·论著.

SF-36 健康调查量表中文版的研制 及其性能测试

李鲁 王红妹 沈毅

【摘要】目的 研制 SF-36 健康调查量表中文版并验证量表维度建立及记分假设、信度和效度。方法 采用多阶段混合型等概率抽样法,用 SF-36 健康调查量表中文版对 1 000 户家庭的居民进行自评量表式调查;参照国际生命质量评价项目的标准程序,进行正式的心理测验学试验。结果 在收回的 1 985 份问卷中,18 岁以上的有效问卷 1 972 份,其中应答者 1 688 人(85.6%),1 316 人回答了所有条目,372 人有 1 个或以上的缺失答案,无应答者中文盲、半文盲占 65.5%。等距假设在活力(VT)和精神健康(MH)维度被打破了,按重编码后值计算维度分数;条目集群的分布接近源量表及其他 2 个中文译本;除了生理功能(PF)、躯体疼痛(BP)、社会功能(SF)维度,其余维度有相似的标准差;除了SF、VT维度,其余6个维度条目维度相关一致;除了 SF 维度,7 个维度集合效度成功率范围为 75%~100%,区分效度成功率范围为 87.5%~100%。一致性信度系数除了 SF、VT维度,其余6维度变化范围为 0.72~0.88,满足群组比较的要求。两周重测信度变化范围为 0.66~0.94。因子分析产生了 2个主成分,分别代表生理健康和心理健康,解释了 56.3%的总方差。结论 为 SF-36 健康调查量表适用于中国提供了证据,已知群效度试验将为量表效度提供更有意义的证据。

【关键词】 生活质量; 心理学试验; SF-36 量表

Development and psychometric tests of a Chinese version of the SF-36 Health Survey Scales ULu^* , WANG Hongmei, SHEN Yi.* Department of Social Medicine, Zhejiang University School of Medicine, Hangzhou 310031, China

[Abstract] Objective To develop and evaluate scaling and scoring assumptions, and the reliability and the validity of a Chinese version of the SF-36 scales. Methods A multi-stage mixed sampling procedure was used to select a representative sample of the general population. The sample size was 1 000 households. All family members of a selected household, aged 18 and older, completed a survey by self-administration. Formal psychometric methods for testing assumptions underlying item scoring and scale construction were used according to the standard procedure of the IQOLA Project. Results Of the 1 985 collected questionnaires, 1 972 were qualified. Of them, 1 688 (85.6%) were respondents. 1 316 respondents answered all 36 items, while the remaining (372 respondents) answered with one or several missing responses. Among the non-respondents, 65.5% were illiteracy or quasi-illiteracy. The assumption of equal intervals was violated for the VT and MH scales. The recoded item values were used to calculate scale scores. The clustering and ordering of the item means were approximately the same as that of the source version and other two Chinese versions. The items in each scale had similar standard deviations except those in the PF, BP, SF scales. The correlations between an item and its hypothesized scale were identical for all except the SF and VT scales. The scaling success rates of convergent validity were 0% for the SF scale, 75% for the VT scale, and 100% for the other six scales. The scaling success rates of discriminat validity ranged from 87.5% to 100% for all scales except for the SF scale. The Cronbach' a coefficients of internal consistency reliability ranged from 0.72 to 0.88, which were satisfactory for group comparison except 0.39 for the SF scale and 0.66 for the VT scale. The two-weeks test-retest reliability coefficient ranged from 0.66 to 0.94. Factor analysis identified two principal components: a "physical" factor and a "mental" factor. Taken together, these two factors could be used to explain 56.3% of the total variance. Conclusion The Chinese version of the SF-36 Health Survey Scale has achieved conceptual equivalence and satisfied the psychometric scaling assumptions well enough to warrant wide use in China. Known-groups validity will give more meaningful evidences of the validity of the Chinese SF-36 scales.

[Key words] Quality of life; Psychological tests; SF-36 health survey

基金项目:浙江省科技计划项目资助(991104209)

作者单位;310031 杭州,浙江大学医学院社会医学教研室(李鲁、王红妹),卫生统计学教研室(沈毅)

"健康相关生命质量" (health related quality of life, HRQoL)的概念自 70 年代引入国外医学界以 来,产生了许多生命质量测评量表。SF-36 健康调查 量表 (the Mos 36-item Short Form Health Survey)成为 全球应用最广的生命质量测评工具。SF-36 量表评 价 HRQoL 的 8 个方面,即生理功能(physical functioning, PF)、生理职能(role-physical, RP)、躯体疼 痛(bodily pain, BP)、总体健康(general health, GH)、活 力(vitality, VT)、社会功能(social functioning, SF)、情 感职能(role-emotional, RE)、精神健康(mental health, MH)。另外还有健康变化(health transition, HT),用 于评价过去1年内健康改变。1991年,国际生命质 量评价项目将 SF-36 量表列为测评工具[14]。在此 介绍 SF-36 量表中文版的研制过程,并用杭州市区 人群健康调查资料验证量表维度建立及记分假设、 信度和效度。

资料与方法

一、SF-36 健康调查量表的研制

采用美国波士顿新英格兰医学中心健康研究所的标准版 SF-36 健康调查量表^[1],由 2 名本专业研究生各完成一中译本,经比较讨论产生初稿。译员对问卷翻译有经验,但不熟悉 SF-36 量表。初稿经本校 2 名英语教师评价翻译质量,再经公共卫生、临床医疗、心理学 8 位对量表调查设计有经验的专业人员讨论,产生修改稿。在普通人群的方便样本中作预试,再作改动后产生终稿。

二、研究背景

采用多阶段混合型等概率抽样法,按街道、居民区、户三级抽样。第1阶段分别从下城区(属中心城区)和拱墅区(属次中心城区)抽3个街道,第2阶段再从每个街道各抽取3个居民区。抽样方法采用等距抽样法,最后阶段即户数的抽样数目,则根据每户被抽中的概率 P = n/N 来确定(其中 n 为样本量1000户,N 为2区总户数173765户)。对抽中户进行人户自评量表式调查。被调查者要求18岁以上,有阅读能力。在第1天调查的居民中随机抽取57人,2周后重测。

三、资料质量

若一维度半数以上条目缺失,则维度分数置为缺失;半数或半数以下条目缺失,则用未缺失条目的平均分代替缺失条目分数。计算 Myer's 指数,该指数取值范围为 0~99,若大于 60,表明存在严重的年龄堆积现象。计算每一条目的完成情况,每一维度

所有条目全部完成的比例;维度分数能被计算的比例。列联表卡方比较应答者与无应答者社会人口学特征的差别,logistic 回归分析诸因素对应答率的综合影响。

四、心理测验学试验

- 1. 条目记分和维度建立的假设:
- (1)是否为等距变量:选择项之间的距离应一致。如果该假设被打破,应重编码。这一假设在含2个以上条目且每一条目含2个以上水平的维度中得到检验——GH、PF、VT、MH维度。我们对每一条目每一水平,用所属维度其他条目分数和的平均值来记分,然后分配经验分数:最低反应水平置为1,最高反应水平置为 *K*(一共 *K* 个水平),中间值按平均分的间距来置分^[2]。
 - (2)每一维度各条目的方差是否一致。
 - (3)每一维度条目维度的相关是否一致。
- (4)集合效度条目与所属维度显著相关(估计条目与所属维度其余条目总和的相关性 r≥0.40)。
- (5)区分效度条目与所属维度的相关性高于与其他维度的相关性(2个标准差或以上,采用相关系数的假设检验 $^{[5]}$)。
- 2.信度和效度:本研究采用重测信度及内部一致性信度(Cronbach'α)。Cronbach'α≥0.7 的信度系数用于群组比较令人满意^[1]。维度间相关系数应低于 Cronbach'α信度系数。效度检验除了集合效度和区分效度,还采用因子分析法。考虑这样一种理论假设:量表测量 2 个概念,分别代表生理和心理健康。检查维度间相关以验证这种假设。

样本资料使用 SPSS 7.0 for Windows 软件进行分析。

结 果

一、SF-36 健康调查量表中文版

中文版在尊重源量表概念的基础上对个别条目根据中国国情作了修正。在 PF 维度,推真空吸尘器,打保龄球、高尔夫球在中国不是一项普及的活动。中国人很熟悉拖地板,打太极拳,但是否在各自文化中同属中等度活动无法确知,故都放在量表中以增加明晰度。1 英里等于1609 m,如果精确翻译,并不表达源量表所希望表达的精确度,故译成1500 m。在中国无街区(block)一词,用"路口"来代替。

在 VT、MH 维度的翻译中, "full of pep", "down in the dumps", "downhearted and blue"在中文中找不到

对应的习语,分别译成"干劲十足","情绪低落,沮丧","闷闷不乐,心情忧郁",仍值得推敲。另外,VT01"full of pep"与 VT02"have a lot of energy"在中文中意义很接近,最后分别译为"干劲十足"、"精力充沛";前者强调精力好而且愿意努力工作,后者偏重于躯体方面。

二、资料完成情况

在收回的 1 985 份问卷中, 18 岁以上的有效问 卷为1972份,其中应答者1688人(85.6%),平均完 成时间为 10 min, Myer's 指数为 7.94。无应答者中 文盲、半文盲占 65.5%, 应答者中 1 316 人(78.0%) 回答了所有条目,372人(22.0%)≥1个答案缺失。 平均每个条目的缺失率为 3.8% (0.3% ~ 6.6%)。 缺失呈现的趋势:题意接近的连续几个问题的缺失 率高,如 PF04"爬数层楼梯"缺失率为 3.2%, PF05 "爬一层楼梯"缺失率为 5.1%;成组问题高于独立 问题,如 MH01~ MH05 平均缺失率为 5.5%,而 GH01 为 0.3%, HT 为 0.7%。各维度全部回答的百 分比范围为88.9%~97.9%,维度分数能被计算的 百分比范围为 95.4%~98.3%。无应答者年老、女 性者多,文化程度低;多元 logistic 回归结果表明:文 化程度越高,家庭关系越紧密,应答率越高(P< 0.05,表1)。

表 1 多因素对应答率的多元 logistic 回归结果

		-	
变量	偏回归系数(β)	Wald 值	P 值
年龄	0.047	0.210	0.647
性别	0.328	2.660	0.103
文化程度	1.951	195.628	0.000
职业	-0.023	0.051	0.822
家庭类型	-0.107	4.699	0.030
婚姻状况	- 0.257	3.736	0.053

三、心理测验学检验结果

1.等距假设检验: GH01 各选择项的值用 GH 维度其余 4条目题值和的平均值来计算。4条目题值和的平均值依次升高"糟糕"(29.6),"过得去"(56.8),"好"(69.9),"很好"(76.6),"棒极了"(82.6)。在此基础上计算 GH01 各选择项的经验分数依次为 1.0,3.0,4.0,4.5,5.0。等距假设在 GH02~GH05 也得到支持,指示健康状态佳的 2 个选择项之间的距离最大,各选择项的经验分数依次为 1.0,1.5,2.5,3.5,5.0。

在 PF 维度中,各选择项的平均经验分数为 1.0,2.0,3.0,支持等距假设。但在 VT 维度,指示较低健康状态的两个选择项的位置互换了,按健康状态从差到好,平均经验分数为 1.4,1.0,1.8,3.6,4.7,6.0。MH 维度也出现同样的情况,平均经验分数的模式为 2.7,1.0,1.2,2.8,4.2,6.0。

- 2.各条目的分布:条目分按从小到大,形成不同的集群。在同一集群内,条目分没有什么差别。条目平均分的顺序除了 PF02、PF03、GH01 以外,与源量表和 2 个中文版研究一致^[6-8]。除了 PF、BP、SF 维度,其余维度的条目有相似的标准差。
- 3.集合效度和区分效度:条目维度相关超过0.40,计1个集合效度定标试验成功。除了SF、VT维度,其余6维度集合效度试验成功率均为100%。在35个集合效度定标试验中,32个试验成功,成功率为91.4%。条目与所属维度相关高于与其他维度的相关,差异显著,计1个区分效度定标试验成功。SF维度的条目维度相关高于条目与HT的相关,差异显著,与其他7维度的相关差别均不显著,该维度的区分效度试验成功率为12.5%,其余7维度成功率范围为87.5%~100%。在280个区分效度定标试验中,259个试验成功,成功率为92.5%(表2)。

表 2 SF-36 健康调查量表中文版的定标特征(n=1 316)

维度 条目		相关系数范围		集合效度 ^d			
	条目数* ~	条目集合效度b	条目区分效度 ^c	成功数	成功率(%)	成功数	成功率(%)
PF	10	0.42 ~ 0.72	0.01 ~ 0.41	10/10	100.0	79/80	98.8
RP	4	0.70 ~ 0.78	$0.09 \sim 0.45$	4/4	100.0	32/32	100.0
BP	2	0.72	0.24 ~ 0.43	2/2	100.0	16/16	100.0
GH	5	$0.43 \sim 0.57$	$0.14 \sim 0.49$	5/5	100.0	39/40	97.5
VT	4	0.39 ~ 0.49	0.11 ~ 0.49	3/4	75.0	28/32	87.5
SF	2	0.28	$0.09 \sim 0.37$	0/2	0.0	2/16	12.5
RE	3	$0.72 \sim 0.78$	0.04 ~ 0.48	3/3	100.0	24/24	100.0
MH	5	0.43 ~ 0.59	0.04 ~ 0.46	5/5	100.0	39/40	97.5

a 条目数和每一维度条目集合效度试验数目, b 条目和假设维度的相关性(经修正), c 条目与其他维度的相关性, d 条目和假设维度的相关性(经修正)≥0.40, e 试验总数中相关系数显著高的数目(≥2 s)

样本资料基本满足定标假设,按最后题值计算 原始分数(表3),并转换成0~100分数^[1]。

表 3 SF-36 健康调查量表维度分计算公式

维度	条目最后题值的累加 (按重编码后值)	最低和最高 可能分数	可能分数 范围
PF	3a + 3b + 3c + 3d + 3e + 3f + 3g + 3h + 3i + 3j	10,30	20
RP	4a + 4b + 4c + 4d	4,8	4
BP	7 + 8	2,11	9
GH	1 + 11a + 11b + 11c + 11d	5,25	20
VT	9a + 9e + 9g + 9i	4,24	20
SF	6 + 10	2,10	8
RE	5a + 5b + 5c	3,6	3
MH	9b + 9c + 9d + 9f + 9h	5,30	25

4.信度和效度:内部一致性信度用 Cronbach'α 统计量来描述。如表 4 所示, SF 维度的内部一致性信度最低, VT 维度为 0.66, 其余 6 维度满足群组比较的要求。8 个维度间观察到显著的正相关。因子分析产生 2 个主成分,解释了 56.3%的总方差(表5)。主成分之一生理健康在 PF 维度上呈中度相关,在 RE 上的因子负荷较高。主成分之二心理健康在 VT 维度的负荷高于 MH 维度,在 SF 维度上的负荷偏低,呈中度相关,与假设不一致。

表 4 SF-36 健康调查量表中文版的信度和维度间相关

维 度	重测信度 Pearson'r n = 57	内部一致 性信度 Cronbach'(α) n=1316	PF	RP	BP	GH	VT	SF	RE
PF	0.94	0.87							
RP	0.82	0.88	0.38						
BP	0.66	0.80	0.34	0.37					
GH	0.76	0.72	0.43	0.36	0.45				
VT	0.80	0.66	0.31	0.28	0.40	0.57			
SF	0.70	0.39	0.31	0.37	0.36	0.33	0.35		
RE	0.92	0.87	0.27	0.50	0.29	0.27	0.26	0.39	
MH	0.80	0.75	0.16	0.18	0.28	0.39	0.52	0.40	0.30

讨 论

一、SF-36 量表的翻译及文化调适

过去 10 年中,大量文献涉及了以英语为母本的健康测试工具的跨文化翻译和测试。将西方量表译成中文更是不同寻常的挑战^[9-10]。为数不多的中文版 SF-36 量表的研究结果令人鼓舞^[6-7]。本研究中

文版量表中个别用词仍需进一步推敲。如源量表中"社会活动"是指与亲密朋友的接触,但中国人对社会活动的理解更多倾向于较高社会阶层的社交活动。

表 5 中文版 SF-36 健康调查量表因子负荷的模型及 实际样本的因子负荷(n = 1.688)

维度	假设联系		因子分	相对效度*			
	生理	心理	与生理 相关	与心理 相关	解释 方差	生理	心理
PF	强	弱	0.59	0.25	0.42	0.49	0.09
RP	强	弱	0.84	0.07	0.70	1.00	0.01
BP	强	弱	0.48	0.45	0.43	0.33	0.29
GH	中	中	0.35	0.68	0.59	0.17	0.67
VT	中	中	0.16	0.83	0.72	0.04	1.00
SF	中	强	0.52	0.42	0.45	0.38	0.26
RE	弱	强	0.74	0.11	0.56	0.78	0.02
MH	弱	强	0.06	0.79	0.63	0.00	0.90

* 主成分解释给定维度与最佳维度的方差比

二、关于维度假设和效度

本研究观察到最低反应水平之间距离接近(GH 维度)或倒置(MH 和 VT 维度)。与法国的早期研究结果相似^[2]。选择项之间的距离是否应均等并无公论^[11-13]。有研究表明不等间距的分析结果与等间距的分析结果一致^[11,13]。

条目按其得分高低会呈现不同的集群分布。测量从事简单活动的条目因较少人报告受限,平均分会较高;而测量从事困难活动的条目因较多人报告受限,平均分会较低。不同国家人群健康水平不同,条目平均分会有所不同,如果翻译理想,条目均数的顺序和大致差异应一致。

同一维度中,条目维度相关应一致,当条目维度在 0.40~0.70 之间变动,可以被认为达到这个要求。加权增加了记分的复杂性,而且只有极少数加权改善了维度表现^[14]。本样本资料除了 SF 维度,均符合这一假设。除了 SF 维度,集合效度和区分效度令人满意。因子分析的结果与日本早期研究结果一致^[15]。在日本,研究发现应答者对条目选择项的理解有偏差,如在 PF 维度,将"活动受限制"理解成"受医生限制"。

综上,本研究的结果为 SF-36 健康调查量表适用于中国提供了证据,已知群效度试验(knowngroups validity)分析量表是否对已知不同健康状况的人群有良好的区分度,会给量表的效度提供更有意义的证据。中文版量表与源量表的性能差异提示应

对量表的概念框架在文化通用性方面作进一步的改进,以便更好地用于国际人群健康状况研究。

参考文献

- 1 Ware JE Jr, Snow KK, Kosinski M, et al. SF-36 health survey manual and interpretation guide. Boston: New England Medical Center the Health Institute, 1993. 1-12.
- 2 Perneger TV, Leplege A, Etter JF, et al. Validation of a French-language version of the MOS 36-Item Short Form Health Survey(SF-36) in young healthy adults. J Clin Epidemiol, 1995,48;1051-1060.
- 3 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol, 1993,46:1417-1432.
- 4 Gandek B, Ware JE Jr. Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol, 1998, 51:953-959.
- 5 顾杏元,金丕焕,直线回归与相关.见:金丕焕,主编.医用统计方法.上海:上海医科大学出版社,1993.115-119.
- 6 Lam CL, Gandek B, Ren XS, et al. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol, 1998, 51;1139-1147.
- 7 Ren XS, Amick B 3rd, Zhou L, et al. Translation and psychometric evaluation of a Chinese version of the SF-36 Health Survey in the United States. J Clin Epidemiol, 1998,51:1129-1138.

- 8 Ware JE Jr, Keller SD, Gandek B, et al. Evaluating translations of health status questionnaires. Methods from the IQOLA project. Int J Technol Assess Health Care, 1995,11:525-551.
- 9 Muller JH, Desmong B. Ethical dilemmas in a cross-cultural context. A Chinese example. West J Med, 1992, 157:323-327.
- 10 Mo B. Modesty, sexuality, and breast health in Chinese-American women. West J Med, 1992, 157; 260-264.
- 11 van Campen C, Sixma H, Friele RD, et al. Quality of care and patient satisfaction: a review of measuring instruments. Med Care Res Rev, 1995, 52:109-133.
- 12 Till JE, Osoba D, Pater JL, et al. Research on health-related quality of life; dissemination into practical applications. Qual Life Res., 1994, 3: 279-283.
- 13 Mathias SD, Fifer SK, Patrick DL. Rapid translation of quality of life measures for international clinical trials: avoiding errors in the minimalist approach. Qual Life Res, 1994,3:403-412.
- 14 Armor DJ. Theta reliability and factor scaling. In: Costner HL, ed. Sociological Methodology: 1973-74. San Francisco, CA: Jossey-Bass, 1974.17-50.
- 15 Fukuhara S, Bito S, Green J, et al. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. J Clin Epidemiol, 1998, 51: 1037-1044.

(收稿日期:2001-04-12) (本文编辑:邵隽一)

·短篇报道·

沈阳市育龄妇女死亡情况调查

于君 刘淑丽 李兵 赵雅梅 冯丹

为了解沈阳市育龄妇女的死亡原因,我们对 1996 ~ 2000 年育龄妇女的死亡情况进行了调查。

- 1.对象与方法:资料来源于 1996 年 1 月 1 日 ~ 2000 年 12 月 31 日卫生防疫部门的生命统计。根据死亡资料,通过户口注销、医院死亡报告和街道调查反复核对。
 - 2.结果:
- (1)死亡变化:城市死亡率 1997~1999 年以 11.45%~ 26.47%速度上升;农村 1997~2000 年则呈下降趋势。除 1997年,其余 4 年城乡死亡率的差异有非常显著性,城市非常显著高于农村。
- (2)死亡原因:1996~1999 年≤35 岁的伤害死亡呈增加趋势,>35 岁因疾病死亡逐年增加。
- (3)疾病死因:参照内科学分类和孕妇死亡共 10 类死因。城乡死因顺位基本相同,为循环系统、呼吸系统、消化系统、泌尿生殖系统及神经系统疾病,城市第 3 位为消化系统,农村为泌尿生殖系统疾病。城市死亡率明显高于农村。农村孕产妇死亡率约为城市的 2 倍。

作者单位:110032 辽宁省,沈阳市妇儿医疗保健中心围产保健科(于君),B超检查室(刘淑丽);沈阳市铁西区妇幼保健所妇女保健科(李兵、赵雅梅、冯丹)

- (4)伤害死因:1996~2000 年城市死因顺位为机动车辆 交通伤害、中毒和自杀;农村为机动车辆交通伤害、自杀和中 毒。城乡的差异无显著性。
- (5)恶性肿瘤死亡: > 35 岁死亡率高于 \leq 35 岁,比值为 5.93:1。城市发生率为 23.44/10 万,农村为 13.61/10 万,城市发生率明显高于农村($\chi^2 = 135, P < 0.001$)。死因中肿瘤的前 3 位城市为肺癌、乳腺癌和胃癌,农村为肺癌、乳腺癌和肝癌。城市恶性肿瘤死亡率高于农村(1.72:1)。
- 3.讨论:沈阳市城市育龄妇女 5年死亡率明显高于农村。1996~1999年死亡率有上升趋势,主要以城市上升为主,死亡中以>35岁因病死亡增加为主。

疾病死因中循环系统疾病占第 1 位,呼吸系统占第 2 位。我市是工业城市,呼吸系统疾病发生率较高。另外,农村孕产妇死亡率为城市的 2 倍,说明农村的围产保健和助产质量仍存在问题,应扶植乡镇医院产科建设,提倡住院分娩。育龄妇女伤害死亡率为 17.67/10 万,城乡无差异。死因中,机动车辆交通伤害为第 1 位,中毒、自杀居其次。育龄妇女恶性肿瘤死亡率为 18.84/10 万,城市高于农村, > 35 岁为高发人群,死因依次为肺癌、乳腺癌。

(收稿日期:2001-04-13) (本文编辑:周星)