

한글판 우울증 선별도구(Patient Health Questionnaire-9, PHQ-9)의 신뢰도와 타당도

울산대학교 의과대학 서울아산병원 정신과학교실
박승진 · 최혜라 · 최지혜 · 김건우 · 홍진표

Reliability and Validity of the Korean Version of the Patient Health Questionnaire-9 (PHQ-9)

Seung-Jin Park, MA, Hye-Ra Choi, MA, Ji-Hye Choi, MA, Kunwoo Kim, MD and Jin Pyo Hong, MD, PhD
Department of Psychiatry, College of Medicine University of Ulsan, Asan Medical Center, Seoul, Korea

ABSTRACT

Objective : The reliability and validity of the Korean version of the Patient Health Questionnaire-9 (PHQ-9) was examined in Korean patients with depressive symptoms.

Methods : Eighty six outpatients diagnosed as major depressive disorder or depressive episode of bipolar I disorder according to the DSM-IV criteria were assessed with the PHQ-9, Hamilton Depression Rating Scale (HDRS), the Quick Inventory of Depressive Symptomatology Self Report (QIDS-SR), and the Center for Epidemiologic Studies Depression Scale (CES-D).

Results : The Cronbach's alpha coefficient from the PHQ-9 was 0.81. And the correlations of each item with the total score were statistically significant ($r=0.28 - 0.70, p<0.01$). The test-retest correlation coefficient ($r=0.89, p<0.01$) was relatively high and correlations of the PHQ-9 with the HDRS, QIDS-SR and CES-D were 0.70, 0.81, and 0.81 respectively.

Conclusion : These results demonstrated that the Korean version of PHQ-9 could be a reliable and valid tool for the screening and assessment of depressive patients. The Korean version of PHQ-9 will be a useful tool for screening depressive symptoms in Korea. (Anxiety and Mood 2010;6(2):119-124)

KEY WORDS : Korean version of Patient Health Questionnaire (PHQ-9) · Reliability · Validity.

서 론

우울장애는 대중에서 매우 흔하면서도 치료 수요가 많은 장애여서 공중보건관점에서 가장 중요한 정신질환 중의 하나이다. 실제로 최근 통계에 의하면 미국의 경우 주요우울증으로 고통 받는 인구가 어느 한 시점에서 측정했을 경우 6.61¹로 추산되고 있으며, 인생에서 한 번이라도 주요우울증으로 고통 받는 사람이 미국 전체 인구의 16%²로 추산되고 있다. 국내의 경우, 주요우울증(major depression)의 유병률이 약 3%, 기분부전장애가 약 2%로 나타나고 있

어서, 전체 인구 가운데 약 5%³ 정도가 우울장애로 고통 받고 있는 것으로 보인다.

우울장애는 환자들의 기능에 장애를 초래하고, 일상적인 활동을 유지하는 데 어려움을 불러일으키며 자살의 위험성을 증가시키기 때문에, 조기 발견을 통한 적절한 치료가 매우 중요한 질환이다. 이와 관련해 The 2002 U.S. Preventive Services Task Force(USPSTF) 보고에서는 우울증을 파악하고 환자의 예후를 향상시키기 위해 성인에서 우울증 선별검사를 시행할 것을 권고하고 있다.⁴

우울장애의 증상을 자기보고를 통하여 측정하는 도구가 다수 개발되어 있지만, 국내외 일차의료영역에서 널리 사용되고 있는 Beck 우울척도, 해밀턴 우울척도 등 대부분의 척도는 항목수가 많아 실시하는 데 시간이 오래 걸리고, 현대적 우울증상과는 다른 측면이 많다는 한계가 있다. 따라서, 좀 더 실사가 간편하고 DSM-IV 진단기준에 부합하는 측정도구의 필요성이 제기되고 있다.

접수일 : 2010년 9월 15일 / 수정일 : 2010년 10월 15일
게재확정일 : 2010년 10월 18일

Address for correspondence

Jin Pyo Hong, M.D., Ph.D., Department of Psychiatry, College of Medicine University of Ulsan, Asan Medical Center, 388-1 Pungnap-2dong, Songpa-gu, Seoul 138-736, Korea
Tel : +82.2-3010-3408, Fax : +82.2-485-8381
E-mail : jphong@amc.seoul.kr

1999년에 Spitzer 등⁵이 개발한 Patient Health Questionnaire (PHQ)는 일차적 임상장면에서 접하기 쉬운 정신 질환들을 감지하고 진단에 도움을 주기 위해 개발된 자기 보고식 설문지이다. 그 중 주요우울장애의 진단을 위해 9 가지 문항으로 이루어진 Patient Health Questionnaire-9 (PHQ-9)은 DSM-IV의 우울삽화의 진단기준과 일치하게 고안되었고, 각 항목당 0점에서 3점까지 증상의 정도에 따라 선택을 하게 한 후, 그 합을 구하여 총 27점 중 10점을 우울증상에 대한 절단점 (cut-off point)으로 설정하였다. 우수한 민감도(88%)와 특이도(88%)를 가진다고 보고되었으며, 기존의 우울증 선별척도보다 문항 수가 적고 검사실시하는 데 시간이 적게 들어, 환자에게 긴 시간을 할애할 수 없는 일차적 임상장면에 적합한 도구로 여겨진다. 2010년 6월 전세계 10여 개 이상의 나라에서 PHQ-9을 자국어로 번안하여 사용하고 있다.

한국의 경우, 최홍석 등⁶이 번역한 PHQ-9이 몇몇 연구에서 사용되고 있지만, PHQ-9의 신뢰도와 타당도에 대해서는 체계적인 연구가 다소 부족한 실정이다.

이에 저자들은 한글로 번역한 PHQ-9을 개발하여 신뢰도와 타당도를 살펴보았다.

연구 방법

연구대상

본 연구는 2009년 3월부터 8월까지 서울에 소재한 일개 대학병원 정신과 외래방문객을 대상으로 5개월간 실시되었다. 정신과 전문의의 면담과 의무기록지 내용을 토대로 DSM-IV에 근거하여 주요우울장애 또는 양극성 장애의 우울 삽화로 진단된 환자 86명(주요우울증 환자 84명, 양극성장애 우울삽화 환자 2명)을 연구대상으로 하였다. 명백한 신체질환을 가지고 있는 환자들은 제외하였고, 한글을 읽고 쓸 수 있는 정도의 인지 수준을 가지고 있는 사람으로 국한하였다.

연구도구

한글판 Patient Health Questionnaire-9(PHQ-9)의 제작

2001년 Kroenke 등⁷이 발표한 논문의 부록에 수록된 PHQ-9을 번역하였다. 원저자에게 e-mail을 통해 원문의 사용에 대한 허락을 얻었다. 우리나라의 사회문화적인 특성을 감안하되, 가능하다면 변형을 최소화하여 원척도의 체제와 의미를 그대로 전달한다는 원칙을 세우고, 번역위원

회(2명의 정신과 전문의, 1명의 임상심리사)에서 함께 검토하고 수정하면서 번역본을 만들었다. 이렇게 만들어진 번역본을 전문 번역기관에서 역번역하였고, 이를 다시 번역 위원회에서 원본과 비교 감수하였다.

비교 측정도구

Hamilton Depression Rating Scale(HDRS)

HDRS는 현재 우울증에 대한 임상적 평가 및 치료 효과 검증을 위하여 가장 널리 이용되고 있는 평가도구로, 처음에는 21개의 문항으로 만들어졌으나, 현재는 17개 문항의 척도가 많이 쓰이고 있다. 각 항목의 총점으로 증상의 정도를 평가하며, 총점의 범위는 0점에서 52점까지이고 점수가 높을수록 우울증의 심각도가 높아짐을 나타낸다. 본 연구에서는 2005년 이중서 등⁸이 표준화한 HDRS를 사용하였다.

Quick Inventory of Depressive Symptomatology-Self Report(QIDS-SR)

QIDS-SR은 IDS(Inventory of Depressive Symptomatology)의 단축형으로, 자기보고식 설문지이다. 16개 문항으로 이루어져 있고, 16개 문항의 반응을 DSM-IV의 9 가지 우울 증상의 기준에 맞춰 채점하게 된다. 각각의 항목은 0점에서 3점까지 채점되고, 총점의 범위는 0점에서 27점까지이다. 본 연구에서는 홍진표 등⁹이 번역한 QIDS-SR을 사용하였다.

Center for Epidemiologic Studies Depression Scale (CES-D)

우울증에 대한 역학 조사에서 주요 측정도구로 사용되고 있으며, 20문항으로 구성되어 있는 자기보고식 설문지이다. 척도의 문항은 4점 척도로 반응하도록 구성되었으며, 총점이 높을수록 우울증상의 심각도가 높아짐을 나타낸다. 국내에서는 전검구 등¹⁰이 2001년에 표준화하였으며, 본 연구에서는 이를 사용하였다.

PHQ-9 및 비교측정도구의 평가 실시

먼저 의무기록지를 검토하고, 주치의 또는 보호자를 통해 환자에 대한 정보를 수집한 후, 경험 많은 임상가가 환자와 면담을 해서 HDRS 평가치를 기록하였다. 이후 평가자가 동석한 자리에서 환자로 하여금 PHQ-9, QIDS-SR, CES-D를 완성하도록 하였다.

이후, 전체 환자 중 30명을 대상으로 5~6일 간격으로 다시 외래를 방문하였을 때, PHQ-9를 다시 실시하도록

해, 검사-재검사 신뢰도를 구하였다.

자료 분석

신뢰도 검증

PHQ-9의 내적 일관성 신뢰도 검증에는 Cronbach's alpha와 교정된 총점-항목간 상관계수를, 검사 재검사 신뢰도 검증에는 Pearson 상관계수를 각각 이용하였다.

타당도 검증

PHQ-9과 HDRS, CES-D, QIDS간의 상관관계를 Pearson 상관계수로 조사하여 공존타당도를 검증하였다. 모든 자료는 SPSS(version 15.0)로 분석하였다.

결 과

연구대상의 인구학적 특징

대상군은 남자가 17명으로 전체의 19.8%를, 여자가 69명으로 전체의 80.2%를 차지하였다. 평균 연령은 50±14.6세(20세~74세)였다. 교육정도에서는 무학이 7명(8.4%), 초등학교 중퇴 또는 졸업자가 12명(14.5%), 중학교 중퇴 혹은 졸업이 10명(12.0%), 고등학교 중퇴 혹은 졸업이 24명(28.9%), 대학교 중퇴 이상이 30명(36.1%)이었다.

PHQ-9의 총점 평가치는 15.3±6.22의 분포를 보였다. PHQ-9 총점은 남녀에 따른 차이를 보이지 않았으며(t=-1.18, df=84, p=.24), 각 항목에서도 남녀간 점수 차이는 발견되지 않았다(Table 1).

신뢰도

내적 일관성 신뢰도

Cronbach's alpha로 살펴본 내적 일관성 신뢰도는 0.81

Table 1. Sociodemographic and clinical characteristics of study subjects

	Male	Female	p
Number (%)	17 (19.8)	69 (80.2)	
Age (M±SD)	52.5±13.7	49.4±14.9	ns*
Education			ns†
0	1 (5.9)	6 (9.1)	
1-6	1 (5.9)	11 (16.7)	
7-9	2 (11.8)	8 (12.1)	
10-12	2 (11.8)	22 (33.3)	
>13	11 (64.7)	19 (28.8)	
Unknown		3	
PHQ-9 total score	13.7± 7.03	15.7±5.60	ns*

* : independent t-test, † : Fisher's exact test

이었고, 각 항목을 차례로 제거시켰을 경우에도 0.760~0.811 정도의 값을 보여 그 값은 크게 변하지 않았다. PHQ-9의 교정된 항목-총점간 상관관계(corrected item-total correction)는 대부분 .50을 넘었으나(0.28~0.70), 수면 감소/증가 항목과 식욕 감소/증가 항목은 다소 낮은 상관성을 보였다(Table 2).

검사-재검사 신뢰도

전체 환자 중 30명을 대상으로 5~6일 간격으로 PHQ-9를 다시 실시하도록 했을 때 검사-재검사 신뢰도는 0.89였다. 각각의 문항간 상관도 모두 .60을 넘었다(0.64~0.93)(Table 3).

타당도

PHQ-9의 공존 타당도

PHQ-9의 총점과 HDRS, QIDS, CES-D의 총점과의 상관관계를 조사하였을 때, 각 척도들과 PHQ-9의 총점 사이에 모두 유의미한 상관성이 관찰되었다(Table 4). 자기 보고형 척도인 CES-D나 QIDS에 비해 관찰자 평가척도

Table 2. Correlations of each item score with the total score of the PHQ-9

Items	Corrected Item-total correlation
Depressed mood	0.70*
Little interest	0.60*
Reduced/Increased sleep	0.28*
Reduced/Increased appetite	0.41*
Psychomotor retardation/agitation	0.47*
Feeling tired	0.64*
Guilty feeling	0.45*
Concentration problem	0.49*
Suicidal ideation	0.51*

* : p-value<0.01

Table 3. Test-retest reliability of the PHQ-9 items

Items	Corrected Item-total correlation
Depressed mood	0.64*
Little interest	0.84*
Reduced/Increased sleep	0.73*
Reduced/Increased appetite	0.89*
Psychomotor retardation/agitation	0.81*
Feeling tired	0.69*
Guilty feeling	0.89*
Concentration problem	0.83*
Suicidal ideation	0.93*

* : p-value<0.01

Table 4. Correlations of the total scores of each scale with the total score of the PHQ-9

Scales	PHQ-9	HDRS	QIDS	CES-D
PHQ-9	1			
HDRS	0.70*	1		
QIDS	0.81*	0.78*	1	
CES-D	0.81*	0.72*	0.76*	1

* : p-value<0.01

인 HDRS와의 상관성이 약간 낮았다.

고 찰

본 연구에서는 일차의료장면에서 접하기 쉬운 정신질환 중 우울증을 감지하고 진단에 도움을 주기 위해 개발된 자기 보고식 질문지인 PHQ-9의 한국판을 개발하고 그 신뢰도와 타당도를 측정하고자 하였다.

PHQ-9의 내적 일관성 신뢰도(Cronbach' alpha=0.84)는 높게 나타났는데, 이는 Kroenke 등⁷의 연구에서 관찰된 것과 비슷한 수치이다. 국내 연구를 살펴보면, 최홍석 등⁶의 연구에서도 내적 일관성 신뢰도가 0.852로 나와 본 연구의 결과와 유사한 것으로 보인다. Cronbach's α 를 사용하는 경우, 집단 수준인 경우에는 α 값이 0.6 이상, 개인 수준에서는 0.9 이상이면 높다고 할 수 있다. 본 연구에서는 Cronbach's α 값이 0.84로 나와 개인적 수준에는 못 미치지만 집단수준에는 높은 신뢰도를 보여주었고, 각 항목을 차례로 제거시켰을 경우에도 내적 신뢰도는 0.806~0.841 정도의 값을 보였다.

개개 항목과 PHQ-9 총점과의 상관계수 역시 전 항목에 걸쳐 통계적으로 유의미한 상관성을 보여, 구성항목의 동질성(homogeneity)을 확인할 수 있었다. 각각의 항목에 대해서 좀 더 자세히 보면, 다른 항목에 비해 식욕 저하/증가나 수면 저하/증가 항목이 총점-항목간 상관관계가 낮은 편인데, 이는 다른 우울증 척도를 사용한 국내외 여러 연구^{11,12}에서도 비슷하게 나타나는 결과이다.

약 5~6일 간격을 두고 PHQ-9을 재시행하였을 때, PHQ-9의 전체 점수에 대한 검사-재검사 신뢰도는 0.89로 높았으며, 각 문항들도 모두 의미 있는 상관관계를 나타내어서, PHQ-9의 구성문항들이 시간이 지나더라도 개인의 우울 증상을 매우 안정되게 측정하는 것으로 나타났다. 따라서 PHQ-9은 이후 개인의 우울증상의 변화 정도를 측정하는 도구로도 유용하게 쓰일 수 있을 것으로 보인다.

PHQ-9의 공존타당도를 보면, 관찰자 평가척도인 HDRS와는 0.70, 자기보고형 척도인 CES-D와 QIDS와는 0.81의 높은 상관성을 보였다. 일반적으로 관찰자 평가척도와

자기보고형 척도는 상관관계가 높지 않다고 알려져 있는데,¹³ PHQ-9의 경우 HDRS와의 상관이 다른 자기보고형 척도에 비해 다소 낮은 하지만, 기존의 다른 자기보고형 우울척도에 비해서는 관찰자 평정척도와 더 높은 상관관계를 보였다.

본 연구 대상자의 80% 이상이 여성이고 연구 대상 인원이 다소 적었다는 점과, 본 연구에 참여한 인구의 대다수가 40대와 50대로서 다소 연령이 높은 편이며, 학력이 비교적 높은 층이 많이 포함된 점 등이 본 연구의 제한점으로 고려할 수 있다. 물론 이러한 인구통계학적 변인의 영향성은 한국판 PHQ-9의 신뢰도와 타당도를 근본적으로 변화시키지는 않을 것이나, 보다 양호한 검사의 개발을 위해서는 고려해야 할 점으로 볼 수 있다.

결 론

주요우울장애 또는 양극성장애의 우울삽화 환자를 대상으로 한 한국판 PHQ-9은 만족할 만한 내적 일관성 신뢰도와 높은 검사-재검사 신뢰도를 보였다. 또한 HDRS, QIDS-SR 및 CES-D 등 다른 우울증 평가척도와 유의한 공존타당도를 나타내어 우울증을 평가하고 선별하는 척도로서의 신뢰도와 타당도가 입증되었다. 한국판 PHQ-9의 경우 기존 우울증 평가도구보다 문항수가 적어 작성 시간이 적게 걸리고 점수 산정이 용이하여 일차 진료 과정에도 적용하기 쉬워, 우울증상을 선별하고 평가하는 데 유용한 도구로 사용될 수 있을 것으로 보인다.

앞으로 증상의 종류와 심각도의 면에서 더욱 다양한 환자군 및 일반인을 대상으로 한 연구가 필요할 것으로 보인다.

중심 단어 : 한글판 PHQ-9 · 신뢰도 · 타당도.

REFERENCES

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters, EE. The epidemiology of major depressive disorder. JAMA 2003;289: 3095-3105.
2. Andrade L, Caraveo-Anduaga JJ, Berglund P, Biji RV, De Graaf R, Vollebergh W. et al. The epidemiology of major depressive episodes: Results from the International Consortium of Psychiatric Epidemiologist (ICPE) Surveys. International Journal of Methods in Psychiatric Research 2003;12:3-21.
3. 이정균, 이희, 광영숙, 김용식, 한진희, 최진옥. 한국정신장애의 역학적 조사연구. 서울의대정신의학 1986;11 부록.
4. Pignone MP, Gaynes BN, Rushton JL, Burchell CM, Orleans CT, Mulrow CD, et al. Screening for depression in adults: a summary of the evidence for the U. S. Preventive Services Task Force. Ann Intern Med 2002;136:765-776.
5. Spitzer RL, Kroenke K, Williams JB. Valiation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA

- 1999;282:1737-1744.
6. Choi HS, Choi JH, Park KH, Joo KJ, Ga H, Ko HJ, et al. Standardization of the Korean Version of Patient Health Questionnaire-9 as a Screening Instrument for Major Depressive Disorder. *J Korean Acad Fam Med* 2007;28:114-119.
 7. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606-613.
 8. Yi JS, Bae SO, Ahn YM, Park DB, Noh KS, Kim YS, et al. Validity and Reliability of the Korean Version of the Hamilton Depression Rating Scale (K-HDRS). *J Korean Neuropsychiatr Assoc* 2005;44:456-465.
 9. Hong JP, Park SJ, Choi JW, Kim KW, Choi JH. Reliability and Validity of the Korean Version of the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR). *Anxiety and Mood* (unpublished observation).
 10. Chon KK, Choi SC, Ynag BC. Integrated Adaptation of CES-D in Korea. *Korean Journal of Health Psychology* 2001;6:59-76.
 11. Ahn YM, Lee KY, Yi JS, Kang MH, Kim DH, Kim YS, et al. A validation Study of the Korean-Version of the Montgomery-Asberg Depression Rating Scale. *J Korean Neuropsychiatr Assoc* 2005;44:466-476.
 12. Davidson J, Turnbull CD, Strickland R, Miller R, Graves K. The Montgomery-Asberg Depression Scale: Reliability and validity. *Acta Psychiatr Scand* 1986;73:544-558.
 13. Hamilton M. Mood disorders: clinical features. In: Kaplan HI, Freedman AM, Sadock BJ (eds): *Comprehensive Textbook of Psychiatry*. Vol. 5. Baltimore, MD: Williams & Wilkins;1989. p.892-913.

□ 부 록 □

한글판 우울증 선별도구(Patient Health Questionnaire-9)

본 검사는 우울한 정도를 스스로 알아보기 위한 것입니다. 이 질문들이 확정된 진단을 위한 것은 아니지만 높은 점수가 나왔을 경우에는 우울증의 가능성이 높으므로, 더 정확한 평가를 위해서 병원에서 진료를 받아볼 것을 추천합니다.

지난 2주간, 얼마나 자주 다음과 같은 문제들로 곤란을 겪으셨습니까?
 지난 2주 동안에 아래와 같은 생각을 한 날을 헤아려서 해당하는 숫자에 표시하세요.

지난 2주 동안에	없음	2, 3일 이상	7일 이상	거의 매일
1. 기분이 가라앉거나, 우울하거나, 희망이 없다고 느꼈다	0	1	2	3
2. 평소 하던 일에 대한 흥미가 없어지거나 즐거움을 느끼지 못했다	0	1	2	3
3. 잠들기가 어렵거나 자주 깬다/혹은 너무 많이 잤다	0	1	2	3
4. 평소보다 식욕이 줄었다/혹은 평소보다 많이 먹었다	0	1	2	3
5. 다른 사람들이 눈치 챌 정도로 평소보다 말과 행동이 느려졌다/ 혹은 너무 안절부절 못해서 가만히 앉아있을 수 없었다	0	1	2	3
6. 피곤하고 기운이 없었다	0	1	2	3
7. 내가 잘못 했거나, 실패했다는 생각이 들었다/ 혹은 자신과 가족을 실망시켰다고 생각했다	0	1	2	3
8. 신문을 읽거나 TV를 보는 것과 같은 일상적인 일에도 집중할 수가 없었다	0	1	2	3
9. 차라리 죽는 것이 더 낫겠다고 생각했다/혹은 자해할 생각을 했다	0	1	2	3

각 칸별로 점수를 더해주세요

	+		+	
--	---	--	---	--

총점=

--

총점에 따른 결과보기

1-4	5-9	10-19	20-27
우울증 아님	가벼운 우울증	중간정도 우울증	심한 우울증