

한글판 우울증 선별 도구 (Patient Health Questionnaire-2)의 표준화 연구

계명대학교 의과대학 정신건강의학교실,¹ 영남대학교 심리학과,² 경북대학교 심리학과³
신재현¹ · 김희철¹ · 정철호¹ · 김정범¹ · 정성원¹ · 조현주² · 정성훈³

The Standardization of the Korean Version of the Patient Health Questionnaire-2

Jae-Hyun Shin, MD¹, Hee-Cheol Kim, MD, PhD¹, Chul-Ho Jung, MD, PhD¹,
Jung-Bum Kim, MD, PhD¹, Sung-Won Jung, MD, PhD¹, Hyun-Ju Cho, PhD² and
Sung-hoon Jung, MA³

¹Department of Psychiatry, Keimyung University School of Medicine, Daegu, Korea

²Department of Psychology, Youngnam University, Daegu, Korea

³Department of Psychology, Kyungpook National University, Daegu, Korea

Objectives An early diagnosis to improve the outcome of depression demands a briefer and more reliable measure than any other diagnosis. The Patient Health Questionnaire (PHQ) is reliable measure made in 1999. In this study, we tried to verify the reliability and validity of two types of PHQ-2, and to compare these two types.

Methods From July of 2011 to September of 2012, a total of 74 outpatients and inpatients in the Department of Psychiatry, Keimyung University Dongsan Medical Center performed the depression scales. As a normal control group, 58 individuals working in the hospital performed the PHQ-2 and PHQ-9.

Results The internal consistency and test-retest reliability were favorable. Two types of the PHQ-2 were statistically significant in correlation between item and total scores. In comparison of mean scores of the PHQ-2 between two groups, the differences between groups were statistically significant. Correlation of the total scores of other depression scales with those of the PHQ-2 was statistically significant both in scales. Examined according to the severity of depression by other scales, the total mean scores of the PHQ-2 tended to increase according to the severity of depression. The optimal cut-off score was 3 in Scaled PHQ-2 (Sensitivity : 91.9%, Specificity : 100%).

Conclusion The assessed reliability and validity of the PHQ-2 made it an attractive measure for depression screening over other scales. Scaled PHQ-2 was more reliable and valid than Yes or No PHQ-2. Therefore, the PHQ-2 is a briefer and more useful measure for screening depression.

J Korean Neuropsychiatr Assoc 2013;52:115-121

KEY WORDS PHQ-2 · Patient Health Questionnaire · Validity · Reliability.

Received January 28, 2013
Revised March 17, 2013
Accepted March 25, 2013

Address for correspondence

Hee-Cheol Kim, MD, PhD
Department of Psychiatry, Keimyung University Dongsan Medical Center,
56 Dalseong-ro, Jung-gu,
Daegu 700-712, Korea
Tel +82-53-250-7813
Fax +82-53-250-7810
E-mail mdhck@dsmc.or.kr

서 론

우울증은 정신과적 질환 중 흔히 진단되는 질환이며 치료가 필요한 질환이다. 미국의 경우 주요우울증의 시점 유병률이 6.61%,¹⁾ 평생 유병률이 16%²⁾로 추산되며, 국내의 경우 2011년 정신질환 실태조사에서 주요우울증의 평생 유병률과 일년 유병률은 각각 6.7%, 3.0%로 나타났다.³⁾

우울증은 개인적, 사회적으로 현저한 기능 장애와 활동 장애를 유발하여 큰 부담이 되는 질환이다. 실제 우울증으로 인한 부담을 추산해 볼 때 미국의 경우 2000년대 들어 우울증

은 8.3억 달러 가량의 경제적 부담을 준다고 하였다.⁴⁾ 우울증 환자는 약 15% 가량에서 자살로 생을 마감하는데 국내의 경우 자살률은 2010년에 인구 10만 명당 31.2명으로 매년 증가하고 있으며 경제협력개발기구(Organization for Economic Cooperation Development) 회원 국가의 평균 자살률보다 약 2.5배로 높은 수준이다.⁵⁾

일반적으로 우울증을 가지고 있는 환자들은 심리적인 문제로 도움을 청하는 경우가 많지 않아 초기 면담에서 우울증으로 진단할 수 있는 환자는 그리 많지가 않다.⁶⁾ 우울증 환자의 30~50%에서는 우울증의 정확한 진단이 내려지지 않고

있다.^{7,8)} 세계보건기구(World Health Organization)에 따르면 일차 진료기관에 내원한 우울증 환자들 중 42%만이 진료하는 의사에게 발견된다고 알려져 있다.⁹⁾

많은 수의 우울증 환자는 일차 진료의를 찾는 경우가 많고 최근 들어 항우울제의 처방도 일차 진료의에서 증가하고 있는 추세이므로 일차 의료기관에서 우울증 선별이 꼭 필요하다.¹⁰⁾ 그러나 국내의 진료 환경 상에서 다양한 증상을 호소하는 우울증 환자에게 충분한 시간을 할애하기 어려우며 일차 진료의는 정신건강의학과 전문의에 비하여 충분한 지식을 가지고 있지 못할 수도 있다.

그러므로 우울증을 조기에 적절하게 진단하기 위하여 간편하고 소요 시간이 적으면서도 신뢰도와 타당도가 높은 선별검사도구가 필요하다. 이와 관련하여 미국에서 2009년 발표된 U.S. Preventive Services Task Force에서는 우울증에 대한 선별검사를 실시하여 우울증 환자의 발견율을 높여야 한다고 공표하였다.¹¹⁾

현재까지 우울증을 선별하기 위한 많은 척도가 개발되어 왔다. 대표적인 것으로 Patient Health Questionnaire-2(이하 PHQ-2),¹²⁾ Children's Depression Inventory,¹³⁾ Center for epidemiological Studies Depression,¹⁴⁾ Beck Depression Inventory(이하 BDI),¹⁵⁾ Geriatric Depression Scale 등이 있다.^{16,17)}

Patient Health Questionnaire-2는 우울증을 선별하기 위하여 1999년 Spitzer 등¹⁸⁾에 의해 개발된 도구이다. PHQ-2는 기존에 개발된 우울증 선별도구인 Patient Health Questionnaire-9(이하 PHQ-9)¹⁹⁾의 9가지 문항 중 우울한 기분과 무쾌감증 두 가지의 항목만을 포함시켜 평가한다. PHQ-2는 평가 방법에 따라 단답형과 점수형의 두 가지 방식이 있다. 단답형 PHQ-2는 문항에 대하여 '예'와 '아니오' 중에서 답을 선택하고¹⁹⁾ 점수형 PHQ-2는 심각도에 따라 총 4가지의 답 중에서 하나를 선택한다.²⁰⁾ 기존의 연구에서, 우울증을 선별하는 데 있어 단답형 PHQ-2를 사용하여 한 가지 항목에만 '예'라고 답하였을 경우 85~90% 정도의 민감도를 보였으며 두 가지 항목 모두에 '예'라고 답하였을 경우에는 95%의 민감도를 보였다.^{19,21)} 또한, 파킨슨병이나 후천성 면역결핍증후군(Acquired Immune Deficiency Syndrome, 이하 AIDS)을 가진 환자를 대상으로 PHQ-2를 시행하여 그 유용성이 입증되기도 하였다.^{22,23)} 국내에서는 2000년 Myung 등²⁴⁾이 청소년을 대상으로 한 '두 문항 질문검사의 표준화 연구'를 시행하여 이 검사가 기존 검사보다는 시간이 덜 걸리는 유용한 도구라고 하였다. 하지만 이 논문에서는 청소년만을 대상으로 연구를 시행하였으며 정신건강의학과가 아닌 가정의학과에서 우울증을 진단하고 연구를 진행한 연구방법론상의 제한점이

있다. 기존 연구들에서는 두 가지 유형의 PHQ-2 중 한 가지 방식(점수형 PHQ-2)만을 사용하여 연구하였기에 단답형 PHQ-2에 대한 연구는 부족하다.

따라서 이 연구에는 일반 성인 인구를 대상으로 두 가지 방식의 PHQ-2를 표준화하고 두 가지 방식의 PHQ-2(단답형과 점수형)를 비교하고자 하였다.

방 법

대 상

이 연구는 2011년 7월부터 2012년 10월까지 일개 대학병원 정신건강의학과 외래 및 입원 환자를 대상으로 정신건강의학과 전공의가 Mini-International Neuropsychiatric Interview-Plus(이하 MINI-Plus)를 시행하여 주요우울증으로 진단된 자 74명을 대상으로 하였다. 정상 대조군은 동 대학병원에 재직 중인 정신질환 병력이 없는 건강한 직원 58명이었다.

연구대상을 선정하는 과정에서 18세 미만의 경우 주요우울증의 진단 기준이 다르므로 제외하였다. 정상 대조군의 경우 주요우울증군의 연령대를 고려하여 대부분 30대 후반에서 40대 초반까지의 연령대를 대상으로 하였다.

모든 연구 절차에 대하여 대학병원 내 의료윤리심의위원회(Institutional Review Board, 이하 IRB)의 승인을 받았다(IRB No.11-271).

방 법

외래 및 입원 환자들에게 MINI-plus를 시행하였고 그 중 주요우울증으로 진단된 자를 대상으로 다음과 같은 우울증 평가 척도를 시행하였다. 두 가지 유형의 PHQ-2, PHQ-9, BDI는 연구대상자에게 설문지를 주고 직접 완성하도록 하였고, Hamilton Rating Scale for Depression(이하 HRSD)는 정신건강의학과 전공의 혹은 임상심리사가 시행하였다. PHQ-2에 대한 검사-재검사 신뢰도를 측정하기 위해 첫 검사 후 2주 후에 재검사를 시행하였다. 정상 대조군에게는 두 가지 유형의 PHQ-2, PHQ-9를 작성하게 하였다.

대상자에게 척도를 시행한 후 단답형 PHQ-2의 경우 '예'라고 응답하면 2점, '아니오'라고 응답하면 1점으로, 점수형 PHQ-2의 경우 "전혀 없다"가 0점, "머칠 동안"이 1점, "일주일 이상"이 2점, "거의 매일"을 3점으로 하여 평가하였다.

Patient Health Questionnaire-2(PHQ-2)

Patient Health Questionnaire-2는 일차 진료 영역 및 다른 일반건강문제를 다루는 환경에서 적은 문항수로 간단하게

우울증을 선별하기 위한 자기보고형 검사이다.¹²⁾ Diagnostic and Statistical Manual of Mental Disorders, fourth edition(이하 DSM-IV)의 주요우울증의 진단기준 중 핵심증상에 해당하는 우울한 기분과 무쾌감증 두 가지의 항목으로 구성되어 있으며, 최근 2주 동안 얼마나 이러한 문제를 자주 겪었는지를 알아보는 검사이다. PHQ-2는 평가 방법에 따라 단답형과 점수형의 두 가지 방식이 있다. 점수형 PHQ-2는 “전혀 없다”, “며칠 동안”, “일주일 이상”, “거의 매일”의 4점 척도로 점수의 범위가 0~6점으로 구성되고, 점수형 PHQ-2, “예” 혹은 “아니오”의 반응으로 평가한다.

Patient Health Questionnaire-2의 문항번안은 번역위원회(2명의 정신과 전문의와 1명의 정신과 전공의, 2명의 임상심리사)가 토의를 거쳐 번역하였다. 이를 미국 정신과 전문의 이면서 한국 정신과 전문의인 1인이 감수하여 완성하였다.

Mini-International Neuropsychiatric Interview-Plus (M.I.N.I.-Plus)

Mini-International Neuropsychiatric Interview는 DSM-IV(정신장애의 진단 및 통계편람 4판)와 ICD-10(국제질병분류 10판)의 주요 제 1축 정신과 질환의 진단을 위하여 1998년 미국과 유럽에서 개발된 간단하고 구조화된 면담도구로 다기관 임상연구나 역학조사 등에서 사용되고 있다.²⁵⁾ 국내에서는 2006년 Yoo 등²⁶⁾이 국내에서 사용하기 위하여 한국판으로 표준화하였다. M.I.N.I.-Plus는 기존의 M.I.N.I.보다 자세하게 편집된 것으로 기질적 원인이나 알코올 또는 약물로 인한 증상에 대한 조사도 포함되어 있다. 본 연구에서는 주요 우울증 대상군으로 진단하기 위하여 M.I.N.I.-Plus를 사용하였다.

Patient Health Questionnaire-9(PHQ-9)

Patient Health Questionnaire-9은 간단하게 우울증을 선별하고, 심각도를 평가하기 위한 목적으로 만들어진 자기보고형 검사이다.¹⁸⁾ 각 문항들은 DSM-IV의 주요우울증의 진단기준에 해당하는 9가지 항목으로 구성되어 있으며 최근 2주 동안 얼마나 이러한 문제를 자주 겪었는지를 알아본다. 반응은 “전혀 없음”, “며칠 동안”, “1주일 이상”, “거의 매일”의 4점 척도로 평가되며 점수의 범위는 0~27점으로 구성되어 있다.

Beck Depression Inventory(BDI)

Beck Depression Inventory는 우울증상의 유무와 증상의 심각성 정도를 평가하기 위한 목적으로 제작된 자기보고형 척도로서 우울증의 인지적, 정서적, 동기적, 신체적 증상을

포함하는 21문항으로 되어 있다. 각 문항마다 0점에서 3점으로 채점되어 총점의 범위는 0점에서 63점까지이다. BDI에 대한 많은 연구결과들은 이 척도가 다양한 집단에 걸쳐 신뢰도와 타당도가 높은 척도임을 보여주고 있으며 우리나라에서도 BDI에 대한 표준화 연구가 다양하게 이루어진 바 있다.^{27,28)} 본 연구에서는 1991년 Lee와 Song²⁹⁾이 번역한 BDI를 사용하였다.

Hamilton Rating Scale for Depression(HRSD)

Hamilton Rating Scale for Depression은 현재 우울증에 대한 임상적 평가 및 치료 효과 검증을 위하여 널리 이용되고 있는 평가도구로서 반구조화된 면담 후 평가자가 증상의 심각도를 기술하는 문장 중 하나를 선택하여 평가한다. 현재는 17개 문항의 척도가 많이 쓰이고 있으며 총점의 범위는 0점에서 52점까지이고 점수가 높을수록 우울증의 심각도가 높아짐을 나타낸다. 본 연구에서는 2001년 Kim³⁰⁾이 번역한 HRSD를 사용하였다.

통계 분석

본 연구에서 통계적인 분석은 Statistical Package for the Social Sciences Windows ver. 18.0(SPSS Inc., Chicago, IL, USA)을 사용하였다. 인구통계학적 분석은 빈도 분석 및 기술 통계, Fisher's exact test를 사용하였다. 검사도구의 신뢰도를 측정하기 위하여 내적 일관성 검증을 위한 Cronbach's alpha 계수를 구하였다. 총점-항목 간 상관계수를 구하는 데 있어서 점수형 PHQ-2는 Pearson 상관계수를, 단답형 PHQ-2의 경우 Spearman rank 상관계수를 구하였다. 검사-재검사 신뢰도 검증에는 Pearson 상관계수를 구하였다.

타당도를 측정하기 위하여 주요우울증군과 정상 대조군의 PHQ-2 척도의 총점 차이를 인구통계학적 특성을 보정하여 공분산분석(analysis of covariance)을 이용하여 비교하였고(ANCOVA) PHQ-2와 PHQ-9, BDI, HRSD 간의 공존타당도를 검증하였다. BDI와 HRSD 점수로 주요우울증의 심각도에 따라 세 군으로 나누어 심각도에 따른 각 군의 PHQ-2 총점 변화 양상을 구하였다. 마지막으로 우울증을 진단하는 데 있어 점수형 PHQ-2의 민감도와 특이도를 평가하여 점수 총합의 최적의 절단점을 산정하였다.

결 과

인구 통계학적 특성

인구 통계학적 특성은 표 1과 같다.

대상군 74명 중 남자가 29명으로 39.2%를, 여자가 45명으

로 60.8%를 차지하였다. 평균 연령은 45.41±14.14세(18~72세)였다. 교육 정도에서는 고졸이 22명(31.6%)으로 가장 많았다. 직업의 경우 가정주부를 포함한 무직이 52명(70.2%)으로 과반수를 차지하였다. 결혼 상태는 기혼이 34명(44.2%)으로 가장 많았다. 정상 대조군은 58명으로 남자가 12명(19.7%), 여자가 46명(80.3%)이었으며, 평균 연령은 39.12±10.1세(23~55세)였다. 교육 정도에서는 대졸이 52명으로(85.2%) 대부분이었다. 직업은 정규직이 38명(65.5%)으로 가장 많았으며 결혼 상태는 기혼이 36명(63%)으로 가장 많은 비중을 차지하였다.

연구 대상군과 정상 대조군을 비교하였을 때 연령과 성별에 있어서는 유의한 차이가 없었다. 그러나 학력과 직업, 중

교에서 유의한 차이가 나타났다(p<0.01).

신뢰도 : PHQ-2의 내적 일치도

Patient Health Questionnaire-2 두 가지 방식의 내적 일치도를 검증하기 위하여 Cronbach's alpha 계수를 산출하였다(표 2). 단답형 PHQ-2의 경우 전체 alpha 계수가 0.775였고, 점수형 PHQ-2의 경우 alpha 계수는 0.938이었다. 두 가지 유형 모두에서 문항 간 동질성이 높았다.

척도의 각 항목과 총점 간의 상관관계

각 유형의 항목-총점 간 상관관계를 산출하였을 때 두 가지 방식의 PHQ-2의 문항 모두가 높은 상관관계를 나타내었다(표 3).

Table 1. Demographic characteristics of the study population

Characteristics	Depressed group : n (%)	Control group : n (%)	p
Age			0.08
18-20	1 (1.3)	0	
21-40	29 (36.4)	30 (51.2)	
41-60	38 (51.5)	28 (48.8)	
61-80	8 (9.8)	0	
Sex			0.23
Male	45 (60.8)	46 (80.3)	
Female	29 (39.2)	12 (19.7)	
Education			<0.01
No educated	3 (3.9)	0	
Elementary school	18 (21.4)	0	
Middle school	7 (9.1)	0	
High school	22 (31.6)	0	
>University	24 (34.0)	58 (100)	
Marriage			0.18
Unmarried	23 (33.9)	18 (31.5)	
Married	34 (44.2)	36 (63.0)	
Remarriage	0	2 (3.3)	
Seperation	2 (2.8)	0	
Divorced	11 (14.5)	2 (3.3)	
Widowed	4 (5.6)	0	
Religion			<0.01
Christianity	11 (14.3)	26 (44.8)	
Catholicism	10 (13.5)	6 (10.3)	
Buddhism	24 (32.9)	24 (41.3)	
None	25 (6.2)	2 (3.4)	
Others	4 (5.4)	0	
Occupation			<0.01
Regular	10 (13.5)	38 (65.5)	
None-regular	3 (4)	18 (31.0)	
Unemployed	52 (70.2)	0	
Others	9 (12.1)	2 (3.5)	

n : Number

검사-재검사 신뢰도

대상군 중 14명에서 2주 간격으로 시행한 검사-재검사 신뢰도의 상관 계수를 산출하였다. 산출된 상관계수는 두 가지 유형에서 각각 0.69와 0.77로 모두 양호한 상관관계를 보였다.

타당도 : 주요우울증군과 정상 대조군의 평균차이 검증

주요우울증군과 정상 대조군에서 두 가지 방식의 PHQ-2의 평균치를 비교하였다. 인구학적 특성(학력, 종교, 직업)을 보정하여 ANCOVA로 분석하였을 때 단답형 PHQ-2의 경우 주요우울증군이 3.64±0.59로 정상 대조군(2.20±0.48)보다 평균 점수가 유의하게 높았고, 점수형 PHQ-2 또한 주요우

Table 2. Internal consistency of the PHQ-2

	Cronbach's alpha coefficient
Yes or No PHQ-2	0.775
Scaled PHQ-2	0.938

PHQ-2 : Patient Health Questionnaire-2

Table 3. Correlation of each item score with the total score of the PHQ-2

	Item	Item-total correlation
Yes or No PHQ-2	Depressed mood	0.902*
	Anhedonia	0.905*
Scaled PHQ-2	Depressed mood	0.970*
	Anhedonia	0.971*

* : p<0.01. PHQ-2 : Patient Health Questionnaire-2

Table 4. Comparison of the PHQ-2 scores between the depressed group and the control group

Scales	Depressed group (mean±SD)	Control group (mean±SD)	p
Yes or No PHQ-2	3.64±0.59	2.20±0.48	<0.01
Scaled PHQ-2	4.81±1.31	0.90±0.93	<0.01

SD : Standard deviation, PHQ-2 : Patient Health Questionnaire-2

Table 5. Correlation of the total scores of each scale with the total score of the PHQ-2

Scales	Yes or No PHQ-2	Scaled PHQ-2	PHQ-9	BDI	HRSD
Yes or No PHQ-2	1				
Scaled PHQ-2		1			
PHQ-9	0.346*	0.671*	1		
BDI	0.439*	0.526*	0.356*	1	
HRSD	0.375*	0.416*	0.411*	0.477*	1

* : $p < 0.01$. PHQ : Patient Health Questionnaire, HRSD : Hamilton Rating Scale for Depression, BDI : Beck Depression Inventory

Table 6. The total mean score of the PHQ-2 according to the severity of the depression

Scales	Severity (score/n)	Mean of total score of Yes or No PHQ-2	Mean of total score of Scaled PHQ-2
BDI	Mild (31-36/5)	2.80	2.80
	Moderate (37-44/12)	3.50	4.17
	Severe (45-84/56)	3.77	5.18
HRSD	Mild (7-17/16)	3.06	3.69
	Moderate (18-24/31)	3.84	5.06
	Severe (25-50/26)	3.73	5.19

n : Number, PHQ-2 : Patient Health Questionnaire-2, HRSD : Hamilton Rating Scale for Depression, BDI : Beck Depression Inventory

Table 7. Sensitivity and specificity of the PHQ-2

Cut-off score of the scaled PHQ-2	1	2	3	4	5	6
Sensitivity (%)	100	100	91.9	82.4	63.5	43.2
Specificity (%)	51.7	37.9	100	100	100	100

PHQ-2 : Patient Health Questionnaire-2

울증군이 4.81 ± 1.31 로 정상 대조군(0.90 ± 0.93)보다 유의하게 높았다(표 4).

다른 주요우울증 척도와의 상관성

Mini-International Neuropsychiatric Interview-Plus를 통해서 주요우울증으로 진단된 집단군에서 두 가지 유형의 PHQ-2와 PHQ-9, BDI, HRSD와의 상관성을 살펴보았다(표 5). 단답형 PHQ-2의 경우 PHQ-9와 BDI, HDRS와의 상관계수는 각각 0.346, 0.439, 0.375로 유의한 양의 상관관계가 있었다($p < 0.01$). 점수형 PHQ-2의 경우에도 PHQ-9, BDI, HRSD와의 상관 계수는 각각 0.671, 0.526, 0.416으로 유의한 양의 상관관계가 있었다($p < 0.01$).

주요우울증 심각도에 따른 총점 평균점수의 변화

Patient Health Questionnaire-2의 우울증 심각도 수준의 반영 정도를 살펴보기 위하여 주요우울증으로 진단된 군에서 시행한 BDI와 HRSD 상의 우울증 심각도에 따라 경도, 중등도, 심함의 3가지 군으로 나누었고,^{24,25} 각 군에 해당되는 대상들에서 시행한 PHQ-2의 점수 평균을 비교하였다. 단답형 PHQ-2의 경우 우울증의 심각도가 심해짐에 따라

대체적으로 PHQ-2의 총점 평균점수도 높아졌다(표 6). 점수형 PHQ-2에서도 대상군의 우울증 심각도가 심해짐에 따라 PHQ-2의 총점 평균점수가 높아졌다(표 6).

PHQ-2 총점의 최적 절단점

연구 대상군에게 실시한 점수형 PHQ-2의 응답에 따른 점수의 총합을 구하여 민감도와 특이도를 고려한 최적의 절단점을 구하고자 하였다(표 7). 단답형 PHQ-2는 점수 총합의 범위가 좁아(2~4점) 제외하였다.

점수형 PHQ-2의 경우 “전혀 없다”가 0점, “머칠 동안”이 1점, “일주일 이상”이 2점, “거의 매일”을 3점으로 하여 평가하였고 총점의 범위는 0~6점이었다. 민감도와 특이도를 고려하였을 때 3점의 경우 민감도가 91.9%, 특이도가 100%로 가장 높아 최적 절단점으로 볼 수 있다.

고 찰

이 연구에서는 우울증 선별도구로서 널리 알려진 PHQ-2의 한국판을 개발하고 신뢰도와 타당도를 측정하여 그 유용성을 밝히고자 하였다.

먼저 인구통계학적 특성에서 성별과 나이에서는 주요우울증군과 정상 대조군 사이에 유의한 차이가 없었지만 학력과 종교, 직업에 있어서는 집단 간의 차이가 나타났다. 이는 정상 대조군을 모집하는 데 있어 대학병원 내의 직원들을 대상으로 하였기 때문에 생기는 것으로 보이며 향후 연구에서는 대상군의 인구통계학적 특성에 맞춘 대조군의 선정이 필요하다.

Patient Health Questionnaire-2 설문 항목들의 신뢰도를 측정하기 위하여 Cronbach's alpha 계수를 사용하였다. 집단 수준의 경우에는 alpha값이 0.6 이상, 개인 수준의 경우에는 0.9 이상이면 높다고 할 수 있다.²⁴ 이 연구에서 Cronbach's alpha 계수는 단답형 PHQ-2의 경우 0.775로 개인 수준에는 미치지 못하지만 집단 수준에서는 비교적 높은 신뢰도를 보여주며, 점수형 PHQ-2의 경우 0.938로 문항 간 동질성이 매우 높게 평가되었다. 방식에 따라 Cronbach's alpha 계수의 차이가 나는 이유는 두 가지 방식에서의 평가 방법과 점수

범위가 다르기 때문인 것으로 보인다. 기존의 청소년을 대상으로 단답형 PHQ-2를 사용한 연구에서 Cronbach's alpha 계수가 0.663으로 산출된 바 있어 여러 집단에서 비교적 높은 신뢰도를 보여줌을 알 수 있다.²⁴⁾ 또한 문항과 총점 사이의 상관관계를 알아보았는데 두 가지 방식에서 각각 0.902~0.971로 유의미한 상관관계를 보여 구성 항목 간의 동질성을 확인할 수 있었다. 대상군 중 14명에서 시행한 검사-재검사 신뢰도는 각각 0.69와 0.77로 이를 통해 두 가지 유형의 PHQ-2 모두 시간의 경과에 대해서 비교적 안정적인 검사임을 알 수 있다. 재검사의 경우 대상군 중 통원치료 중단 등으로 인한 변수 때문에 많은 수를 대상으로 재검사를 실시하지는 못하였으며 향후 많은 통제집단에서 체계적인 연구를 통한 신뢰도의 검증이 더 필요할 것으로 보인다.

Patient Health Questionnaire-2의 타당도를 측정하기 위하여 두 가지 방식의 PHQ-2와 다른 우울증 평가 도구인 PHQ-9, BDI, HRSD들과의 공존타당도를 구해보았다. 단답형 PHQ-2의 경우 PHQ-9, BDI, HRSD와의 상관관계는 각각 0.346, 0.439, 0.375로 유의한 양의 상관관계가 있었다. 점수형 PHQ-2의 경우에도 PHQ-9, BDI, HDRS와의 상관관계는 각각 0.671, 0.526, 0.416으로 유의한 양의 상관관계가 있었다. 두 가지 방식 중에서 점수형 PHQ-2가 단답형 PHQ-2보다 각각의 검사 도구들과 더 높은 상관성을 보였다. 또한 주요우울증군과 정상 대조군 간에 PHQ-2의 평균 점수를 비교해 보니 학력과 종교, 직업 등의 인구학적 특성을 보정한 평균 점수가 주요우울증군이 정상 대조군보다 유의한 수준으로 높게 나타났다.

Patient Health Questionnaire-2 점수가 우울증의 심각도 수준을 잘 반영하는지를 알아보기 위해 주요우울증군에서 시행한 BDI와 HRSD 상의 우울증의 심각도에 따라 경도, 중등도, 심함의 3가지 군으로 나누어서 각 군에 해당되는 대상들의 PHQ-2 점수 평균을 구해보았다. 단답형 PHQ-2와 점수형 PHQ-2 모두에서 대상군의 우울증 심각도가 심해짐에 따라 대체로 PHQ-2 총점 평균점수가 높아졌다. 이는 PHQ-2가 다른 검사 도구들과 비교하여 간단하고 쉬운 방법이며서도 우울증의 심각도 정도를 대략적으로 파악할 수 있다는 것을 보여준다.

점수형 PHQ-2의 최적 절단점을 구하기 위해 PHQ-2 총점을 바탕으로 민감도와 특이도를 조사하였다. 단답형 PHQ-2의 경우 문항 수 자체가 적고 총합의 점수 범위가 좁아 제외하였다. 점수형 PHQ-2의 경우 최적 절단점은 민감도 91.9%, 특이도 100%인 3점으로 산정하였다. 기존의 연구에서는 파킨슨병을 가진 환자들을 대상으로 점수형 PHQ-2를 시행하여 민감도 및 특이도를 고려해 최적 절단점을 2점

으로 산정한 바 있다(민감도 92.9%, 특이도 69.5%).²²⁾ 국내에서는 2000년 Myung 등²⁴⁾이 PHQ-2를 이용하여 청소년의 우울성향 선별검사도구로서 표준화 연구를 시행하였고 한 가지 질문 이상에 긍정적으로 답하는 경우를 검사 양성으로 보았을 때 민감도 100%, 특이도 54.6%로 유용성을 증명한 바 있다. 절단점이 차이나는 이유는 기존 연구의 경우 청소년을 대상으로 하였고 점수형이 아닌 단답형 PHQ-2로 결과를 산출하였으며 연구한 대상 수의 차이 때문으로 사료된다. 향후 PHQ-2를 이용한 우울증의 선별에 있어 대규모를 대상으로 하여 최적 절단점을 연구할 필요성이 있다.

본 연구의 제한점은 다음과 같다. 첫째, 주요우울증군과 정상 대조군의 수가 적었다. 둘째, 연구 설계상 정상 대조군에 대해서는 PHQ-2와 PHQ-9만 검사를 시행하였고 다른 진단적인 검사나 면담을 시행하지 않았다. 셋째, 정상 대조군의 경우 대부분 비슷한 환경에서 근무하는 자들로 다양한 대상을 상대로 연구하지 못하였다. 넷째, 검사-재검사를 실시한 대상자의 수가 총 대상자의 수에 비하여 적었기 때문에 대표성에 문제가 있을 수 있다. 또한, PHQ-2의 문항 번안 과정에서 역번역이 이루어지지 않았으므로, 원문의 의도를 정확하게 표현하지 못할 수도 있다.

결론적으로, 이러한 제한점들은 있으나 본 연구를 통해 한국판 PHQ-2가 우울증을 선별하는 검사도구로서 신뢰도가 높고 타당하다는 것을 알 수 있었다. 따라서 한국판 PHQ-2의 경우 작성 시간이 짧게 소요되어 우울증을 선별하는 데에도 유용한 것으로 나타나서 비용과 효과 면에서 큰 이점이 있다는 것을 보여준다. PHQ-2 평가 방식에 있어서는 점수형 PHQ-2가 단답형 PHQ-2보다는 신뢰도와 타당도 면에서 더 좋은 것으로 나타났다.

결론

이 연구에는 대표적인 우울증 선별도구인 PHQ-2를 표준화하고 현재 사용되고 있는 두 가지 방식의 PHQ-2(단답형과 점수형)를 표준화하고 비교하고자 하였다. 일개 대학병원 정신건강의학과 외래를 방문한 사람 및 입원 환자를 대상으로 MINI-plus를 시행하여 주요우울증으로 진단된 자 74명에게 PHQ를 비롯한 우울증 선별 도구들을 시행하였고 정상 대조군으로 동 대학병원 근무자 58명을 대상으로 검사를 시행하였다. PHQ-2의 내적일치도와 검사-재검사 신뢰도는 양호하였다. 두 가지 유형의 PHQ-2는 항목-총점 간의 상관관계에서 통계적으로 유의하였다. 주요우울증군과 정상 대조군 PHQ-2의 점수 평균을 비교하였을 때 주요우울증군이 유의하게 높았다. 우울증 검사도구들과 총점 간의 상관관계

는 단답형 PHQ-2와 점수형 PHQ-2 방식 모두에서 유의한 상관관계를 보였으며 두 가지 방식 중에서 점수형 PHQ-2가 단답형 PHQ-2보다 각각의 평가 도구들과 더 높은 상관성을 보였다. 우울증 심각도의 분류에 따른 PHQ-2의 평균 점수 변화를 구하였을 때 두 가지 유형의 PHQ-2 모두 우울증의 심각도가 증가할수록 점수가 높아지는 경향을 보였다. 민감도와 특이도를 고려하여 산정한 최적 절단점은 점수형 PHQ-2에서 3점(민감도 : 91.9%, 특이도 : 100%)이었다.

이러한 결과는 두 가지 방식의 PHQ-2 모두가 신뢰도와 타당도가 높은 검사이며 이 중 점수형 PHQ-2가 더 신뢰도가 높고 타당하다는 것을 의미한다. 따라서 PHQ-2는 실제 임상에서 간편하게 사용할 수 있고 우울증을 선별하는 유용한 검사 도구임을 시사한다.

중심 단어 : PHQ-2 · 신뢰도 · 타당도 · 표준화.

Conflicts of Interest

The authors have no financial conflicts of interest.

REFERENCES

- Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003; 289:3095-3105.
- Andrade L, Caraveo-Anduaga JJ, Berglund P, Bijl RV, De Graaf R, Vollebergh W, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res* 2003;12:3-21.
- Ahn J. Depression, suicide, and Korean society. *J Korean Med Assoc* 2012;55:320-321.
- Greenberg PE, Kessler RC, Birnbaum HG, Leong SA, Lowe SW, Berglund PA, et al. The economic burden of depression in the United States: how did it change between 1990 and 2000? *J Clin Psychiatry* 2003;64:1465-1475.
- Organization for Economic Cooperation and Development: Health at a glance [serial online] 2011. OECD Indicators Paris 2011 [cited 2012 Mar 5];4(2). Available from: http://dx.doi.org/10.1787/health_glance-2011-en.
- Simon GE, VonKorff M. Recognition, management, and outcomes of depression in primary care. *Arch Fam Med* 1995;4:99-105.
- Gerber PD, Barrett J, Barrett J, Manheimer E, Whiting R, Smith R. Recognition of depression by internists in primary care: a comparison of internist and "gold standard" psychiatric assessments. *J Gen Intern Med* 1989;4:7-13.
- Susman JL, Crabtree BF, Essink G. Depression in rural family practice. Easy to recognize, difficult to diagnose. *Arch Fam Med* 1995;4: 427-431.
- Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen Hosp Psychiatry* 2006;28:71-77.
- Song SW. The Evaluation of Depression in Primary Care. *J Korean Acad Fam Med* 1999;20:1392-1399.
- 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.
- Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003;41: 1284-1292.
- Saylor CF, Finch AJ Jr, Spirito A, Bennett B. The children's depression inventory: a systematic evaluation of psychometric properties. *J Consult Clin Psychol* 1984;52:955-967.
- Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385-401.
- Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behav Res Ther* 1997;35:785-791.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1982-1983;17:37-49.
- Sharp LK, Lipsky MS. Screening for depression across the lifespan: a review of measures for use in primary care settings. *Am Fam Physician* 2002;66:1001-1008.
- Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA* 1999;282:1737-1744.
- Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med* 1997;12:439-445.
- Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res* 2005;58: 163-171.
- Williams JW Jr, Mulrow CD, Kroenke K, Dhanda R, Badgett RG, Omori D, et al. Case-finding for depression in primary care: a randomized trial. *Am J Med* 1999;106:36-43.
- Chagas MH, Crippa JA, Loureiro SR, Hallak JE, Meneses-Gaya Cd, Machado-de-Sousa JP, et al. Validity of the PHQ-2 for the screening of major depression in Parkinson's disease: two questions and one important answer. *Aging Ment Health* 2011;15:838-843.
- Monahan PO, Shacham E, Reece M, Kroenke K, Ong'or WO, Omollo O, et al. Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. *J Gen Intern Med* 2009;24:189-197.
- Myung SK, Jeong B, Lee WJ, Koh HJ, Suh SY, Yoo T, et al. Standardization of the two-question case-finding instrument as a screening instrument for the adolescent's depression. *J Korean Acad Fam Med* 2000;21:100-106.
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59 Suppl 20:22-33;quiz 34-57.
- Yoo SW, Kim YS, Noh JS, Oh KS, Kim CY, Namgung K, et al. Validity of Korean Version of the Mini-International Neuropsychiatric Interview. *Anxiety and Mood* 2006;2:50-55.
- Chung YC, Rhee MK, Lee YH, Park SH, Sohn CH, Hong SK, et al. A standardization study of Beck Depression Inventory 1-Korean version (K-BDI): reliability and factor analysis. *Korean J Psychopathol* 1995; 4:77-95.
- Rhee MK, Lee YH, Jung HY, Choi JH, Kim SH, Kim YK, et al. A standardization study of Beck Depression Inventory 2-Korean version (K-BDI): validity. *Korean J Psychopathol* 1995;4:96-104.
- Lee YH, Song JY. A study of the reliability and the validity of the BDI, SDS, and MMPI-D scales. *Korean J Clin Psychol* 1991;10:98-113.
- Kim CY. The Evaluation Tool for the Depressive Disorder. 1st ed. Seoul: Hana Medical Publisher;2001. p.72-98.