

脑卒中 PRO 量表纸质化和电子化测量方式的等同性分析*

罗艳虹¹ 张 瑛² 张 耀³ 张岩波^{1△}

【摘要】目的 对于自行研发的电子化脑卒中患者报告结局(electronic patient-reported outcome, ePRO)量表,考核量表的等同性,即条目功能差异(differential item functioning, DIF),旨在比较纸质化和电子化的测量方式是否影响受试者对 PRO 量表的应答。**方法** 使用自行研发的 App 采集患者信息,使用均值协方差结构(mean and covariance structure, MACS)模型考查 DIF,分析 PRO 量表管理方式由纸质化转变为手机 App 后,量表是否保持了良好的测量等同性。**结果** 量表的 DIF 考核结果如下:条目 PHD(physical domain, 生理领域)7、PHD8、PHD14、PSD(psychological domain, 心理领域)3、PSD9、SOD(social domain, 社会领域)1、SOD6 均存在一致性 DIF($P_1 < \alpha_1$),条目 PHD9、PHD10、PHD11、PHD12、SOD5、THD(therapeutic domain, 治疗领域)1 存在非一致性 DIF($P_2 < \alpha_2$),条目 SOD7 和 THD2 同时存在两种形式 DIF。以上 15 个条目存在测量方式间的 DIF,但尚在可接受范围。**结论** 本文研发的移动的患者报告结局(mobile-based patient-reported outcomes, mPRO)随访系统,通过对量表不同数据收集方式间的 DIF 进行考核,证实了电子化的数据收集方式并不影响受试者的应答,测量方式的转变未影响量表测量等同性。该系统得到的测量结果准确可靠,可在临床实践中使用,并在应用过程中对其进行优化。

【关键词】 脑卒中 电子化 PRO 纸质化 PRO 条目功能差异

【中图分类号】 R195 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.05.011

Analysis of Equivalence between Paper-based and Electronic Patient-reported Outcome Measures for Stroke

Luo Yanhong, Zhang Ying, Zhang Yao, et al (Department of Health Statistics, School of Public Health, Shanxi Medical University, Taiyuan 030001)

【Abstract】Objective For the development of the electronic patient-reported outcome measures(ePRO) for stroke, the scale's equivalence, i.e., differential item functioning(DIF) was assessed, aiming to compare whether paper- and mobile-based collection methods affect the participants' responses. **Methods** The patient information was collected using the developed App, and the DIF was detected using the mean and covariance structure(MACS) model to analyse whether the scale maintained a good measurement equivalence after changing to mobile-based from paper-based. **Results** Items PHD7, PHD8, PHD14, PSD3, PSD9, SOD1, and SOD6 exhibited uniform DIF($P_1 < \alpha_1$), PHD9, PHD10, PHD11, PHD12, SOD5, and THD1 exhibited non-uniform DIF($P_2 < \alpha_2$), and SOD7 and THD2 exhibited both uniform and non-uniform DIF. The DIF between the different measurement modalities existed for the above 15 items; however, it was within the acceptable limit. **Conclusion** The mPRO patient follow-up system developed in this study by assessing the DIF between the scale's different data collection methods confirmed that mobile-based data collection did not affect the participants' response, and the measurement method shift did not affect the scale measurement equivalence. The system yielded accurate and reliable measurements that can be used in clinical practice and optimized in its application course.

【Key words】 Stroke; Electronic patient-reported outcome measures; Paper-based patient-reported outcome measures; Differential item functioning

疾病谱和死亡谱的改变使得我国慢性病管理面临重大压力,PRO 作为慢性病评价的有效工具已被患者和专家认可。在技术进步的推动下,PRO 已不仅仅以纸质化形式存在,还能够借助智能手机完成电子化患者报告结局(electronic patient-reported outcomes, ePRO)的测量和数据收集。基于智能手机的 ePRO,又称为移动的患者报告结局(mobile-based patient-reported outcomes, mPRO)。PRO 由纸质化向电子化的转变提升了 PRO 测量的科学

性^[1],降低了应答压力,提高了满意度,改善了易用性,有更少的缺失数据^[2]。

美国食品药品监督管理局发布的 PRO 指南明确指出:测量工具管理方式发生改变,必须对量表特性(如信度、效度)进行重新评价^[3]。PRO 数据收集由纸质化向电子化方式转变之后,数据收集等同性即条目功能差异(differential item functioning, DIF)也是量表评价的重要内容。DIF 是指潜在特质相同但所处组别不同的受试者对同一测量条目的反应不同,一般表现为条目测试得分不同,此时我们就说测试条目存在 DIF。当某个测量工具中的多数条目都有功能差异,有理由认为该测量工具在不同组人群中测量的潜在特质不同,因而不同组别受试人群在该量表上的测量得分不能相互比较。

* 基金项目:国家自然科学基金(81872714;82173631)

1.山西医科大学公共卫生学院卫生统计学教研室,重大疾病风险评估山西省重点实验室,煤炭环境致病与防治教育部重点实验室(030001)

2.国药同煤总医院医务科

3.山西中医药大学健康服务与管理学院

△通信作者:张岩波, E-mail: sxmuzyb@126.com

Griffiths Jones W 等表明使用基于网页的系统进行 ePRO 远程收集与纸质化 PRO 收集相比显示了极好的等同性^[4],良好的等同性是否同样适用于以智能手机应用程序为载体的 mPRO 测量,尚未见相关文献报道。考核等同性的方法很多,均值协方差结构模型(mean and covariance structure, MACS)相对于其他方法有一定优势^[5-6],因此,本文采用 MACS 对自行研发的脑卒中 mPRO 和纸质化 PRO 数据的等同性进行考核,以期 ePRO 的推广提供科学依据。

资料来源

1. 样本来源

本文旨在比较电子化和纸质化测量方式是否影响受试者对测量工具的应答,因此将电子化组作为目标组,纸质化组作为参照组。

本文的电子化组样本为调查期间(2016年12月10日—2017年1月12日)山西医科大学第一医院、山西医科大学第二医院、太原市中心医院所有脑卒中住院患者,坝陵桥社区、大北门东社区中有脑卒中病史的居民。病例纳入标准:①经头颅 CT 或 MRI 检查,诊断为脑卒中的任何一种类型(包括短暂性缺血发作和腔隙性脑梗)的任何疾病阶段,若病情不允许或无条件进行相应检查,根据临床表现可确诊为脑卒中的患者;②社区卫生服务中心居民电子健康档案记录且经本人证实确有脑卒中病史;③志愿参与调查。排除标准:①患有精神疾病,意识不清;②病情严重,不允许进行测试;③存在语言、智力障碍等问题,影响量表理解;④独立或在家人的帮助下都不能使用手机。

纸质化组随机选取了2012年9月初至2013年1月底研制纸质版 PRO 时收集的230例样本。患者来自于山西医科大学第一医院、山西医科大学第二医院、山西省人民医院、太化医院、太原万柏林区下元社区卫生服务中心、临汾市人民医院、吕梁柳林县、兴县和交口镇。病例纳入和排除标准与电子化组相同。

2. 样本量确定

Nunnally^[7]建议被试人数是研究变量的5~10倍。采用 Nunnally 的建议,本文至少需要样本量230例。

3. 数据收集

课题组将通过测试的 mPRO 随访系统移动端 App 上传至安卓市场,同时提供 App 二维码。调查开展前,调查者预先熟悉量表内容、系统使用及常见问题解决方法,并编写系统使用手册。课题组与样本来源医院和社区取得联系,安排专门调查员负责数据收集,对受试者理解有困难的条目做适当解释,必要时指导系统使用。患者自行或在家人帮助下下载安装 App 进行测量,若受试者不愿安装系统,可在课题组提供的手机上完成测试。PRO 量表一般由患者本人根据自

身实际情况独立填写,但对于文化程度低、年龄偏大或偏瘫患者,可由调查员将题目和选项客观地转述患者,根据患者意愿由家属帮助患者完成作答。

患者填写的数据自动储存在后台数据库中,由于系统开发过程中的设置,患者不能跳跃测试题目,故而数据收集不会出现缺失,收集到的数据从后台导出,经过整理后直接使用。

原理及方法

严格意义上来说,DIF 属于结构效度,它是指具有相同潜在特质水平的不同特征(如年龄、性别等)受试者对同一测量的反应不同,出现 DIF 即认为测量条目在不同特征受试者中测量的潜在结构是不同的,这些受试者在该测量工具上的得分不能相互比较。DIF 分为一致性条目功能差异(uniform DIF)和非一致性条目功能差异(non-uniform DIF),前者为受试者的能力水平与其分组特征不存在交互作用,后者则是被测的潜在特质水平与组别特征有交互作用。

考核 DIF 的方法很多,本文采用 MACS 模型^[5-6]。MACS 模型在 1974 年由 Sörbom 提出,是一种发现 DIF 灵活有用的方法^[5]。MACS 模型相对于其他方法有一定优势:同时适用于一致性 DIF 和非一致性 DIF;MACS 模型根据修正指数(modified index, MI)发现 DIF,可以减小 I 类错误^[6]。

在 MACS 模型中,对 Likert 评级条目的应答被近似地视为连续尺度的应答^[8]。MACS 模型如下^[5-6]:

$$x_{ij}^{(g)} = \mu_j^{(g)} + \lambda_j^{(g)} \xi_i^{(g)} + e_{ij}^{(g)}$$

$$e_{ij}^{(g)} \sim [0, \sigma_j^{2(g)}]$$

x_{ij} 是个体 i 对条目 j 的应答,是连续尺度上的一个数字, ξ 表示潜在特质,该模型反映 x 对 ξ 的线性回归关系。 μ_j 是当 ξ 为 0 时对条目 j 的平均应答水平;回归系数或斜率 λ_j 是潜在特质每改变一个单位,应答的期望变化量; e_{ij} 是随机误差项; g 表示组别。在 MACS 条目应答模型中,截距对应条目难度参数,因子载荷对应条目区分度参数。当组间截距和斜率都相等时,则不存在 DIF;若截距不同,表明存在一致性 DIF;若斜率不同,表明存在非一致性 DIF。

由于 MACS 模型考查 DIF 时要求单维性,故本文以维度为单位分别检验 DIF。

第一,在两种测量方式间拟合“完全限制模型”,即将该维度上所有对应条目的截距和因子载荷在两组间都设置为相等,纸质化组的潜在特质均值固定为 0,电子化组的潜在特质均值自由估计。第二,分别考查条目截距和因子载荷中最大的 MI,该指标表示相应参数自由估计而不加等同性限制时模型卡方值的降低量,因此在给定 α 水平($\alpha = 0.05/n$, n 为维度所包含的条目数)判断该卡方值的统计显著性即可确定是否存

在 DIF。若 MI 值均不显著,则有理由认为该维度的条目不存在 DIF;以 MI₁ 表示截距相关的最大修正指数,若与截距相关的 MI₁ 有意义,表明该条目存在一致性 DIF;以 MI₂ 表示因子载荷相关最大修正指数,若与因子载荷相关的 MI₂ 有意义,表明存在非一致性 DIF。第三,将以上标记条目的载荷或截距组间等同性限制去除,重新拟合模型,考查剩余仍有等同性限制条目中载荷或截距的最大 MI,依次迭代,直至所有条目的 MI 不显著。最后,如果维度中所有条目均不存在 DIF,则拟合组间的“完全限制模型”,即所有条目参数均自由估计,并与“完全限制模型”比较,若两模型卡方值无差别,则证实该维度中所有条目等同性良好。如果个别条目显示 DIF,则拟合“部分限制模型”,即自由估计相关条目参数,其余条目仍保持组间相等,该模型与“完全限制模型”有差异表明条目确实存在 DIF。

结 果

1. 患者基线情况

表 1 230 例脑卒中患者基线情况

变量	$\bar{x}\pm s/n(\%)$		t/χ^2 值	P	变量	$\bar{x}\pm s/n(\%)$		t/χ^2 值	P
	电子化组	纸质化组				电子化组	纸质化组		
性别			0.394	0.693	吸烟			-0.051	0.959
男	151(65.7)	155(67.4)			从不	108(47.0)	114(49.6)		
女	79(34.3)	75(32.6)			已戒烟	44(19.1)	31(13.5)		
年龄(岁)	65.50±12.53	63.41±14.99	1.628	0.104	吸烟	78(33.9)	85(37.0)		
身高(cm)	166.67±7.41	167.00±7.84	-0.471	0.638	饮酒			-0.652	0.514
体重(kg)	67.38±11.70	68.31±9.99	-0.919	0.358	从不	120(52.2)	126(54.8)		
饮食习惯			1.363	0.174	戒酒	33(14.3)	13(5.7)		
素食偏多	96(41.7)	77(33.5)			偶尔	50(21.7)	56(24.3)		
肉食偏多	53(23.0)	74(32.2)			经常	27(11.7)	35(15.2)		
均衡	81(35.2)	79(34.3)			卒中类型			1.189	0.235
					出血性	38(16.5)	29(12.6)		
					缺血性	192(83.5)	201(87.4)		

表 2 躯体症状维度 DIF 考核模型比较结果

躯体症状	χ^2	df	GFI	RMR	NFI	NNFI	CFI	IFI	$\Delta\chi^2$	Δdf	P
完全限制模型	227.876	48	0.854	0.158	0.833	0.881	0.864	0.863			
部分限制模型	217.995	47	0.856	0.159	0.840	0.884	0.871	0.870	9.881	1	<0.05

* :GFI: goodness-of-fit index; RMR: root mean square residual; NFI: normed fit index; NNFI: non-normed fit index; CFI: comparative fit index; IFI: incremental fit index.

考查其余维度截距相关和因子载荷相关最大修正指数结果见表 3,模型比较结果见表 4。

表 3 各维度最大修正指数

维度	MI ₁	P ₁	α_1	MI ₂	P ₂	α_2
认知能力	16.568(PHD8) *	<0.005	0.013	7.326(PHD10) *	0.007	0.017
	5.829(PHD13)	0.018	0.017	6.928(PHD9) *	0.020	0.025
言语交流	7.255(PHD14) *	0.007	0.013	1.204(PHD8)	0.324	0.050
				7.262(PHD12) *	0.007	0.017
	4.463(PHD15)	0.037	0.017	9.499(PHD11) *	<0.005	0.025
自理能力	6.452(PHD17)	0.012	0.010	2.965(PHD14)	0.089	0.050
				2.503(PHD16)	0.122	0.013
焦虑	7.140(PSD3) *	0.008	0.010	3.775(PSD5)	0.053	0.013
	1.996(PSD5)	0.177	0.013			
抑郁	19.546(PSD9) *	<0.005	0.010	0.546(PSD6)	0.472	0.013
	6.261(PSD8)	0.083	0.013			
回避	4.011(PSD13)	0.046	0.013	1.456(PSD11)	0.235	0.017

续表 3

维度	MI ₁	P ₁	α ₁	MI ₂	P ₂	α ₂
社会交往	12.764(SOD1)*	<0.005	0.017	2.494(SOD2)	0.123	0.025
	0.100(SOD2)	0.750	0.025			
家庭支持	8.455(SOD6)*	<0.005	0.013	11.142(SOD7)*	<0.005	0.017
	8.907(SOD7)*	<0.005	0.017	12.328(SOD5)*	<0.005	0.025
	4.182(SOD4)	0.043	0.025	3.270(SOD4)	0.075	0.050
满意度	15.119(THD2)*	<0.005	0.010	11.540(THD1)*	<0.005	0.013
	5.490(THD3)	0.021	0.013	11.684(THD2)*	<0.005	0.017
				2.434(THD4)	0.130	0.025

注：“*”表示有统计学意义，加黑条目存在一致性 DIF，画横线条目存在非一致性 DIF；PHD, physical domain(生理领域)；PSD, psychological domain(心理领域)；SOD, social domain(社会领域)；THD, therapeutic domain(治疗领域)。

综上所述，条目 PHD7、PHD8、PHD14、PSD3、PSD9、SOD1 和 SOD6 所对应的 P 值均小于对应的 α，所以这 7 个条目均存在一致性 DIF，PHD9、PHD10、PHD11、PHD12、SOD5 和 THD1 存在非一致性 DIF，SOD7 和 THD2 同时存在两种形式 DIF。“自理能力”和“回避”两个维度的条目 MI 值表明不存在 DIF。

表 4 各维度 DIF 考核模型比较结果

维度		χ ²	df	GFI	RMR	NFI	NNFI	CFI	IFI	Δχ ²	Δdf	P
认知能力	完全限制模型	59.880	15	0.966	0.124	0.919	0.951	0.939	0.938			
	部分限制模型 1	34.743	13	0.972	0.083	0.953	0.973	0.970	0.970	25.137	2	<0.05
	部分限制模型 2	27.726	12	0.977	0.045	0.963	0.978	0.978	0.978	7.017	1	<0.05
言语交流	完全限制模型	206.424	15	0.833	0.161	0.653	0.737	0.671	0.671			
	部分限制模型 1	190.903	13	0.852	0.137	0.679	0.718	0.695	0.694	15.521	2	<0.05
	部分限制模型 2	180.794	12	0.864	0.121	0.696	0.710	0.710	0.710	10.109	1	<0.05
自理能力	完全限制模型	241.867	24	0.792	0.105	0.905	0.928	0.914	0.914			
	完全不限模型	183.176	10	0.807	0.057	0.928	0.863	0.932	0.932	58.691	14	<0.05
焦虑	完全限制模型	137.436	24	0.947	0.087	0.891	0.924	0.909	0.908			
	部分限制模型	130.246	23	0.949	0.085	0.897	0.925	0.914	0.913	7.190	1	<0.05
抑郁	完全限制模型	172.789	24	0.907	0.101	0.862	0.899	0.879	0.879			
	部分限制模型	152.555	23	0.911	0.099	0.878	0.909	0.895	0.895	20.234	1	<0.05
回避	完全限制模型	29.119	15	0.974	0.087	0.960	0.984	0.980	0.980			
	完全不限模型	7.818	4	0.989	0.031	0.989	0.984	0.995	0.995	21.301	11	<0.05
社会交往	完全限制模型	31.672	8	0.980	0.033	0.957	0.976	0.967	0.967			
	部分限制模型	18.675	7	0.979	0.033	0.975	0.986	0.984	0.984	12.997	1	<0.05
家庭支持	完全限制模型	152.896	15	0.830	0.262	0.795	0.850	0.812	0.811			
	部分限制模型 1	130.209	13	0.863	0.207	0.825	0.853	0.840	0.840	22.660	2	<0.05
	部分限制模型 2	109.042	11	0.891	0.138	0.854	0.854	0.866	0.867	21.167	2	<0.05
满意度	完全限制模型	155.163	24	0.867	0.193	0.905	0.932	0.918	0.918			
	部分限制模型 1	129.544	22	0.882	0.159	0.920	0.939	0.933	0.933	25.619	2	<0.05
	部分限制模型 2	117.199	21	0.897	0.111	0.928	0.943	0.940	0.940	12.345	1	<0.05

从表 4 可知，认知能力、言语交流、家庭支持和满意度维度分别拟合了两个部分限制模型，这是由于每次模型拟合只能去除一个截距和一个斜率组间等同性限制，也就是模型每次只考察一个 MI₁ 和一个 MI₂，由于每个维度有统计学意义的 MI₁ 和 MI₂ 都不超过两个，因此这四个维度只需要分别再拟合两个部分限制模型即可。以认知能力维度 DIF 考核模型比较为例，拟合的部分限制模型 1 去除掉 PHD8 的截距组间等同性限制和 PHD10 的斜率组间等同性限制，拟合的部分限制模型 2 去除 PHD9 的斜率组间等同性限制，由于部分限制模型 1 和部分限制模型 2 与完全限制模型差异显著(P<0.05)，因此，PHD8 确实存在一致性 DIF，PHD9 和 PHD10 确实存在非一致性 DIF。言语交流、家庭支持和满意度维度的两个部分限制模型构建以此

类推。这四个维度的结果表明，PHD8、PHD14 和 SOD6 确实存在一致性 DIF，PHD9、PHD10、PHD11、PHD12、SOD5、THD1 确实存在非一致性 DIF。SOD7 和 THD2 存在两种形式 DIF。

从表 3 可知，焦虑、抑郁和社会交往维度各自都有一个存在一致性 DIF 的条目，因此只拟合一个部分限制模型，由焦虑、抑郁和社会交往维度的部分限制模型拟合优度结果可知，这 3 个部分限制模型与对应的完全限制模型差异显著，证明 PSD3、PSD9 和 SOD1 确实存在一致性 DIF。

从表 2 可知，“自理能力”和“回避”两个维度的条目 MI₁ 和 MI₂ 值表明条目不存在 DIF，拟合组间的“完全不限模型”。自理能力和回避维度的完全不限模型与完全限制模型比较结果为 P<0.05，表明条目存

在 DIF, 该矛盾的结果可能与 PHD17、PSD13 两个条目的 P 值(0.012, 0.046) 与对应的检验水准(0.010, 0.013) 接近有关, 它们是否存在 DIF 需进一步验证。

讨 论

在慢性病压力日益增大的当下, 有效的慢病管理对患者、家庭、医疗卫生机构乃至整个社会都有重大意义。本研究基于以上背景, 以脑卒中为起点, 研发了患者报告结局移动随访系统, 根据预先设定的纳入排除标准, 使用该系统对 230 例脑卒中住院患者和院后患者进行测量。本研究根据预先设定的纳入排除标准选择符合条件的各种类型(缺血性和出血性) 各个疾病阶段(未急性发病、急性期、恢复期、后遗症期、康复) 的脑卒中患者作为受试者, 保证样本多样性。

本文研究结果表明, 测量工具的管理方式由纸质化转变为电子化形式, 对量表不同管理方式间的 DIF 进行考核, 整个量表中 15 个条目存在 DIF, 但尚在可接受范围, 证实了电子化的收据收集方式并不影响受试者的应答方式, 但这些条目是否存在偏倚尚需讨论。总体来说, 电子化数据收集方式是可行的, 测量结果真实可靠。

本文发现存在 DIF 的条目较多, 且各维度的拟合指标不甚理想, 究其原因可能与以下几方面有关: 其一, 两组受试者并非同一时期、同一来源样本, 纸质组是 2012 年 9 月到 2013 年 1 月收集的患者样本, 分析过程中控制了人群年龄、性别两个因素, 但不同时期及不同的社会环境、诊疗水平等很容易导致受试者基本人口学特征不一致, 影响 DIF 分析; 其二, 无论采用何种方法考查 DIF, 首先必须选择一个或一组参照条目, 将其因子载荷设为 1, 随后进行其他参数估计, 但目前对参照条目的选择尚无理方法。本文以维度为单位分析 DIF, 将每个维度的第一个条目做为参照, 这样选择的参照条目可能并不合适, 从而影响 DIF 的判断; 其三, 研究样本存在偏倚。尽管如此, 模型经过调整, 拟合优度指标整体上均得到改善, 且有文献表明 40% 的 DIF 条目数并不太大^[8], 据此本文表现 DIF 的条目数尚可接受。对于 DIF 的分析, 发现 DIF 仅仅是第一

步, DIF 并不一定由条目偏倚产生, 受试者间确实存在的潜在特质差异也会造成 DIF, 因此考虑 DIF 的来源在今后的研究中非常必要^[9]。此外, 有研究表明, 仅有少数条目不等同, 测量结果的组间比较仍然有意义^[6], 因此对于表现 DIF 的条目处理成为研究面临的实际问题, 直接删除条目还是调整测量工具计分方式来解释该差异需要考量^[6]。

此外, 脑卒中疾病本身会对患者认知能力造成影响, 该随访系统在脑卒中患者中可行, 则其他慢性疾病数据的电子化收集应当能被受试者接受。本研究研发的患者报告结局移动随访系统是国内首个以 PRO 为主体的移动医疗 App, 该系统的投入使用可改善国内 PRO 应用现状的不规范, 并且该系统灵活的手机终端形式, 使得患者和临床工作者从中受益。

参 考 文 献

- [1] 姚倩. 基于移动平台的高血压健康管理系统的设计与实现[D]. 西安: 西北大学, 2014.
- [2] Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice[J]. CA: A Cancer Journal for Clinicians, 2012, 62(5): 336-347.
- [3] US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims [EB/OL]. <https://www.fda.gov/media/77832/download.pdf>, 2009-2012.
- [4] Griffiths-Jones W, Norton MR, Fern ED, et al. The equivalence of remote electronic and paper patient reported outcome (PRO) collection[J]. The Journal of Arthroplasty, 2014, 29(11): 2136-2139.
- [5] Sörbom D. A general method for studying differences in factor means and factor structure between groups[J]. British Journal of Mathematical and Statistical Psychology, 1974, 27(2): 229-239.
- [6] Chan D. Detection of differential item functioning on the Kirton Adaptation-Innovation Inventory using multiple-group mean and covariance structure analyses[J]. Multivariate Behavioral Research, 2000, 35(2): 169-199.
- [7] Nunnally JC. Psychometric theory[M]. New York: McGraw-Hill, 1967.
- [8] Ferrando PJ. Calibration of invariant item parameters in a continuous item response model using the extended Lisrel measurement submodel[J]. Multivariate Behavioral Research, 1996, 31(4): 419-439.
- [9] Stark S, Chernyshenko OS, Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy[J]. Journal of Applied Psychology, 2006, 91(6): 1292-1306.

(责任编辑: 张悦)

(上接第 685 页)

- [4] Lyu JJ, Hou YW, Chen Z. The use of restricted mean time lost under competing risks data[J]. BMC Medical Research Methodology, 2020, 20(1): 197.
- [5] 吕晶晶, 巫宏基, 侯雅文, 等. 竞争风险存在时基于限制平均损失时间的统计分析[J]. 中国卫生统计, 2021, 38(5): 661-664.
- [6] Lyu J, Hou Y, Chen Z. Combined Tests Based on Restricted Mean Time Lost for Competing Risks Data[J]. Statistics in Biopharmaceutical Research, 2023, 15(2): 332-339.
- [7] Hashimoto H, Kada A. A note on confidence intervals for the restricted mean survival time based on transformations in small sample

size[J]. Pharmaceutical Statistics, 2022, 21(2): 309-316.

- [8] Chen J, Hou Y, Chen Z. Statistical inference methods for cumulative incidence function curves at a fixed point in time[J]. Communications in Statistics-Simulation and Computation, 2020, 49(1): 79-94.
- [9] 陈金宝, 侯雅文, 陈征. 竞争风险数据中累积发生率置信区间的估计研究[J]. 中国卫生统计, 2018, 35(1): 22-25.
- [10] 黄兴辉, 陈金宝, 杨紫荆, 等. 基于两条生存曲线间面积的非参数统计推断方法研究[J]. 中国卫生统计, 2019, 36(1): 8-12.

(责任编辑: 张悦)