

# 病人健康问卷抑郁量表(PHQ-9)在青少年中应用的信效度检验

胡星辰 张迎黎 梁 炜 张红梅 杨世昌

**【摘要】**目的 检验病人健康问卷抑郁量表(PHQ-9)在青少年中应用的信度和效度。方法 方便抽取河南省济源、新乡两地共 3 所中学,再以班级为单位采用整群抽样方法抽取初中一年级到高中三年级的 14~18 岁在校学生 471 名进行 PHQ-9 调查,同时进行美国《精神障碍诊断与统计手册(第 4 版)》(DSM-IV)临床定式访谈(SCID)抑郁量表评定和贝克抑郁量表(BDI)评定。在样本中选取 88 人 4 周后完成 PHQ-9 重测。结果 PHQ-9 总分的 Cronbach  $\alpha$  系数为 0.85,条目间相关系数为 0.289~0.560,各条目与 PHQ-9 总分相关系数为 0.616~0.730,4 周后重测信度为 0.88。PHQ-9 总分与 BDI 总分 Pearson 相关系数为 0.77,主成分因子分析显示因子(躯体化-情感因子)突出。PHQ-9 评估抑郁情绪的最佳划界分为 10 分,灵敏度为 93.33%,特异度为 96.83%;ROC 曲线下面积(AUC)为 0.984(95% CI 0.969~0.993)。结论 PHQ-9 应用于青少年抑郁情绪评估中具有良好的信度与效度,10 分为该量表的最佳诊断划界分。

**【关键词】** PHQ-9; 信度; 效度; 青少年; 抑郁

中图分类号: R749.4

文献标识码: A

doi: 10.3969/j.issn.1007-3256.2014.04.021

## Reliability and validity of the patient health questionnaire - 9 in Chinese adolescents

HU Xing - chen , ZHANG Ying - li , LIANG Wei , ZHANG Hong - mei , YANG Shi - chang

The Second Affiliated Hospital , Xinxiang Medical University , Xinxiang 453002 , China

**【Abstract】Objective** To test the reliability and validity of the Patient Health Questionnaire - 9 ( PHQ - 9 ) in adolescents.

**Methods** A total of 471 adolescents with age 14 - 18 years completed the PHQ - 9 and the Beck Depression Inventory ( BDI ) questionnaire. The Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders ( SCID ) was used to diagnose depression. A total of 88 participants completed PHQ - 9 test - retest after 4 week. **Results** The factor analysis of PHQ - 9 identified a unifactorial structure ( somatization - emotional factor ) . Cronbach  $\alpha$  coefficient of PHQ - 9 was 0.85. The mean inter - item correlation coefficients ranged from 0.289 ~ 0.560. The correlation coefficients of the nine items with the total scale ranged from 0.616 ~ 0.730. The test - retest reliability value of PHQ - 9 was 0.88. The scores of PHQ - 9 was significantly associated with that of BDI (  $r = 0.77$  ) . PHQ - 9 had an optimal cutoff score of 10 , which indicated a sensitivity of 93.33% and a specificity of 96.83% , with an area under the curve of 0.984 ( 95% CI 0.969 ~ 0.993 ) . **Conclusion** The PHQ - 9 has satisfactory reliability and validity and can be used in adolescents. For screening purposes , cutoff score of 10 was recommended.

**【Key words】** PHQ - 9; Reliability; Validity; Adolescents; Depression

近年来,青少年抑郁症的发生率呈逐步上升趋势。最新的数据显示,中国青少年抑郁症状总检出率为 14.81%<sup>[1]</sup>。青少年的抑郁症常常会导致自杀、辍学、行为问题和物质滥用。国内已有的儿童和青少年抑郁症筛查工具或较为复杂或需要专业人员进行评定,因此寻找适合筛查工具对提高青少年抑郁症的检出率就显得尤为重要。

病人健康问卷抑郁量表(Patient Health Questionnaire - 9, PHQ - 9)是以《精神疾病诊断与统计

手册(第 4 版)》(Diagnostic and Statistical Manual of Mental Disorders Fourth edition, DSM - IV)的抑郁症(Major depressive disorder, MDD)为标准编制成的九条目自评工具,长度只有其它抑郁筛查工具的一半。在国外,PHQ - 9 已成为基层卫生中心筛查抑郁症的首选工具之一<sup>[2]</sup>。PHQ - 9 在中国成人和大学生群体中应用的有效性已经得到证实<sup>[3-5]</sup>,但在青少年人群中尚未见到相关报道。本研究旨在检验 PHQ - 9 在青少年人群中的信度与效度,同时探讨其应用于该人群的最佳诊断划界分,为该量表在我国青少年中的应用提供心理测量学依据。

作者单位: 453002 新乡医学院第二附属医院

通信作者: 张迎黎, E-mail: yinglizhang2000@163.com

## 1 对象与方法

1.1 对象 方便抽取河南省济源、新乡两地共 3 所中学,再以班级为单位采用整群抽样方法抽取初中一年级到高中三年级的学生。共发放 500 份问卷,回收有效问卷 471 份,有效问卷回收率 94.2%。其中 88 名被试 4 周后完成 PHQ-9 问卷的重测。

1.2 方法 以班级为单位统一发放问卷,采取不记名方式自愿完成 PHQ-9 和贝克抑郁量表(Beck Depression Inventory, BDI),当场收回。由 4 名精神科主治医师对完成问卷的被试进行 DSM-IV 临床定式访谈(SCID)抑郁量表的评定。访谈以面对面的形式进行。完成问卷的被试中抽取 88 名 4 周后进行 PHQ-9 重测。在调查的同时,如发现被试中自杀条目(如自杀观念或尝试)为阳性,或者表示有精神上的痛苦希望得到帮助者,给予及时的心理辅导和帮助,并向监护人和校方提出相应的处理建议。

### 1.3 工具

1.3.1 PHQ-9 PHQ-9 是拥有 9 个条目的抑郁自评问卷<sup>[6]</sup>。评估被试过去两周的感受。9 个条目反映的是 DSM-IV 标准中的 MDD,每个条目分为 0~3 分 4 级评分。问卷总分为 9 个条目的评分总和,得分 0~27 分,得分越高表示抑郁症状越严重。PHQ-9 已经被翻译成中文,在中国成年人中的信度和效度已经得到证实<sup>[3-5]</sup>。

1.3.2 BDI 为抑郁症自评工具,本研究用 13 条目版本,每条目的得分按强度分 0~3 分,总分 0~39 分,得分越高表明抑郁症状越严重。其在中国青少年人群的有效性已得到验证<sup>[7]</sup>。

1.3.3 DSM-IV 临床定式访谈(SCID)抑郁量表 SCID 是多种精神疾病诊断的金标准。该工具是以 DSM-IV 为标准用于诊断精神障碍的结构式临床访谈。本研究应用 SCID 抑郁量表进行抑郁症

的临床诊断。访谈在 PHQ-9 和 BDI 测试结束后独立进行。访谈人员均接受过 SCID 培训。

1.4 统计方法 使用 SPSS 19.0 统计软件对数据进行分析,包括描述性统计分析、*t* 检验、 $\chi^2$  检验、内部一致性检验、Pearson 相关检验和探索性因素分析;使用 MedCalc 11.0 统计软件对统计数据进行分析,包括灵敏度、特异度、阳性预测值、阴性预测值以及受试者工作特征曲线(Receiver Operating Characteristic curve, ROC 曲线)。

## 2 结果

2.1 一般资料 500 名被试接受问卷调查,29 名填写信息不完整未能纳入研究,最后回收有效问卷 471 份(占 94.20%)。其中男性 235 名(49.89%),女性 236 名(50.11%)。年龄 14~18 岁,平均年龄(15.84±1.37)岁。PHQ-9 得分 0~23 分,平均得分(5.11±3.67)分。BDI 得分 0~24 分,平均得分(4.56±4.14)分。以 SCID 访谈为诊断标准,抑郁症患者 30 人(6.37%),其中男生 13 人(5.53%),女生 17 人(7.20%),两组比较差异无统计学意义( $\chi^2=0.552, P=0.458$ )。男生和女生分别在 PHQ-9 的平均得分 [(4.93±3.71) vs (5.28±3.63)  $t=-1.040, P=0.299$ ] 在 BDI 的平均得分 [(4.47±4.18) vs (4.66±4.11)  $t=-0.505, P=0.613$ ] 差异均无统计学意义,提示抑郁的发生和严重程度无性别差异。

### 2.2 信度检验

2.2.1 内部一致性信度 采用 Cronbach's  $\alpha$  系数对 PHQ-9 的内部一致性进行估计,总量表的 Cronbach's  $\alpha$  系数为 0.85。采用 Pearson 相关系数进行分析,各条目与总分的相关系数为 0.616~0.730,各条目间相关系数为 0.289~0.560,均具有相关关系( $P<0.01$ )。见表 1。

表 1 PHQ-9 条目间及各条目与量表总分间的相关性(*r*)

条目	愉快感丧失	心情低落	睡眠障碍	精力缺乏	饮食障碍	自我评价低	集中注意力困难	烦躁不安	消极观念
愉快感丧失	1	-	-	-	-	-	-	-	-
心情低落	0.560 <sup>a</sup>	1	-	-	-	-	-	-	-
睡眠障碍	0.390 <sup>a</sup>	0.343 <sup>a</sup>	1	-	-	-	-	-	-
精力缺乏	0.456 <sup>a</sup>	0.361 <sup>a</sup>	0.395 <sup>a</sup>	1	-	-	-	-	-
饮食障碍	0.314 <sup>a</sup>	0.347 <sup>a</sup>	0.381 <sup>a</sup>	0.396 <sup>a</sup>	1	-	-	-	-
自我评价低	0.507 <sup>a</sup>	0.438 <sup>a</sup>	0.328 <sup>a</sup>	0.438 <sup>a</sup>	0.367 <sup>a</sup>	1	-	-	-
集中注意力困难	0.427 <sup>a</sup>	0.366 <sup>a</sup>	0.305 <sup>a</sup>	0.416 <sup>a</sup>	0.344 <sup>a</sup>	0.408 <sup>a</sup>	1	-	-
烦躁不安	0.354 <sup>a</sup>	0.362 <sup>a</sup>	0.329 <sup>a</sup>	0.289 <sup>a</sup>	0.332 <sup>a</sup>	0.346 <sup>a</sup>	0.434 <sup>a</sup>	1	-
消极观念	0.460 <sup>a</sup>	0.479 <sup>a</sup>	0.396 <sup>a</sup>	0.464 <sup>a</sup>	0.441 <sup>a</sup>	0.536 <sup>a</sup>	0.422 <sup>a</sup>	0.420 <sup>a</sup>	1
PHQ 总分	0.730 <sup>a</sup>	0.693 <sup>a</sup>	0.644 <sup>a</sup>	0.695 <sup>a</sup>	0.642 <sup>a</sup>	0.720 <sup>a</sup>	0.668 <sup>a</sup>	0.616 <sup>a</sup>	0.717 <sup>a</sup>

注:<sup>a</sup> $P<0.01$ 。

2.2.2 重测信度 采用 Pearson 相关系数对 PHQ-9 初测与 4 周后重测进行分析,相关系数为 0.88 ( $P < 0.01$ )。

### 2.3 效度检验

2.3.1 校标效度 PHQ-9 总分与 BDI-13 总分作相关分析,显示两者呈正相关( $r = 0.77$ ,  $P < 0.01$ )。

2.3.2 结构效度检验 进行探索性因素分析,结果显示 KMO 检验值为 0.90,球形检验 Bartlett 检验统计量 1346.00 ( $P < 0.001$ ),表明条目间有共同因子存在,可进行因子分析。采用主成分因子分析提取一个特征值 4.21,该因子的方差累计占总数的 46.75%,该因子在各条目中的载荷系数均大于 0.618 (0.618 ~ 0.761)。由此表明,该因子(躯体化-情感因子)在所有条目中均突出。

2.4 ROC 曲线 采用 DSM-IV 临床定式访谈(SCID)对所有完成问卷的被试( $n = 471$ )进行定式访谈,并以此诊断结果作为 ROC 曲线的标准,以最大 Youden 指数(灵敏度 + 特异度 - 1)作为临界值的甄选标准,以阳性预测值和阴性预测值作为参考值。

PHQ-9 的 ROC 曲线下面积(AUC)为 0.984 (95% CI: 0.969 ~ 0.993) (见图 1)。PHQ-9 在不同划界分上的准确性,PHQ-9 得分为 10 时灵敏度和特异度为 93.33% 和 96.83%,Youden 指数最大,且得分从 9 增至 10 时阳性预测值由 0.483 增至 0.667,提示 10 分作为诊断划界分更为均衡。见表 2。

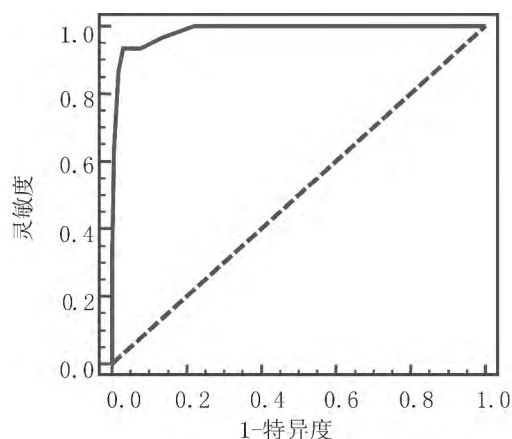


图 1 PHQ-9 的 ROC 曲线图

注: - 代表 PHQ-9 的 ROC 曲线; - - 代表诊断参考线

表 2 PHQ-9 不同划界分的灵敏度、特异度、阳性和阴性预测值

划界分	灵敏度% (95% CI)	特异度% (95% CI)	阳性预测值(95% CI)	阴性预测值(95% CI)
PHQ-9 总分				
>8	93.33(0.779-0.992)	92.29(0.894-0.946)	0.452(0.325-0.583)	0.995(0.982-0.999)
>9	93.33(0.779-0.992)	93.20(0.904-0.954)	0.483(0.348-0.619)	0.995(0.983-0.999)
>10	93.33(0.779-0.992)	96.83(0.947-0.983)	0.667(0.502-0.806)	0.995(0.983-0.999)
>11	86.67(0.693-0.962)	98.19(0.965-0.992)	0.765(0.585-0.894)	0.991(0.977-0.998)
>12	63.33(0.439-0.801)	99.32(0.980-0.999)	0.864(0.651-0.971)	0.97.6(0.957-0.988)
>13	43.33(0.255-0.626)	99.55(0.984-0.999)	0.867(0.583-0.985)	0.963(0.941-0.978)
>14	33.33(0.173-0.528)	99.77(0.987-1.000)	0.909(0.587-0.998)	0.957(0.934-0.973)

注: CI: 可信区间(Confidence Interval)。

## 3 讨 论

本研究首次将 PHQ-9 作为抑郁症筛查工具在中国大陆青少年人群中进行心理测量学的检验。结果显示,PHQ-9 是一个可靠的、有效的抑郁症筛查工具,与近期国外的研究结果基本一致<sup>[8-9]</sup>。

研究并未发现青少年抑郁症的检出率存在性别差异,这与国外普遍支持的青春期抑郁症“女性优势论”不相符。但国内最新的两项有关中国儿童青少年抑郁的研究结果也发现中国青少年抑郁的发生无显著性别差异<sup>[1,10]</sup>,这可能与我国特有的文化背

景有关,也可能与中外学校家庭教育模式的差异有关<sup>[11]</sup>,这些假设需要进一步的研究证实。

本研究 PHQ-9 的 Cronbach's  $\alpha$  系数为 0.85,与近期在成年人中的研究基本一致<sup>[12]</sup>,提示其具有良好的内部一致性信度。此外,条目间及各条目与 PHQ-9 总分间均显著相关,进一步提示该量表有良好的同质信度。重测信度 0.88,提示具有良好的测试稳定性。

校标效度分析显示,PHQ-9 总分与 BDI 总分显著相关,这与先前成人样本的研究结果一致<sup>[13]</sup>,提示该问卷在青少年群体中也有较好的效度。结构

效度分析显示,单因子累计贡献率 47%,各条目的载荷系数均大于 0.618,提示具有较好的结构效度,与国内外同类研究一致<sup>[14-15]</sup>。

PHQ-9 的 ROC 曲线下面积(AUC) 0.984,提示其作为筛查工具特性良好。本研究得出 PHQ-9 最佳诊断划界分为 10,与该量表在成人中推荐的 10 分一致。近期的一项关于 PHQ-9 多语种版本的研究显示,划界分在 8~11 分之间均具有较好的筛查和诊断特性<sup>[16]</sup>;国外近期的一项在青少年群体中的研究推荐最佳划界分为 11 分<sup>[9]</sup>。这些发现也提示,不同的地域和人群可能会有不同的诊断划界分。

本研究也存在一些局限性。如研究样本来自中国北方某个省份,如需广泛应用还需要更多地区的样本加以验证;此外,样本中的抑郁症患者比率较低,可能会影响到灵敏度分析的稳定性。因此,以后的研究有必要将取样范围扩大,并在临床患者中取样,进一步验证其在青少年群体中使用的心理测量学特征和临床价值。

总之,PHQ-9 用于青少年人群时,是一个可靠的、有效的抑郁症筛查工具。鉴于其准确性,推荐总分 10 分作为该量表在青少年中抑郁筛查的最佳划界分。

### 参 考 文 献

- [1] 王熙,孙莹,安静,等. 中国儿童青少年抑郁症状性别差异的流行病学调查[J]. 中华流行病学杂志, 2013, 34(9): 893-896.
- [2] Zuithoff NP, Vergouwe Y, King M, et al. The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a cross-sectional study[J]. BMC Fam Pract 2010, 11: 98-104.
- [3] Zhang YL, Liang W, Chen ZM, et al. Validity and reliability of Patient Health Questionnaire-9 and Patient Health Questionnaire-2 to screen for depression among college students in China[J]. Asia-Pacific psychiatry, 2013, 5(4): 268-275.
- [4] 徐勇,吴海苏,徐一峰,等. 病人健康问卷抑郁量表(PHQ-9)在社区老年人群中的应用——信度与效度分析[J]. 上海精神医学 2007, 19(5): 257-259.
- [5] Lai BP, Tang AK, Lee DT, et al. Detecting postnatal depression in Chinese men: a comparison of three instruments[J]. Psychiatry Res 2010, 180(2-3): 80-85.
- [6] Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure[J]. Psychiatr Ann 2002, 32(9): 509-515.
- [7] 张志群,郭兰. Beck 抑郁问卷在成都市中学生中的试用[J]. 中国心理卫生杂志 2004, 18(7): 486-487.
- [8] Richardson LP, McCauley E, Grossman DC, et al. Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents[J]. Pediatrics, 2010, 126(6): 1117-1123.
- [9] Ganguly S, Samanta M, Roy P, et al. Patient health questionnaire-9 as an effective tool for screening of depression among Indian adolescents[J]. J Adolesc Health 2013, 52(5): 546-551.
- [10] Chen J, Li X, Natsuaki MN, et al. Genetic and Environmental Influences on Depressive Symptoms in Chinese Adolescents[J]. Behav Genet 2014, 44(1): 36-44.
- [11] 王君,张洪波,胡海利,等. 儿童抑郁量表在合肥市中小学生的应用[J]. 中国学校卫生, 2009, 30(4): 336-338.
- [12] 李振华,肖亚洲,谢知,等. 病人健康问卷抑郁量表(PHQ-9)在农村社区老年人群中的应用[J]. 中国临床心理学杂志, 2011, 19(2): 171-174.
- [13] Cameron IM, Cardy A, Crawford JR, et al. Measuring depression severity in general practice: discriminatory performance of the PHQ-9, HADS-D, and BDI-II[J]. Br J Gen Pract, 2011, 61(588): 419-426.
- [14] Merz EL, Malcarne VL, Roesch SC, et al. A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas[J]. Cultur Divers Ethnic Minor Psychol, 2011, 17(3): 309-316.
- [15] Yu X, Tam WW, Wong PT, et al. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong[J]. Compr Psychiatry 2012, 53(1): 95-102.
- [16] Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis[J]. CMAJ 2011, 184(3): E191-196.

(收稿日期: 2014-08-17)