基于条目反应理论、计算机自适应测试及数据模拟的胃痞患者报告结局量表简短版的研制与评价

侯政昆[1]　刘凤斌[1]　陈新林[2]　李筱颖[3]　罗　迪[4]　谢　地[5]　罗　芳[6]　车晓璐[6]

摘要　目的　采用现代测量理论技术研制并评价胃痞患者报告结局量表(简短版),以期为相关临床和科研提供质量较高且管理和回答负担较小的结局测评工具。**方法**　在前期已成功构建胃痞患者报告结局量表全条目库版的基础上,2011 年 9 月—2012 年 3 月采用描述性研究设计,应用社会学资料调查表及胃痞患者报告结局量表全条目库版进行现场和网络调查并收集数据。确定量表终止测试参数后通过条目区分度、拟合残差值、条目信息函数(item information curve, IIC)、条目特征曲线(item characteristic curve, ICC)及计算机自适应测试(computerized adaptive testing, CAT)选择排名等参数筛选确定量表条目;在进行条目反应理论(item response theory, IRT)前提假设评估后,采用 IIC、ICC、难度分布、条目—回答者相关图、阈值顺序等参数评估量表属性。**结果**　最终有 331 例患者记录进入分析,平均(31.99 ±10.29)岁,男性 56.3%(186/331)。量表测试终止标准为 Max SE =0.2 或最大条目数达到 16 条;经条目筛选得到 15 个条目的量表简短版,包含胃痞症状(8 个条目)及胃痞影响(7 个条目)共 15 个条目。在属性评估中,量表两个方面的单维度性、局部独立性、单调性较好,IRT 参数评价显示量表有较好的 IIC 和 ICC;难度分布向中心集中,难易跨度合理;条目分布和回答等级与受试者有良好的交互作用;大多数条目的阈值顺序满足测评要求。**结论**　胃痞患者报告结局量表简短版具有良好的方法学质量和报告质量,量表可用于成人胃痞患者症状和疾病影响的临床和科研评价。

关键词　胃痞;患者报告结局量表;简短版;条目反应理论;计算机自适应测试;数据模拟

Development and Evaluation of GS-SF for Gastric Stuffiness (Wei Pi) Patients Based on Item Response Theory, Computerized Adaptive Testing, and Data Simluation

HOU Zheng-kun[1], LIU Feng-bin[1], CHEN Xin-lin[2], LI Xiao-ying[3], LUO Di[4], XIE Di[5], LUO Fang[6], and CHE Xiao-lu[6]　*1 Department of Gastroenterology, First Affiliated Hospital, Guangzhou University of Chinese Medicine, Guangzhou (510405); 2 Department of Preventive Medicine and Health Statistics, College of Basic Medical Sciences, Guangzhou University of Chinese Medicine, Guangzhou (510006); 3 Department of Gastroenterology, Jiangmen Hospital of Traditional Chinese Medicine, Guangdong (529000); 4 Department of Gastroenterology, Huizhou Hospital of Traditional Chinese Medicine, Guangdong (516001); 5 Department of Science and Education, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou (450007); 6 Postgraduate College, Guangzhou University of Chinese Medicine, Guangzhou (510000)*

ABSTRACT　Objective　To develop and evaluate the short version of patient reported outcomes (PROs) questionnaire for gastric stuffiness (Wei Pi) patients with modern test theory and technologies, hoping to provide testing tools for related clinical practice and scientific researches with higher quality and less administrative and response burdens. Methods　Using descriptive study design, clinical data

基金项目:国家自然科学基金青年项目(No. 81303148);国家自然科学基金面上项目(No.81373786);广东省中医药管理局科研课题资助项目(No.20132177);中央财政支持地方高校发展专项资金中医内科学特色重点学科建设项目(粤财教[2014]276 号);广东省中医药建设中医药强省专项资金中医优势病种突破项目(粤中医函[2015]19 号)

作者单位:1. 广州中医药大学第一附属医院脾胃病科(广州 510405);2. 广州中医药大学基础医学院预防医学和卫生统计学教研室(广州 510006);3. 暨南大学附属江门中医院脾胃病科(广东 529000);4. 惠州市中医医院脾胃病科(广东 516001);5. 郑州大学附属郑州中心医院科教科(郑州 450007);6. 广州中医药大学研究生院(广州 510000)

通讯作者:刘凤斌,Tel:020 -36591109,E-mail:liufb163@163.com

DOI: 10.7661/CJIM.2015.07.0806

中国中西医结合杂志

were collected with sociological questionnaire and previous developed full items version of PROs instrument for gastric stuffiness (Wei Pi) patients via field and online surveys between Sep 2011 and Mar 2012. The statistical analysis group identified the termination parameters firstly, and then selected items with discrimination, fitting residual, item information curve (IIC), item characteristic curve (ICC), and the rank of computerized adaptive testing (CAT) select proportion, etc. After assumption evaluation of item response theory (IRT), IIC, ICC, difficulty coefficient distribution, items-response relation and thresholds, etc. were used for psychometric evaluation of instrument. Results A total of 331 patients [Ages: 31.99 ±10.29 yrs; Male: 186 (56.3%)] were enrolled in statistical analysis. The test termination criterion was Max SE =0.2 or Max items number =16. After items selection, a 15-item short version of instrument, which contains symptoms facet (8 items) and impact facet (7 items) was generated. With good unidimensionality, local independence, and monotonicity, the IIC and ICC in IRT analysis showed good working capability of the questionnaire. The difficulty coefficient distribution and items-response relation were also rational, as well as response thresholds. Conclusions The short version of PROs instrument for adult gastric stuffiness (Wei Pi) patients was successfully developed and assessed. The instrument with good methodological and reporting quality could be used in clinical and scientific evaluating their symptoms and impact.

KEYWORDS gastric stuffiness; patient reported outcomes; questionnaire; short version; item response theory; computerized adaptive testing; data simulation

胃痞(gastric stuffiness, GS)以患者自觉胃脘部闷满胀塞,触按之无形柔软,压之不痛为主要表现,西医可对应功能性消化不良、急/慢性胃炎、胃神经官能症、胃肠功能紊乱、慢性肝炎、胃电频率降低或胃肠动力不足等一系列疾病[1]。GS 在临床上有较高发病率[2,3],虽较少带来严重身体伤害,但在较多情况下仍对患者的生理、心理状态、生活工作和社会活动等造成影响,故针对 GS 的研究有较大的临床价值和社会意义。本病的临床特点使其在临床干预和结局评价中需更多关注患者自身内在的主观感受,但既往临床评价仍主要局限在传统的"有效率"指标(痊愈/治愈、显效、有效、无效),且在其分级、组成(症状体征、胃镜检查、钡餐检查等)、标准等方面存在差异[4,5]。因此,完善 GS 临床结局评价体系,提高结论真实性和可重复性已成为本病临床研究的关键所在,而患者报告结局量表是其重要方向。

目前,在科学研究和临床实践中极少量表直接测评本病的症状、体征和生存质量,且来自 GS 患者自身的报告显示其更强调胃脘胀闷及由此带来的生理、心理、社会活动及持久的不良影响,有更高的自身感觉权重和健康诉求。因此有必要针对本病研制特异性量表。传统的量表研制方法数据挖掘能力欠佳,无法反映真实的客观世界和受试者的个体化特征,且存在较大的管理和回答负担。条目反应理论(item response theory, IRT)可对受试者特质和条目难度同时做出评价,并通过等值处理使条目具有可比性;计算机自适应测试(computerized adaptive testing, CAT)可在同样的测评精度下将测试时间缩短 50% ~ 90%,降低测试费用和负担,并可提供更高的测试安全性和更好的测试感受[6-8]。本研究旨在研制基于 IRT 和 CAT 的 GS 患者报告结局量表,并建立基于经典和现代测试理论、西医和中医学理论共同指导下的患者报告结局量表研制方法学。在同期研究中,项目小组已通过成立研究小组、预设基本特征、定义基本概念术语、构建初步理论框架、构建条目库和条目初筛及条目定性评阅等步骤,建立了 GS 患者报告结局量表的全条目库版以备 CAT 测试。为进一步扩大量表的临床和科研实用性,拓展其在传统纸张版测评领域的应用范围,本研究采用 IRT、CAT 和数据模拟方法技术研制 GS 患者报告结局量表的简短版,拟为 GS 临床结局评价提供具有良好科学属性的测评工具。

资料与方法

1 诊断标准 本项目组参照相关教材和共识标准等论述,确定 GS 诊断需同时满足以下条件:(1)症状:患者自觉胃脘部(心和剑突下,脐以上)闷满胀塞,上述感觉可单独或合并出现;(2)体征:触按之无形柔软,压之不痛[9,10]。对应西医疾病中功能性消化不良(参照罗马Ⅲ标准[11]),慢性胃炎(参照《中国慢性胃炎共识意见》标准[12])。

2 纳入与排除标准 纳入标准:需同时满足以下条件:(1)年龄≥18 岁;(2)符合 GS 诊断标准;(3)能够说和阅读汉语;(4)同意签署知情同意书。

排除标准:(1)调查员认为依据参与者目前的医疗或心理状态不适合进行研究;(2)存在认知或其他损害(如视力等)而影响参与者完成自我报告。脱落标准:(1)患者自行要求退出;(2)调查者有充分理由认为受试者数据存在明显偏倚。剔除标准:(1)疾病误诊者;(2)研究期间新发出现排除标准者;(3)回答调查表的条目缺失比例达到 20% 以上者。

3　研究方法　采用描述性研究设计,2011 年 9 月—2012年 3 月采用现场调查及网络调查两种方式收集数据。

3.1　现场调查　由经过统一培训合格的临床调查员对中医诊断为 GS 的患者,采用筛选表进行筛选,并与符合纳入标准者逐人沟通,确定参与者,随即发放病例报告表进行调查。患者填写完毕后现场回收调查表,并检查缺失项,若有缺失,则提醒患者补充完整。现场调查地点为广州中医药大学第一附属医院脾胃病科住院部和门诊部或患者认为合适的地点。

3.2　网络调查　首先建立研究的网络版本,然后通过网络搜索,寻找可能与本研究相关的人群(QQ 病友群;医学网站如爱爱医、医学考试在线),通过电子邮件、群发信息、公告板等形式发送邀请信,邀请患者通过网络完成调查。网络调查无地点限制。

4　测评工具

4.1　GS 患者社会学资料调查表　包括年龄、性别、民族、居住地、婚姻、教育程度、职业等信息。

4.2　GS 患者报告结局量表全条目库版[13]　本库共包含 2 个维度共 129 个条目,其中胃痞症状(GSS)方面 111 个条目,胃痞影响(GSPI)方面 18 个条目(可向作者索取)。

上述两种工具均有纸张版和网络版两种,分别用于现场调查和网络调查。数据收集方式采用工具管理,由患者自我报告,备用方式为调查员采访(只负责记录患者意见),但所有方式的数据均直接来源于参与者,由代理人评估、补充或解释者将被剔除。

5　数据分析及指标评价

5.1　数据描述分析　患者的社会学特征以均数 ± 标准差、频数(率)等表示。量表内部一致性信度分析采用克朗巴哈系数表示,结构效度分析以条目间相关阵、条目 - 总分相关阵、敏感性分析等表示。描述分析用 SPSS 11.0 实现。

5.2　量表简短版研制步骤及方法　首先,基于真实数据进行 *Post-hoc* 模拟以确定测量的终止标准,包括标准误(SE)值和最大条目数,前者通过 Firestar 和 R 软件实现,后者通过 CATSim 软件实现。再对各个条目的基本参数和运行能力进行评估,主要指标包括条目区分度、条目粗分均值和在 CAT 中被选择的比例排名;次要指标包括拟合残差值、条目信息函数(item information curve, IIC)、条目特征曲线(item characteristic curve, ICC)和条目内容。将条目区分度、条目粗分均值和 CAT 测试选择比例排名按照由高到低的顺序排列取最优者,其中条目区分度为核心指标。条目区分度、IIC 和 ICC 通过 Multilog 7.03 实现,拟合残差值通过 RUMM 2030 实现,CAT 测试选择比例通过 Firestar 和 R 软件实现,条目粗分均值通过 SPSS 11.0 实现。

5.3　条目反应理论的前提假设评价　单维度性(unidimensionality)评价采用证实性因子分析(confirmatory factor analysis, CFA),指标包括比较拟合指数(comparative fit index, CFI)、近似误差均方根(root mean square error of approximation, RMSEA)、路易斯指数(non-normed fit index, NNFI)等,若模型拟合严重欠佳则执行探索性因子分析;CFA 采用 LISREL 8.70 实现。局部独立性(local independence)评价采用因子分析,指标包括局部独立指数(partial separation index, PSI)和内部一致性信度,分别通过 RUMM 2030 和 SPSS 11.0 实现。单调性(monotonicity)评价采用局部加权回归散点平滑法(locally weighted scatterplot smoothing, LOWESS)观测各层面、方面、领域和维度二维变量之间的关系进行分析,采用 SPSS 11.0 软件实现。

5.4　条目反应理论的模型参数分析　对各条目的 ICC 和整体的 IIC 进行分析评价量表的条目功能,分析以 MULTILOG 7.03 软件实现。同时分析量表的条目难度分布、条目 - 回答者相关、阈值顺序等,综合评价条目自身、回答选项、回答者关系特征,分析通过 RUMM 2030 实现。

结　果

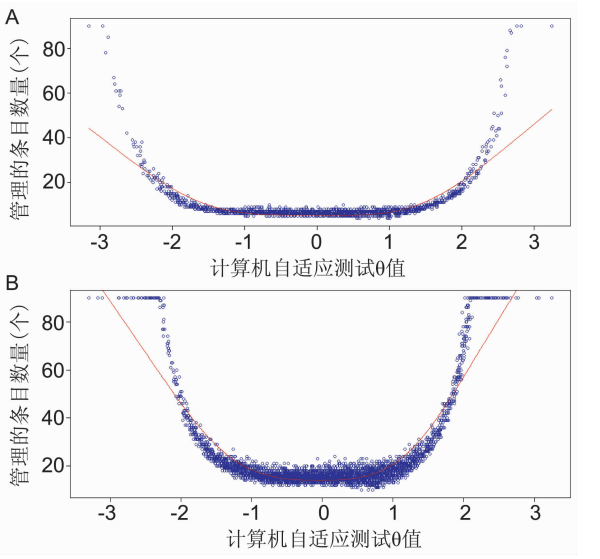
1　纳入患者的一般资料　共 554 例患者进入现场调查筛选,排除 204 例,脱落 48 例,剔除 31 例,共 271 例完成调查。网络调查通知发布于 19 个 QQ 病友群(合计 2 474 人,具体受众未知)和 2 个医学网站(爱爱医、医学考试在线,具体受众未知),回收 142 条研究记录,其中脱落 69 例,剔除 13 例,共 60 例完成调查。最终有 331 例患者的记录进入分析,患者年龄 18 ~66 岁,平均(31.99 ± 10.29)岁,男性 186 例(56.3%),汉族 316 例(95.5%)。在完成现场调查的 271 例患者中,功能性消化不良患者 149 例

(54.98%)、慢性胃炎患者 85 例(31.37%),其他患者 37 例(13.65%)。

2 确定测量终止标准参数

2.1 SE 值确定(图 1) 首先使用 Firestar 软件,以"均数为 0、标准差为 1、条目选择方法为 Max Posterior Weighted Info (MPWI)法、终止标准为 Max SE =0.3(或 0.2,对比分析以确定)"进行参数设定,产生 R 运算命令后再使用 R 软件运算分析,进而生成 5 000 例模拟数据及其测试信息、标准误和 theta(θ)值。分析显示 Max SE =0.3 时,模拟数据与真实数据的 θ 相关系数为 0.964,CAT 测试能力与条目关系图显示有 5 ~7 个条目组合可满足测量能力要求,并且在 θ 值为 -2 ~2 时相对稳定,平均管理条目数量为 7.229 个;Max SE =0.2 时,模拟数据与真实数据的 θ 相关系数为 0.988,CAT 测试能力与条目关系图显示有 14 ~15 个条目组合可满足测量能力要求,并且在 θ 值为 -1.5 ~1.5 时相对稳定,平均管理条目数量为 21.103 个。由于 Max SE =0.2 时 θ 相关系数相对较高,且 Max SE =0.3 存在使疾病严重程度偏高或偏低的患者难以真实反映自身信息而测试结束的局限,因此确定以 Max SE =0.2 为测试终止的标准之一。

2.2 最大条目数的确定(表 1) 采用 CATSim 软件进行 *post-hoc* 数据模拟,以 Max SE =0.2 为终止标准确定最大条目数量,结果显示均数为(29.798 ± 21.437)个,最少为 12 个,累计条目和比例分析显示条目组合以 12 ~20 最多。在此范围内对逐个条目组合(12 ~20)的 θ 和 SE 值与条目总库进行对比,结果显示两者分别在 14 ~20 个、15 ~20 个条目时相对稳定,其他各项指标变异不大。为保证最低的管理负担,综合各项参数,将简短版条目数量确定为 16 条,但可根据分析结果适当变化。



注:A 为 Max SE =0.3;B 为 Max SE =0.2

图 1 基于 CAT *Post-hoc* 模拟的 theta(θ)与管理条目数量关系

2.3 简短版量表条目的确定(表 2,图 2) 首先,界定条目筛选方法为 MPWI,将量表分为 GSS 和 GSPI 两个方面独立模拟运算,单个数据库模拟终止规则均为 Max Items No. =8 或 Max SE =0.2,可得到条目使用图及条目 CAT 选择比例顺序。结合条目的区分度、粗分均值、拟合残差值、*P* 值、条目信息函数和条目特征曲线等参数进行条目筛选,确定 16 个条目简表。但证实性因子分析显示 FC4GSS109 更多偏向 GSPI 方面(原属 GSS 方面),结合其内容效度、拟合残差值等特征予以删除。因此得到 15 个条目的胃痞患者报告结局量表的简短版(Short Form of Patients Reported Outcomes Questionnaire for Gastric Stuffiness Patients, GS-SF),包含 GSS 8 个条目,GSPI 7 个条目。

表 1 基于 CAT 技术的 *Post-hoc* 模拟不同条目组合与全 90 条目库的信息比较

条目数	θ 值					SE 值				
	r	ASD	AAD	MSD	RMSD	*r*	ASD	AAD	MSD	RMSD
12	0.876	-0.089	0.397	0.246	0.496	0.663	-0.123	0.125	0.003	0.051
13	0.889	-0.044	0.365	0.212	0.461	0.660	-0.115	0.116	0.003	0.052
14	0.914	-0.073	0.330	0.161	0.401	0.751	-0.107	0.108	0.002	0.043
15	0.945	-0.036	0.251	0.101	0.318	0.857	-0.084	0.084	0.001	0.033
16	0.929	-0.051	0.289	0.134	0.365	0.792	-0.096	0.096	0.002	0.040
17	0.930	-0.036	0.276	0.129	0.358	0.770	-0.091	0.091	0.002	0.043
18	0.939	-0.035	0.263	0.109	0.331	0.822	-0.086	0.086	0.001	0.037
19	0.932	-0.043	0.263	0.134	0.366	0.856	-0.082	0.083	0.001	0.033
20	0.935	-0.033	0.255	0.126	0.355	0.855	-0.079	0.079	0.001	0.034

注:*r*:相关系数;ASD:平均分配差异;AAD:平均绝对差异;MSD:差异的方差;RMSD:差异的标准差;SE:标准误

中国中西医结合杂志

表 2　GS-CAT 量表简表研制中条目筛选

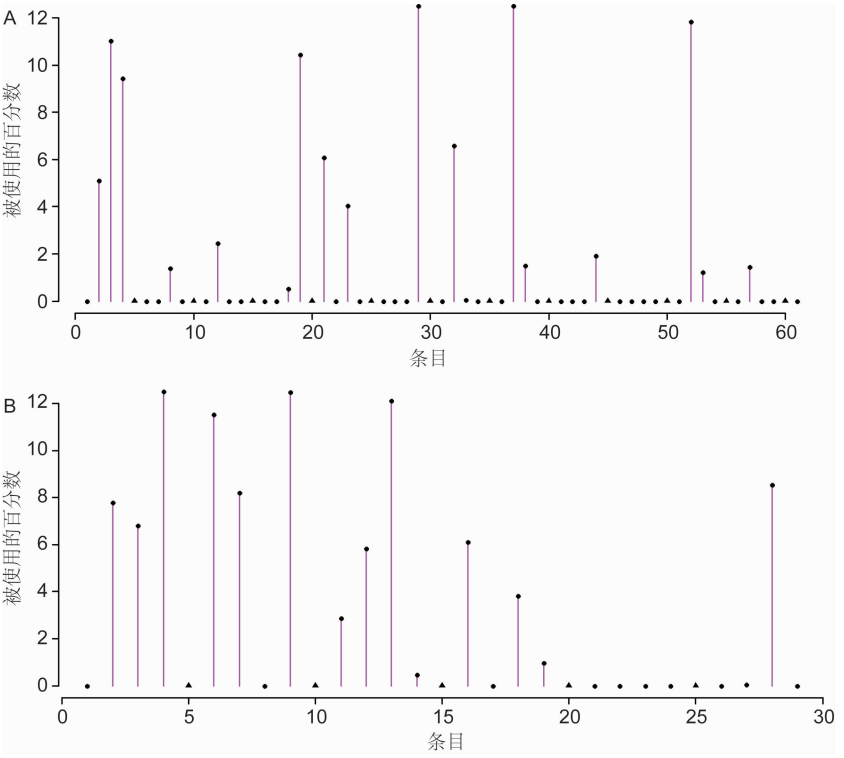
条目	a	RS	fit ris	*P*	IRT	CAT	条目	a	RS	fit ris	*P*	IRT	CAT
FC1GSS1	1.680	2.628	0.131	0.338			FC7GSS34	1.300	1.583	0.784	0.002		
FC1GSS2*	1.770	2.517	-0.269	0.119		9	FC7GSS35	1.390	2.148	0.987	0.592		
FC1GSS3*	1.920	2.187	-0.558	0.456		4	FC7GSS36	1.410	2.039	0.535	0.959		
FC1GSS4	1.980	2.036	-0.486	0.318		6	FC7GSS32	1.250	2.547	2.661	0.013	I	
FC1GSS6	1.570	2.221	0.649	0.161			FC7GSS41	1.630	1.695	1.026	0.669	I	
FC1GSS7	1.480	2.511	1.956	0.017	I		FC7GSS42	1.770	1.728	-0.424	0.110	I	
FC1GSS8	1.670	1.789	-0.852	0.403			FC7GSS43	1.840	1.873	-0.857	0.080		3
FC1GSS9*	1.660	1.958	-0.835	0.051		15	FC7GSS44*	1.760	1.900	-0.561	0.052		16
FC1GSS10	1.180	2.133	5.671	<0.001	I,C		FC7GSS29	1.100	1.770	2.677	0.251	I,C	
FC5GSS53	1.410	1.943	1.115	0.611	I		FC7GSS30	0.890	1.378	2.746	0.021	I,C	
FC5GSS60	1.300	1.710	0.674	0.577			FC7GSS37	1.410	1.782	0.351	0.624	I	
FC5GSS62*	1.580	2.190	1.146	0.589	I	11	FC7GSS40	1.780	1.692	-0.556	0.160		14
FC5GSS63	1.120	1.341	1.167	0.144	I,C		FC7GSS13	1.260	1.991	1.356	0.112		
FC5GSS64	1.190	1.405	0.688	0.371	I		FC7GSS31	1.320	2.079	1.462	0.082	I	
FC5GSS99	1.060	1.396	0.853	0.087	I,C		FC7GSS12	1.500	1.429	-0.369	0.566	I	
FC5GSS66	1.130	1.879	2.980	0.056	C		FC7GSS15	1.230	2.163	1.084	0.176	I	
FC5GSS68	1.460	1.689	0.147	0.089	I		FC3GSPI7	1.980	1.849	-2.813	<0.001	I	
FC5GSS71	1.760	1.798	-1.011	0.058	I	17	FC3GSPI8*	1.480	1.441	-1.172	0.007	I	7
FC5GSS73	1.820	1.634	-1.007	0.140		5	FC3GSPI9	2.090	1.885	-2.218	<0.001	I	8
FC5GSS74	1.340	1.810	0.389	0.806			FC3GSPI10	1.680	1.553	-1.680	0.107	I	1
FC5GSS75*	1.800	2.121	-0.425	0.712		8	FC4GSS106	1.250	1.740	-0.824	0.005	I,C	
FC5GSS78	1.360	2.323	1.360	0.042	I		FC4GSS108	2.400	1.644	-3.116	<0.001	I	4
FC5GSS80	1.740	1.764	-0.987	0.407		10	FC4GSS109*	2.310	1.659	-3.446	<0.001		6
FC5GSS82	1.300	1.879	1.503	0.755	I		FC4GSS110	1.410	2.021	-0.460	0.001		
FC5GSS84	1.430	1.698	0.017	0.406	I		FC4GSS111	2.040	1.782	-2.074	<0.001	I	2
FC5GSS86	1.510	1.589	-0.204	0.883			FC8GSS54	1.390	2.221	-0.278	<0.001		
FC6GSS45	1.560	2.248	1.379	0.434			FC8GSS65	1.090	1.761	0.665	<0.001	I	12
FC6GSS46	1.480	2.136	1.068	0.827			FC8GSS69	0.980	1.792	0.980	0.001	I	10
FC6GSS47*	2.200	2.257	-1.539	0.001		2	FC8GSS87*	2.340	1.900	-2.008	<0.001	I	
FC6GSS48	1.640	1.770	-0.052	0.258			FC8GSS88*	2.400	2.517	-2.081	<0.001		
FC6GSS49	1.460	1.583	0.374	0.715			FC8GSS89	1.850	2.124	-1.350	<0.001	I	
FC6GSS50	1.870	1.858	-1.199	0.460	I	7	FC8GSS90	2.580	1.716	-3.104	<0.001		9
FC6GSS51	1.770	1.795	0.666	0.941			FC8GSS96	2.650	1.955	-2.317	<0.001	I	
FC6GSS52	1.730	1.746	-0.903	0.113	I		FC8GSS97*	3.000	2.332	-2.087	<0.001		11
FC6GSS55	1.520	2.018	0.328	0.006	I		FC8GSS98	2.880	2.236	-1.977	<0.001		13
FC6GSS56	1.040	1.677	1.883	0.034	C		FC8GSS100	1.230	1.695	-0.465	0.024	I,C	
FC6GSS57*	2.340	2.172	-2.150	0.002			FC8GSS101	1.500	1.544	-1.454	0.003		
FC6GSS58	1.560	2.263	0.165	0.298		13	FC8GSS102	1.270	1.369	-1.772	0.004	I	
FC6GSS59	1.780	1.900	-0.224	0.013	I		FC8GSS103	1.190	1.511	-0.173	0.005		
FC7GSS11	1.370	2.100	1.579	0.178			FC8GSS104*	1.470	2.517	-0.265	0.003	I	
FC7GSS38	1.580	2.369	-0.359	0.551	I		FC8GSS92*	2.260	1.921	-1.423	<0.001		
FC7GSS39	1.860	2.006	-1.041	0.530	I		FC8GSS93*	2.520	1.964	-1.769	<0.001		
FC7GSS77	1.450	1.752	1.485	0.938			FC8GSS94	2.140	1.846	-1.643	<0.001		
FC7GSS91	1.570	2.296	0.025	0.137		12	FC8GSS95	2.260	1.849	-2.110	<0.001		
FC7GSS33	1.520	2.112	-0.208	0.790	I		FC8GSS105	1.080	1.671	0.504	0.069	I,C	5

注：a：区分度；RS：粗分均值；fit ris：拟合残差值；I：条目信息函数欠佳；C：条目特征曲线欠佳；CAT：计算机自适应测试的选择比例排名顺序；* 号表示从条目总库中筛选出的简短版量表条目

3　简短版量表的科学属性评价

3.1　经典测试理论下的量表属性评估

GS-SF 的条目整体均数为 2.126（方差为 0.085），其中 GSS 方面均数为 2.163（方差为 0.035），GSPI 方面均数为 2.085（方差为 0.154）。一致性信度分析显示总克朗巴哈系数为 0.886，其中 GSS 方面为 0.845，GSPI 方面为 0.828，敏感性分析显示无条目删除后量表的系数大于 0.886（范围：0.874 ~0.886）。结构效度分析



注:A 为胃痞症状;B 为胃痞影响

图 2　CAT *Post-hoc* 模拟使用的量表条目

显示校正条目 - 总分相关系数范围为 0.357 ~ 0.661,呈中等强度相关,条目间相关阵、条目 - 总分的相关阵分析显示条目在工具中有良好的运行能力。

3.2　条目反应理论的前提假设评价

(1)单维度性评价　探索性因子分析显示量表总的 KMO = 0.895,特征根比值 = 3.551,其中 GSS 方面 KMO = 0.875,特征根比值 = 4.426,GSPI 方面 KMO = 0.846,特征根比值 = 4.049。证实性因子分析显示量表的 NNFI = 0.954、CFI = 0.961、调整拟合优度指数(adjusted goodness of fit index, AGFI) = 0.966、RMSEA = 0.063,且条目结构无修正建议。以上分析显示简表有良好单维度性。

(2)局部独立性评价　分析显示量表总的 PSI 为 0.864,其中 GSS 方面为 0.795,GSPI 方面为 0.774,结合各克朗巴哈系数综合认为量表局部独立性较好。

(3)单调性评价　LOWESS 法观测显示量表的二维变量之间呈良好的同向关系,提示有较好的单调性。

3.3　条目反应理论的主要参数分析(图 3 ~ 5)

量表各条目的 ICC分别如图 3 和图 4 所示。整体测试信息显示 GSS 方面在 θ = 0.6 或 0.8 时信息最大为 9.262,GSPI 方面在 θ = 1.2 时信息最大为 10.370(图 5),简表整体在 θ = 0.6 时信息最大为 13.329。

3.4　条目反应理论的其他参数分析(图 6 ~ 8)

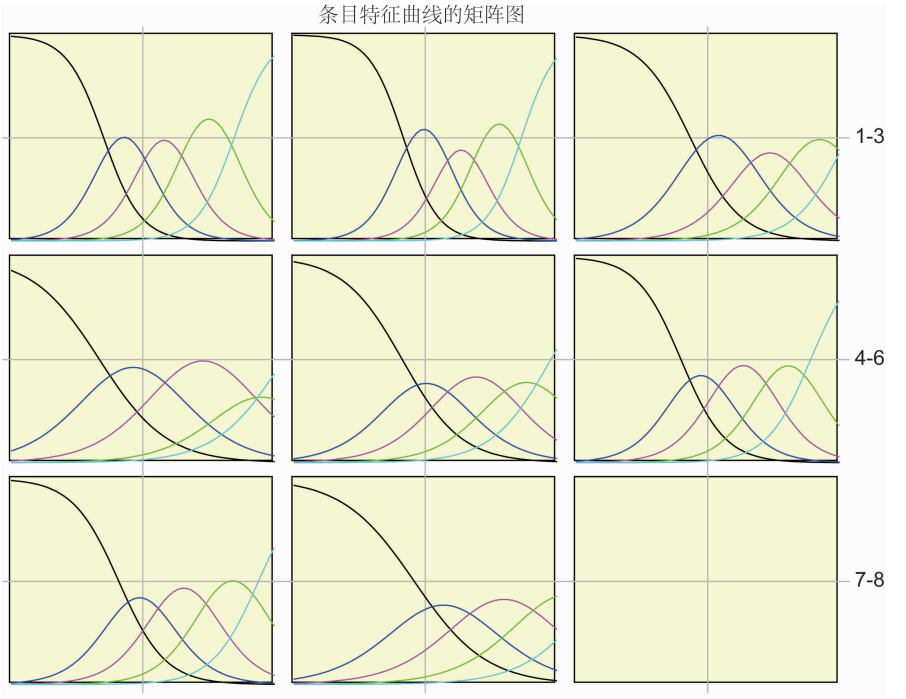
(1)量表的条目难度分布见图 6。

(2)量表的条目 - 回答者交互分析见图 7。

(3)量表的条目阈值顺序分析见图 8。

讨　论

本研究的整体结果和量表特征:本研究在前期建立胃痞患者报告结局量表全条目库版的基础上,为进一步降低其管理和回答负担,采用 IRT、CAT 和数据模拟技术等,并结合经典测试理论,成功研制并评价 GS-SF。本简表最终包含 2 个维度 15 个条目:胃痞症状方面 8 个条目,胃痞影响方面 7 个条目。本量表为疾病特异性量表,主要测评胃痞患者的症状和影响,适用于成人患者,无疾病亚型、中医证型、性别、民族、教育程度等限制,各条目、方面权重相同,回答环境无特殊要求,回答选项类型为 5 级等级尺度,计分为条目得分总和平均分,回忆时间为“在过去 2 周内”,临床管理时间无明确限制,载体为纸张、笔,数据收集方式为患者独立填写,调查员不辅助且不解释,主要信息提供者为患者,替代管理方式需明确记录,且本研究始终有双语背景研究者参与,具有良好的跨文化跨语言调适可行性。



注:曲线颜色(自左至右):黑色、蓝色、红色、绿色、青色分别表示各条目下回答选项 1、2、3、4、5 的特征;图 4 同

图 3 胃痞患者报告结局量表“胃痞症状”方面 8 个条目的条目特征曲线

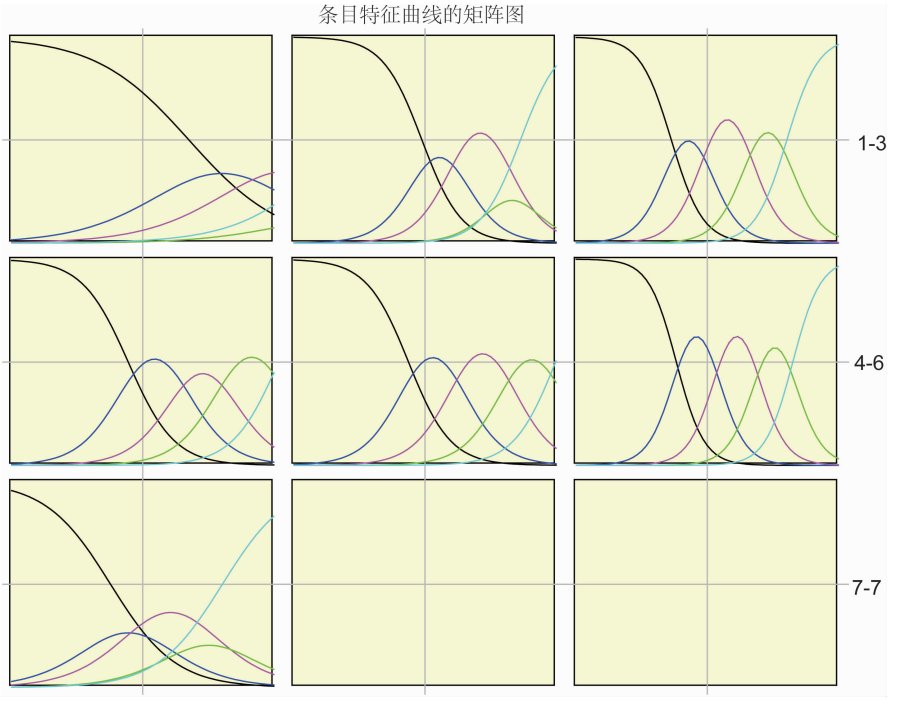
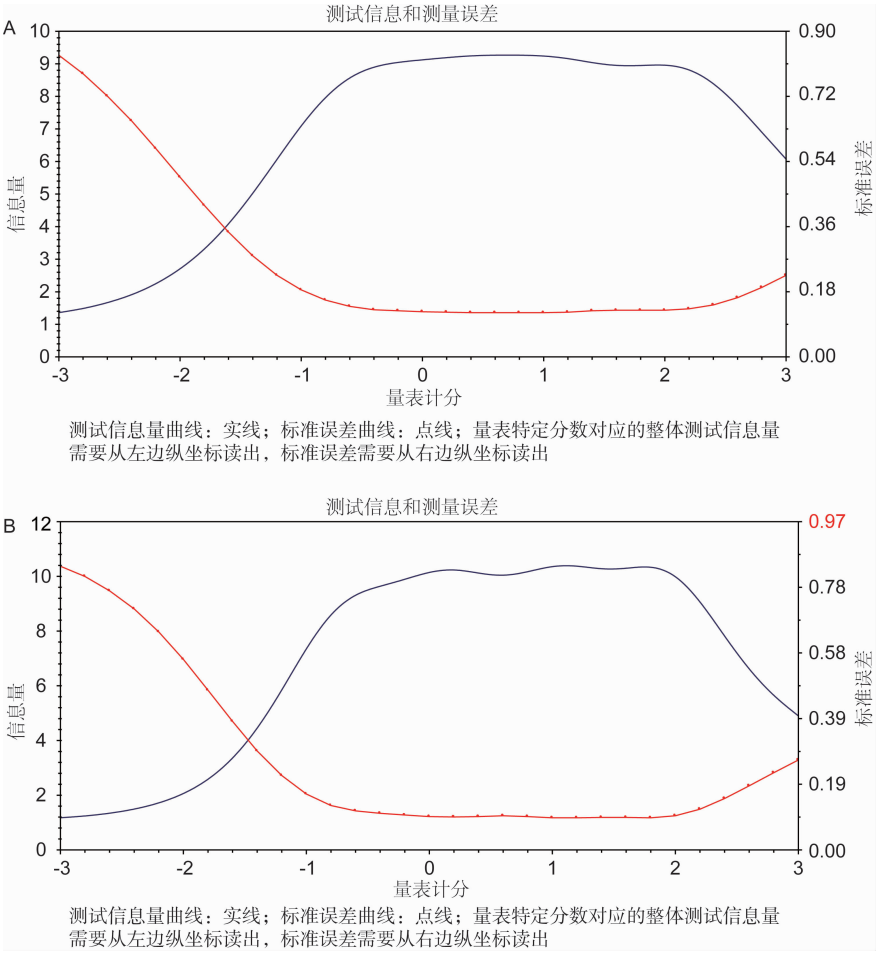


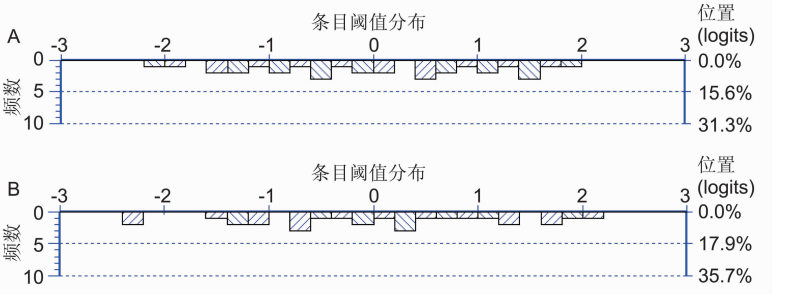
图 4 胃痞患者报告结局量表“胃痞影响”方面 7 个条目的条目特征曲线

本量表的属性和方法学评价:在内容效度方面,本量表的条目库采用文献回顾、临床病例回顾、患者调查、专家调查评阅、仓储法、精选法、核心小组研究、认知采访研究、信息定性评阅等多种定性研究方法,充分对数据进行挖掘解释,较好地避免了客观的信息偏倚。在结构效度方面,各条目与自身方面均具有良好的相关性。由于本研究主要采用现代测试理论分析,故未进行重测信度、效标效度、区分度、反应度等检测。



注:A 为胃痞症状;B 为胃痞影响

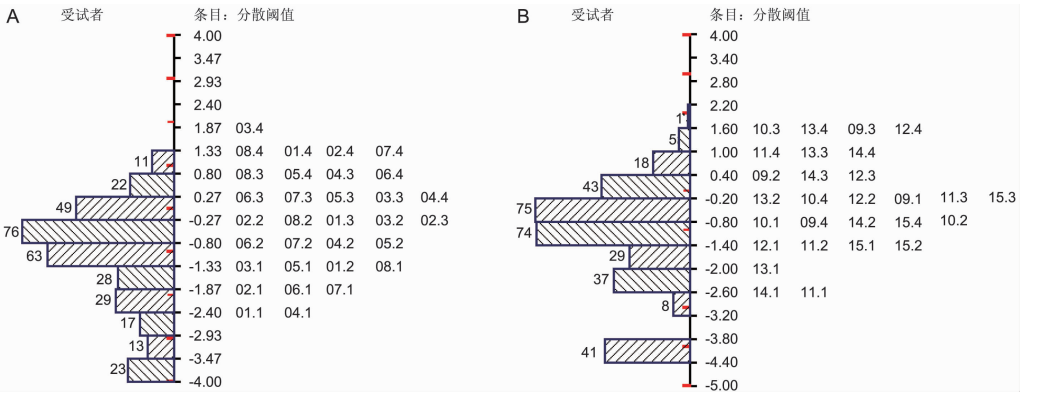
图 5 胃痞患者报告结局量表的信息函数和标准误



注:A 为胃痞症状;B 为胃痞影响

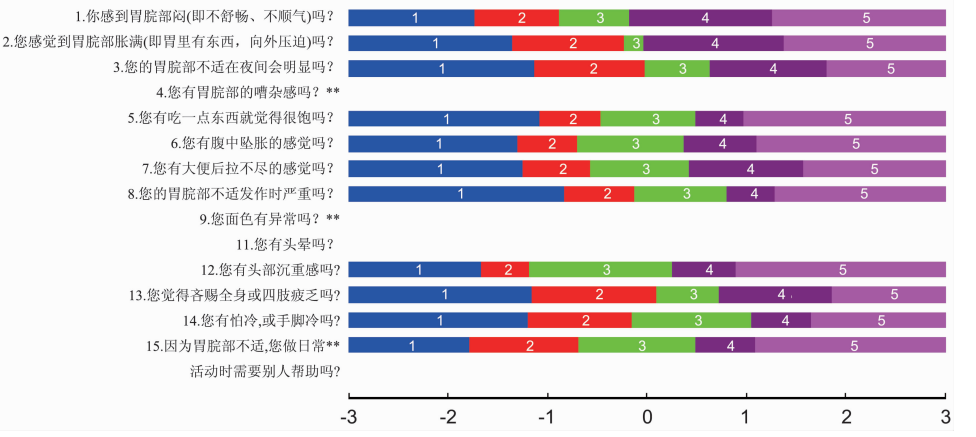
图 6 胃痞患者报告结局量表的条目难度分布

在现代测量属性方面,单维性、局部独立性和单调性评价均显示量表有良好的条目反应理论分析前提条件,提示分析结果有良好的稳定性和可信性。条目的模型拟合性评估中 RMSEA、NNFI、CFI 等指标显示量表在各级结构之间和不同级别之间均有良好的拟合性。IRT 参数估计也显示虽然有极少数条目的区分度和难度系数稍大,但大多数条目的区分度系数在 1.5 ~3.0,难度系数在 -0.5 ~0.3,提示量表整体有良好的区分度和难度评测能力。在条目特征曲线分析中,大多数条目具有良好的工作能力,虽然胃痞影响方面少数条目(如:因为胃脘部不适,您做日常活动时需要别人帮助吗?)显示曲线过度偏移,但其 CAT 测试的选择比例和条目区分度等特征较好,且阈值顺序分析显示合理,结合专家意见予以保留。在其他参数分



注：A 为胃痞症状；B 为胃痞影响

图 7　胃痞患者报告结局量表的条目－回答者相关图



注：＊＊表示阈值紊乱

图 8　胃痞患者报告结局量表的条目阈值顺序图

析中，条目的难度分布向中心集中，难易跨度合理；条目—回答者相关分析显示条目分布和回答等级与受试者有良好的交互作用。而条目阈值参数分析中，基于部分评分模型（partial credit model，PCM）估计获取的 5 分类 4 个阈值理论上存在 $\beta1 < \beta2 < \beta3 < \beta4$ 关系，表示越来越难获得“好”的健康状态[14]。虽然有 4 个条目的阈值顺序并未完全满足以上条件，但 IRT 的条目特征曲线分析显示各条目整体阈值顺序仍基本满足测评要求。因此，量表具有相对较好的现代测量属性。

本量表和研究的质量评价：量表的研究质量对相关的公共卫生、医学研究及医疗实践有重要意义。为评价量表研究的方法学质量，Mokkink LB 等[15,16]于 2010 年发布了健康测量工作选择的共识标准（COSMIN）清单（4 个步骤 12 个评分模块）。结合其中的 IRT 类别评价条目，本研究清晰地报道了运算模型、前提假设评价和分析软件等，而且在内容效度、结构效度等分析中显示量表具有良好的局部独立性和单维度性等。虽然未执行定量化的跨文化效度分析及条目功能差异分析等，但本量表在研制过程中一直有双语研究者参与，其翻译可行性较高，因此量表的整体 IRT 方法学质量仍较可靠。在报告质量方面，研究小组前期研制了“患者报告结局测评量表研制推荐报道条目”（Preferred Reporting Items for Patient-Reported Outcome Instrument Development，PRIPROID）（7 个维度 23 个条目）[17]，本研究采用上述清单不同研究部分组合报告的模式，详细提供了量表的研究背景、研制流程、方法学细节、尽可能全面的结果和多角度的讨论，具有良好的报道质量。

本量表的临床应用注意事项：本量表为疾病特异性患者报告结局量表，测评病种为胃痞（中医第一诊断），西医第一诊断为慢性胃炎和功能性消化不良者可直接使用（中医或西医病名对测评目标无本质影响），以其他西医疾病为第一诊断者建议使用对应疾病量表；测评目标是胃痞的症状和带来的生理影响；管理方式为患者独立填写，调查员不辅助且不解释，主要

信息提供者为患者,若无法实现时必须明确记录替代管理方式;计分为胃痞症状、胃痞影响两个方面单独计分,分数越低表示健康状态越好。本量表为免费使用,研究者和临床医师可向通讯作者索取完整量表,无需签订使用协议。

本研究的局限:本研究主要采用 IRT、CAT 及数据模拟等现代测试理论技术,虽然有传统测试不可比拟的优点,但由于自身技术特点限制,在某些方面还无法完全真实反映潜在数据特征,对部分参数的界值、选择依据、适用范围等还存在一定争论,尤其是健康医学领域的现代测试理论研究更少,给相关研究的标准化带来困惑,而本研究也将可能受到一定影响。但项目组也同时认为定量化的现代测试理论虽然存在一定学术争论,但基本的核心思想固定且一致,各争论只是对现行方法的优化和更严谨的论证各模型工作能力,本研究使用的各参数目前已取得广泛共识,故不会对研究结论产生决定性影响。

综上所述,本研究最终研制的 GS-SF 包含胃痞症状 8 个条目,胃痞影响 7 个条目,2 个方面均独立积分,主要测评成人 GS 患者的症状和疾病影响,现代测量属性评价显示具有良好的条目难度、区分度、工作信息量和个体化测评特征。但由于现代测试理论的发展仍不完全成熟,所以本研究也不可避免存在特定局限,故还需进一步展开后续研究以提升本量表质量。

[致谢 特别感谢 Prof. Sanjiv Mahadeva(Malaya)、Prof. Foteini Anastasiou(Greece)、Nathan A. Thompson, Ph.D.(USA)、David J. Weiss, Ph.D.(USA)、Seung W. Choi, M.D.(USA)、Kimberly Bennett, Ph.D.(USA)等专家对本研究给予热情、耐心的支持指导!致谢名录已获得授权]。

参 考 文 献

[1] 国家中医药管理局医政司. 中医临床路径—22 个专业 95 个病种[M].北京:中国中医药出版社, 2011: 409.

[2] Tanih NF, Dube C, Green E, et al. An African perspective on *Helicobacter pylori*: prevalence of human infection, drug resistance, and alternative approaches to treatment[J]. Ann Trop Med Parasitol, 2009, 103(3): 189 -204.

[3] Mahadeva S, Goh KL. Epidemiology of functional dyspepsia: a global perspective[J]. World J Gastroenterol, 2006, 12(17): 2661 -2666.

[4] 李春林. 六味安消胶囊联合多潘立酮片治疗脾虚气滞型胃痞 214 例疗效观察[J]. 中国医院用药评价与分析, 2011, 11(9): 824 -826.

[5] 戎光,陈姝,易亮. 自拟消痞汤治疗胃痞疗效观察[J]. 辽宁中医药大学学报, 2011, 13(10): 206 -207.

[6] Hornke LF. Item response times in computerized adaptive testing [J]. Psicológica Revista De Metodología Y Psicología Experimental, 2000, 21: 175 -190.

[7] Wang P, Dai HQ, Ding SL. Computerized adaptive testing in mainland China: a review[J]. Int J Digital Content Technol Its Applications, 2010, 4(6): 94 -99.

[8] 高怀勇,金桂林. 项目反应理论及其在计算机自适应测试中的应用[J]. 西华师范大学学报(自然科学版). 2008,29(1): 84 -88.

[9] 周仲瑛主编.中医内科学[M]. 北京: 中国中医药出版社, 2007: 197 -204.

[10] 中华中医药学会. 中医内科常见病诊疗指南·中医病证部分[M]. 北京: 中国中医药出版社, 2008: 76 -79.

[11] Drossman DA.The functional gastrointestinal disorders and the Rome Ⅲ process[J]. Gastroenterology,2006,130(5):1377 -1390.

[12] 房静远, 刘文忠,李兆申,等. 中国慢性胃炎共识意见[J].胃肠病学, 2013, 18(1): 24 -36.

[13] 侯政昆. 基于计算机自适应测试的胃痞患者报告结局量表研制[D]. 广州: 广州中医药大学, 2012.

[14] Masters GN. A rasch model for partial credit scoring[J]. Psychometrika, 1982, 47(2): 149 -174.

[15] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study[J]. Qual Life Res, 2010, 19(4): 539 -549.

[16] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes [J]. J Clin Epidemiol, 2010, 63(7): 737 -745.

[17] Hou ZK, Liu FB, Fang JQ, et al. Advising on Preferred Reporting Items for patient-reported outcome instrument development: the PRIPROID [J]. Chin J Integr Med, 2013, 19(3): 172 -181.

(收稿:2014 -08 -31 修回:2015 -03 -12)

中国中西医结合杂志