平。横排灰度相对应的值随着阈值的改变而改变。两个最容易的项目是“你感到过疲劳吗(SF-36 9i)”和“我觉得身体状况非常好”。在这里, 术语“最容易”意味着一个非常低的能量和驱力损伤水平的个体适合该项目。换句话说, 容易的项目可以在能量和驱力从无到轻度损伤的个人之间有差异。

两个最难的项目是“我不能持续行走”和“需要花费太多的力气而不能出去与人见面”(RAQoL 25)。术语“困难”意味着只有在能量和驱力上有严重损伤的个体才适用该项目。换句话说, 困难项目是在能量和驱力上那些把个体区分为中度到重度损伤的测量项目。困难项目通常其 logit 为正值, 而容易项目的 logit 为负值。

基于具体的例子, 图 2 可以解释如下: “你感到过疲劳吗?” 最容易的测量项目被重新评分(它的反应选择被排除), 从最初的 6 个反应选择到 3 个反应选择(1=从来没有, 2=偶尔, 3=一直)。在排除反应选择后项目有 2 个阈值(阈值 1=7.9, 阈值 2=78.8), 这与图 2 中最底行灰度的改变相对应。它是最容易的项目因为即便是一个能量和驱力功能损伤程度非常轻(7.9% 损伤)的人也能够适合这个测量项目(例如选择反应选择 2=偶感疲劳)。然而, 仅仅那些能量和驱力功能严重损伤几乎达到 80%

损伤的个体, 会选择下一个反应选择 3=几乎所有的时间均感疲劳(阈值 2=78.8)。

ICF 类目等距量表中每一个反应选择的 ICF 限定值位置通过纵向的箭头被表示如图 2 所示。

ICF 类目等距量表分值达到 5 分的个体在 ICF 类目 b130 能量和驱力功能的限定值为 0, 得分在 5~25 的限定值为 1, 得分在 25~50 的限定值为 2, 得分在 50~95 的限定值为 3, 得分在 95~100 的限定值为 4。

通过将测量 16 个项目的分数相加所得到的原始分被转换为 logit 量表、ICF 类目等距量表和 ICF 限定值。转换表的一部分如表 7 所示。例如, 如果某人获得的原始分是 4, 他在 logit 量表中的位置将是 -2.94, 他的 ICF 类目等距量表将是 29 分, 他的限定值是 2, 也就是说, 在能量和驱力功能上存在中度障碍。

表 7 从原始分到 logit 量表、ICF 类目等距量表, 和 ICF 限定值的部分转换表格

原始分 (16 个项目)	logit 量表	ICF 类目等距量表 (0 到 100)	ICF 限定值
0	-6.36	0	0
1	-4.72	14	
2	-3.81	21	1
3	-3.29	26	
4	-2.94	29	
...	2
21	-0.54	49	
22	-0.45	49	
23	-0.35	50	
24	-0.25	51	
...	3
49	4.1	88	
50	4.72	93	
51	5.59	100	4

原始分的计算:将 RA 患者对 16 个项目的回答相加。完整的转换表格可从作者处获得。

4 讨论

我们已经说明了如何基于等距量表来评估 ICF 限定值的测试值,特定的 ICF 类目等距量表的开发从一系列量表工具的测量项目中整合而来。用来构建 ICF 类目等距量表的项目最初形式保持不变。因此,有可能使用由最初测量工具的项目以及 ICF 类目提供的信息。考虑到在记录和报告功能和残疾时越来越多地使用 ICF 和 ICF 限定值作参考,这项应用是十分有用的^[38-39]。针对能量和驱力功能的 ICF 类目等距量表的开发能应用于任何 ICF 类目。

ICF 类目等距量表的构建取决于测量工具的项目与 ICF 的类目之间建立联系^[15,18]。连接过程的准确度可由 kappa 统计评估两者之间的可靠度来验证^[27]。本文中提到的 Rasch 分析,弥补了 kappa 统计的不足并且有助于研究内容连接是否为确定测量项目的合适方法用于说明相同的 ICF 类目。通过内容连接拟合 Rasch 模型确定的 19 个项目中的 16 个项目代表单一维度的事实支持了这样一种假说,即所有这 16 个项目都说明了挑选出来的 ICF 类目能量和驱力功能的内容。另外三个随后被排除的不拟合 Rasch 模型的项目是:“我得比我希望的时间更早上床睡觉”(RAQoL_1),“我不想吃东西”(CES-D_2)以及“我感觉身体状况很差”(MFI_14)。我们可能会认为以上三个项目分别指活动或感觉,与能量和驱力水平没有必然直接的关系。换言之,一个人比他所希望的时间更早上床睡觉的事实与其第 2 天早上必须很早去工作的事实有关。并且,一个人不想吃东西的事实可能更多的与其对能量水平的担心有关。至于第三个被排除的项目,我们可能会凭直觉认为有很多人知晓自己的身体状况很差,但他们并没有感觉到与之相关的能量和驱力处于低水平。

因子分析也可应用于测量项目是否说明共同的单一维度的研究中。然而我们选择的 Rasch 分析,不仅获得了单维度信息,而且获得了这些项目的附加属性信息,例如反应选择的表现,目标以及项目难度等。

我们通过计算 kappa 系数来评估关联过程,该系数显示了关联一致性的满意结果。kappa 系数是一个解释几率经常使用的而且简单的一致性指标。我们可以认为由于几率而产生的非系统性误差似乎是关联过程的次要相关性因素。因此,使用多面的,潜在分类^[40]和潜在特质^[41]分析等模型方法,在将来可用于解释关联者之间的任何不一致(即由于经验和专业)。

该模型的整体拟合度统计和个体项目拟合度,支持构建的能量和驱力功能的 ICF 类目等距量表的结构效度。高达 0.94 的 Person 分离信度指数表明通过使用该量表可达到测量方法的高精确度。因此使用通过量表获得的分数,在能量和驱力功能维度人们能可靠区分至少 5 个独立层面。

只有 4 个测量项目呈现出非等级的阈值。当人们的反应选择与他们评估能量和驱力功能水平的期望值不相符时,产生了阈值逆转。这可能是由于察觉到的反应选择的不确定或反应选择的狭窄范围所致^[36]。有趣的是,4 个表现出非等级阈值项目中的 3 个来自 MFI,因为在 MFI 中答案没有特指而是用下列用语来限定的,从“是的,那是真的”到“不,那不是真的”。研究这个测量工具的心理属性是有价值的,该测量工具使用 Rasch 或者其他 IRT 模型来调查预期反应选择表现的内容。

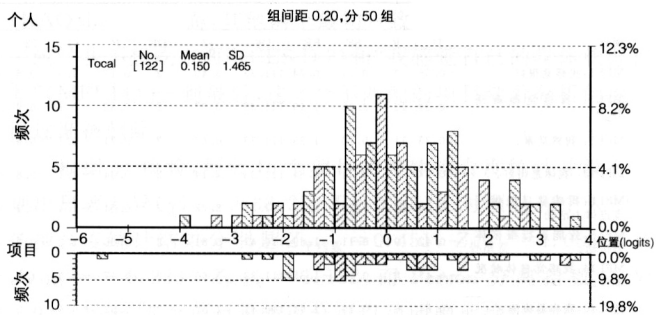


图 1 个人-项目阈值分布

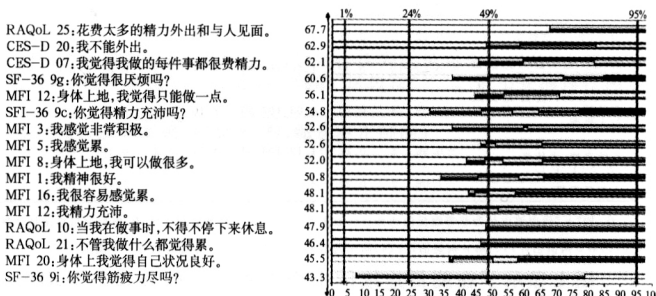


图 2 能量和驱力损伤的 ICF 类目等距量表

X 轴和 Y 轴代表能量和驱力连续损伤的 ICF 类目等距量表,取值范围从 0 到 100。由于空间的限制,并非所有从 0 到 100 的数值都在 Y 轴上表示。16 个项目按难易程度从最容易的项目(底部)到最难的项目(顶部)在 Y 轴上排列。对应于每一项目位置的分值标示在其旁边。测量项目的反应选择阈值位置标示在图中的横条中。不同的横条灰度代表每个项目的不同反应选择。垂直箭头代表 ICF 限定值的每个反应选择的位置。

第四个显示非等级阈值的项目来自 SF-36。采用 Rasch 分析和项目-反应-理论模型,SF-36 的心理属性已经在不同的调查中得到研究。说明心理健康的测量项目已经受到特别关注^[42-45]。然而就我们所知,还没有调查对包括在分量表效力中的项目反应选择功能进行研究。

创建的 ICF 类目等距量表也证明能很好地适用于在能量和驱力功能方面表现出中重度障碍的大部分类风湿关节炎患者。然而它也显示出地板效应,对于在能量和驱力功能方面有低水平损伤的类风湿关节炎患者,测量量值的低端不能被量表中包含的测量项目所涵盖。

ICF 类目等距量表的开发可以作为能量和驱力功能的自评工具开发的一个起点。然而要强调本文的目的是说明方法学而非建立一种新的测量工具是很重要的。我们知道,为建立一个普遍应用的能量和驱力功能的等距量表,还有很多的问题需要解决。首先,在进行研究之前,要考虑从大量的测量工具包括通用的、针对状况的和特定领域的测量工具中选择量表的测量项目(而且在本研究中并非研究之后才选择的)^[47]。这将有助于选

择一些容易的项目以恰当地代表了可测量到的能量和驱动力功能整个变化过程中的低端范围。已经发表的关于使用 ICF 测量健康状况的内容比较方面的研究可能是达成这一目标非常有价值的信息来源^[26, 48-50]。从事构建 ICF 类目等距量表以测量特定 ICF 类目的研究人员也可能使用国立卫生研究所(NIH)患者报告结局测量信息系统(PROMIS)项目的研究结果,开发、验证和标准化项目库去测量患者的报告结果(PROs)^[51-52]。将来,患者报告结果测量信息系统将是备选项目最综合性的资源。

第二,需要有更大的样本量来进行分析,尤其在包含了其他的项目时。尽管没有标准值去界定多大的样本量能获得有效稳定的项目校准,由 Linacre 估计(2002)^[36]提出的评估方法,提供了一个参照架构。按照这个评估方法,每个反应选择和项目至少需要 10 次观察。而且,个人-项目残差应该用主成分分析(PCA)^[53]进行检验,以确保部分独立性的假设成立。采用与 Rasch 模型充分拟合的个人-项目差异残差的独立性指标来支持单一维度性。

第三,需要对来自不同国家的数据进行测量效度评估。由于 ICF 的开发是用于国际应用的^[54],因此这一考虑具有特殊意义。

第四,采用传统测验理论(TTT)的方法可以进行额外的分析,如因子分析、辐合和区别效度研究分析。

最后但并非最不重要的是,如果应用这种方法创建新的 ICF 类目等距量表,必须特别注意每个测量项目的内容和难度水平。为避免冗余,具有相同内容和难度水平的测量项目不应该包含在同一等距量表中。

本文提到的方法应用作为创建新测量工具的起点,特别适用于那些 ICF 类目中还没有测量工具存在的,广泛和独特涵盖单个 ICF 类目内容的工具开发。基于 ICF 类目等距量表开发的工具有现实的优势,得到的分值非常直观,例如,从 0 到 100 并且能基于分值估计出 ICF 限定值。

总之,本研究说明了来自不同的患者导向的测量工具中的项目如何整合为一种心理测量有效的 ICF 类目等距量表,用于单独的 ICF 类目评估。本研究也说明了该量表得到的患者分数能很容易地转换成 ICF 限定值。它代表了将 ICF 的操作化和今后实施应用的重要步骤。

致谢:感谢 Heinrich Gall、Alicia Garza、Andrea Gl 分类的能量和 Michaela Kirschnack 对本研究提供的帮助,感谢 Pieter Lozekoot 和 Jos Ramaker 收集的数据。

本研究受到欧洲抗风湿联盟(EULAR)的部分资助。

[参考文献]

[1] Zola IK. Towards the necessary universalizing of disability policy. Part 2: disability policy: restoring socioeconomic independence[J]. The Milbank Quarterly, 1989, 67(suppl 2): 401-428.

[2] Bickenbach JE, Chatterji S, Bradley EM, Üstün TB. Models of disablement, universalism and the international classification of impairments, disabilities and handicaps[J]. Soc Sci Med, 1999, 48:1173-1187.

[3] Cieza A, Stucki G. New approaches to understanding the impact of musculoskeletal conditions[J]. Best Pract Res Clin Rheumatol, 2004, 18(2):141-154.

[4] Lollar DJ. Public health and disability: emerging opportunities[J]. Public Health Rep, 2002, 117(2):131-136.

[5] Stucki G. International Classification of Functioning, Disability, and

Health (ICF): a promising framework and classification for rehabilitation medicine[J]. Am J Phys Med Rehabil, 2005, 84(10):733-740.

[6] Stucki G, Stier-Jarmer M, Grill E, Melvin J. Rationale and principles of early rehabilitation care after an acute injury or illness[J]. Disabil Rehabil, 2005, 27(7-8):353-359.

[7] Stucki G, Cieza A, Melvin J. The International Classification of Functioning, Disability and Health (ICF): a unifying model for the conceptual description of the rehabilitation strategy[J]. J Rehabil Med, 2007, 39(4):279-285.

[8] World Health Organization. International Classification of Functioning, Disability and Health; ICF[M]. Geneva: WHO, 2001.

[9] Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine[J]. Disabil Rehabil, 2002, 24(17):932-938.

[10] Finger ME, Cieza A, Stoll J, Stucki G, Huber EO. Identification of intervention categories for physical therapy, based on the international classification of functioning, disability and health: a Delphi exercise[J]. Phys Ther, 2006, 86(9):1203-1220.

[11] Rentsch HP, Bucher P, Dommen Nyffeler I, Wolf C, Hefti H, Fluri E, Wenger U, Wälti C, Boyer I. The implementation of the 'International Classification of Functioning, Disability and Health' (ICF) in daily practice of neurorehabilitation: an interdisciplinary project at the Kantonsspital of Lucerne, Switzerland[J]. Disabil Rehabil, 2003, 25(8):411-421.

[12] McDowell I. Measuring health: a guide to rating scales and questionnaires[M]. 3rd ed. Oxford University Press, 2006.

[13] Cieza A, Geyh S, Chatterji S, Kostanjsek N, Üstün B, Stucki G. ICF linking rules: an update based on lessons learned[J]. J Rehabil Med, 2005, 37(4):212-218.

[14] Wallerstein SL. Scaling clinical pain and pain relief[M]. // Bromm B. Pain measurement in man: neurophysiological correlates of pain. New York: Elsevier, 1984.

[15] Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, disability and health[J]. J Rehabil Med, 2002, 34:205-210.

[16] Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? [J]. Med Care, 2004, 42(1 Suppl):17-16.

[17] Tjijhuis GJ, de Jong Z, Zwiderman AH, Zuijderduin WM, Jansen LM, Hazes JM, Vliet Vlieland TP. The validity of the Rheumatoid Arthritis Quality of Life (RAQoL) questionnaire[J]. Rheumatology, 2001, 40(10):1112-1119.

[18] Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis[J]. Arthritis Rheum, 1980, 23:137-145.

[19] Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). A. Conceptual framework and item selection[J]. Med Care, 1992, 30: 473-483.

[20] The Euroqol Group. Euroqol - a facility for the measurement of health? related quality of life[J]. Health Policy, 1990, 16: 199-208.

[21] Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue[J]. J Psychosom Res, 1995, 39(3):315-325.

[22] Center for Epidemiologic Studies, National Institute of Mental Health. Center for Epidemiologic Studies Depression Scale (CES-D) [M]. Rockville, MD: National Institute of Mental Health, 1971.

[23] Cieza A, Stucki G. Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF)[J]. Qual Life Res, 2005, 14(5):1225-1237.

- [24]Cohen J. A coefficient of agreement for nominal scales[J]. *Educational and Psychological Measurement*, 1960, 20:37-46.
- [25]Efron B. *The Jackknife, the bootstrap and other resampling plans* [M]. Philadelphia, Pa: SIAM, 1982.
- [26]Vierkant RA. (2000). A SAS macro for calculating bootstrapped confidence intervals about a Kappa coefficient. *SAS Users Group International Online Proceedings*[OL]. Retrieved July 23, 2004, from <http://www2.sas.com/proceedings/sugi22/STATS/PAPER295.PDF>.
- [27]Andrich D. Application of a psychometric rating model to ordered categories, which are scored with successive integers[J]. *Applied Psychological Measurement*, 1978, 2:581-594.
- [28]Andrich D. *Rasch Models for Measurement*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-068 [M]. Newbury Park, CA: Sage, 1988.
- [29]Styles I, Andrich D. Linking the standard and advanced forms of the Raven's progressive matrices in both the pencil-and-paper and computer-adaptive-testing formats[J]. *Educational and Psychological Measurement*, 1993, 53:905-925.
- [30]Andrich D. An index of person separation in latent trait theory, the traditional KR. 20 index, and the Guttman scale response pattern[J]. *Education Research and Perspectives*, 1982, 9:95-104.
- [31]Fisher WP. Reliability statistics[J]. *Rasch Measurement Transactions*, 1992, 6:238.
- [32]Wright BD, Masters GN. *Rating Scale Analysis*[M]. Chicago: MESA, 1982.
- [33]Linacre JM. Optimizing rating scale category effectiveness[J]. *J Appl Meas*, 2002, 3:85-106.
- [34]Andrich, D. The Rasch model explained[M]. //Sivakumar Alagumalai, David D Durtis, and Njora Hungi. *Applied Rasch Measurement: A book of exemplars*. Springer-Kluwer, 2005: 308-328.
- [35]Andrich D, Sheridan BS, Luo G. *RUMM2020: Rasch Unidimensional Models for Measurement* [M]. Perth, Western Australia: RUMM Laboratory, 2002.
- [36]SPSS Inc. *SPSS Release 14.0.2*. Wacker Drive[Z]. Chicago: Illinois
- [37]SAS Institute Inc. *The SAS System for Windows, Version 8.2*[Z]. Cary, NC: SAS Institute Inc; 2001.
- [38]Stucki G, Melvin J. The International Classification of Functioning, Disability and Health: a unifying model for the conceptual description of physical and rehabilitation medicine[J]. *J Rehabil Med*, 2007, 39(4):286-292.
- [39]Jette AM. Toward a common language for function, disability, and health[J]. *Phys Ther*, 2006, 86(5):726-734.
- [40]Linacre JM. *Many-Facet Rasch Measurement*[M]. Chicago: MESA Press, 1992.
- [41]Dillon WR, Mulani N. A probabilistic latent class model for assessing inter-judge reliability [J]. *Multivariate Behavioral Research*, 1984, 19: 438-458.
- [42]Uebersax JS, Grove WM. A latent trait finite mixture model for the analysis of rating agreement[J]. *Biometrics*, 1993, 49: 823-835.
- [43]Stucki G, Daltroy L, Katz JN, Johannesson M, Liang MH. Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts[J]. *J Clin Epidemiol*, 1996, 49(7):711-717.
- [44]Haley SM, McHorney CA, Ware JE. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale[J]. *J Clin Epidemiol*, 1994, 47(6): 671-684.
- [45]McHorney CA, Haley SM, Ware JE. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods[J]. *J Clin Epidemiol*, 1997, 50(4):451-461.
- [46]Taylor WJ, McPherson KM. Using Rasch analysis to compare the psychometric properties of the Short Form 36 physical function score and the Health Assessment Questionnaire disability index in patients with psoriatic arthritis and rheumatoid arthritis [J]. *Arthritis Rheum*, 2007, 57(5):723-729.
- [47]Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life[J]. *Ann Intern Med*, 1993, 118: 622-629.
- [48]Geyh S, Cieza A, Kollerits B, Grimby G, Stucki G. Content comparison of health-related quality of life measures used in stroke based on the international classification of functioning, disability and health (ICF): a systematic review[J]. *Qual Life Res*, 2007, 16(5):833-851.
- [49]Stucki A, Stucki G, Cieza A, Schuurmans MM, Kostanjsek N, Ruof J. Content comparison of health-related quality of life instruments for COPD[J]. *Respir Med*, 2007, 101(6):1113-1122.
- [50]Grill E, Stucki G, Scheuringer M, Melvin J. Validation of International Classification of Functioning, Disability, and Health (ICF) Core Sets for early postacute rehabilitation facilities: comparisons with three other functional measures[J]. *Am J Phys Med Rehabil*, 2006, 85(8):640-649.
- [51]Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes [J]. *Clin Exp Rheumatol*, 2005, 23(5 Suppl 39):S53-S57.
- [52]Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years[J]. *Med Care*, 2007, 45(5 Suppl 1):S3-S11.
- [53]Smith EVJ. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals[J]. *J Appl Meas*, 2002, 3(2):205-231.
- [54]World Health Organization. *Towards a Common Language for Functioning, Disability and Health, ICF*[M]. Geneva: World Health Organization, 2002.
- [55]Nunnally JC, Bernstein I. *Psychometric Theory*[M]. 3rd ed. New York: McGraw Hill, 1994.
- [56]Stucki, G, Liang MH, Stucki S, Brühlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research[J]. *Arthritis & Rheumatism*, 1995, 38: 795-798.
- [57]Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research [J]. *Arthritis Rheum*, 2003, 49:156-163.

(收稿日期:2011-01-17)

(原文载:J Clin Epidemiol, 2009, 62(9): 912-921.)