

• 心理护理 •
• 论 著 •

2 型糖尿病患者自我报告抑郁项目库计算机 自适应测试版的构建及验证

邵梦瑶^{1,2}, 陆敏², 王艳梅²

摘要: **目的** 构建 2 型糖尿病患者自我报告抑郁项目库计算机自适应测试(CAT)版,并评价其测量性能。**方法** 采用便利抽样法,在上海市浦东新区公利医院及金桥医联体内 4 家社区卫生服务中心选取符合纳排标准的 2 型糖尿病患者进行面对面纸质问卷调查,共发放问卷 2 139 份。评估中文版 PROMIS 抑郁项目库在 2 型糖尿病患者中的测量学特性,并据此筛选条目,随后利用 FlexCAT 在线平台构建 T2DM 患者自我报告抑郁项目库 CAT 版。通过事后模拟及蒙特卡罗模拟,从测量效率、测量准确性和测量精度 3 个方面进行评价。**结果** 共调查 2 062 例患者。中文版 PROMIS 抑郁项目库符合等级反应模型假设,难度参数(−0.50~3.40)与区分度参数(2.05~4.11)均在合理范围。不同性别、年龄及文化程度组中均未见显著的项目功能差异。基于条目质量分析,在 FlexCAT 平台构建患者自我报告抑郁项目库 CAT 版,测试者可在线完成个性化测评,研究者可在后台获取评估结果。事后模拟及蒙特卡罗模拟显示,CAT 版应答条目数 4~12 条,均值分别为 4.77 和 4.57,超 90% 的测试者应答 4~6 个条目即可完成评估。模拟所得抑郁潜在特质水平与全条目库估计值的 Pearson 相关系数达 0.942 和 0.953,Cohen's d 效应量 0.03。**结论** 中文版 PROMIS 抑郁项目库的条目均具有较好的测量学特性,可用于 CAT 版的构建。构建的 2 型糖尿病患者自我报告抑郁项目库 CAT 版在测量效率、精度和准确性方面表现良好,具有个性化评估的特点,能够减少患者答题负担、减轻医护人员工作压力并提高临床工作效率。

关键词: 抑郁; 计算机自适应测试; 等级反应模型; 中文版 PROMIS 抑郁项目库; 事后模拟; 蒙特卡罗模拟; 测量性能; 系统误差

中图分类号: R473.5; R395.6 **DOI:** 10.3870/j.issn.1001-4152.2025.13.093

Development and validation of computer adaptive testing version of self-reported depression item bank for type 2 diabetes patients Shao Mengyao, Lu Min, Wang Yanmei. Nursing Department, Shihezi University School of Medicine, Shihezi 832003, China

Abstract: **Objective** To develop a computer adaptive testing (CAT) version of the self-reported depression in patients with type 2 diabetes Mellites (T2DM) and evaluate its measurement performance. **Methods** Convenience sampling was employed to select T2DM patients who met the inclusion criteria from Shanghai Pudong New Area Gongli Hospital and four community health service centers in Jinqiao Medical Federation, and face-to-face paper questionnaire surveys were conducted, with 2,139 questionnaires distributed. The measurement characteristics of the Chinese version of the PROMIS depression item bank were evaluated among T2DM patients, and items were selected accordingly. Then, a CAT version of the self-reported depression for T2DM patients was constructed using the FlexCAT online platform. Post-hoc simulation and Monte Carlo simulation were used to evaluate from three aspects: measurement efficiency, measurement accuracy, and measurement precision. **Results** A total of 2,062 patients were investigated. The Chinese version of the PROMIS depression item bank met the assumptions of the graded response model, the difficulty parameters was ranging from −0.50 to 3.40 and discrimination parameters from 2.05 to 4.11, which indicated the reasonable range. Significant item function differences were not observed among different genders, ages, and educational levels. Based on item quality analysis, a CAT version of self-reported depression for patients was constructed on the FlexCAT platform. Testers could complete personalized assessments online, and researchers could obtain evaluation results in the background. Post-hoc simulation and Monte Carlo simulation showed that the CAT version had 4–12 response items, with mean values of 4.77 and 4.57, and more than 90% of the testers could complete the assessment by responding to 4–6 items. The Pearson correlation coefficients between the simulated latent depression trait levels and the full item bank estimates were 0.942 and 0.953, with Cohen's d effect size of 0.03. **Conclusion** The items of the Chinese version of the PROMIS depression item bank have good measurement characteristics and can be used for the construction of the CAT version. The constructed CAT version of self-reported depression for T2DM patients performed well in terms of measurement efficiency, precision, and accuracy, with personalized assessment characteristics, which can reduce the burden of patient responses, alleviate the work pressure of medical staff, and improve clinical work efficiency.

Keywords: depression; computer adaptive testing; graded response model; Chinese version PROMIS depression item bank; post-hoc simulation; monte carlo simulation; measurement performance; systematic error

作者单位: 1. 石河子大学医学院护理系(新疆 石河子, 832003);
2. 上海市浦东新区公利医院护理部
通信作者: 王艳梅, 877927981@qq.com
邵梦瑶: 女, 硕士在读, 护士, 3160328790@qq.com
科研项目: 浦东新区卫生健康委员会重点学科建设计划(PWZxk2022-14)
收稿: 2025-02-09; 修回: 2025-04-21

我国糖尿病患病率高达 12.8%, 其中 2 型糖尿病(Type 2 Diabetes Mellites, T2DM) 占 95% 以上^[1]。T2DM 可引起多个重要组织、器官的不可逆性损害, 加之患者需要进行终生的饮食管理、药物治疗等, 加重了患者的心理负担。据调查, 超过 40% 的 T2DM

患者存在不同程度的抑郁症状^[2]。抑郁会增加并发症的发生风险、加快疾病进展,显著影响患者生活质量和护理依从性^[3]。因此,及时有效地识别 T2DM 患者的抑郁症状对于及早干预至关重要。患者报告结局测量信息系统(Patient-Reported Outcomes Measure Information System,PROMIS)是当前国际公认的最精确可信的信息化测量工具系统^[4],计算机自适应测试(Computer Adaptive Testing,CAT)是 PROMIS 的一种动态测量形式。相较于传统测量工具要求所有受测者回答相同的题目,基于项目反应理论(Item Response Theory,IRT)的 PROMIS CAT 能够根据受测者特征动态调整测评内容,实现个性化评估,在保持测量精确度的同时显著减少题目数量,智能化地为被试者量身定制测试,实现高效的个性化评估^[5]。因此本研究采用中文版 PROMIS 抑郁项目库^[6],构建 T2DM 患者自我报告抑郁 CAT 版,以期为临床护理实践提供一种个性化、快速、精准化的抑郁测评方式。

1 对象与方法

1.1 对象 采用便利抽样法,于 2024 年 3—7 月选取上海市浦东新区公利医院及金桥医联体内 4 所社区卫生服务中心的 T2DM 患者作为调查对象。纳入标准:①年龄 ≥ 18 岁并经医生诊断符合《中国 2 型糖尿病防治指南(2020 版)》^[7]标准;②意识清醒,有书写和表达能力,无沟通交流障碍;③自愿参与且签署知情同意书。排除标准:①精神疾病或认知障碍等无法配合;②病情危重、有生命危险;③生活无法自理。条目参数估计(Item Parameter Estimation)的最低样本量要求为至少 1 000 例,样本量越大,条目的参数估计越准确,测量的标准误越小^[8]。除此之外,考虑到调查中存在约 20% 的无效样本,因此本研究至少获取 1 250 例样本。本研究已通过医院伦理委员会批准(GLYY1s2024-006)。

1.2 研究方法

1.2.1 研究工具 ①一般资料调查表:查阅相关文献后自行设计,问卷内容包括人口学资料(性别、年龄、文化程度、婚姻状况、职业、收入水平、医疗保障类型)、临床疾病相关资料(病程、糖尿病家族史、并发症等)。②中文版 PROMIS 抑郁项目库^[6]:项目库共 28 个条目,采用 Likert 5 级评分法评估个体在过去 1 周内的抑郁状况。1~5 分代表“从不”至“总是”,总分 28~140 分。该项目库的使用已获得复旦大学护理学院患者体验研究中心患者报告结局中国中心(PNC-China)负责人袁长蓉教授的授权。

1.2.2 研究步骤与统计学方法

本研究基于全部样本数据构建 CAT 系统,以确保参数估计的稳定性与代表性。为避免评估过程中的过拟合现象,采用 R 软件 sample() 函数将完整样本随机分为两半,其中一半用于事后模拟,初步评估 CAT 在真实作答数据中的测量效率与精度。此外,

基于 IRT 模型已计算的参数进行蒙特卡罗模拟,生成不同潜在特质的模拟样本,以进一步检验 CAT 系统的测量稳定性,增强 CAT 模拟评价的科学性与可信度。

1.2.2.1 项目反应理论模型的选择 等级反应模型(Graded Response Model,GRM)由 Samejima^[9] 于 1969 年针对多级评分的 Likert 量表提出。该模型属于 IRT 中的双参数概率模型,主要用于测量潜在特质,并适用于有多个等级类别的条目。因中文版 PROMIS 抑郁项目库采用了 Likert 5 级递进式计分方式,因此本研究使用 GRM,以确保条目参数估计的准确性、可靠性和科学性。

1.2.2.2 描述分析 计数资料采用频数和百分比表示。项目库的内部一致性采用 Cronbach's α 系数表示;条目性能分析包括各条目及整个项目库的均数、标准差、上限效应和下限效应;结构效度包括条目与总分相关系数、敏感性分析,后者通过计算剔除当前条目后量表的 Cronbach's α 系数,以评估各条目对内部一致性的影响。

1.2.2.3 基于 GRM 的三大假设检验 ①单维性:采用 Reeve 等^[10] 推荐的探索性因子分析(Exploratory Factor Analysis,EFA)方法,第 1 因子的特征值 $> 20\%$ 且第 1 因子特征根与第 2 因子特征根的比值 > 4 ^[11],则能够解释模型 50% 以上的方差,可认为基本满足单维性假设。②局部独立性:被试者在回答某题目时不被其他题目影响。如果残差相关值小于 0.7^[10],说明局部独立性假设成立。③单调性:可通过各条目的项目特征曲线(Item Characteristic Curve,ICC)判断,若从左向右曲线呈现为“S”型递增趋势,则支持单调性假设^[11]。

1.2.2.4 基于 GRM 的条目质量分析 依据 PROMIS 标准对符合三大假设的抑郁项目库进行条目质量分析并计算条目参数,确保最终保留在题库中的条目具有较高的质量^[10]。条目质量分析包含①项目拟合度检验:基于 GRM 对项目库中每个条目进行拟合,采用 $S-X^2$ 统计量进行项目拟合检验,当 $S-X^2$ 检验对应的 P 值 < 0.01 时,表示项目拟合较差,应删除对应条目^[10]。②区分度参数(a)和难度参数(b): a 和 b 的取值范围在 $[-5,5]$ 为宜^[12]。③选项特征曲线(Option Characteristic Curves,OCC):是以函数曲线的方式来描述被试者潜在特质水平、区分度及难度参数和项目作答概率之间的关系。各条目均可画出对应的 OCC。OCC 中第 1、5 条曲线呈单调变化,中间 3 条曲线呈现先上升后下降的正态或近似正态趋势。若出现明显的项目曲线“平缓”或项目曲线被覆盖或扎堆的情况,则应删除题目^[13]。④测验信息量分析:本研究通过计算平均信息量评估条目,平均信息量指信息函数在特定潜在特质水平($\theta = -2, -1, 0, 1, 2$)上的均值^[14]。研究采用 25 和 16 作为量表总体信息量的评判标准:平均信息量 $> (25/\text{总条目数})$ 评为较

好,(25/总条目数)~(16/总条目数)评为一般,<(16/总条目数)评为较差^[15]。抑郁项目库共 28 个条目,依据此标准,平均信息量>0.893 评为较好,0.571~0.893 评为一般,<0.571 评为较差,拟删除。⑤项目功能差异(Differential Item Functioning, DIF):将不同年龄、性别、教育水平的被试者进行分组,观察条目组间差异性。当 McFadden's pseudo R^2 变化量<0.02 时,表示该条目不存在 DIF;当 R^2 变化量在 0.02~0.13 时,表明条目存在轻度 DIF,但对测量结果的影响较小^[16]。

1.2.2.5 CAT 的算法设置 在 FlexCAT 平台题库管理模块导入具体条目内容、条目参数以及条目得分情况。在配置模块,设置相应的 CAT 算法,包括①初始条目选择:选取信息量最高的条目,以提高潜在特质水平估计的精度。②特质水平估计(θ):在本研究中, θ 表示被试者在抑郁心理健康维度上的潜在状态或严重程度,采用期望后验法进行计算。③条目选择策略:采用最大 Fisher 信息量法^[17]。④终止规则:结合固定长度原则与固定测量误差原则。根据相关研究经验,本研究设定 $SE(\theta) \leq 0.3$,并限定测试条目数为 12。即当测试条目数达到 12 或受试者能力估计的标准差超过 0.30 时,测试终止^[18]。在用户管理模块导入临时账户的序号、用户名和密码,即可完成 T2DM 患者自我报告抑郁 CAT 版的构建。

1.2.3 T2DM 患者自我报告抑郁 CAT 版的模拟评价 CAT 评价方式包括蒙特卡罗模拟、事后模拟及混合模拟。本研究采用混合模拟,即结合事后模拟与蒙特卡罗模拟,评估所构建的 T2DM 患者自我报告抑郁 CAT 版的适用性及测量稳定性。事后模拟基于真实条目参数数据,模拟受试者在真实测试环境中的表现^[5]。蒙特卡罗模拟是一种基于随机数的统计抽样实验方法,由计算机生成随机观测值^[19]。该过程涉及 4 个主要步骤,同 1.2.2.5 CAT 算法部分。

1.2.4 资料收集与数据分析 本研究采用一对一现场纸质问卷的调查方式,由 3 名经过培训的调查员按照纳入排除标准筛选患者。向患者介绍此次调查的目的、内容及问卷填写方法,在患者签署知情同意后发放问卷,填写完成后当场回收并核对,若有缺漏,及时补充。本研究共调查 2 139 例患者,其中 13 份信息缺失,64 份规律作答,剔除后有效问卷 2 062 份,有效回收率为 96.4%。利用 SPSS 26.0 软件进行描述性分析及单维性检验,其他分析均利用 R Studio 软件进行,其中使用了 lavaan、Mokken、mirt、lordif 和 cat R 包。

2 结果

2.1 患者一般资料及抑郁得分 2 062 例患者中,男 1 120 例,女 942 例;年龄 18~80(57.4±12.8)岁;初中及以下文化程度 890 例,高中及中专 593 例,大专 369 例,本科及以上 210 例。抑郁得分 28~119(43.41±15.79)分,天花板效应为 0.05%(最高得分 1 例)、地

板效应为 8.0%(最低得分 164 例),均低于 10%。抑郁项目库中条目-总分相关系数为 0.559~0.831(均 $P<0.05$)。本研究中抑郁项目库的 Cronbach's α 系数为 0.967,敏感性分析中去掉当前条目后量表的 Cronbach's α 系数范围为 0.951~0.953。

2.2 基于 GRM 的假设检验 中文版 PROMIS 抑郁项目库的 KMO 检验结果为 0.956,Bartlett 球形度检验, $\chi^2=36\ 116.209$, $P<0.001$,表明该项目库的数据适合进行探索性因子分析。探索性因子分析结果显示,第 1 因子特征根值为 15.197,第 2 因子特征根值为 1.110,第 1 因子特征根与第 2 因子特征根的比值为 13.691,且其方差解释率为 54.276%。各条目在第 1 因子上的因子负荷 0.430~0.749,说明中文版 PROMIS 抑郁项目库符合单维性假设。第 2 主成分处出现明显拐点,进一步支持单维性假设(碎石图见附件 1)。验证性因子分析分析结果显示,中文版 PROMIS 抑郁项目库的残差相关在 0.286~0.678,符合局部独立性假设。观察条目的 ICC(以 D14 为例,其单调性曲线见附件 1),发现所有条目的特征曲线均满足从左至右“S”形单调递增特点,说明该项目库符合单调性假设。

2.3 基于 GRM 的条目质量分析 抑郁项目库中所有条目的拟合度检验结果性显示, $S-X^2$ 检验对应的 P 值为 0.012~0.963。条目的区分度范围为 2.05~4.11,难度范围为-0.50~3.40,且各条目难度参数均符合 $b_1<b_2<b_3<b_4$ 的规律,见表 1。以 D14 为例,条目的选项特征曲线(OCC)见图 1,从左到右 5 条曲线分别代表条目的 5 个选项,基本符合“两侧单调分布,中间呈正态或近似正态分布”的形态且 5 条曲线的峰值相互独立、疏密适中。条目的平均信息量范围为 1.221~1.887(见表 1)。项目功能差异检验结果见表 1,在各年龄段(18~<45 岁、45~<60 岁、60~<75 岁、75~80 岁)上的 R^2 变化范围为 0.005~0.026,其中 6 个条目的 R^2 超过 0.020。不同性别(男/女)的 R^2 变化范围为 0~0.012、不同文化程度(初中及以下、高中及中专、专科、本科及以上)中的项目功能差异性检验 R^2 变化范围为 0.000~0.009,均不超过 0.020(见表 1)。

2.4 CAT 版本的呈现 上述条目质量分析结果表明,中文版 PROMIS 抑郁项目库在 T2DM 患者中的心理测量学属性较好,项目库无条目删减,可用于构建 CAT 版。通过 FlexCAT 在线平台,在题库管理、CAT 配置和用户管理模块中导入相应内容并完成设置,成功构建 T2DM 患者自我报告抑郁 CAT 版。被测试者可使用临时账户登录 FlexCAT 平台完成个性化测试,研究者则可登录后台获取该被测试者的评估结果。

2.5 T2DM 患者自我报告抑郁 CAT 版的评价 为避免过度拟合,本研究采用随机分半的事后模拟评价 T2DM 患者自我报告抑郁 CAT 版的适用性。事后模

拟的潜在特质水平(θ)与整个项目库的 θ 相关系数为 0.942($P<0.001$)。Cohen's d 效应量为 0.03。事后模拟中应答条目数在 4 条目、5 条目和 6 条目的占比

分别为 48.88%、39.19%和 5.72%，总和超过 90.0%（见表 2），CAT 模拟后平均作答题目数为（4.77±1.11）题。

表 1 各条目区分度、难度参数以及项目功能差异

条目	区分度	难度				平均 信息量	项目功能差异(McFadden's pseudo R^2)		
	(a)	(b_1)	(b_2)	(b_3)	(b_4)		年龄	性别	文化程度
D1 我觉得自己没有用	2.94	-0.11	0.61	1.41	2.53	1.833	0.026*	0.006	0.001
D2 我觉得我没有什么可期待的	2.64	0.01	0.69	1.51	2.61	1.608	0.023*	0.003	0.000
D3 我感到无助	2.82	0.04	0.77	1.53	2.66	1.887	0.015	0.002	0.003
D4 我会从他人身边离开	2.40	-0.11	0.65	1.51	2.77	1.347	0.020*	0.001	0.002
D5 我感觉任何事都无法使我高兴起来	3.49	0.09	0.75	1.47	2.32	1.362	0.010	0.002	0.001
D6 我感觉自己不如别人那么好	2.67	0.07	0.71	1.42	2.37	1.452	0.016	0.004	0.001
D7 我觉得难过	2.70	-0.50	0.22	1.20	2.31	1.871	0.011	0.001	0.001
D8 我感到想要放弃一切	2.70	0.20	0.89	1.71	2.75	1.334	0.013	0.001	0.004
D9 我感到对事情很自责	2.69	0.25	0.93	1.61	2.49	1.298	0.018	0.000	0.006
D10 我觉得自己很失败	3.49	0.43	0.98	1.58	2.36	1.283	0.018	0.000	0.005
D11 我很难和别人亲近	2.05	0.04	0.87	1.71	2.84	1.697	0.013	0.001	0.001
D12 我对自己感到失望	3.27	-0.08	0.56	1.34	2.33	1.446	0.025*	0.004	0.009
D13 我觉得自己不被需要	2.64	0.12	0.81	1.55	2.52	1.611	0.018	0.001	0.001
D14 我感到孤独	2.70	-0.06	0.63	1.40	2.33	1.636	0.008	0.002	0.004
D15 我感到沮丧	3.45	0.11	0.71	1.30	2.22	1.569	0.007	0.006	0.006
D16 我很难做决定	2.43	-0.25	0.60	1.49	2.62	1.740	0.005	0.005	0.003
D17 我对未来感到灰心	3.32	-0.11	0.51	2.44	2.10	1.426	0.016	0.001	0.000
D18 生活中的事情让我感到不堪重负	2.56	0.27	1.00	1.83	2.83	1.501	0.023*	0.002	0.001
D19 我感到不快乐	4.11	-0.10	0.56	2.66	2.07	1.853	0.008	0.012	0.007
D20 我感觉我没有活着的理由	2.58	0.91	1.35	2.08	2.87	1.394	0.014	0.009	0.004
D21 我感到绝望	3.94	0.48	1.02	1.69	2.52	1.221	0.014	0.003	0.005
D22 我觉得被人忽视了	2.14	0.44	1.25	2.20	3.40	1.392	0.012	0.002	0.003
D23 我无缘无故地感到不安	2.72	0.38	0.98	1.67	2.82	1.553	0.018	0.001	0.001
D24 我觉得没什么有趣的	3.08	0.32	0.96	1.67	2.68	1.598	0.021*	0.003	0.001
D25 我感到悲观	3.06	0.05	0.67	1.38	2.35	1.532	0.009	0.000	0.005
D26 我感觉我的生活是空虚的	3.47	0.27	0.84	1.44	2.33	1.524	0.014	0.001	0.002
D27 我感到内疚	2.68	0.20	0.81	1.50	2.43	1.518	0.019	0.002	0.003
D28 我感到情感耗竭	3.50	0.03	0.62	1.17	2.11	1.333	0.009	0.004	0.002

注：* 存在项目功能差异。

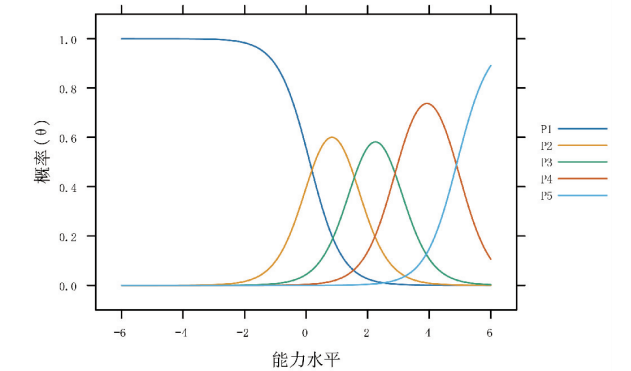


图 1 D14 的选项特征曲线

表 2 事后模拟应答条目数及百分比

应答条目数	频数(n)	百分比(%)
4	504	48.88
5	404	39.19
6	59	5.72
7	25	2.42
8	18	1.75
9	10	0.97
10	6	0.58
11	4	0.39
12	1	0.10

力值相近的虚拟数据进行蒙特卡罗模拟,蒙特卡罗模拟的 θ 值与被试者真实值之间的相关系数为 0.953 ($P<0.001$),平均作答题目数为 4.57,说明 CAT 版本测验明显减少了被试者 83.7%应答题目量。被试的测量信息曲线(蓝色)与测量误差曲线(红色)如图 2 所示,当 θ 为 -0.2 时项目库测试信息最大为 53.387,最小测量标准误差为 0.021; θ 在 -1.27~1.35 时,测验信息量均在 16 以上, θ 在 -1.08~1.22 时,测验信息量在 25 以上。计算潜在特质水平标准误 SE(θ)小于 0.3 的被试比例为 92.3%。

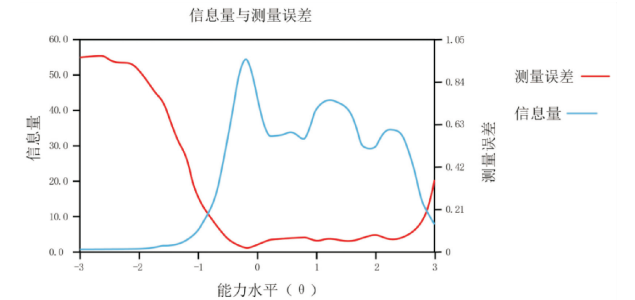


图 2 测量误差、测量信息、潜在特质水平关系图

3 讨论

3.1 T2DM 患者自我报告抑郁 CAT 版的构建 高质量的测量条目库是确保 CAT 科学性与优势发挥的

使用计算机随机生成 5 000 例与真实作答者能

基础^[20]。其中,PROMIS 条目库涵盖多个领域,并经过严格的开发流程,为 CAT 构建提供了坚实基础^[21]。本研究的检验结果进一步证实,所采用的中文版 PROMIS 抑郁项目库具有良好的心理测量学特性。描述性分析中,中文版 PROMIS 抑郁项目库的天花板效应和地板效应均低于 10%,影响较小,表明其测量效能良好,适用性较高,能够有效区分不同能力水平的被试者。本研究结果与国外两版本^[14,22]的 PROMIS 抑郁项目库结论一致。项目库的条目-总分相关系数范围均大于 0.4,Cronbach's α 系数为 0.967,敏感性分析显示,删除条目后量表的 Cronbach's α 系数变化幅度极小(<0.05),表明项目库具有较高的内部一致性和稳定性,该结论与荷兰语版本的结论相似^[14]。在假设检验分析中,中文版 PROMIS 抑郁项目库满足等级反应模型的三大假设,与其他版本的 PROMIS 抑郁项目库的研究结果一致^[14,23]。探索性因子分析结果显示,该项目库的第 1 因子特征根远大于第 2 因子,二者比值超过 4,且第 1 因子方差解释率显著高于 20%;同时,碎石图在第 2 主成分处出现明显拐点,进一步支持该项目库的单维性假设。这一结论与荷兰语版本的研究结果相似^[14],表明该项目库主要测量单一的抑郁潜在特质,具有良好的结构效度,能够准确反映被试者的抑郁水平。此外,条目间残差相关值均低于 0.70,表明条目之间不存在严重的局部依赖问题。这意味着,中文版 PROMIS 抑郁项目库的条目能够有效避免测量冗余,确保测试效率的最大化。进一步分析发现,所有条目的项目特征曲线均呈现单调递增趋势,符合单调性假设,即随着得分增加,被试者抑郁的严重程度也相应增加。在条目质量分析中,中文版 PROMIS 抑郁项目库的所有条目均表现出良好的测量学特性,表明该项目库所有条目均可保留,该项目库可以作为 T2DM 患者自我报告抑郁 CAT 版构建的条目来源。首先,本研究基于 GRM 进行条目库测量学分析,提供了不同于经典测量理论的信息,能够独立于样本估计条目的核心测量参数,即条目区分度和难度,为后续构建 T2DM 患者自我报告抑郁 CAT 提供了测量依据。其次,GRM 拟合检验结果表明,所有条目均与模型拟合良好,进一步验证了 GRM 在 PROMIS 工具中的适用性。此外,本研究采用项目功能差异检验评估项目库各条目在不同性别、年龄和文化程度患者中的适用性。结果显示,所有条目在不同性别组和文化程度组间均未表现出显著差异。在年龄组比较中,条目伪决定系数 R^2 范围为 0.005~0.026,其中 6 个条目 R^2 超过 0.020,但最高 R^2 仅为 0.026,远低于 0.130,表明年龄相关项目功能差异影响极小,可忽略不计。结论与国内外的部分研究果^[24-28]相似,即使部分条目存在项目功能差异,由于其对量表总得分的影响较小,研究者仍可选择保留,以维持量表的完整性和适用性。本研究基于 FlexCAT 在线平台完成了 T2DM

患者自我报告抑郁 CAT 版的构建,整体测验流程便捷简明,已初步具备用户测试、研究者数据查看和管理的功能,包括用户潜在特质水平、测验长度、测验信息量、边际信度、测量误差、总用时以及具体应答条目和内容等关键信息,研究者也可根据需求对条目库管理模块、CAT 配置模块和用户管理模块进行调整。

3.2 T2DM 患者自我报告抑郁 CAT 版的模拟评价

本研究的 CAT 模拟评价结果表明,T2DM 患者自我报告抑郁 CAT 版具有较高的测量效率、精度和准确性。本研究沿用了 PROMIS CAT 的标准化算法,为未来 CAT 在国际范围的推广应用提供了实践依据。事后模拟结果表明,抑郁 CAT 版本的潜在特质水平(θ 值)与完整项目库估计的 θ 值显著相关,Cohen's d 效应量低于 0.2,表明两者误差较小。CAT 版本的平均应答条目数为 4.77 个,说明其测量准确性高、测试负担低,可有效替代完整项目库进行评估。Flens 等^[14]同样采用分半数据事后模拟的方法对荷兰语版抑郁 CAT 版本进行评价,结果显示该版本测量效率、精度和准确性均较高,平均应答条目数为 8.64 个,模拟运行与完整项目库的潜在特质水平估计相关系数为 0.98,Cohen's d 效应量为 0.01,与本研究结果相似。此外,蒙特卡罗模拟($n=5\ 000$)结果表明,模拟 θ 值与完整项目库 θ 值高度相关,CAT 评估时减少了 83.7%的应答条目,测量误差小于 0.3 的被试比例达 96.3%。这说明,相较于完整项目库的纸笔测试,PROMIS 抑郁 CAT 版本具有更高的测量效率、精度和准确性。姚思梦^[29]采用相同方法评估七情量表 CAT 版本,研究结果与本研究相似,进一步支持了本研究 CAT 版本在 T2DM 患者抑郁测评中的应用价值。

4 小结

中文版 PROMIS 抑郁项目库在 T2DM 患者中具有良好的心理测量学属性,所有条目均可用于 CAT 版的构建。基于该项目库构建的 T2DM 患者自我报告抑郁 CAT 版不仅具备个性化评估优势,还能在提高测量效率的同时,确保测量的准确性和精度。该工具缩短了测评时间,减轻了医护人员及患者的负担,具有较高的应用潜力,有助于推动智能化测评在心理健康评估中的应用。但受时间限制,本研究尚未在真实临床环境中应用 CAT 版并收集应答数据进行评价,课题组下一步计划在临床环境中验证其适用性。研究者可利用 FlexCAT 平台管理数据并获取被试者的潜在特质水平(θ 值),但解读抑郁严重程度时,仍需参考 PROMIS 官网,将 θ 值转换为 T 分数。未来可在平台中集成 T 分数自动换算功能,以提高测评便捷性和临床适用性。

附件 1:碎石图和单调性曲线扫描
二维码查看。



碎石图



单调性曲线

参考文献:

- [1] Li Y, Teng D, Shi X, et al. Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study[J]. *BMJ*, 2020, 369: m997.
- [2] 安淑敏, 高月乔, 付东藩, 等. 强化心理干预对首诊 2 型糖尿病患者负性情绪、治疗依从性及血糖控制效果的影响[J]. *中国健康心理学杂志*, 2021, 29(9): 1306-1309.
- [3] Chen V C, Wang T N, Hsieh M C, et al. Depression, diabetes mellitus and mortality in older adults: a national cohort study in Taiwan[J]. *Neuropsychiatr Dis Treat*, 2022, 18: 2639-2648.
- [4] 潘华英, 任佳丽, 张洲, 等. 患者报告结局工具在脑肿瘤病人中的应用进展[J]. *全科护理*, 2023, 21(24): 3346-3349.
- [5] 李心钰, 王超, 陆宏. 计算机自适应测验有效性检验的探索与优化[J]. *现代教育技术*, 2024, 34(6): 123-132.
- [6] 复旦大学护理学院患者体验研究中心患者报告结局中国中心. 成人 PROMIS-心理健康[EB/OL]. (2022-09-21)[2025-03-21]. <https://pnc-china.fudan.edu.cn/workprogress/scale?ifChild=2&type=1>.
- [7] 中华医学会糖尿病学分会. 中国 2 型糖尿病防治指南(2020 年版)(上)[J]. *中国实用内科杂志*, 2021, 41(8): 668-695.
- [8] 闫宇翔, 王洪源. 项目反应理论在医学量表测量误差估计中的应用[J]. *中国卫生统计*, 2004(3): 33-35.
- [9] Samejima F J P. Estimation of latent ability using a response pattern of graded scores[J]. *Psychometrika*, 1969, 34(1): 1-97.
- [10] Reeve B B, Hays R D, Bjorner J B, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS)[J]. *Med Care*, 2007, 45(5 Suppl 1): S22-31.
- [11] Hung M, Voss M W, Bounsanga J, et al. Psychometrics of the Patient-Reported Outcomes Measurement Information System Physical Function instrument administered by computerized adaptive testing and the Disabilities of Arm, Shoulder and Hand in the orthopedic elbow patient population[J]. *J Shoulder Elbow Surg*, 2018, 27(3): 515-522.
- [12] Samuel D B, Bucher M A. Translating Item Response Theory findings for clinical practice[J]. *J Pers Assess*, 2019, 101(4): 452-453.
- [13] Zhang L, Lin J, Liu K, et al. Factor structure and psychometric properties of the purpose in life test (PIL) in a sample of Chinese college students: an application of confirmatory factor analysis and item response theory[J]. *Curr Psychol*, 2021, 42: 11049-11068.
- [14] Flens G, Smits N, Terwee C B, et al. Development of a computer adaptive test for depression based on the Dutch-Flemish Version of the PROMIS Item Bank[J]. *Eval Health Prof*, 2017, 40(1): 79-105.
- [15] 刘叶, 鲁杰, 李顶春, 等. 基于经典测量理论和项目反应理论对慢性病毒性肝炎患者生命质量量表的评价[J]. *临床肝胆病杂志*, 2022, 38(11): 2470-2477.
- [16] Choi S W, Gibbons L E, Crane P K. lordif: an R Package for detecting differential item functioning using iterative hybrid ordinal logistic regression/Item Response Theory and Monte Carlo simulations[J]. *J Stat Softw*, 2011, 39(8): 1-30.
- [17] Baker F B, Kim S H. Item Response Theory: parameter estimation techniques[M]. 2nd Ed. Boca Raton: CRC Press, 2024: 528.
- [18] Green B F, Bock R D, Humphreys L G, et al. Technical guideline for assessing computerized adaptive tests[J]. *J Educ Meas*, 2010, 21(4): 347-360.
- [19] 王超, 陆宏. 计算机自适应测验的研究策略与应用实践[J]. *现代教育技术*, 2017, 27(12): 44-49.
- [20] 涂冬波, 郑蝉金, 戴步云, 等. 计算机化自适应测验理论与方法[M]. 北京: 北京师范大学出版社, 2017: 37-38.
- [21] Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks; 2005-2008[J]. *J Clin Epidemiol*, 2010, 63(11): 1179-1194.
- [22] Jakob T, Nagl M, Gramm L, et al. Psychometric Properties of a German Translation of the PROMIS® Depression Item Bank[J]. *Eval Health Prof*, 2017, 40(1): 106-120.
- [24] Flens G, Smits N, Terwee C B, et al. Development of a computerized adaptive test for anxiety based on the Dutch-Flemish Version of the PROMIS Item Bank[J]. *Assessment*, 2019, 26(7): 1362-1374.
- [23] de Castro N F C, de Melo Costa Pinto R, da Silva Mendonça T M, et al. Psychometric validation of PROMIS® Anxiety and Depression Item Banks for the Brazilian population[J]. *Qual Life Res*, 2020, 29(1): 201-211.
- [25] Lameijer C M, van Bruggen S G J, Haan E J A, et al. Graded response model fit, measurement invariance and (comparative) precision of the Dutch-Flemish PROMIS® Upper Extremity V2.0 item bank in patients with upper extremity disorders[J]. *BMC Musculoskelet Disord*, 2020, 21(1): 170.
- [26] Carlberg Rindestig F, Wiberg M, Chaplin J E, et al. Psychometrics of three Swedish physical pediatric item banks from the Patient-Reported Outcomes Measurement Information System (PROMIS)®: pain interference, fatigue, and physical activity[J]. *J Patient Rep Outcomes*, 2021, 5(1): 105.
- [27] Peersmann S H M, Luijten M A J, Haverman L, et al. Psychometric properties and CAT performance of the PROMIS pediatric sleep disturbance, sleep-related impairment, and fatigue item banks in Dutch children and adolescents[J]. *Psychol Assess*, 2022, 34(9): 860-869.
- [28] Zhou T, Wang Y, Chen J, et al. Psychometric properties of the Chinese version of the PROMIS-Cancer-Anxiety item bank assessed using a graded response model[J]. *Asia Pac J Oncol Nurs*, 2023, 10(12): 100312.
- [29] 姚思梦. 基于混合方法学的七情量表及计算机自适应版本的研制[D]. 广州: 广州中医药大学, 2023.