

慢性病生命质量量表共性模块项目反应理论分析*

杨铮¹ 戚艳波² 万崇华³ 李晓梅² 赵芝焕⁴ 段丽萍⁴ 潘家华⁴ 张晓馨²

摘要:目的 运用项目反应理论(IRT)对慢性病患者生命质量测定量表共性模块(QLICD-GM)条目进行分析,筛选信息量较高条目。方法 应用QLICD-GM测评7种慢性病患者620例,采用塞姆吉玛等级反应模型计算每个条目的难度、区分度系数和信息量,绘制项目特征曲线;根据平均信息量筛选条目;采用MULTILOG 7.0软件进行计算和作图。结果 QLICD-GM 共性模块29个条目的区分度均为1.2~1.9;难度(程度)均呈严格单调递增,取值范围为-3.05~2.18;依据平均信息量结合条目特征筛选保留24个条目。结论 QLICD-GM 各条目区分度均较好、选项设置合理、难度合适,分析模型选择正确;项目反应理论可筛选出信息量较高条目,弥补经典测量理论(CTT)的不足。

关键词: 慢性病; 生命质量; 项目反应理论(IRT); 信息函数

中图分类号: R 541.3

文献标志码: A

文章编号: 1001-0580(2012)11-1477-04

Analysis on items of general module of quality of life instrument for chronic diseases by item response theory

YANG Zheng*, QI Yan-Bo, WAN Chong-Hua, et al (* Department of Epidemiology and Health Statistics, School of Public Health, Guangdong Medical College, Dongguan, Guangdong Province 523808, China)

Abstract: Objective To analyze items of the general module of the quality of life instrument for chronic diseases (QLICD-GM) for selection of items with more information using item response theory (IRT). **Methods** A total of 602 inpatients with one of the 7 different chronic diseases were measured by QLICD-GM. With MULTILOG, Samejima's Graded Response Model was used to calculate parameters of discrimination and difficulty, information function, and to get item characteristic curve. **Results** Among the 29 items, the discrimination parameters of all items were between 1.2-1.9 and the difficulty parameters of each item increased monotonically for all items with a range from -3.05 to 2.18. Twenty-four items were selected as good ones by average information and item characteristics. **Conclusion** Each item of the QLICD-GM has good discrimination, suitable item choice and moderate difficulty. The item analysis based on item response theory can select good items with more information and refined characteristics to overcome weakness of classical test theory.

Key words: chronic disease; quality of life; item response theory; information function

随着疾病谱与健康观的改变,生命质量的研究日益受到重视。本课题组在国家自然科学基金资助下开发了慢性病患者生命质量测定量表体系(Quality of Life Instrument for Chronic Diseases, QLICD)^[1-4]。该体系(第1版)由适用于各种慢性病的共性模块QLICD-GM及10多种慢性病特异测定量表组成,其中的共性模块起着基础和核心的作用。为了进一步分析该体系共性模块各条目的难度、区分度、信息量等,为慢性病患者生命质量测定量表体系第2版的开发提供理论基础,本研究运用项目反应理论(item response theory, IRT)模型^[5-8],

对该体系共性模块3个维度29个条目进行了分析和筛选。现将结果报告如下。

1 对象与方法

1.1 对象 昆明医学院第一附属医院确诊并住院治疗的高血压、冠心病、慢性胃炎、消化性溃疡、肠易激综合征、肺源性心脏病、慢性阻塞性肺病7种慢性疾病患者620例,进行生命质量测评。所有调查对象均具备一定读写能力,并自愿参加测评;排除文盲、入院时病情危重及神志不清患者。

1.2 方法

1.2.1 测评调查 采用本课题组研制的慢性病患者生命质量测定量表体系共性模块(QLICD-GM, V1.0)^[1]对慢性病患者进行问卷调查和量表测评。该模块包括躯体功能、心理功能和社会功能3个维度共30个条目。调查由统一培训的调查员说明量表要求后,将量表发给患者,由患者根据自己的实际情况选答。

* 基金项目: 国家自然科学基金(30860248); 广东医学院“高级人才科研启动项目”(XG1004)

作者单位: 1. 广东医学院公共卫生学院流行病学与卫生统计教研室, 广东 东莞 523808; 2. 昆明医学院公共卫生学院; 3. 广东医学院人文与管理学院; 4. 昆明医学院第一附属医院

作者简介: 杨铮(1974-),女,云南个旧人,副教授,硕士,主要从事流行病学与卫生统计学工作。

通讯作者: 万崇华 E-mail: wanchh@hotmail.com

1.2.2 项目反应理论分析 (1) 项目反应理论^[5-8]的基本原理:应用数学模型分析被测试者潜在特质(能力水平)与其对测验项目(条目)反应之间的关系,即以数学形式表达其能力与对测验项目正确作答概率之间的关系,从而综合分析项目(条目)的难度、区分度等特征,筛选条目和编制测验分数等。通常采用测验分数作为这种潜在特质(能力水平)的估算。(2) 分析方法:根据慢性病患者测评结果,采用等级多分类资料的塞姆吉玛(Samejima)等级反应模型(其基本模型为 logistic 双参数模型)^[5-6],建立每个条目的特征曲线,以信息函数的大小并结合各条目区分度、难度等特征进行条目筛选。特征曲线基本公式为: $Px_i(\theta) = [1 + e^{-Da_i(\theta - b_{xi})}]^{-1} - [1 + e^{-Da_i(\theta - b_{xi})}]$ 。其中, θ 为被测试者的潜在特质(能力水平); x_i 为测试分数($x_i = 1, 2, 3, \dots, m$); $Px_i(\theta)$ 为能力水平为 θ 的被测试者在项目 i 上正确作答的概率; a_i 为第 i 个项目的区分度,指测试条目对被测试者潜在特性的区分能力。 a_i 越大,则特征曲线起伏越大,表明项目与特质 θ 之间关联越密切。 b_{xi} 为第 x_i 个等级的难度,它表示被测试者以 50% 概率选择第 i 个及第 i 个以上等级答案时所对应特质 θ 的值,其平均值越大,表明该条目难度越小,反之则该条目难度越大。在心理测评和生存质量领域的等级资料,难度相当于条目不同等级的平均得分程度。当 θ 取标准分数量表时, b_i 和 θ 的取值在 $-2 \sim +2$ ^[9]。(3) 信息函数及测验信息函数:第 i 个条目的信息函数及测验信息函数计算公式为: $I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$, $I(\theta) = \sum I_i(\theta) = \sum P'_i(\theta)^2 / P_i(\theta) [1 - P_i(\theta)]$ 。其中 $I_i(\theta)$ 为条目信息函数, $P'_i(\theta)$ 是条目反应函数对 θ 的导数, $Q_i(\theta) = 1 - P_i(\theta)$; 测验信息函数 $I(\theta) = 1/S_x^2(\theta)$, 它是测量标准误的倒数,表示测验信息函数值越大,对被测试者能力水平所做估计的误差越小。本研究选取 5 个 θ 参数分别为 -2 、 -1 、 0 、 1 、 2 的代表性位置点,分别计算各条目各点的信息函数值及其平均值。估计出各参数后,即可得到特征曲线公式,从而按上述公式求得条目信息函数 $I_i(\theta)$ 。

1.2.3 条目筛选判定标准 参照相关文献[8, 10] 整个测验(量表)的信息量 > 25 表明测评条目质量良好; 信息量 $16 \sim 25$ 表明测评条目有待改进; 信息量 < 16 表明测评条目很差。本研究预计 QLICD-GM 第 2 版最多包含 34 个条目,分别以 16

和 25 除以 34,得到每个条目的平均信息量,其中条目平均信息量 > 0.74 ($25/34$) 的条目判定为优, < 0.47 ($16/34$) 的条目判定为差, 0.47 ($16/34$) ~ 0.74 ($25/34$) 的条目判定为好。

1.3 分析工具与策略 应用 MULTILOG 7.0 软件进行统计分析。采用塞姆吉玛等级反应模型计算每个条目难度、区分度系数和信息量,绘制项目特征曲线。分别按 3 个维度进行拟合与分析,其中,第 30 个条目“患病或治疗影响您的性生活吗”因同时涉及躯体功能、心理功能和社会功能 3 个维度未进行 IRT 分析。

2 结果

2.1 一般情况 620 例慢性病患者中,男性 379 例,占 61.13%,女性 241 例,占 38.87%; 年龄 16 ~ 192 岁,平均年龄(53.4 ± 16.8) 岁; 其中慢性阻塞性肺病患者 43 例,占 6.93%,慢性肺原性心脏病患者 35 例,占 5.65%,消化性溃疡患者 109 例,占 17.58%,慢性胃炎患者 117 例,占 18.87%,肠易激综合征患者 102 例,占 16.45%,高血压患者 117 例,占 18.87%,冠心病患者 97 例,占 15.65%。

2.2 各条目得分及难度与区分度 分析结果表明,29 个条目的区分度均为 1.2 ~ 1.9,表明各条目区分度均较好。其中躯体功能各条目的区分度系数为 1.2,心理功能各条目区分度系数为 1.9,社会功能各条目的区分度系数分别为 1.45 和 1.56。难度系数从 1 ~ 4 级均呈正向递增,不存在逆反阈值,取值范围为 $-3.05 \sim 2.18$,表明 QLICD 量表共性模块的各条目选项设置合理、难度合适,分析模型选择正确。

2.3 各条目平均信息量与筛选(表 1) 在进行项目反应分析的 29 个条目中,躯体功能维度 8 个条目的平均信息量为 0.351 ~ 0.426,均未达到 0.47,均判定为差。心理功能维度 11 个条目的平均信息量均较高,其中 2 个条目分别为 0.712、0.716,其他条目为 0.744 ~ 0.909,表明 11 个条目均处于好或优的水平。社会功能维度 10 个条目中,有 2 个条目的平均信息量分别为 0.448、0.453,未超过 0.47; 其他 8 个条目为 0.526 ~ 0.644,处于较好水平。上述结果提示,心理功能 11 个条目和社会功能 8 个条目可直接进入 QLICD 第 2 版的共性模块。选择躯体功能维度平均信息量较高的 4 个条目(PH2、PH6、PH7、PH8) 作为保留条目。

表 1 QLICD-GM 各条目得分及不同位置点的信息量、平均信息量及条目筛选

维度/条目	条目得分		θ					平均信 息量	筛选 结果
	\bar{x}	s	-2	-1	0	1	2		
躯体功能(PH1-PH8)									
料理自己的日常生活	3.84	1.32	0.446	0.461	0.432	0.311	0.145	0.359	差
感到容易疲乏	3.08	1.19	0.417	0.447	0.439	0.422	0.404	0.426	差
走 ≥ 800 m 的路程	3.81	1.43	0.425	0.463	0.433	0.299	0.134	0.351	差
爬楼梯	3.62	1.35	0.428	0.459	0.449	0.390	0.230	0.391	差
依赖药物维持日常活动	3.49	1.33	0.416	0.459	0.455	0.407	0.255	0.398	差
胃口	2.88	1.24	0.380	0.445	0.452	0.447	0.392	0.423	差
自己的睡眠	2.66	1.28	0.309	0.431	0.453	0.453	0.416	0.412	差
疼痛或不适	3.18	1.11	0.434	0.440	0.437	0.429	0.381	0.424	差
心理功能(PS1-PS11)									
疾病影响您的脑力活动	3.86	1.11	1.047	1.083	1.028	0.807	0.226	0.838	优
疾病使您在精神上感到痛苦	3.30	1.22	0.774	1.101	1.074	0.981	0.617	0.909	优
感到孤独无助	4.19	1.14	1.088	1.115	0.990	0.293	0.074	0.712	好
感到悲观失望	4.20	1.10	1.079	1.101	1.005	0.453	0.089	0.745	优
对自己的疾病感到担忧	3.45	1.21	0.784	1.113	1.047	0.979	0.559	0.896	优
感到烦躁或容易发脾气	3.55	1.18	0.960	1.094	1.040	0.979	0.452	0.905	优
感到紧张焦虑	3.78	1.17	0.995	1.095	1.018	0.885	0.288	0.856	优
担心药物的副作用	4.11	1.04	1.094	1.081	1.043	0.547	0.115	0.776	优
自己是家庭的负担	4.15	1.13	1.081	1.121	1.008	0.428	0.082	0.744	优
觉得自卑	4.29	1.04	1.109	1.087	0.959	0.360	0.066	0.716	好
将情绪压在心底不表现出来	3.94	1.09	1.078	1.060	1.043	0.756	0.197	0.827	优
社会功能(SO1-SO10)									
疾病或治疗影响您工作	3.42	1.28	0.665	0.759	0.743	0.678	0.362	0.641	好
承担相应的家庭角色	3.22	1.30	0.546	0.656	0.646	0.602	0.398	0.570	好
对家人的关怀和照顾	3.81	1.17	0.740	0.736	0.738	0.549	0.201	0.593	好
和家人的关系	4.22	1.03	0.660	0.629	0.562	0.297	0.092	0.448	差
获得物质和情感上的帮助	4.23	1.02	0.658	0.645	0.570	0.301	0.093	0.453	差
影响参加喜欢的业余活动	3.21	1.32	0.574	0.752	0.756	0.712	0.425	0.644	好
积极乐观地看待疾病	3.22	1.27	0.573	0.645	0.652	0.613	0.398	0.576	好
接受的医疗诊治对疾病帮助	3.57	1.17	0.609	0.646	0.636	0.540	0.257	0.538	好
经济问题影响了您的生活	3.34	1.35	0.565	0.761	0.752	0.684	0.376	0.628	好
得到亲戚朋友的关心和支持	3.69	1.22	0.639	0.652	0.634	0.497	0.210	0.526	好

3 讨论

项目反应理论在心理与教育测量中广泛应用,具有样本自由性与结果准确性等优点^[5-6]。它是包含项目参数和被试参数的反应函数,不仅考虑了项目参数间关系,而且考虑了项目参数与被试能力间的关系,是一个关于项目质量的、包含难度和区分度等各方面作用的统一的综合指标,这是经典测量理论(classical test theory, CTT)所不具有的优越性^[11-12]。本研究采用(IRT)分析 QLICD 量表共性模块,以信息函数概念代替了信度概念,通过建立并比较被测试者对项目的反应与潜在特质之间的数量关系,获得各条目信息量,估计其可靠程度,为编制新版本 QLICD 量表筛选更为科学合理的条目。结果表明 29 个测试条目对被测试者潜在特性的区分度均为 1.2~1.9,表明各条目区分度均较好;难度取值范围为 -3.05~2.18,表明 QLICD 量表共性模块各条目设置合理、难度合适。

项目反应理论应用的一个前提条件就是单维性检查,可采用因子分析来进行。按 Hambleton 的标准^[13],若第 1 特征值与第 2 特征值之比 > 3 ,则认为符合单维性假设,可以进行 IRT 分析。本文若整个量表一起进行 IRT 分析则不满足单维性(第 1 特征值与第 2 特征值之比为 $7.74/2.88 = 2.69$),为此,分别按生理功能、心理功能和社会功能 3 个维度进行 IRT 分析均可满足单维性。

对于条目的区分度与难度的取值大小尚无统一的标准,但有一些普遍采用的原则^[8,10]。一般来说如果 $a_i < 0.3$ 则认为题目的区分度过小,应在 1~2 为宜; $bi > 2.95$ 或者 $bi < -2.95$ 可以认为题目过难或过易。本文除 1 个条目 SO5 难度稍低(-3.05)外,其余均符合要求,其原因可能是大部分被调查者出现在一个等级(329/620),形成一定的“地板”效应。

条目筛选结果提示,心理功能 11 个条目和社会功能 8 个条目可直接进入新版本 QLICD 量表共性

模块。而躯体功能维度信息量均偏低,没有直接入选的条目。分析其原因,可能与这些条目各等级的回答人数比较居中,条目得分变异不大,因而测评信息量较小有关。为了保证共性模块各维度的完整性,在区分度适宜、条目分级较为合理的前提下,结合其项目特征曲线和信息函数曲线,选择其中平均信息量较高的 4 个条目,作为保留条目进入新版本 QLICD 量表共性模块。此外,第 30 个条目虽没有参加 IRT 分析,但经专家审评和预调查后,认为对慢性病患者的生活质量有影响,因此也予以保留。共选出 24 个条目进入 QLICD 第 2 版的共性模块。但项目反应理论对样本含量的要求较高,一般要求 > 1 000 例。本研究样本量还不够大,需进一步观察和分析。

参考文献

[1] 万崇华,杨铮,杨玉萍,等.慢性病患者生命质量测定量表体系共性模块的考评[J].中国行为医学科学,2007,16(6):559-561.
 [2] 杨铮,李晓梅,万崇华,等.慢性阻塞性肺病患者生命质量测定量表 QLICD-COPD 的研制与考评[J].中国全科医学,2007,10(13):1080-1083.

[3] 杨瑞雪,潘家华,万崇华,等.高血压患者生命质量量表研制及评价[J].中国公共卫生,2008,24(3):266-269.
 [4] 高丽,万崇华,周曾芬,等.慢性病生命质量量表中慢性胃炎量表研制[J].中国公共卫生,2008,24(12):1447-1449.
 [5] 漆书青,戴海崎,丁树良.现代教育与心理测量学原理[M].北京:高等教育出版社,2002:79-91.
 [6] 顾海根.心理与教育测量[M].北京:北京大学出版社,2008:110-134.
 [7] Sébille V,Hardouin JB,Le Neel T,et al.Methodological issues regarding power of classical test theory(CTT) and item response theory(IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients - a simulation study[J].BMC Medical Research Methodology,2010,10(1):1186-1191.
 [8] 涂冬波,蔡艳.信息函数在标准参照测验中的应用研究[J].江西师范大学学报,2005,29(2):167-171.
 [9] 傅德荣,章慧敏.教育信息处理[M].北京:高等教育出版社,2001:122.
 [10] 刘炳伦,郝伟,杨德森,等.网络依赖诊断量表初步编制[J].中国临床心理学杂志,2006,14(3):227-232.
 [11] 杨静.三种教育与心理测量理论比较研究[J].中国考试,2006(6):33-35.
 [12] 熊江玲.经典测量理论、概化理论及项目反映理论比较研究[J].求索,2004(4):99-100.
 [13] Hamilton M. Development of a rating scale for primary depressive illness[J].Br J Soc Clin Psychol,1967,6(4):278-296.

收稿日期:2012-02-16

(孔繁学编辑 刘铁校对)

• 调查报告与分析 •

氟砷污染对暴露人群骨代谢交互作用*

曾奇兵,喻仙,杨望,洪峰

摘要:目的 探讨氟砷对暴露人群骨代谢的交互作用。方法 2009 年选择贵州省清镇市流长乡 198 例氟砷联合暴露者作为研究对象,分别检测尿氟、尿砷及骨代谢效应标志尿羟脯氨酸(UHYP)和尿 I 型胶原交联氨基末端肽(UNTX)、骨强度指数(STI),以尿氟(UF)、尿砷(UAs)分组。结果 UNTX 含量(26.82 nmol/mmol Cr)在低氟低砷水平低于氟、砷单独作用水平之和 [(8.97+20.91) nmol/mmol Cr],UNTX 含量(45.19、54.79 nmol/mmol Cr)在低氟高砷、高氟低砷水平高于氟、砷单独作用水平之和 [(20.51+20.91)、(8.97+33.68) nmol/mmol Cr],且氟、砷对 UNTX 的交互作用差异有统计学意义 (P<0.01);氟、砷对 UHYP、STI 的交互作用差异均无统计学意义 (P>0.05)。结论 在该暴露水平下,氟、砷对 UNTX 在低氟低砷水平主要表现为拮抗作用,在低氟高砷、高氟低砷水平主要表现为协同作用;氟、砷对 UHYP、STI 不存在交互作用。

关键词:氟;砷;骨代谢;交互作用

中图分类号:R 599.1

文献标志码:A

文章编号:1001-0580(2012)11-1480-03

Interactive effect of fluoride and arsenic pollution on bone metabolism in exposed population

ZENG Qi-bing, YU Xian, YANG Jun, et al (Department of Occupational Health and Industry Medicine, School of Public Health, Guiyang Medical University, Guiyang, Guizhou Province 550004, China)

Abstract: Objective To explore the interactive effect of fluoride and arsenic pollution on bone metabolism in exposed population. Methods Totally 198 people with both fluoride and arsenic exposure were selected from Liuchang town of Qinzhen city, Guizhou province. Urinary fluoride (UF), urinary arsenic (UAs), urinary hydroxyproline (UHYP), urinary cross-linked N-telopeptides of type I collagen (UNTX) and bone strength index (STI) of the people were detected. The people were divided into groups according to the contents of UF and UAs. Results The content of UNTX in low

* 基金项目:国家自然科学基金(30660163,81072248);贵州省优秀青年科技人才培养计划专项资金(黔科合人字[2007]112号);贵州省科学技术基金(黔科通[2006]63号);贵州省教育厅自然科学科研基金(黔教高发[2005]328号)

作者单位:贵阳医学院公共卫生学院职业卫生与职业医学教研室,贵州 贵阳 550004

作者简介:曾奇兵(1985-),男,贵州遵义人,硕士在读,研究方向:氟砷联合毒作用机制。

通讯作者:洪峰, E-mail: hongfeng-73@163.com