

한국판 Patient Health Questionnaire-9 (PHQ-9)의 신뢰성, 타당성 및 임상적 유용성

박 경 연

신라대학교 간호학과

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

Park, Kyung-Yeon

Department of Nursing, Silla University, Busan, Korea

Purpose: This study aimed to verify the reliability, validity and clinical usefulness of the Korean version of the 9-item depression module of the Patient Health Questionnaire-9 (PHQ-9) among adult Korean participants. **Methods:** The PHQ-9 questionnaire was administered to 257 participants older than 18 years, chosen using stratified random sampling. Data from the 2014 community health survey conducted by the Korea Centers for Disease Control and Prevention were examined and analyzed by SPSS 23.0 with an independent t-test, factor analysis and receiver operating characteristic curve analysis. **Results:** The PHQ-9 showed significant positive internal consistency ($r=.80$) and the discriminant validity with two group comparison by depression and non-depression group was significantly different ($t=-7.19, p<.001$). Two factors were extracted through a principal component analysis and the total variance explained was 51.67%. **Conclusion:** The Korean version of the PHQ-9 is an appropriate screening tool for detection of depression among adult Koreans, and would be valuable in a primary health care setting.

Key Words: Depression, Mental health, Reliability and validity

서 론

1. 연구의 필요성

우울장애는 환자들의 기능 장애를 초래하고, 일상적인 활동을 유지하기 어렵게 하며[1] 조기 치료 시 완치될 수 있지만, 치료되지 않은 상태로 지속될 경우 자살 행동으로 연결될 수 있다

[2]. 한국의 심각한 사회문제로 대두된 자살률의 감소를 위해 정부차원에서도 자살예방대책추진위원회 설치[2] 등 다양한 노력을 하여왔지만, 한국의 자살률은 OECD 국가 중 여전히 최고 수준의 자살률을 보이며[3] 특별한 개선을 보이지 않고 있다. 반면, 하루 항우울제 소비량은 28개국 중 두 번째로 낮아 우울증 환자 가운데 치료를 받는 사람의 비중이 낮음을 추정할 수 있다. 높은 자살률에 비해 자살의 주된 선행요인으로 알려진 우

주요어: 우울, 정신건강, 신뢰성 및 타당성

Corresponding author: Park, Kyung-Yeon

Department of Nursing, Silla University, 140 Baegyang-daero (Blvd) 700 beon-gil, Sasang-gu, Busan 46958, Korea.
Tel: +82-51-999-5461, Fax: +82-51-999-6237, E-mail: kypark@silla.ac.kr

- 이 논문은 질병관리본부의 국민건강영양조사 제6기 2차년도(2014) 원시자료의 이차분석연구임.

- This manuscript is secondary analysis study written based on some of the data from the Sixth Korea National Health and Nutrition Examination Survey(KNHANES VI-2), 2014, Korea Centers for Disease Control and Prevention.

투고일: 2016년 12월 17일 / **심사완료일:** 2017년 3월 14일 / **게재확정일:** 2017년 7월 25일

울증의 치료를 위한, 항우울제 소비량이 낮은 것은 한국인들의 정신과 치료에 대한 부정적 인식[4]이 관련된 것으로 보고된다. 이는 2011년 전국 만 18~74세 성인을 대상으로 한 정신질환실태조사 결과 정신질환 경험자 중 정신과 전문의, 기타 정신건강전문가를 통한 상담과 치료를 받은 비율은 15.3%에 불과하였으며, 85% 정도가 정신의료서비스를 이용한 적이 없었다[5]는 것과 흐름을 같이하는 것이다. 우울증은 조기 진단을 통해 치료가 가능하지만 신경정신의학과의 방문을 통한 정신과적 평가에 대한 저항감이 있어 조기진단에 어려움이 있으므로 선별검사 활용을 고려할 수 있다. 특히, 선진국과 달리 정신과 방문 및 치료에 대한 한국인들의 부정적 인식(Kim, 2015)을 고려할 때, 우울증의 조기 발견을 위해서는 정신과 방문자에 한정된 환자발견에 국한하지 않는 다양한 의료기관에서 신체사정과 함께 정신건강 사정을 정규적으로 하는 것이 효율적일 것이다.

한국판 우울 평가 척도는 성별[6], 연령[7,8], 산모[9] 등 특정 위험 집단을 대상으로 한 도구가 있으며, 그 외 널리 임상 활용도가 높은 것은 해밀턴 우울증 평가척도[1]와 Beck 우울 척도[10], Center for Epidemiologic Studies-Depression Scale (CES-D) [11]이 있다. 해밀턴 우울증 평가척도는 주요우울장애로 진단된 환자의 질병 심각도를 평가하기 위해 개발된 도구로 치료효과 평가를 위한 관찰자 평가척도이다. 또한 우울증 본래의 증상 외에도 건강염려증, 성적인 증상, 병식, 위장관계 신체증상, 체중감소 등 다양한 차원의 증상이 포함되어 내적일관성 신뢰도가 낮다는 문제가 있었다[12]. 자기보고식 검사로 널리 알려진 21문항의 한국판 Beck 우울 척도 II는 증상의 심각도와 인지적 증상이 강조되며 개별적이고 구체적 문항으로 평가에 장시간이 소요되거나 미묘한 변별과 선택을 요하는 질문형식으로 인해 적절히 반응하는데 어려움이 있다[8]. 조기 진단이 중요한 우울장애는 스크리닝 단계에서 진단적 도구로서의 유용성이 있어야 하는데[13] CES-D는 지역사회 역학연구를 위해 개발되었고, 지난 1주간의 삽화에 대해 응답하여 주요우울장애 진단에 사용하는 DSM-IV의 평가시기와 차이가 있으며 자살을 포함하고 있지 않아 진단적 도구로서의 아쉬움이 있다.

Patient Health Questionnaire (PHQ)-9은 Spitzer et al.[14]이 일차건강관리센터(primary health care center)에서 정신건강을 평가하기 위해 개발한 것으로 Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)의 우울삽화의 진단기준과 일치하도록 구성된 강점이 있으며 도구의 민감도와 특이도가 모두 우수하고 9문항으로 간결하여 측정에 시간소요가 적어 일상적 screening의 실무 적용 가능성이 높다는 강점이 있다[13].

앞서 살펴 본 바와 같이 우울은 방지될 경우 자살로 이어질 수 있는 심각한 건강문제이지만 초기 단계에서는 질병이나 치료를 받아야 한다는 인식이 부족한 현실이다. 따라서 우울 증상으로 신경정신과를 방문하는 자만을 대상으로 우울 검진을 하기보다는, 일차건강센터나 의료기관 방문자 모두를 대상으로 우울선별검사를 시행할 필요가 있다[14]. 신경정신과 방문자 뿐 만이 아닌 의료기관 방문자 모두에게 일상적으로 우울을 사정하기 위해서는, 그 평가도구는 문항 수가 적어 측정에 소요되는 시간이 적게 걸리고 개인의 특성에 영향을 받지 않는 우울 선별을 위한 일반적 문항으로 구성되는 것이 유용할 것이다. 이러한 특성에 맞게 개발된 PHQ-9[14]은 전 세계적으로 다수 언어로 변환되어 사용하고 있으며 한국어판도 보고되었으나 국내연구대상자는 노인만을 대상[13]으로 했거나 정신과 외래 방문객만을 대상으로 한 것[1]이다. 정신과 방문자가 아닌, 일차건강센터나 의료센터 방문자들을 대상으로 임상적 우울 선별검사로 활용할 수 있도록 개발된[14] PHQ-9 원 도구의 대상자 특성을 그대로 반영하고 한국에서의 일차 우울 선별 검사로서의 타당성을 확인하기 위해 편중없이 표집된 확대된 표본으로 PHQ-9 한국어판을 재검정해 볼 필요가 있다. 또한 우울증 선별 도구로서의 타당성은 임상적 유용성 검증을 통해 실무활용에 대한 지지를 받을 수 있으나, PHQ-9에 대한 선행연구에서는 일 개 중증도시 노인 대상 연구[13]에서만 이를 보고하고 있다.

이에 본 연구는 Spitzer 등[14]이 개발하고 Han 등[13]이 변환한 한국어판 PHQ-9을 확률표집에 의한 전국 자료인 국민건강영양조사 원시자료를 이용하여 우울선별도구로서의 신뢰도와 타당도를 검증하고자한다. 본 연구결과는 한국어판 PHQ-9 우울 측정도구의 임상적 유용성을 확인하여 대상자들이 정신과 진료에 대한 부담없이 지역사회 일차의료기관 단계에서 선별하여 우울을 조기에 발견하고 치료 관리하는데 기여할 수 있을 것이다.

2. 연구목적

본 연구는 국민건강영양조사 제6기 2차년도(2014) 원시자료의 일부를 이용하여 한국 성인을 대상으로, 우울증 선별 검사 도구로서의 한국어판 PHQ-9의 신뢰도와 타당도를 검증하는 것이며 구체적인 연구목적은 다음과 같다.

- 한국어판 PHQ-9의 신뢰도를 검증한다.
- 한국어판 PHQ-9의 타당도를 검증한다.
- 한국어판 PHQ-9의 우울증 선별 검사 도구로서의 임상적 유용성을 확인한다.

연구 방법

1. 연구설계

본 연구는 Spitzer et al.[14]이 일차 건강관리센터에서 우울을 평가하기 위해 개발한 Patient Health Questionnaire-9 (PHQ-9)을, Han et al.[13]이 임상에서 우울증 선별도구로 활용하기 위해 번안한 한국어판 PHQ-9을 18세 이상의 성인을 대상으로 신뢰도와 타당도 및 임상적 유용성을 검정한 것으로, 질병관리본부의 국민건강영양조사 제6기 2차년도(2014) 원시자료의 이차분석연구이다.

2. 연구대상

본 연구 분석의 모집단은 한국의 만 19세 이상의 성인이다. 질병관리본부의 국민건강영양조사 제6기 2차년도(2014) 자료의 조사기간은 2014년 1월부터 12월까지이며, 표본은 인구주택총조사를 추출틀로 하였으며, 모집단의 층화지표는 행정구역(시도, 동·읍면)과 주택유형(일반주택, 아파트)을 기준으로 추출틀을 층화하고, 주거면적 비율, 가구주 학력 비율 등을 내재적 층화 기준으로 한 것이다. 본 분석은 2014년 조사일 현재 우울증 유병 여부에 대해 ‘없음’ 혹은 ‘있음’으로 응답한 성인 265명 중, 한국판 PHQ-9에서 결측값을 보인 8명을 제외한 총 257명의 자료를 이용한 것이며, 우울증이 있다는 응답은 의사의 진단에 근거한 것이다. 즉, 우울증이 있는 자와 그렇지 않은 자가 모두 포함된 것으로 우울증의 진단이 있는 자들은 우울군으로 그렇지 않은 자들은 정상군으로 적용하였다. 분석방법별 최소 표본 수를 검토한 결과, 양측검정, 유의수준 .05, 검정력 .95로 설정한 G*Power 3.1 프로그램에서, 중간효과크기 .30의 상관분석에서 최소 134명이 필요하였고, 본 연구에서의 allocation ratio인 .71을 대입하여 t-test에서의 중간효과크기 .50을 적용하였을 때 218명이 필요하여 본 분석의 표본 수 257명은 적절한 것으로 나타났다.

3. 연구도구

본 연구는 대상자의 일반적 특성 7항목(성별, 연령, 결혼, 교육수준, 직업, 주관적 건강상태), 한국어판 PHQ-9문항 및 현재 우울유병여부 1문항을 분석 자료로 사용하였다. 일반적 특성은 원 자료의 연령을 29세 이하, 30~39세, 40~49세, 50~59세, 60~69세 및 70세 이상으로 구분하였다. 결혼은 ‘OOO님은 결혼

한 적이 있습니까?’에 대한 질문에 ‘예’, ‘아니오’로, 교육수준은 초졸 이하, 중졸, 고졸, 대졸 이상으로 응답한 제6기 2차년도 국민건강영양조사 자료를 그대로 분석하였다. 직업은 주부, 학생으로 응답한 자는 직업이 없는 것으로, 그 외 항목인 관리자, 전문가 및 관련 종사자, 사무종사자, 서비스 및 판매종사자, 농림어업 관련 종사자, 기능원, 장치 기계조작 및 조립종사자, 단순노무종사자는 직업이 있는 것으로 분류하였다. 주관적 건강상태는 ‘평소에 OOO님의 건강은 어떻다고 생각하십니까?’라는 질문에 ‘매우 좋음’, ‘좋음’, ‘보통’, ‘나쁨’, ‘매우 나쁨’의 국민건강영양조사에 대한 응답자료를 본 연구에서 그대로 분석하였다. 한국어판 PHQ-9[13]은 국민건강영양조사 자료수집기관에서 자료수집 전 각각의 도구 사용에 대한 승인을 받고 조사가 진행된 것이다.

한국판 PHQ-9는 Spitzer et al.[14]이 일차 건강관리센터에서 우울을 평가하기 위해 개발한 Patient Health Questionnaire-9 (PHQ-9)을, Han et al.[13]이 임상에서 우울증 선별도구로 활용하기 위해 번안한 PHQ-9 한국어판을 사용한 것이다. 번안 당시 노인 대상 Han et al.[13]의 연구에서 내재일관성 신뢰도($r=.88$)와 Geriatric Depression Scale ($r=.74$)을 이용한 수렴타당도의 지지를 받았다. 9개 문항에 대해 “지난 2주 동안, 아래 나열되는 증상들에 얼마나 자주 시달렸습니까?”라는 질문에 각 문항별 ‘전혀 아니다’ 0점, ‘여러 날 동안’ 1점, ‘일주일 이상’ 2점, ‘거의 매일’ 3점으로 평정하고, 총 점수는 0~27점의 범위를 가진다. Han et al.[13]은 PHQ-9를 이용하여 선별할 수 있는 우울장애의 임상적 분할점을 5점으로 제시하였다.

4. 자료수집 및 윤리적 고려

본 연구에 활용된 자료는 질병관리본부의 ‘2014년도 국민건강영양조사’ 건강설문조사자료의 일부이다. 질병관리본부의 2014년 국민건강영양조사를 위한 자료수집은 이동진집센터에서 이루어졌으며 훈련된 조사원에 의해 면접방법 혹은 자기기입식으로 조사된 것이다. 국민건강영양조사에 사용된 도구와 조사과정은 질병관리본부 연구윤리심의위원회 승인(2013-12EXP-03-5C)을 받아 수행한 것이며 모든 조사대상자에게서 조사에 대한 사전동의서를 받고 진행하였다. 본 연구는 「개인정보보호법」, 및 「통계법」에 따라 조사대상 가구 및 개인을 추정할 수 있는 정보를 제거한 상태의 자료를 받아서 이차분석한 것이다.

5. 자료분석

수집된 자료를 SPSS/WIN 23.0 프로그램을 이용하여 통계

적 유의수준 .05에서 양측검정 하였다.

- 대상자의 일반적 특성은 서술적 통계를 이용하였다.
- 도구의 신뢰도는 문항 제거시 신뢰도 계수와 Cronbach's α 값으로 내적일관성을 분석하였다.
- PHQ-9의 우울증 선별도구로서의 타당도는 집단비교법 (known group technique)과 요인분석을 활용한 구성타당도를 확인하였다. 집단비교법은 측정하고자 하는 개념에 대해 차이가 있을 것으로 기대되는 집단들에게 측정값을 서로 비교하여 구성타당도를 확인하는 것이다[15]. 본 연구에서는 임상적 진단에 근거한 우울군과 정상군을 대상으로, 독립 t-test를 이용하여 우울을 측정한 PHQ-9 값의 평균차이가 있는지를 확인하였다. 우울군은 의사의 진단에 따른 우울증을 현재 앓고 있는지, 여부를 묻는 1개의 문항에 대해 '그렇다'고 반응한 대상자는 우울군으로 '그렇지 않다'고 응답한 대상자는 비우울군으로 분류하였다.
- PHQ-9 도구의 우울개념 측정 여부 및 도구의 변수들이 동일한 요인으로 묶이는지를 확인하기 위하여 탐색적 요인분석을 하였으며, 요인분석 시 적절한 요인 적재값을 추출하기 위해 고유치(eigen value) 1 이상인 값을 선택하였으며 요인 판정을 위하여 직교(Varimax) 회전법을 사용하였다.
- PHQ-9 도구의 임상적 유용성을 확인하기 위해 수용자 작업특성 곡선(Receiver Operator Characteristics, ROC curve) 및 ROC curve에 관한 곡선 아래 면적(Area Under the Curve, AUC)을 구하였다.

연구 결과

1. 대상자의 일반적 특성

대상자들은 여자가 79.8%로 남자보다 높은 분포를 보였고, 연령은 50~59세가 25.3%, 기혼자가 88.3%, 직업이 없는 자가 56.0%, 월 수입이 하위 1/4인 자가 30.0%, 주관적 건강상태 보통이라는 응답을 보인 자가 44.4%로 그렇지 않은 집단보다 다수를 차지하였다. 대상자의 12.8%가 기초생활수급자이었다 (Table 1).

2. 도구의 신뢰도

도구의 각 문항 및 문항 제거 시 신뢰도 계수와 도구의 내적 일관성 분석을 통한 신뢰도 검정결과는 Table 2와 같다. PHQ-

Table 1. Sociodemographic Characteristics among Participants (N=257)

Characteristics	Categories	n (%)
Gender	Male	52 (20.2)
	Female	205 (79.8)
Age (year)	≤ 29	18 (7.0)
	30~39	28 (10.9)
	40~49	33 (12.8)
	50~59	65 (25.3)
	60~69	59 (23.0)
	≥ 70	54 (21.0)
Marital status	Married	227 (88.3)
	Unmarried	30 (11.7)
Education	≤ Elementary school	97 (37.7)
	Middle school	34 (13.2)
	High school	86 (33.5)
	≥ College	40 (15.6)
Occupation	Yes	113 (44.0)
	No	144 (56.0)
Recipient of basic living	Yes	33 (12.8)
	No	224 (87.2)
Monthly income	The lower first quartile	77 (30.0)
	The lower second quartile	67 (26.1)
	The lower third quartile	58 (22.6)
	The lower fourth quartile	53 (20.6)
	Missing	2 (0.8)
Perceived health status	Very poor	22 (8.6)
	Poor	81 (31.5)
	Fair	114 (44.4)
	Good	37 (14.4)
	Very good	3 (1.2)

9 도구 전체의 내적일관성 Cronbach's α 는 .80이었고, 각 문항-전체 상관계수 r은 .40 (지난 2주 동안 식욕저하 혹은 과식에 얼마나 자주 시달렸습니까?)부터 .66 (지난 2주 동안 가라앉은 느낌, 우울감 혹은 절망감에 얼마나 자주 시달렸습니까?)의 범위에 있었으며 해당 항목이 삭제된 경우의 Cronbach's α 는 .75~.79의 범위에 있어 도구의 신뢰도는 적절한 것으로 나타났다.

3. 도구의 타당도

PHQ-9 도구의 타당도를 검정하기 위하여 우울군과 정상군 간의 PHQ-9 점수 비교(Table 3), 및 탐색적 요인분석 (Table 2)을 이용하였다. 우울군과 정상군 간의 집단비교법을 이용한 구성타당도를 확인한 결과, PHQ-9 우울점수가 우울군이 8.41 ± 5.62 점으로 정상군의 3.93 ± 4.37 점보다 더 높았다.

이들 PHQ-9 값은 우울군과 정상군 모두 Han 등(2008)이 제시한 우울증에 대한 임상적 분할점인 5점 기준에도 부합하는 점수이었다. 두 집단 간 PHQ-9 우울점수는 통계적으로 의미있는 차이가 있어 PHQ-9은 성인 우울증 선별 검사 도구로서의 타당도를 지지받았다($t=-7.19, p<.001$).

PHQ-9의 요인분석을 실시한 결과(Table 2), 표본 적절성의 Kaiser-Meyer-Olkin 측도 값은 .80으로 기준치 .50보다 높았고, Bartlett의 구형성 검정 결과 전반적으로 변수들 간의 상관관계는 유의하여($p<.001$) 본 분석 자료는 요인분석을 하기에 적합하였다. Anti-image Correlation 행렬상의 모든 Anti-image Matrices (MSA)는 .73~.89의 범위 내에 있어 기준인 .50보다 낮은 값이 없었으므로 제거할 항목이 없는 것으로 나타났다. 주성분 분석방식으로 요인을 추출하고 직교(Varimax) 회전법을 적용하여, 고유치(eigen value) 1 이상

인 값을 선택한 결과 2개의 요인이 추출되었다. 회전 이후 요인 1의 eigen value는 2.59, % 분산은 28.74였으며 요인 2의 eigen value는 2.06, % 분산은 22.93이었고 이들 두 요인들이 전체 분산의 51.67%를 설명하였다. 요인 1은 문항 9, 문항 2, 문항 1, 문항 8, 문항 6, 문항 7로 사회심리적 요인이 포함되었고, 요인 2는 문항 5, 문항 4, 문항 3으로 신체적, 생리적 증상 항목이 포함되었다.

4. 임상적 유용성

PHQ-9가 임상적으로 유용한 선별도구인지를 확인하기 위한 ROC curve 및 ROC curve에 관한 AUC를 구한 결과는 Figure 1과 같다. Figure 1과 같이 ROC curve가 그림의 대각선의 위쪽에 위치해 있어 PHQ-9는 ROC curve에서 임상적으로 유의

Table 2. Item Analysis and Factor Analysis of the PHQ-9

(N=257)

Over the last 2 weeks, how often have you been bothered by any of the following problems? Items*	M±SD	A	B	C	Factor loadings		Com.	EV (%)
					1	2		
9. Thoughts that you would be better off dead or of hurting yourself in some way	0.41±0.80	.46	.78	.76	.75		.57	28.74
2. Feeling down, depressed, or hopeless	0.89±1.07	.66	.75		.72		.64	
1. Little interest or pleasure in doing things	0.75±1.07	.45	.78		.66		.46	
8. Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	0.31±0.74	.44	.78		.66		.45	
6. Feeling bad about yourself or that you are a failure or have let yourself or your family down	0.60±0.96	.51	.77		.51		.41	
7. Trouble concentrating on things, such as reading the newspaper or watching television	0.43±0.88	.47	.78		.49		.36	
5. Poor appetite or overeating	0.73±1.05	.40	.79	.66		.79	.63	22.93
4. Feeling tired or having little energy	1.26±1.16	.60	.76			.74	.65	
3. Trouble falling or staying asleep or sleeping too much	1.16±1.21	.44	.79			.68	.50	
Total	0.73±0.62			.80				51.67

KMO=.80; Bartlett $\chi^2=598.03$; $df=36$; $p<.001$

*The items maintain literal fidelity to the original tool quoted from Brief Patient Health Questionnaire by Spitzer et al. (1999). A=corrected item-total correlation; B=Cronbach's α if item deleted; C=Cronbach's α ; Com=communality; EV=explained variance.

Table 3. Difference of PHQ-9 between Two Groups

Variable	Non-depression group		Depression group		t	p
	n (%)	M±SD	n (%)	M±SD		
PHQ-9	107 (41.6)	3.93±4.37	150 (58.4)	8.41±5.62	-7.185	<.001

PHQ-9=Patient Health Questionnaire-9.

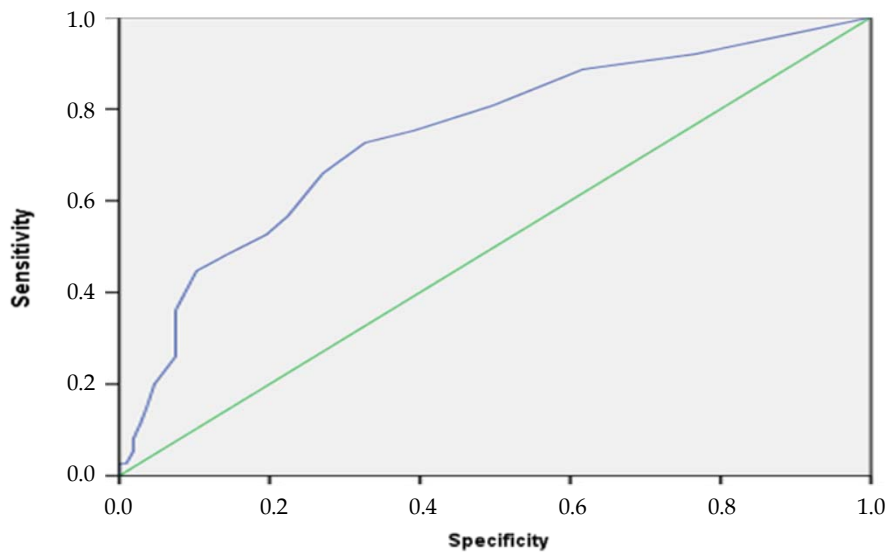


Figure 1. Receiver operating characteristics curve of PHQ-9.

미한 결과를 나타내었고 곡선아래 면적 Area Under the Curve (AUC)=.74, SE=.03 ($p < .001$, 95% Confidence Interval=.68~.80)으로 나타나 유의한 임상적 유용성이 있는 것으로 나타났다. 그려진 curve가 좌측 위쪽 모서리에 가까울수록 진단에 대한 실제 양성률은 높고 위양성률은 낮음을 의미하여 AUC=1은 완벽한 검사로 분류할 수 있는데 본 결과는 중등도의 정확성 ($0.7 < \text{AUC} \leq 0.9$)을 보임을 알 수 있었다[16].

논 의

본 연구는 일차 건강관리센터에서 임상적 우울의 선별도구로 활용하기 위해 번안한 한국어판 PHQ-9[13]을 19세 이상의 전국 규모의 확률표본자료인 국민건강영양조사 제6기 2차년도 자료를 이용하여 신뢰도, 타당도 및 임상적 유용성을 확인한 것으로 검정결과를 중심으로 논의하고자 한다.

본 분석 대상자 전체의 PHQ-9 평균평점은 0.73점으로 Park 등[1]의 1.7점 보다 낮았다. 이는 본 연구대상자들은 임상적인 우울증 진단을 받은 자와 받지 않은 정상군 모두가 포함된 지역 사회 재가 성인을 대상으로 수집한 자료인 반면, Park 등[1]은 정신과 외래 방문자가 연구대상이었으므로 나타난 결과로 해석된다. PHQ-9의 내적일관성 신뢰도 Cronbach's α 는 .80이었는데, 이는 국내 일개 시의 노인들을 대상으로 한 연구에서의 .88[13], 대학병원 정신과 외래 방문자를 대상으로 한 Park 등의 .84[1] 및 국외 연구인 Kroenke 등[17]의 일차의료센터 방문자들의 .89와 산부인과 방문 대상자의 .86과 함께 PHQ-9의

내적일관성 신뢰도가 양호함을 지지하는 결과이다. 각 문항-전체 상관계수 r 은 .40부터 .66의 범위 내에 있었다. 이는 선행연구[1,13,17]에 비해서는 낮은 상관성이다. 측정하고자 하는 개념의 공통성 수준인 .30 이상[18]을 충족하여 도구를 구성하고 있는 항목의 동질성은 유지됨을 확인할 수 있었다. 도구 총점과 개별 항목 간의 상관성을 본 결과 식욕의 저하 혹은 증가 항목이 .40으로 가장 낮은 관련성을 보였고 가라지는 느낌, 우울감 혹은 절망감 항목이 .66으로 가장 높은 관련성을 보였는데 이러한 최소, 최대 관련성 항목은 대학병원 정신과 외래 방문객을 대상으로 한 연구[1]와 같은 것이다. Park 등[1]의 연구에서 상관계수도 각각 식욕의 저하 혹은 증가 항목 .41, 가라지는 느낌, 우울감 혹은 절망감 항목이 .70로 본 분석과 유사하였다. PHQ-9 도구 전체의 내적일관성 Cronbach's α 는 .80이었던 데 반해, 각 항목이 삭제된 경우의 Cronbach's α 는 .75~.79의 범위에 있어서 구성된 9개 항목 모두가 적합함을 알 수 있었다. 이상의 신뢰도 평가를 통해 본 도구는 19세 이상의 성인을 대상으로 반복 측정 시 일관된 평가결과를 도출할 수 있음을 알 수 있었다.

우울군과 정상군 간의 독립 t-test를 이용한 변별타당도를 확인한 결과, PHQ-9 우울점수는 우울군이 정상군보다 더 높아 우울증 선별 검사 도구로서의 변별력이 있음을 보여주었다. 이 같은 PHQ-9의 타당도 지지는 8곳의 일차의료기관 및 7곳의 산부인과 클리닉 방문자 6,000명을 대상으로 하여 임상과 연구에서의 우울 측정도구로서의 타당성을 입증한 Kroenke 등[17]과 일관된 결과이며, 국내의 노인 대상연구[13]와 정신

과 외래 방문객 대상[1] 연구결과와도 같은 것이다. 이러한 결과는 본 도구와 동일한 한글판이 노인만을 대상으로 검증한 것이나 임상 및 지역사회에서의 다양한 곳에서 우울 선별평가 도구로 활용 가능함을 추론했던 Han 등의 추론[13]을 검증하여 보여주었다고 본다.

PHQ-9의 요인분석 결과 Anti-image Correlation 행렬상의 모든 Anti-image Matrices (MSA)는 .73~.89의 범위 내에 있어 전체 9문항 중 제거할 항목이 없었으며 주성분 분석방식에 의한 요인 추출 결과 신체적 요인과 사회심리적 요인 모두가 포함되어 우울의 증상을 대체적으로 포괄하여 선행연구[11]의 하부 영역과 공통된 것이며 이들 요인은 전체 분산의 51.67%를 설명하여 받아들일만한 수준의 설명력을 보였다. 본 결과는 신체화 요인, 우울증 요인, 불면요인 및 기타요인 등의 전체 4개 영역 17개 항목으로 구성되고 우울증에 대한 치료효과를 평가하는 관찰자 평가척도의 표준으로 여겨지고 있는 해밀턴 우울 평가 척도의 설명력인 50.4%[19] 보다 조금 더 높은 것으로, 9개 문항으로 구성된 본 도구의 임상적용의 용이성과 함께 우울 평가에 대한 타당도를 지지받을 수 있음을 보여 준 결과라고 본다. Spitzer 등[14]과 일관된 것으로 한국어판 PHQ-9은 한국 성인의 우울선별도구로 활용함에 있어서 문화적 특이성을 반영하기 위한 별도의 삭제문항은 없어도 됨을 보여주는 것이다. ROC curve 및 ROC curve에 관한 AUC를 구한 결과 PHQ-9은 진단적 유용성이 있는 것으로 나타나 Han 등[13]과 같은 결과를 보여 임상상황에서도 우울선별도구로서의 타당성을 지지 받음을 재확인할 수 있었다. 이러한 결과는 한국의 노인들을 대상으로 한 Han 등[13]의 연구에서 제시한 PHQ-9 도구의 우울증 절단점이 연령에 제한을 받지 않고 18세 이상의 성인에서는 동일하게 적용할 수 있음을 보여 준 것이다. 다만 본 도구가 9개 문항으로 구성되었다는 것은 우울증 조기 발견을 위한 선별도구 활용 확대에 기여할 수 있다는 장점이 있지만 우울이라는 개념의 모든 요소가 문항으로 포함되지 못하였을 수 있다는 문제도 배제할 수 없으므로 실무 적용과 결과해석에서 주의 필요로 한다고 본다.

본 결과에서 나타난 신뢰성과 타당성은 9문항으로 시간이 많이 소요되지 않는다는 강점과 함께 우울 선별도구로서의 임상적 실무활용 가능성을 보여준 것이다. 즉, 의료기관 방문 시 활력징후 측정을 하여 육체적 건강사정의 기본자료로 활용하는 것과 같이, 우울증상이 없어도 한글판 PHQ-9을 일차의료기관 및 건강관리센터 방문하는 모든 대상자에게 적용하여 정신건강을 사정하는데 활용할 수 있을 것이다.

본 연구는 이차자료분석 연구라는 한계가 있다. 따라서 일

회의 횡단조사 자료만 존재하여 Pearson 상관계수 추정 공식에 의한 도구의 검사-재검사 신뢰도와 평가자 간 신뢰도를 확인하지 못한 한계가 있다. 뿐만 아니라, 평가에 소요되는 시간이나 문항에 대한 이해도나 용어의 가독성 수준 등의 도구 접근성과 활용 용이성은 다소 부족하지만 우울의 속성을 모두 반영할 수 있도록 우울의 구체적인 증상이나 여러 요소가 모두 문항으로 포함된, 또 다른 우울측정도구와의 준거타당도를 확인하지 못하였다는 아쉬움이 있다. 이와 같은 점에서 도구의 타당도와 신뢰도 검증에서의 미흡함이 있다고 보며 이는 이차 자료 분석 연구가 가지는 한계라고 할 것이다. 추후 이를 보완할 수 있는 연구설계를 통한 반복연구를 통해 한국어판 PHQ-9의 신뢰도와 타당도를 재검정할 필요가 있다.

이상과 같은 제한에도 불구하고, 기존의 임상적 우울 선별도구의 타당도 및 신뢰도 검증 연구가 우울 삽화가 있는 대상이거나 특정 지역, 연령, 혹은 소수 의료기관 등에서 표본을 추출한 것인데 반해, 우울장애 환자와 정상인 모두를 포함한 전국 자료를 이용하여 분석하여, 한국 지역사회 일차 의료기관 혹은 건강센터에서의 임상적 우울 선별도구로서의 PHQ-9의 신뢰도, 타당도 및 진단적 유용성을 검증하였다는데 본 연구의 의의가 있다고 할 것이다.

결론 및 제언

전국자료인 국민건강영양조사 자료를 이차분석한 본 연구에서, 한국어판 PHQ-9 도구는 9문항 개별 해당 항목이 삭제된 경우의 Cronbach's α 가 .75~.79의 범위에 있고 전체 문항의 Cronbach's α 가 .80으로 나타나 신뢰도가 지지되었다. 우울군과 정상군 간의 비교를 활용한 집단비교법과 요인분석을 통해 타당도의 적합성이 확인되었고 AUC를 통한 임상적 유용성이 입증되었다. 결론적으로 PHQ-9은 지역사회 일차건강센터에서 성인을 대상으로 한 우울연구 및 임상 우울선별도구로 활용할 수 있는 신뢰롭고 타당한 도구임을 알 수 있었다. 추후 이차자료분석이 아닌 확률표집에 의한 연구설계로 표본의 대표성을 유지하면서 신뢰도와 타당도가 널리 확보된 준거를 이용한 구성타당도 및 수렴타당도 재검정을 제언한다.

REFERENCES

1. Park SJ, Choi HR, Choi JH, Kim KW, Hong JP. Reliability and validity of the Korean version of the Patient Health Questionnaire-9 (PHQ-9). *Anxiety and Mood*. 2010;6(2): 119-24.

2. Kahng SK. Does depression predict suicide?: Gender and age difference in the relationship between depression and suicidal attitudes. *Korean J Soc Welfare Stud.* 2010;41(2):67-100.
3. Ministry of Health & Welfare. OECD health at a glance 2013 [Internet]. Daejeon: Ministry of Health & Welfare; 2016 [cited 2016 November 20]. Available from: http://www.mohw.go.kr/front_new/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&page=1&CONT_SEQ=293922
4. Kim HJ. The article of the hankyoreh newspaper 2015. 11. 18. [Internet]. Seoul: the hankyoreh newspaper; 2016 [cited 2016 November 20]. Available from <http://www.hani.co.kr/arti/society/health/717951.html>
5. Ministry of Health & Welfare. 2011 Mental disease epidemiology survey [Internet]. Daejeon: Ministry of Health & Welfare; 2016 [cited 2016 November 20]. Available from: http://www.mohw.go.kr/front_new/sch/index.jsp?sflag=1&coll=MWNEWS_FILE&query=정책보고서 2011년 정신질환 실태 역학조사 &before_query=정책보고서 2011년 정신질환 실태 역학조사&SortType=1&dateGbn=all&boolType=and&cat=ALL&nowPage=2#preView
6. Woo SB, Yang EJ. The validation of the Korean Masculine Depression Scale. *The Korean J Human Develop.* 2014;21: 115-32.
7. Jo HS, Kwon YM. Validation of the DEQ-A. *The Korean J Develop Psychol.* 2009;22(3):145-167.
8. Jung IK, Kwak DI, Shin DK, Lee MS, Lee HS, Kim JY. A reliability and validity study of Geriatric Depression Scale. *J Korean Neuropsychiatr Assoc.* 1997;36(1):103-112.
9. Kim JI. A validation study on the translated Korean version of the Edinburgh postnatal depression scale. *Korean J Women Health Nurs.* 2006;12(3):204-9.
10. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* 1961; 4(6):561-71. <https://doi.org/10.1001/archpsyc.1961.01710120031004>
11. Radloff LS. The CES-D scale: A self-reported scale for research in the general population. *Appl Psychol Meas.* 1977;1(3):385-401.
12. Park HW, Lee EH, Kim DK, Yu BH, Lee DS, Kim JH. The factor structure of the Korean Hamilton Depression rating Scale (K-HDRS): a confirmatory factor analysis. *J Korean Neuropsychiatr Assoc.* 2009;48(1):21-8.
13. Han CS, Ahnjo SM, Kwak JH, Pae CU, Steffens D, Jo IH, et al. Validation of the patient health questionnaire-9 Korean version in the elderly population: the Ansan geriatric study. *Comprehensive Psychiatry.* 2008;49(2):218-23. <https://doi.org/10.1016/j.comppsy.2007.08.006>
14. Spitzer RL, Kroenke K, Williams JB. Patient Health Questionnaire Primary Care Study Group. Validation 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(18):1737-44. <https://doi.org/10.1001/jama.282.18.1737>
15. Lee HK, Yang YH, Gu MO, Eon Y. Introduction to nursing research. Seoul: Hyunmoonsa; 2009. p. 164-5.
16. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine.* 2000;45(1-2): 23-41. [https://doi.org/10.1016/S0167-5877\(00\)00115-X](https://doi.org/10.1016/S0167-5877(00)00115-X)
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9 validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9): 606-13. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
18. Kang JY, Kang HY, Kim NY, Lee MJ, Kim YJ, Kim JH, et al. Validity and reliability test of the Korean version of the oncology patients' perception of the quality of nursing care scale. *Korean J Adult Nurs.* 2016;28(2):191-201. <https://doi.org/10.7475/kjan.2016.28.2.191>
19. Yi JS, Bae SO, Ahn YM, Park DB, Noh K S, Shin HK, et al. Validity and reliability of the Korean version of the Hamilton depression rating scale (K-HDRS). *J Korean Neuropsychiatr Assoc.* 2005;44(4):456-62.