·调查报告与分析·

中国公众视角卫生体系评价工具的开发及信效度验证

苟媛洁[1]，赵红梅[1,2]，骆萌[1,3]，梅家瑞[1]，申希平[1]，张彤宇[4]，徐东[5]，王晓辉[1]

1. 兰州大学公共卫生学院，兰州 730000；
2. 普洱市人民医院，普洱 665000；
3. 孝感市疾病预防控制中心，孝感 432100；
4. 兰州大学第二临床医学院，兰州 730030；
5. 南方医科大学卫生管理学院，广州 510515

通信作者：徐东，E-mail：romanxu@i.smu.edu.cn；王晓辉，E-mail：wangxiaohui@lzu.edu.cn

【摘　要】目的 开发兼具国际可比性和本土特征的中国公众视角卫生体系评价工具(PVS-CN)，并检验其信效度。**方法** 在合作开发全球版评价工具的基础上经本土化修订、翻译—回译、跨文化调适和优化调整形成PVS-CN，并对其中的量表条目进行检验。于2023年10月在全国范围内(港澳台地区除外)随机抽取移动电话号码进行电话调查，共收集有效问卷534份，其中42人在间隔2周后进行了重测。通过项目分析、探索性因子分析和验证性因子分析及其他信效度检验方法验证量表信效度。**结果** PVS-CN量表共27个条目，包含自评健康状况、最近就医体验、对公立医疗体系的评价、对卫生体系的整体评价和对卫生体系质量的期望5个维度，可解释总变异的67.663%；量表条目内容效度指数(I-CVI)为0.71～1.00，平均量表水平内容效度指数(S-CVI/Ave)为0.98，全体一致量表内容效度指数(S-CVI/UA)为0.85，内容效度良好；验证性因子分析表明，量表结构效度良好，χ^2/df=2.131，规范拟合指数(NFI)=0.934，比较拟合指数(CFI)=0.964，拟合优度指数(GFI)=0.910，Tucker-Lewis指数(TLI)=0.960，递增拟合指数(IFI)=0.964，近似误差均方根(RMSEA)=0.046；信度检验结果显示，总量表Cronbach's α系数为0.942，折半信度为0.849，重测信度为0.874，表明量表内部一致性及稳定性较好。**结论** 本研究开发的PVS-CN信效度良好，可作为从公众角度评估我国卫生体系的有效工具并可进一步推广应用。

【关键词】 卫生体系；评价；公众；信效度；中国视角；公众视角；效度指数

Development and validity testing of the Chinese Version People′s Voice Survey

GOU Yuanjie[1], ZHAO Hongmei[1,2], LUO Meng[1,3], MEI Jiarui[1], SHEN Xiping[1], ZHANG Tongyu[4], XU Dong[5], WANG Xiaohui[1] (*1. School of Public health*, *Lanzhou University*, *Lanzhou* 730000, *China*; *2. Pu'er People's Hospital*, *Pu'er* 665000, *China*; *3. Xiaogan Municipal Center for Disease Control and Prevention*, *Xiaogan* 432100, *China*; *4. The Second Clinical Medical School*, *Lanzhou University*, *Lanzhou* 730030, *China*; *5. School of Health Management*, *Southern Medical University*, *Guangzhou* 510515, *China*)

Corresponding authors: XU Dong, E-mail: romanxu@i.smu.edu.cn; WANG Xiaohui, E-mail: wangxiaohui@lzu.edu.cn

【Abstract】 Objective To develop the Chinese Version People′s Voice Survey (PVS-CN) that incorporates both international comparability and local characteristics and test its reliability and validity. **Methods** On the basis of collaborative development of the global version of the evaluation tool, the PVS-CN was formed through localization revisions, translation–back translation, cross-cultural adaptation, and adjustments, with the scale items being tested. In October 2023, a mobile phone survey was conducted across China′s mainland (excluding Hong Kong, Macao, and Taiwan), where mobile phone numbers were randomly selected. A total of 534 valid questionnaires were collected, with 42 participants undergoing retesting after a two-week interval. The reliability and validity of the scale were assessed by item analysis, exploratory factor analysis, confirmatory factor analysis, and other methods of reliability and validity testing. **Results** The PVS-CN scale consists of 27 items, covering five dimensions: self-assessed health status, recent healthcare experience, evaluation of the public healthcare system, overall evaluation of the health system, and expectations for the quality of the health system, explaining 67.663% of the total variance. The scale presents the item-level content validity index (I-CVI) ranging from 0.71 to 1.00, the scale-level content validity index/universal agreement (S-CVI/UA) of 0.85, and the average scale-level content validity index (S-CVI/Ave) of 0.98, demonstrating strong content validity. Confirmatory factor analysis shows that the structural validity of the scale is good, with χ^2/df = 2.131,

开放获取：CC BY-NC-ND 4.0 DOI：10.11847/zgggws1146275
基金项目：国家自然科学基金(72274088)
第一作者：苟媛洁(2000—)，硕士在读，研究方向：卫生政策与卫生事业管理。
收稿日期：2024-10-28 修回日期：2025-07-22 录用日期：2025-08-06
利益冲突：不存在 伦理审查：已获取 出版授权：全体作者已与编辑部签署作者贡献声明及版权转让协议

normed fit index (NFI) = 0.934, comparative fit index (CFI) = 0.964, goodness-of-fit index (GFI) = 0.910, Tucker-Lewis index (TLI) = 0.960, incremental fit index (IFI) = 0.964, and root mean square error of approximation (RMSEA) = 0.046. The reliability test results show that the overall scale has the Cronbach′s α coefficient of 0.942, the split-half reliability of 0.849, and the test-retest reliability of 0.874, indicating good internal consistency and stability of the scale.**Conclusion** The PVS-CN developed in this study demonstrates good reliability and validity, serving as an effective tool for assessing China′s healthcare system from the public perspective, with potential for further application.
【Keywords】 health system; evaluation; public; reliability; Chinese perspective; public perspective; validity index

良好的卫生体系是实现全民健康覆盖的重要依托[1]。中国始终将公众健康作为发展的首要任务，构建高质量的卫生体系是实现“人人享有健康”目标的关键。《柳叶刀》全球健康专业委员会指出，高质量的卫生体系（high quality health system, HQHS）应以人民为中心，提供优质且可持续的医疗卫生服务，并具备满足公众不断变化的健康需求与期望的能力，从而赢得公众信任[2]。而卫生体系是否以人民为中心、是否赢得公众信任，需通过公众的视角进行评价。因此，以公众视角评价我国卫生体系建设成效，对构建新时期高质量卫生体系、实现高质量发展目标具有重要意义。目前，国内外开发的卫生体系评价工具多围绕供方及患者视角，针对卫生体系的某一方面开展评估，如医疗服务质量[3−6]、患者满意度[7−11]、卫生系统绩效[12−14] 等，缺乏从公众视角出发的综合性评价。由哈佛大学联合 18 个国家的研究团队基于 HQHS 框架[2] 开发的公众视角卫生体系评价工具（People′s Voice Survey, PVS），可以从公众角度快速评估卫生体系建设成效，为卫生体系向高质量转型提供证据支持[15]。本研究在深度参与 PVS 全球版开发的基础上，采取主位–客位平衡研究方式，开发并验证具有国际可比性且适用于国内情境的中国公众视角卫生体系评价工具（Chinese Version People′s Voice Survey, PVS-CN），以期为评价我国卫生体系提供科学可靠的工具。

1 对象与方法

1.1 对象　邀请相关领域专家对评价工具内容进行评议。函询专家的入选标准为：（1）具有本科及以上学历。（2）拥有≥5 年工作经验。（3）具有卫生体系研究、临床工作、量表研制方面的专业背景。在保证专家意见多样性的同时避免因人数过多而导致的评价重复，本研究最终邀请到 7 名专家。其中从事卫生体系领域 4 人，临床工作 2 人，量表设计与评估领域 1 人；男性 3 人，女性 4 人；本科学历 1 人，研究生及以上学历 6 人；所有专家工作年限均≥5 年。采取便利抽样法，在甘肃省、贵州省、广东省邀请受访者进行认知访谈。为使访谈结果更能代表一般人群，本研究对受访者的年龄、性别、职业等方面进行配额，保证样本多样性。认知访谈过程主要遵循信息饱和原则[16]，即不设样本量上限，持续纳入被访谈对象直至不再出现新的问题，最终共纳入 17 名受访者。纳入标准：（1）年满 18 周岁。（2）中国大陆常住居民。（3）自愿参与且无影响理解能力或认知功能的疾病。采用随机数字拨号（random digit dialing, RDD）法[17] 选取量表施测对象，于 2023 年 10 月在全国范围内（港澳台地区除外）产生和抽取移动电话号码进行电话调查，调查对象纳入标准与认知访谈一致。依据相关标准，信效度验证的样本量应为条目数的 5～10 倍[18]，PVS-CN 量表共计 27 个条目。因此，研究的有效样本量为 135～270 人。为提高分析结果可靠性，本研究在满足最低样本量要求的前提下扩大样本量，最终在全国各省收集到问卷 771 份，剔除作答率<80% 的问卷后，获得 534 份有效问卷，有效率为 69.26%。遵循自愿原则，间隔 2 周后对愿意接受二次访问的受访者进行重测，共获得有效问卷 42 份。本研究均获得所有受访者知情同意，且已通过兰州大学公共卫生学院伦理委员会审批（22030203）。

1.2 评价工具开发　本研究评价工具的开发主要经过 3 个阶段：第 1 阶段客位参与，项目组以客位视角与哈佛大学 PVS 开发团队共同开发 PVS 全球版；第 2 阶段主位引导，项目组从主位视角对 PVS 全球版进行本土化修订、翻译与回翻和跨文化调适，开发形成适应我国情境的 PVS-CN 1.0；第 3 阶段修改优化，在保持核心内容不变的前提下，优化 1.0 版本中的题目顺序和跳转逻辑，并对部分题项进行修改调整，形成 PVS-CN 2.0。

1.2.1 PVS 全球版开发（图 1）　在厘清相关概念内涵的基础上，国际开发团队根据 HQHS 框架提出的卫生体系基础、服务过程和服务质量的影响三大领域[2]，构建出 PVS 框架。通过文献调查和专家咨询，结合各国卫生体系共性和既有卫生体系评价相关条目，筛选形成包含 117 个原始条目的条目池。在此基础上，经各国研究者多轮网络会议及书面意见征集，整合不同国家认知访谈及预调查结果，对初始条目进行精简、完善和补充[15]。最终形成了由 4 个维度、87 个条目组成的 PVS 全球版，供各国参与团队进行翻译、调适和验证。

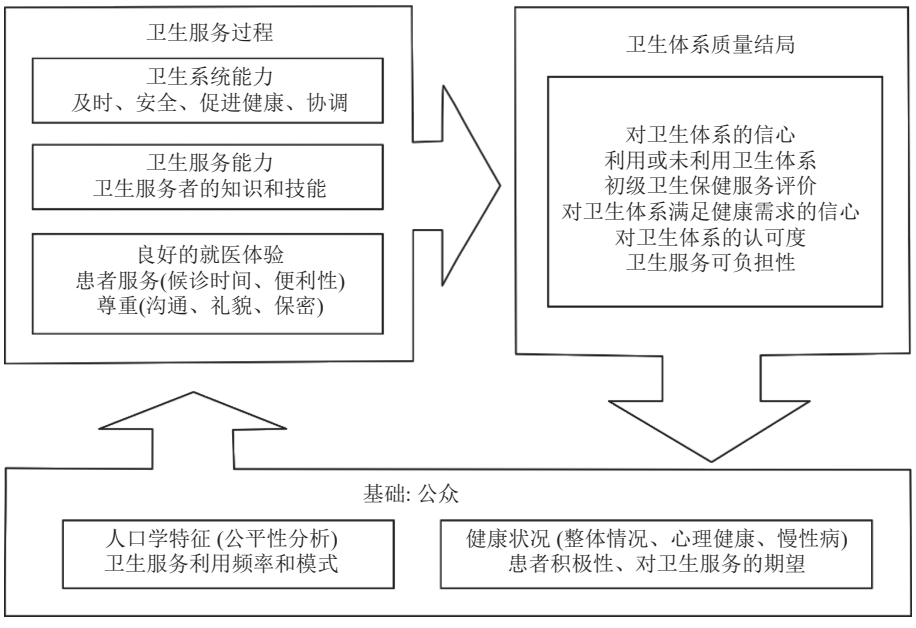


图 1　PVS 框架及关键结构[15]

Fig. 1　Framework and key structures of PVS

1.2.2　PVS-CN 开发(PVS-CN 1.0)　(1)本土化修订：根据中国文化背景和文献检索结果，对全球版的部分条目和回答选项进行删改，使其与中国卫生体系(如医疗机构设置、医疗保险类型等)和人口特征(如教育水平、居住地、家庭收入等)相适配。修订后的 PVS 全球版总计 64 个条目，本研究在此基础上进行汉化。(2)翻译—回翻：遵循翻译—回翻指南和国际量表翻译规范要求[19]，对修订后的版本进行双向翻译。对翻译过程中存在偏差的部分，按“翻译—回翻”流程再次进行翻译，直至翻译人员达成一致。项目组成员对最终版本的条目内容、回答选项及跳转逻辑进行仔细审查，形成 PVS-CN 初稿。(3)跨文化调适：通过线上会议和邮件函询的方式，邀请 7 名专家采用李克特 4 级评分法(1 为不相关，4 为强相关)针对各条目措辞和内容进行评价，并提出本土化调适建议[20]。采用迭代认知访谈的方法，基于信息饱和原则，通过便利抽样在甘肃省、广东省、贵州省选取 17 名受访者进行访谈，询问其对于各条关键词的理解程度。参照专家咨询意见及认知访谈结果，对 PVS-CN 初稿的条目进行修订，形成由基本特征、卫生服务利用及卫生体系能力、就医体验和对卫生体系的信任 4 个维度，共 64 个条目组成的 PVS-CN 1.0[21]。

1.2.3　PVS-CN 优化(PVS-CN 2.0)　为确保 PVS-CN 内容的国际可比性，项目组根据 PVS 开发团队调整要求，结合 1.0 版本的前期测试结果进行了以下优化调整：(1)条目删减与归类：考虑到新冠病毒感染疫情已进入常态化管理阶段，相关问题的突出性已不如疫情初期。为确保工具的时效性和相关性，删除与新冠病毒感染疫情直接相关的 5 个条目，仅保留“您如何评价政府对新冠病毒感染疫情防控措施”的问题，以收集公众对突发公共卫生事件下卫生系统的评价。同时，为提高评价工具简洁性，将内容相似的条目归并至同一题项下。(2)选项更改：对“目前的居住地区”“家庭人均月收入的范围”等条目的回答选项进行了调整。例如，将“目前居住地区”的选项从“城市、农村”扩展为“城市、城市郊区、小城镇、农村”，以便受访者能更准确地选择。(3)条目添加：新增“您虚拟或远程就诊的主要原因是什么”“您如何评价上次接受虚拟或远程诊疗的质量”“过去 12 个月是否有借钱或变卖物品以支付医疗费用的情况”和“您预约这次就诊所花的时间”4 个条目，以更全面了解公众在卫生服务利用与就医体验方面的情况。经过上述调整，最终形成与 1.0 版本维度相同，由问卷和量表共 51 个条目组成的 PVS-CN 2.0。除设置跳转逻辑的 2 个量表条目外，本研究对其余量表条目(以下简称“PVS-CN 量表”)进行信效度检验。

1.3　量表施测

1.3.1　调查工具　(1)一般资料调查表：主要为 PVS-CN 2.0 中的人口学资料条目，如性别、年龄、家庭人均月收入等。(2)PVS-CN 量表：量表初步划分为“自评健康状况”(4 个条目)、“上次就医体验”(11 个条目)、“对公立医疗体系的评价”(4 个条目)、“对卫生体系的整体评价”(6 个条目)和“对卫生体系质量的期望”(2 个条目)5 个维度。条目 3～4、20～22 以 0(完全没信心)～3(非常有信心)计分，其余条目以 0(差)～4 分(非常好)计分。

1.3.2　数据收集与质量控制　委托外部调查公司，

由公司调查员根据调查工具内容，通过计算机辅助电话调查（computer assisted telephone interview, CATI）的方式[22]，在电话中对受访对象进行提问，并将受访者回答实时录入计算机系统中完成数据收集。为确保数据的真实性和有效性，由项目组成员对所收集的样本进行核查，重点排查“数据录入错误”“前后逻辑矛盾”“提问不清”及“追问不到位”等不规范访问行为。

1.4　统计分析　使用 Excel 2016 整理数据并导入 SPSS 26.0 软件和 Amos 24.0 软件进行分析，计量资料以 $\bar{x}\pm s$ 表示，计数资料以构成比（%）表示，$P<0.05$ 为差异有统计学意义。本研究通过临界比（critical ratio, CR）值法和相关系数法筛选量表条目。（1）临界比值法：将总得分按前后 27% 划分为高分组和低分组，通过 t 检验计算 2 组之间的 CR 值。若 CR 值>3.00[23]，则表示该量表条目的鉴别度较好。（2）相关系数法：使用 Pearson 相关系数检验量表各条目与总量表和所在维度的相关程度，若相关系数>0.40[24]，则认为条目与总量表和所在维度的一致性较高，可以保留。本研究通过内容效度和结构效度两方面来评价量表效度。（1）内容效度：根据专家评分结果计算条目内容效度指数（item-level content validity index, I-CVI）、量表内容效度指数（scale-level content validity index, S-CVI）和调整后的 *kappa* 值（K^*）。若 I-CVI≥0.78、全体一致 S-CVI（S-CVI/UA）≥0.80、平均 S-CVI（S-CVI/Ave）≥0.90[25] 及 K^*≥0.60[26]，则表明量表及各条目可以较好反映出测量内容，内容效度较好。（2）结构效度：通过验证性因子分析检验模型的结构效度，以卡方自由度比（χ^2/df）<3.00，比较拟合指数（comparative fit index, CFI）、拟合优度指数（goodness of fit index, GFI）、规范拟合指数（normed fit index, NFI）、递增拟合指数（incremental fit index, IFI）和非规准拟合指数（tucker-lewis index, TLI）>0.90，近似误差均方根（root mean square error of approximation, RMSEA）<0.80 为评价标准，若符合表示模型整体拟合度较好[27]。本研究采用 Cronbach's α 系数、Spearman-Brown 系数评价 PVS 量表信度，采用 Pearson 相关系数评价量表的重测信度，检验结果>0.70 表示量表信度较好[28]。

2 结　果

2.1　一般资料　本研究有效调查 534 名受访者进行信效度验证，其中男性 299 人（56.0%），女性 235 人（44.0%）；年龄为 18～82 岁，平均年龄（37.99±14.05）岁，其中 18～29 岁 165 人（30.9%），30～49 岁 249 人（46.6%），50～69 岁 102 人（19.1%），≥70 岁 18 人（3.4%）；居住在城市的居民 310 人（58.0%），城镇居民 104 人（19.5%），农村居民 120 人（22.5%）；受教育程度初中及以下 130 人（24.3%），高中/中专 118 人（22.1%），大专及以上 286 人（53.6%）；常用医保类型为城镇职工医保 258 人（48.3%），城乡居民医保 232 人（43.4%），无医保 35 人（6.6%），商业医疗保险及其他 9 人（1.7%）；长期患病 86 例（16.1%）；家庭人均月收入在 4 000 元以下的为 344 人（46.4%）；调查对象覆盖 31 个省级行政区（不含港澳台地区），其中华北地区 29 人（5.5%），东北地区 12 人（2.4%），华东地区 233 人（43.8%），华中地区 63 人（11.8%），华南地区 83 人（15.5%），西南地区 80 人（14.9%），西北地区 34 人（6.1%）。每份问卷平均用时（16.64±3.68）min。

2.2　项目分析　使用临界比值法和相关系数法进行项目分析。由于条目 26 和条目 27 分别设置为低质量和高质量的卫生服务情境题（条目内容见表 1），因此所收集到的数据存在偏态分布。对于以上两个条目，仅通过其与所属维度的相关系数进行分析判断。样本数据得分范围为 13～96 分，以高分组（≥64 分）、低分组（≤45 分）为界限，计算高低分组之间的 CR 值。结果显示，前 25 个条目的 CR 值为 7.373～30.218（$P<0.001$）。Pearson 相关系数检验结果显示，前 25 个条目与总分的相关系数为 0.401～0.825（$P<0.001$），27 个条目与各所属维度相关系数为 0.492～0.925（$P<0.001$）。项目分析结果表明，量表各条目鉴别度良好，符合标准，所有条目保留进行下一步检验。

2.3　效度检验

2.3.1　内容效度　经 7 名专家打分，除条目 25 的 I-CVI 为 0.71 外，其余条目的 I-CVI 均为 0.78～1.00，量表 S-CVI/UA 为 0.89，S-CVI/Ave 为 0.98。考虑到专家意见可能会因对选项的随机选择产生偏差。本研究进一步计算调整后的 K^* 值进行校准。结果显示，各条目的 K^* 值为 0.65～1.00，表明量表内容效度理想。

2.3.2　结构效度（表 1）　KMO 检验和 Bartlett 球形检验结果显示，PVS-CN 量表的 KMO 值为 0.953，Barlett 球形检验值 χ^2/df 为 28.45（$P<0.001$），表明适宜进行因子分析。对数据进行主成分正交旋转，在未设定因子数目的前提下提取出特征值>1.00、因子载荷>0.45 的公因子 5 个，累计方差解释率为 67.663%，提示 PVS-CN 量表应为 5 维度结构且划分合理，与预设维度吻合。采用极大似然法，通过验证性因子分析验证以上 5 个因子的模型拟合程度。结果显示，初始模型拟合效果较不理想，根据修正指数对初始模型进行修正后，结果显示，各项指标均达到了适配标准：χ^2/df=2.131，CFI=0.964，GFI=0.910，NFI=0.934，IFI=0.964，TLI=

0.960，RMSEA=0.046，说明修正后的模型整体拟合较好，量表结构效度良好。

2.4 信度检验 考虑到条目 26、27 的条目设置，本研究仅对其所在维度进行重测信度分析。相关分析结果显示，PVS-CN 量表 Cronbach′s α 系数为 0.942，前 4 个维度的 Cronbach′s α 系数分别为 0.823～0.952；总量表的折半信度为 0.849，前 4 个维度的折半信度为 0.710～0.946；总量表的重测信度为 0.874，各维度重测信度为 0.722～0.847。所有信度检验结果均在可接受范围内，量表信度较好。

表 1 各条目内容及探索性因子分析结果

Table 1 Items of the scale and results of the exploratory factor analysis

维度	条目	因子1	因子2	因子3	因子4	因子5
自评健康状况	1 您认为您的健康状况				0.803	
	2 您认为您的心理健康状况				0.820	
	3 您是自己健康的负责人				0.793	
上次就医体验	4 在医护人员没有询问的情况下，您也可以告诉他们您所担心的健康问题				0.734	
	5 整体而言，您所接受的医疗服务质量	0.795				
	6 医护人员的知识和技能	0.816				
	7 机构拥有的设备和医疗用品，如：医疗设备、实验室检测的可及性	0.746				
	8 医护人员对您的尊重程度	0.806				
	9 医生对您之前的就诊或检查结果的了解程度	0.734				
	10 医护人员就您的病情给您提供了简单易懂的解释	0.781				
	11 医护人员按照您本人所希望的参与程度，让您参与了您的医疗决策	0.796				
	12 医护人员为您服务的时间	0.804				
	13 候诊时间	0.813				
	14 除医生外，医疗机构其他工作人员（接待员、管理人员等）对您的礼貌程度和给予的帮助	0.785				
	15 您预约这次就诊所花的时间	0.500				
对公立医疗体系的评价	16 孕期保健，如产前保健			0.740		
	17 儿童保健服务，如健康儿童和患儿保健服务			0.786		
	18 持续性或慢性病，如高血压或糖尿病的医疗卫生服务			0.744		
	19 精神性疾病，如抑郁和焦虑的卫生服务			0.695		
对卫生体系的整体评价	20 如果身患重病，您可以享受到有效的医疗服务		0.783			
	21 如果身患重病，您能负担得起医疗费用		0.754			
	22 政府对医疗系统做出相关决策时，会听取公众意见		0.822			
	23 总体而言，您如何评价我国公立医疗系统的质量		0.620			
	24 总体而言，您如何评价我国非公立医疗系统的质量		0.568			
	25 您如何评价政府对新冠疫情的防控措施		0.499			
对卫生体系质量的期望	26（张三感觉近3天胃痛逐渐加剧，决定去医疗机构看病）在医疗机构医生没有询问他的症状或查体，开了止痛药，没有告诉他诊断结果。您如何评价这样的医疗服务质量					0.610
	27 在医疗机构医生给他做了检查，并抽血化验，医生告诉他病情不严重，给了止痛药，建议清淡饮食，并让他如果情况恶化就来复诊。您如何评价这样的医疗服务质量					0.788
累计方差解释率		28.496	41.562	53.459	63.747	67.663

3 讨 论

随着我国经济社会的不断发展，人民群众对卫生健康的需求呈现出全方位、全周期、多层次、多样化的趋势，对卫生体系建设提出了新要求[29]。通过公众视角进行卫生体系评价，可以系统了解人民健康需要，准确把握卫生体系改革成效，为我国推进优质高效、协同整合的新型卫生体系建设提供重要参考依据。目前，国内外开发的卫生体系评价工具多基于结构—过程—结果模型[30]、服务质量差距模型等理论框架[31]，评价重点多是医疗服务质量和患者体验，难以全面反映人民群

众的健康需求。虽有国内学者开发的全国卫生服务调查[32]、“清华城市健康指数”[33]等扩大了评价内容的覆盖面，但以上工具每次调研所需时间较长，投入较高。因此，有必要基于系统性的评价框架，开发一种具有人群代表性的卫生体系综合评价工具，以准确、迅速反映居民需求。就开发方式而言，国内相关测量工具的开发通常采用 2 种方式：(1)采取客位研究取向，通过引进国际成熟量表进行翻译调适[34–36]。(2)采取主位研究取向，开发适应中国情境的卫生体系评价工具[10–11]。前者效率较高，能够确保条目选择的全面性，但在本土化应用时可能会受到挑战；后者针对性强，但普适性不足，评价结果往往难以进行国际比较。

本研究采取主位—客位平衡研究方式，前期与国际团队合作，从公众视角出发，针对 HQHS 框架关注的卫生体系服务过程与结果构建条目池。在此基础上，结合各国卫生体系共性，采用多种理论方法进行条目筛选，构建出 PVS 全球版，可从微观(如诊疗护理)、中观(如医疗机构)和宏观(如整体卫生系统)3 个层面全面评估公众对卫生体系的看法。后期经过本土化修订、翻译—回翻、专家咨询、认知访谈、优化调整、正式施测等步骤，最终形成兼具主位新颖性和客位普适性，具有中国卫生系统特色且核心内容国际可比的 PVS-CN，以期为我国卫生体系建设、评价等相关研究提供有效支持。

为保证评价工具的可靠性，本研究进行的信效度验证遵循健康测量工具选择的共识标准指南[37]，从项目分析、效度及信度评价 3 个方面对 PVS-CN 量表进行验证。项目分析结果显示，量表各条目 CR 值为 7.574～30.453，与总分和各维度相关系数分别为 0.401～0.825 和 0.492～0.925，具有较好的鉴别度[23]。量表的 I-CVI 为 0.71～1.00，K^*为 0.65～1.00，S-CVI/UA 为 0.88，S-CVI/Ave 为 0.98，提示整体内容效度较好[25]。对于 I-CVI 为 0.71 的条目“您如何评价政府对新冠疫情的防控措施”，考虑到其能反映公众在突发公共卫生事件下对卫生系统的整体评价，且 K^*值在可接受范围内[26]，因而未剔除。效度分析中，本研究先通过探索性因子分析对 PVS-CN 量表划分维度的合理性进行检验，共提取出 5 个公因子，累计方差解释率为 67.663%，表明量表维度划分较为合理。后通过验证性因子分析，验证模型拟合程度。结果显示，修正后的标准化模型各项拟合指标均在可接受的范围内[27]，表明量表具有良好的结构效度。信度检验结果显示，PVS-CN 量表及各维度的 Cronbach′s α 系数、折半信度、重测信度结果均>0.70[28]，表明量表整体内在一致性程度和稳定性较好，测量结果较为可靠[28]。

综上所述，本研究开发的 PVS-CN 具有良好的信效度，可在全国范围内获取居民对我国卫生系统的评价。所得结果一方面有助于了解公众对于卫生体系的看法和评价，为政府部门决策提供参考；另一方面可用于国际比较研究，在展示中国卫生体系建设成效的同时借鉴国外经验和成果，探索卫生体系改革的有益经验。未来还可与投入指标(如卫生资源配置)或其他卫生体系评价工具结合，从多维度综合评估中国卫生体系的建设现状，为我国进行卫生体系体制改革提供科学依据。

本研究仍有优化空间。其一，本研究采用 RDD 法进行电话调查，与传统面对面调查相比，回复率较低，容易产生调查结果偏倚。为更真实地反映调查情况，未来应基于人口普查结果对数据赋值权重后进行分析，以提高其在人群中的代表性。其二，本研究开发的评价工具题项较多，电话调查过程中可能导致受访者因访问时间过长而中途拒访，增加无应答偏倚的产生概率，未来可考虑开发简版工具，以提高配合率和应答率。

数据可用性声明　本研究的原始数据因涉及参与者隐私保护等原因无法公开共享，但经脱敏处理后的数据可在合理请求下向通信作者申请获取。申请需通过伦理委员会审批，并签署数据使用协议。

参考文献

[1] Kruk M E, Gage A D, Joseph N T, et al. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries[J]. The Lancet, 2018, 392(10160): 2203-2212.

[2] Kruk M E, Gage A D, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution[J]. The Lancet Global Health, 2018, 6(11): E1196-E1252.

[3] Lei Y H, Liu J L, Tang Q, et al. Constructing ECMO care quality evaluation index system based on" structure-process-outcome" three-dimensional theoretical model[J]. International Journal of Food Engineering and Technology, 2019, 2(2): 27-35.

[4] Rotar A M, van den Berg M J, Kringos D S, et al. Reporting and use of the OECD Health Care Quality Indicators at national and regional level in 15 countries[J]. International Journal for Quality in Health Care, 2016, 28(3): 398-404.

[5] 滕佳利, 宋桂杭, 李倩, 等. 基于 DIP 的医保服务质量评价指标体系的构建与应用——基于德尔菲法和层次分析法 [J]. 中国卫生政策研究, 2023, 16(2): 29-35.

[6] 方纪元, 张浩成, 陈文琴, 等. 医养结合型养老机构卫生服务质量指标体系构建研究 [J]. 中国卫生政策研究, 2020, 13(4): 24-30.

[7] Beattie M, Murphy D J, Atherton I, et al. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review[J]. Systematic Reviews, 2015, 4(1): 97.

[8] Westbrook K W, Babakus E, Grant C C. Measuring patient-perceived hospital service quality: validity and managerial usefulness of HCAHPS scales[J]. Health Marketing Quarterly, 2014, 31(2): 97-114.

[9] Lee D. HEALTHQUAL: a multi-item scale for assessing healthcare service quality[J]. Service Business, 2017, 11(3): 491-516.

[10] 胡广宇, 刘倩楠, 尹畅, 等. 全国儿童区域医疗中心患者体验评价研究 [J]. 中国卫生政策研究, 2021, 14(7): 11-17.
[11] 匡莉, 李丽娜, 罗卓君, 等. 中文版全科医疗核心特征功能患者体验量表的开发与验证 [J]. 中国全科医学, 2021, 24(13): 1621-1628.
[12] 刘智勇, 姚岚, 徐玲, 等. 中国卫生系统绩效评价指标体系构建 [J]. 中华医院管理杂志, 2016, 32(5): 339-342.
[13] 张晓溪, 宗莲, 王海银, 等. 超大城市卫生体系绩效评价框架构建 [J]. 中国卫生资源, 2018, 21(2): 167-175.
[14] Mesarić J, Bogdan S, Bosanac V, et al. Performance Assessment Tool for Quality Improvement in Hospitals (PATH): first experiences in Croatia[J]. Liječnički Vjesnik, 2011, 133(7/8): 250-255.
[15] Lewis T P, Kapoor N R, Aryal A, et al. Measuring people′s views on health system performance: design and development of the People′s Voice Survey[J]. PLoS Medicine, 2023, 20(10): e1004294.
[16] 谢诗桐. 基于方法对比的中国一般人群健康效用积分体系构建研究——以六维健康调查简表第二版 (SF-6Dv2) 为例 [D]. 天津: 天津大学, 2020.
[17] L'Engle K, Sefa E, Adimazoya E A, et al. Survey research with a random digit dial national mobile phone sample in Ghana: methods and sample quality[J]. PLoS One, 2018, 13(1): e0190902.
[18] 宋俊敏, 万丹婷, 郑中梅, 等. 美国 CDC 社区老人跌倒风险自评表在我国老年人群中测试和信效度 [J]. 中国公共卫生, 2020, 36(4): 592-595.
[19] Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation[J]. Value in Health, 2005, 8(2): 94-104.
[20] Sousa V D, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline[J]. Journal of Evaluation in Clinical Practice, 2011, 17(2): 268-274.
[21] 骆萌. 基于高质量卫生体系框架的公众视角卫生体系质量评价工具的开发与验证 [D]. 兰州: 兰州大学, 2023.
[22] 国家统计局. 计算机辅助电话调查是如何开展的 [EB/OL]. (2023-01-01)[2024-10-20].https://www.stats.gov.cn/zs/tjws/tjdc/202301/t20230101_1903798.html.
[23] 魏艺琳, 张莉, 陈芳菲, 等. 高血压患者主动健康行为量表编制 [J]. 中国公共卫生, 2023, 39(3): 370-374.
[24] 潘舒娅, 周波, 杨程惠, 等. 中国职业紧张反应量表编制 [J]. 中国公共卫生, 2022, 38(2): 157-160.
[25] 史静琤, 莫显昆, 孙振球. 量表编制中内容效度指数的应用 [J]. 中南大学学报 (医学版), 2012, 37(2): 49-52.
[26] Polit D F, Beck C T, Owen S V. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations[J]. Research in Nursing & Health, 2007, 30(4): 459-467.
[27] 付会斌, 潘海燕, 孔丹莉, 等. 二阶验证性因子模型的 AMOS 实现 [J]. 中国卫生统计, 2013, 30(1): 52-56.
[28] Cook D A, Beckman T J. Current concepts in validity and reliability for psychometric instruments: theory and application[J]. The American Journal of Medicine, 2006, 119(2): 166. E7-166. E16.
[29] 刘雨璇, 王军永, 周佳佳, 等. 我国医疗服务水平、居民健康水平与区域经济水平耦合协调时空演变分析 [J]. 医学与社会, 2024, 37(6): 51-58.
[30] Donabedian A. Evaluating the quality of medical care[J]. The Milbank Memorial Fund Quarterly, 1966, 44(3): 166-206.
[31] Parasuraman A, Zeithaml V A, Berry L L. A conceptual model of service quality and its implications for future research[J]. Journal of Marketing, 1985, 49(4): 41-50.
[32] 国家统计局. 全国第七次卫生服务统计调查制度 [EB/OL]. (2023-12-12)[2024-10-20].https://www.stats.gov.cn/fw/bmdcxmsp/bmzd/202312/t20231212_1945458.html.
[33] 郭晓薇. 首份全国城市健康大数据报告——深杭宁京沪并列引领型 [J]. 中国卫生, 2021(2): 108-109.
[34] Yang H, Shi L Y, Lebrun L A, et al. Development of the Chinese primary care assessment tool: data quality and measurement properties[J]. International Journal for Quality in Health Care, 2013, 25(1): 92-105.
[35] 闫温馨, 刘珏. 中文版以人为本的基层卫生服务量表的信效度验证 [J]. 中国全科医学, 2022, 25(25): 3135-3142.
[36] 吴怡雪, 童莺歌, 陈思亦, 等. 患者就医体验调查问卷的汉化及信效度评价 [J]. 中国卫生质量管理, 2022, 29(11): 61-66.
[37] Mokkink L B, Terwee C B, Patrick D L, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes[J]. Journal of Clinical Epidemiology, 2010, 63(7): 737-745.