

우울증 변증도구의 신뢰도, 타당도 평가

이현수, 강위창*, 정인철

대전대학교 한의과대학 신경정신과학교실, 대전대학교 리버럴아츠칼리지 통계학과*

Reliability and Validity Analysis of the Instrument on Pattern Identifications for Depression

Hun-Soo Lee, Wee-Chang Kang*, In-Chul Jung

Department of Oriental Neuropsychiatry, College of Oriental Medicine, Daejeon University, *Department of Statistics, H-Liberal Arts College, Daejeon University

Received: November 30, 2015

Revised: December 20, 2015

Accepted: December 21, 2015

Correspondence to

In Chul Jung

Daejeon Univ. College of Oriental Medicine, Yongun-dong, Dong-gu, Daejeon, Korea.

Tel: +82-42-470-9129

Fax: +82-42-470-9005

E-mail: npjeong@dju.kr

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2011690).

This study was supported by a grant of the Traditional Korean Medicine R&D Project, Ministry of Health & Welfare, Republic of Korea (HI13C0493 & HI15C0006).

Objectives: This study was performed to evaluate the reliability and validity of the instrument on pattern identifications for depression.

Methods: Two assessors carried out an evaluation about the instrument on pattern identifications for depression, targeting 201 participants, who after taking the HAM-D score over 12 or under 7 twice.

Results: Inter-assessor reliability was higher than intra-assessor reliability in a reliability analysis about classification of pattern identification evaluated by the instrument on pattern identifications for depression. Reliability of intra-assessor and inter-assessor showed a moderate to strong agreement when reliability analysis about classification score of the pattern identification had been performed. Reliability analysis to evaluate the validity of the instrument on pattern identifications for depression showed moderate agreement.

Conclusions: The results reveal that reliability analysis of the instrument on pattern identifications for depression showed an over moderate agreement and validity analysis represented a positive correlation.

Key Words: Depressive disorder, Depression, Instrument on pattern identifications, Reliability and validity.

I. 서론

우울증은 우울한 기분, 거의 모든 일상생활 및 취미 활동에 있어서의 흥미나 즐거움의 감소, 때를 부정적으로 생각하게 되는 경향을 갖는 상태로^{1,2)}, 여러 종류의 정신 및 신체의 장애를 나타내는 증후군이라고 할 수 있다³⁾. 우울증의 증상으로는 우울한 기분, 흥미나 즐거움의 감소 외에도 불면 또는 과다수면, 식욕저하 또는 식욕과다, 피로, 긴장, 정신운동 지체, 무가치감, 죄책감, 집중과 결정에서의 어려움, 그리고 죽음에 대한 생각 또는 자살 계획 및 시도 등이 있다^{4,5)}.

우울증은 가장 흔한 정신질환 중 하나로 평생 유병률이 여자의 경우에는 10~25%, 남자의 경우에는 5~12% 정도로 매우 높으며⁶⁾, 국내 유병률 또한 지속적으로 증가하고 있는 추세로 2001년에는 4.6%였던 유병률이 2006년 6.2%, 2011년 7.5%로 10년간 무려 63%나 증가하였다. 또한 우울증 환자의 자살 문제는 심각한 사회적 문제로 대두되고 있는 실정이며, 실제적으로 10~15%에서 자살 시도가 있는 것으로 조사되었다⁷⁾. 2015년 통계청 자료에 따르면 2014년 국내에서 자살로 인한 사망자의 수는 인구 10만 명 당 27.3명으로 2004년의 23.7명에 비해 약 15% 증가하였으며, 2003년부터 10년 이상 경제협력개발기구 OECD (Organization for Economic Cooperation and Development) 회원국 중 자살률 1위를 유지하고 있다⁸⁾.

한의학에서는 우울증이라는 명칭이 존재하지는 않으나 울증(鬱證), 전증(癲證)의 범주에서 견주어 볼 수 있다. 울증은 우울증에서 신경증적인 부분과, 전증은 정신병적인 부분과 유사하며, 우울증의 증상 중 피로감 및 무기력은 허로(虛勞)의 범주에, 수면장애는 불면(不眠) 또는 기면(嗜眠)의 범주에, 식욕감퇴는 불사식(不思食)의 범주에 든다고 볼 수 있다⁹⁾. 울증의 주요 원인으로는 정지불서(情志不舒), 칠정소상(七情所傷), 음식내상(飲食內傷), 외감(外感) 등을 들 수 있는데, 이 중에서 특히 정지소상(情志所傷)이 주된 원인이라고 할 수 있다¹⁰⁾. 전증의 주요 원인으로는 담(痰), 칠정(七情)의 문제, 심혈부족(心血不足), 비위허한(脾胃虛寒) 등을 들 수 있다¹¹⁾.

한의학에서는 진단의 개념으로 변증(辨證)을 사용하는데, 이를 통해 병리의 본질을 보다 명확히 규정하면서 치료의 근거로 사용할 수 있다는 이점이 있다¹²⁾. 변증이란 사진(四

診) 즉, 망(望), 문(聞), 문(問), 절(切)을 통해 수집된 정보들과 신체에서 드러나는 다양한 여러 징후들을 종합적으로 분석하여 진단하는 방법이다¹³⁾. 하지만 변증을 통해서서는 환자의 증상에 대한 객관적인 평가를 내리기 힘들다는 한계점이 있어 변증을 통한 진단 과정의 객관화, 정량화가 필요하다¹⁴⁻¹⁶⁾. 실제 이러한 한계를 극복하기 위한 한의학계에서의 여러 시도 및 연구들이 있었다. 특히 알레르기 비염¹⁷⁾, 기능성 소화불량¹⁸⁾, 폐암¹⁹⁾, 위-식도 역류질환²⁰⁾, 만성 기침²¹⁾ 등 다양한 질환에 대한 변증 도구 개발 시도가 있었으며, 신경정신과 영역에서도 화병^{22,23)}, 경계정증^{24,25)}에 대한 변증 도구 개발 연구가 시행된 바 있다.

지금까지의 한의학계의 우울증 연구를 살펴보면, 우울증에 대한 문헌 연구^{9,26-29)}, 우울증에 관한 임상보고³⁰⁻³⁵⁾ 및 진단 도구에 관한 연구³⁶⁻³⁸⁾, 침 치료 및 한약 치료의 효과에 대한 연구³⁹⁻⁴⁹⁾ 등, 우울증의 개념, 임상양상 및 치료에 대해서는 많은 논의가 이루어졌으나, 진단에 관한 논의는 상대적으로 부족하여 학계에서의 우울증 변증 표준안에 대한 합의는 아직 이루어지지 못한 상태이다.

과거에는 우울증과 관련된 한의학적 진단도구가 부재하여 임상현장에서 우울증 환자를 진단, 변증함에 있어 객관성이 부족하다는 문제가 있었다. 이에 따라 2013년에 우울증의 객관적인 한의학적 진단을 위해 문헌고찰과 전문가합의를 통한 변증도구 개발의 기초연구가 이루어 졌다⁵⁰⁾. 이를 통해 <우울증 변증도구>를 제시하였으나 임상연구를 거치지 않았기 때문에 신뢰도와 타당도가 검증되지 않았다는 한계가 있었다. 이에 본 연구팀은 <우울증 변증도구>의 신뢰도와 타당도를 알아보기 위해 임상연구를 실시하였고, 이를 통해 수집된 자료에 대한 통계적 분석을 시행하여 다소의 지견을 얻었기에 그 결과를 보고하는 바이다.

II. 연구대상 및 방법

1. 연구기관과 대상인원

본 연구는 ○○대학교 ○○한방병원에서 2014년 1월 15일부터 2015년 5월 30일까지 본원 임상연구센터에 방문하여 HAM-D (Hamilton Rating Scale for Depression) 검사를 시행하여 12점 이상인 167명(시험군)과 7점 이하인 34명(대조군), 총 201명을 대상으로 하였다.

2. 연구 방법

1) 임상시험심사위원회 임상시험 승인

본 연구를 시작하기 전 임상시험계획서 등의 서류를 ○○대학교 ○○한방병원 임상시험심사위원회에 제출 후 임상시험 승인을 받았다(승인번호: 둔산한방병원 14-01).

2) 피험자 모집

포스터 및 배너를 통한 병원 내 홍보, 지하철 광고, 생활정보지, 전단지, 지인 소개를 통해 피험자를 모집하였다. ○○대학교 ○○한방병원 임상연구센터에 직접 방문을 하거나 전화를 통해 방문 일정을 정한 후 시험에 참가하도록 하였다.

3) 피험자 동의와 적합성 평가

임상시험에 참여하게 된 지원자에게는 임상시험의 목적, 시행 방법, 시험 중 발생 가능한 위험 또는 불편 사항, 중도포기의 권리, 시험 중 이루어지는 비밀보장, 임상시험 이후 보상 등을 설명하였으며 이 후 자유의사를 통해 피험자 동의서에 서면 동의를 하고 임상시험에 참여할 수 있도록 하였다.

‘대상자 동의 설명서’에 서면 동의를 한 지원자에 관한 인구학적 조사, 과거병력 및 치료력 조사, 선행 및 병용약물 조사, HAM-D 검사 등을 시행한 후, 이를 바탕으로 피험자 선정 및 제외기준에 따라 피험자로 적합한지의 여부를 평가하였다.

4) 선정 제외기준의 판정

(1) 선정 기준

20세 이상 65세 이하의 남녀로 시험군은 우울증상을 호소하면서 HAM-D 점수가 12점 이상인 경우, 대조군은 HAM-D 점수가 7점 이하인 경우로 하였다.

(2) 제외 기준

현재 망상, 환각 등의 정신증적 양상이 있거나 과거력이 있는 경우, 조증 삽화, 경조증 삽화, 또는 흔재성 삽화가 과거에 한번이라도 있었던 경우, 알코올 또는 다른 물질의 남용 또는 의존이 있거나 과거에 있었던 경우, 우울증상 발현 정도에 영향을 줄 것으로 판단되는 물질을 임상시험 참여

전 2주일 이내에 투약을 시작한 경우, 임상시험 참여 시작 시 우울증상 발현 정도에 영향을 줄 것으로 판단되는 의학적 상태에 있는 경우, 간암 또는 간경변증, 만성 신부전, 율혈성 심부전 등으로 치료 중인 경우, 다른 임상시험 참여 중인 경우, 기타 임상시험을 수행하는데 부적합하다고 판단되는 경우이다.

5) 우울증 변증도구⁵⁰⁾

〈우울증 변증도구〉는 구조화된 면담도구로 측정된 항목별 답변을 작성된 엑셀 파일에 입력하면 자동적으로 계산되며 관련된 변증에 증상 점수가 부과되어 최종적으로 심비양허, 간기울결, 음허화왕, 간화상염, 담기울결, 기허담결의 6가지 변증 중 한가지로 자동 분류된다.

6) 평가 지표

환자의 우울증 및 관련 상태를 평가하기 위해 사상체질 분류검사(SCAT)를 시행하고, 단중혈, 양측 승모근 경결점, 이외 피검자가 불편감을 호소하는 기타 1부위의 통증 정도를 Algometer를 사용하여 측정하였다.

7) 임상 연구 실시

대상자가 자의에 의해 임상시험에 참여할 것을 서면으로 동의하면, 첫 방문(선정방문) 시 HAM-D 검사를 시행하여 점수에 따라 시험군과 대조군으로 분류하고, 사상체질 분류 검사, Algometer를 사용한 압통측정을 실시하고 이 후 A, B 2명의 한의사가 〈우울증 변증도구〉를 통한 검사를 시행하였다. 5일(±2일) 후 이루어지는 두 번째 방문(종료방문)에서는 2명의 한의사가 〈우울증 변증도구〉를 통한 검사를 실시하여 첫 방문과 두 번째 방문의 각각의 검사 결과와 변증간의 상관관계를 평가하였다.

8) 통계 및 검정

우울증 변증도구의 신뢰도 분석은 변증 분류의 신뢰도 평가의 경우 검사자 내 또는 검사자 간 카파분석(Kappa analysis)을 사용하여 평가하였다. 변증도구 점수 사이의 검사-재검사법을 이용한 신뢰도 분석의 경우에는 피어슨 상관 계수분석(Pearson correlation coefficient analysis), 독립 표본 t-검정(Independent t-test)을 이용하여 평가하였다.

우울증 변증도구의 타당도 분석에는 피어슨상관계수 분

석을 통해 우울증 변증도구와 HAM-D와의 상관관계를 분석하였고, 독립표본 t-검정을 통해 시험군과 대조군에서 우울증 변증도구의 변증별 평균 점수와 HAM-D 점수와의 관계를 분석하였다. 통계분석은 SPSS (IBM® SPSS® Statistics Ver. 21, USA)를 사용하였으며, 통계적 유의성은 유의수준 5%로 판단하였다.

III. 결과

1. 우울증 변증도구의 신뢰도 분석

1) 변증 분류의 신뢰도 평가

(1) 변증 분류의 평가자 내, 평가자 간 신뢰도 분석
 ‘변증 분류’는 <우울증 변증도구>를 시행하였을 경우 최종적으로 결정된 변증을 의미하는 것으로 평가자 A, B는 대상자들이 총 2회 방문 시, 각 방문마다 변증 분류를 하였다. 각 평가자 별로 카파 분석(Kappa analysis)을 했을 때 평가자 A는 0.28, 평가자 B는 0.26으로 나타났다.

또한 대상자의 각 방문 시마다 평가자 A, B가 판정한 변증 사이의 관계를 평가하였다. 각 방문 별로 카파 분석을

했을 때 방문 1의 경우 0.37, 방문 2의 경우는 0.34로 나타났다(Table 1).

(2) 변증도구의 각 변증별 점수 차이의 크기에 따른 신뢰도 분석

변증 분류에서 1위 변증의 점수와 2위 변증의 점수의 차이가 1점 이상 또는 3점 이상인 각각의 경우에만 1위 변증으로 최종 판별하게 보정하여 평가자 내, 평가자 간의 신뢰도를 평가하였다.

평가자 내 분석 시 평가자 A의 경우에는 점수 차이가 1점 이상인 경우 0.28, 3점 이상인 경우 0.46으로 나타났으며, 평가자 B의 경우에는 점수 차이가 1점 이상인 경우 0.31, 3점 이상인 경우 0.35로 나타났다. 평가자 간 분석 시 방문 1의 경우에는 점수 차이가 1점 이상인 경우 0.38, 3점 이상인 경우 0.57로 나타났으며, 방문 2의 경우에는 점수 차이가 1점 이상인 경우 0.40, 3점 이상인 경우 0.49로 나타났다(Table 2).

Table 1. Reliability Analysis of Classification of the Pattern Identification through the Kappa Analysis

	N	Kappa coefficient
Intra-Assessor		
Assessor A	163	0.28
Assessor B	161	0.26
Inter-Assessor		
Visit 1	166	0.37
Visit 2	162	0.34

Table 2. Reliability Analysis of Classification of the Pattern Identification After Score Gap Correction through the Kappa Analysis

	Score gap	N	Kappa coefficient
Intra-Assessor			
Assessor A	1	134	0.28
	3	83	0.46
Assessor B	1	137	0.31
	3	101	0.35
Inter-Assessor			
Visit 1	1	144	0.38
	3	97	0.57
Visit 2	1	133	0.40
	3	89	0.49

Table 3. Reliability Analysis of Classification Score of the Pattern Identification through the Pearson Correlation Coefficient Analysis

Pattern identifications		Pearson coefficient
Intra-Assessor		
Assessor A	Stagnation of liver Qi	0.77
	Dual deficiency of the heart and spleen	0.68
	Stagnation of phlegm and Qi	0.70
	Qi deficiency mingled with sputum	0.58
	Qi depression transforming into fire	0.72
	Yin deficiency with effulgent fire	0.64
Assessor B	Stagnation of liver Qi	0.79
	Dual deficiency of the heart and spleen	0.74
	Stagnation of phlegm and Qi	0.72
	Qi deficiency mingled with sputum	0.57
	Qi depression transforming into fire	0.72
	Yin deficiency with effulgent fire	0.59
Inter-Assessor		
Visit 1	Stagnation of liver Qi	0.74
	Dual deficiency of the heart and spleen	0.72
	Stagnation of phlegm and Qi	0.72
	Qi deficiency mingled with sputum	0.58
	Qi depression transforming into fire	0.78
	Yin deficiency with effulgent fire	0.64
Visit 2	Stagnation of liver Qi	0.81
	Dual deficiency of the heart and spleen	0.82
	Stagnation of phlegm and Qi	0.81
	Qi deficiency mingled with sputum	0.60
	Qi depression transforming into fire	0.80
	Yin deficiency with effulgent fire	0.77

2) 변증 점수의 신뢰도 평가

(1) 피어슨 상관계수분석을 통한 변증 점수의 신뢰도 분석

‘변증 점수’는 <우울증 변증도구>를 시행하였을 경우 가장 높은 점수를 얻어 최종적으로 결정된 변증의 점수를 의미한다. A, B 각 평가자 별로 측정된 1, 2차 변증 점수 사이의 관계를 피어슨 상관계수분석(Pearson correlation coefficient analysis)을 통해 평가하였을 경우 평가자 A의 경우 0.58~0.77, 평가자 B의 경우 0.57~0.69로 나타났다. 1, 2차 각 방문 시 측정된 변증점수를 평가자 간 피어슨 상관계수분석을 통해 평가하였을 때에는 방문 1의 경우 0.58~0.78, 방문 2의 경우 0.60~0.82로 나타났다(Table 3).

(2) 독립표본 t-검정을 통한 변증 점수의 신뢰도 분석

평가자 A, B를 통틀어 방문 1과 방문 2에서 측정된 변증 점수 차이의 평균값이 간기울결(肝氣鬱結) 6.6, 심비양허(心脾兩虛) 8.7, 담기울결(痰氣鬱結) 7.9, 기허담결(氣虛痰結) 3.3, 기울화화(氣鬱化火) 6.0, 음허화왕(陰虛火旺) 6.0으로 나타났으며, 모든 변증에서 방문별 점수의 차이가 통계적으로 유의성이 있게 나타났다.

방문 1과 방문 2를 통틀어 평가자 A와 평가자 B의 변증

점수 차이의 평균값은 간기울결 2.6, 심비양허 1.8, 담기울결 4.5, 기허담결 -0.6, 기울화화 2.6, 음허화왕 2.6으로 나타났으며, 기허담결을 제외한 모든 변증에서 평가자간 점수의 차이가 통계적으로 유의성이 있게 나타났다(Table 4).

3) HAM-D 점수의 신뢰도 평가

방문 1과 방문 2의 HAM-D 점수 사이의 상관관계는 급내상관계수를 사용하여 평가한 경우 0.67로 나타났고, 피어슨 상관계수분석을 통해 평가한 경우에는 0.69로 나타났다(Table 5).

2. 우울증 변증도구의 타당도 분석

1) 우울증 변증도구와 HAM-D와의 상관분석

우울증 변증도구의 타당도 분석은 HAM-D와의 상관분석을 통해 평가하였다. 대상자의 1, 2차 각 방문 시 평가자 A, B가 측정한 변증별 점수의 평균을 HAM-D의 점수와 피어슨 상관계수분석을 통해 평가한 결과, 방문 1의 경우 0.39~0.57, 방문 2의 경우 0.43~0.61로 나타났고 두 가지의 경우 모두 심비양허, 간기울결, 기울화화, 음허화왕의 순으로 상관관계가 높게 나타났다(Table 6).

Table 4. Reliability Analysis of Classification Score of the Pattern Identification through the Independent t-Test

	Pattern identifications	Mean	p-value
Intra-Assessor	Stagnation of liver Qi	6.6	<0.001
	Dual deficiency of the heart and spleen	8.7	<0.001
	Stagnation of phlegm and Qi	7.9	<0.001
	Qi deficiency mingled with sputum	3.3	<0.001
	Qi depression transforming into fire	6.0	<0.001
	Yin deficiency with effulgent fire	6.0	<0.001
Inter-Assessor	Stagnation of liver Qi	2.6	0.005
	Dual deficiency of the heart and spleen	1.8	0.036
	Stagnation of phlegm and Qi	4.5	<0.001
	Qi deficiency mingled with sputum	-0.6	0.467
	Qi depression transforming into fire	2.6	<0.001
	Yin deficiency with effulgent fire	2.6	<0.001

Table 5. Reliability Analysis of HAM-D Score through the Intra-Class Correlation Coefficient Analysis and Pearson Correlation Coefficient Analysis

	Pearson correlation coefficient analysis			Intra-class correlation coefficient analysis	
	Mean	S.D.*	Pearson coefficient	ICC	95% C.I.†
Visit 1	19.31	4.24	0.69	0.67	0.57~0.74
Visit 2	18.22	4.87			

*Standard deviation, †Confidence interval.

2) 시험군과 대조군 사이의 우울증 변증도구와 HAM-D 점수의 비교

각 방문별로 시험군과 대조군에서의 평가자 A, B가 측정 한 변증별 점수와 HAM-D 점수의 평균을 독립표본 t-검정을 통해 구간 비교하였다. 1, 2차 방문 모두의 경우 시험군과 대조군 사이의 차이는 통계적으로 유의하였으며, HAM-D의 점수 역시 구간 차이가 통계적으로 유의하게 나타났다 (Table 7).

3. 변증도구와 실험 대상자의 특성과의 상관성

각 방문별 평가자 A, B가 측정한 변증별 점수의 평균과 algometer로 측정된 값의 상관성을 피어슨 상관계수분석으로

로 평가한 결과, 방문 1, 방문 2 모두 상관성이 거의 없게 나타났다. 동일한 방법으로 변증별 점수의 평균과 사상체질 분류검사를 통해 분류된 체질과의 상관성을 ANOVA를 통해 평가한 결과 둘 사이의 상관성은 통계적으로 유의성이 없게 나타났다.

IV. 고찰

변증은 한의학의 독특한 진단 방법으로 병리본질을 보다 명확히 규정할 수 있다는 특징이 있으며 치료에 있어서의 근거로 활용이 된다¹²⁾. 그동안 사용되어 왔던 변증은 객관적인 진단 기준이 미흡하여, 환자의 진단에 있어서 진단의 임상적인 지식과 경험에 의해 좌우되는 경향이 있기에 보조적인 진단 기준이 필요한 실정이다⁵¹⁾. 최근 한의학계에서는 여러 질환의 변증도구가 개발 되었거나 혹은 개발 중¹⁷⁻²¹⁾에 있으며, 신경정신과 질환 중에서도 화병^{22,23)}, 경계정충^{24,25)} 등에 대한 변증도구가 개발되어 타당도와 신뢰도 검증이 진행되었다.

우울증은 지속적인 우울한 기분과 일상생활에 있어서의 흥미나 즐거움의 감소를 주소로 하는 질환이다^{1,2)}. 우울증은 신경정신과 질환 중 자살률이 가장 높은 축에 속하며, 특히 OECD 국가 중 1위를 차지하고 있는 우리나라의 자살률은 우울증 유병률의 가파른 증가와 관련이 있다고 할 수 있다⁸⁾. 이러한 점에서 우울증은 한의 임상에서 매우 중요한 신경정신과 질환 중 하나이며, 다양한 임상연구가 진행되어 왔다³⁰⁻³⁵⁾. 하지만 객관적으로 우울증에 대해 변증진단을 할

Table 6. Validity Analysis between Classification Score of the Pattern Identification and HAM-D Score through the Pearson Correlation Coefficient Analysis

	Pattern identifications	Pearson coefficient
Visit 1	Stagnation of liver Qi	0.51
	Dual deficiency of the heart and spleen	0.57
	Stagnation of phlegm and Qi	0.45
	Qi deficiency mingled with sputum	0.39
	Qi depression transforming into fire	0.50
	Yin deficiency with effulgent fire	0.49
Visit 2	Stagnation of liver Qi	0.59
	Dual deficiency of the heart and spleen	0.61
	Stagnation of phlegm and Qi	0.54
	Qi deficiency mingled with sputum	0.43
	Qi depression transforming into fire	0.57
	Yin deficiency with effulgent fire	0.56

Table 7. Validity Analysis between Gap of Classification Mean Score of the Pattern Identification and HAM-D Score through the Independent t-Test

	Pattern identifications	Gap of mean score between experimental group and control group	Gap of HAM-D score between experimental group and control group	p-value
Visit 1	Stagnation of liver Qi	49.57		
	Dual deficiency of the heart and spleen	57.91		
	Stagnation of phlegm and Qi	48.78		
	Qi deficiency mingled with sputum	26.14	18.16	<0.001
	Qi depression transforming into fire	30.78		
	Yin deficiency with effulgent fire	32.49		
Visit 2	Stagnation of liver Qi	42.95		
	Dual deficiency of the heart and spleen	50.08		
	Stagnation of phlegm and Qi	40.47		
	Qi deficiency mingled with sputum	23.13	17.49	<0.001
	Qi depression transforming into fire	25.31		
	Yin deficiency with effulgent fire	27.51		

수 있는 도구가 부재한 상황이었으나 최근 문헌을 근거로 하여 전문가 합의 방식으로 <우울증 변증도구 가안>이 제작되었다⁵⁰⁾. 이에 우울증의 한의학적 진단에 있어서의 과학성 확보와 객관적인 임상적 활용을 위해 우울증 변증도구의 타당도, 신뢰도를 평가하는 연구를 수행하였다.

<우울증 변증도구 가안>은 일부 문항들의 표현이 모호하다는 점과 평가자의 주관에 개입할 여지가 크고, 평가자의 숙달의 정도에 따라 차이를 보일 수 있는 맥진(脈診), 설진(舌診)이 포함되어 있다는 점이 검사 결과에 부정적인 영향을 미쳤을 가능성이 존재하여 이를 보완한 <우울증 변증도구>를 제작하고 보다 엄격한 표준작업지침을 작성한 후 본 임상연구를 시행하였다.

본 연구는 201명의 피험자를 대상으로 수행되었으며, 변증 분류의 신뢰도 분석을 했을 때에는 평가자 내의 신뢰도보다 평가자 간의 신뢰도가 더 높게 나타났는데, 이를 통해 <우울증 변증도구>를 시행함에 있어서 평가자 사이의 변이보다 시간적인 변이가 결과에 더 큰 영향을 미친다는 것을 알 수 있었다. 변증도구의 각 변증별 점수 차이의 크기에 따른 신뢰도 분석을 시행했을 경우에는 평가자 내, 평가자 간 분석 모두 점수 차이가 3점 일 때 가장 높은 상관성을 보였으며 모두 moderate agreement 수준 이상으로 나타났다.

피어슨 상관계수분석을 통해 변증 점수의 신뢰도를 분석했을 경우에는 평가자 내, 평가자 간 모두 moderate agreement에서 strong agreement 사이의 신뢰도를 보여, 두 가지 모두의 경우 신뢰도가 확보되었다고 볼 수 있다. 하지만 평가자 간 분석시의 피어슨 상관계수가 평가자 내의 경우보다 높게 나타나 변증 점수 역시 변증 분류 분석에서와 마찬가지로 평가자 사이의 변이보다 시간적인 변이가 결과에 큰 영향을 미쳤음을 알 수 있었다.

독립표본 t-검정을 통한 변증도구 점수의 신뢰도 분석을 한 결과에서는 평가자 내, 평가자 간 모두 통계적으로 유의성이 있게 나타나 방문 1과 방문 2에서의 결과와 평가자 A와 평가자 B의 결과 사이에 차이가 있음이 나타났다. 하지만 평가자 간 분석시의 평균값의 차이가 평가자 내의 평균값의 차이보다 적게 나타나 독립표본 t-검정을 통한 평가에서도 시간적인 변이가 영향을 미쳤음을 알 수 있었다.

HAM-D 점수의 신뢰도 평가에서는 급내상관계수(ICC: Intra-class correlation coefficient analysis)를 통해 분석

을 하였을 경우와 피어슨 상관계수분석을 통해 평가를 하였을 경우 모두 strong agreement를 보였다⁵²⁾.

우울증 변증도구의 타당도를 분석하기 위하여 각 변증 점수와 HAM-D 점수의 상관성을 피어슨 상관분석을 통해 분석하였을 경우에는 방문 1, 2의 경우 모두 moderate agreement 정도의 상관성을 보였다. 분석 결과 심비양허, 간기울결, 기울양허 순서로 상관성이 높게 나타나 HAM-D 문항과의 연관성이 보다 높은 변증들이 있음을 알 수 있었다. 특히 심비양허의 경우 HAM-D 점수의 신뢰도 평가의 결과와 상관계수가 비슷하게 높게 나타나 변별타당도가 있다고 볼 수 있다. 시험군과 대조군에서의 평가자 A, B가 측정한 변증 점수와 HAM-D 점수의 평균을 t-검정을 통해 군간 비교한 결과는 1, 2차 방문 모두의 경우 두 가지 군의 차이가 통계적으로 유의하게 나타났다. 본 결과로 볼 때 <우울증 변증도구>가 우울증 여부를 판별하고 정도를 평가하는 검사도구로는 의미미하다고 할 수 있으나, 변증 분류에 대한 타당도가 확보되었다고 단정하기는 부족하다.

변증도구와 실험 대상자의 특성과의 상관성을 알아보기 위해 변증별 점수의 평균과 algometer 측정값 및 사상체질 분류검사와의 상관성을 평가한 결과는 상관성이 매우 약하게 나타나, 변증을 분류하는데 있어서의 기초적인 평가 자료로 사용하기에는 다소 한계가 있을 것으로 보인다.

본 연구의 결과를 토대로 우울증 변증도구를 임상에서 활용하는 데 있어서 두 가지 방법을 제시할 수 있다. 첫째, 본 도구를 사용하여 한가지의 변증유형을 결정하는 것으로, 이는 1위 변증점수가 2위 변증점수 보다 적어도 3점 이상 높은 경우로 한정한다. 둘째, 각 변증의 점수를 모두 참고하는 것으로, 하나의 변증으로 결정하는 방식은 결과의 편차가 크기 때문에 각 변증의 점수로 해당 환자가 변증의 경향이 각각 어느 정도인지를 확인하고 전체적인 치료방향의 설정 및 약물치료와 침 치료 등에서의 수증가감에 참고하는 방식이다. 본 연구팀은 언급한 두 가지 방법 중 후자를 추천하는 바이다.

본 연구는 다음과 같은 몇 가지 제한점이 있다.

첫째, 첫 방문 시 직접적인 치료가 시행되지 않았음에도 평가자 A, B 모두의 경우 1차 방문시보다 2차 방문시의 <우울증 변증도구>의 값이 전체적으로 감소하였다. 우울증이라는 정신과적인 질환의 특성 상 1차 방문 시 검사자를 통해 시행된 일련의 설문 과정이 피험자에게 위약 효과를 일으켰

을 가능성이 있다고 생각된다. 향후 이러한 위약효과를 고려한 임상연구 디자인에 대한 검토가 필요할 것으로 보인다.

둘째, 본 연구에서 평가자 내의 신뢰도보다 평가자 간의 신뢰도가 높게 나타나 시간적 변이가 평가자 변이보다 크게 작용했음을 알 수 있다. 1차와 2차 방문 사이의 기간이 약 5일 정도인 것을 감안할 때 보다 정확한 연구 결과를 위해서는 그보다 1차 방문과 2차 방문 사이의 기간을 줄이는 방안을 고려해 볼 수 있을 것이다.

셋째, 6가지의 변증유형과 각각의 변증유형이 몇 개 이상의 증상으로 구성되어 있어 경계정충 변증도구의 연구^{24,25)}와 유사하게 검사결과와 편차가 있을 가능성이 많아서 향후 타 질환의 변증도구 개발 연구에서는 변증유형의 개수, 변증별 증상의 개수에 대한 보다 심도 있는 고민이 필요할 것으로 사료된다.

넷째, <우울증 변증도구>가 변증을 분류하는 도구로서의 타당도가 확보되었다고 단정 지을 수 없다. 타당도를 보다 정확히 평가할 수 있는 추가적인 연구가 필요할 것으로 보인다.

본 연구는 위와 같은 한계에도 불구하고 문헌검토와 전문가 합의를 통해 제작된 <우울증 변증도구 가안>의 신뢰도와 타당도를 평가하여, 우울증의 한의학적 진단을 객관적, 과학적으로 수행할 수 있는 기초를 마련했다는 점에 의의가 있다고 할 수 있다.

V. 결론

<우울증 변증도구>의 신뢰도와 타당도를 평가하고자, HAM-D 검사 상 12점 이상인 시험군 167명과 7점 이하인 대조군 34명, 총 201명에 대해 <우울증 변증도구>를 통한 설문을 2차례씩 2회 실시한 후 결과를 분석하여 다음의 결론을 얻었다.

1. 우울증 변증도구의 변증 분류의 신뢰도 분석을 했을 때, 평가자 내의 신뢰도 보다 평가자 간의 신뢰도가 더 높게 나타났다.

2. 변증도구의 점수 값 크기에 따른 신뢰도 분석을 시행했을 경우, 점수 차이를 3점으로 했을 때 가장 높은 상관성을 보였으며 모두 moderate agreement 수준 이상이었다.

3. 피어슨 상관계수 분석을 통한 변증 점수의 신뢰도 분석을 한 결과에서는, 평가자 내, 평가자 간 모두 moderate

agreement에서 strong agreement 사이의 신뢰도를 보였다.

4. 독립표본 t-검정을 통한 변증 점수의 신뢰도 분석을 한 결과에서는 평가자 내, 평가자 간의 차이가 모두 통계적으로 유의성이 있게 나타났다.

5. 우울증 변증도구의 타당도를 분석하기 위하여 각 변증 점수와 HAM-D 점수의 상관성을 분석한 결과 두 번의 방문 모두 moderate agreement 수준으로 나타났다.

REFERENCES

1. Katona C, Robertson M. Psychiatry at a Glance-Third Edition. Blackwell Publishing Ltd. 2008.
2. Kim HL, Shin HS. Mental Health Problems. Hakjisa. 2001.
3. Hwan WW, Kim JH. Dong Ui Jeong Sin Ui Hag. Seoul: Hyundai Uihag Seojeogsa. 1992;471-87,576-82,609-11.
4. American Psychiatric Association. DSM-IV Diagnosis and Statistical Manual of Psychiatric Disorders. 4th ed. Washington DC. 1994;713-84.
5. American Psychiatric Association. Diagnosis and Statistical Manual of Mental Disorders 4th ed. Washington DC: American Psychiatric Association. 1994;317-29.
6. The Textbook Compilation Committee of Neuropsychiatry of Oriental Schools in Nation. The Neuropsychiatry of Oriental Medicine. Seoul: Jipmoon-Dang. 2007; 256-65,295,472,474-8.
7. 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;289:3095-105.
8. Statistics Korea. The cause of death statistics 2014. Available from: URL: http://kostat.go.kr/portal/korea/kor_ko/5/2/index.board?bmode=read&aSeq=348646
9. Kang HW, Jang HH, Kang IS, Moon HC, Hwang YJ, Lyu YS. A Study on The Oriental-medical Understanding of Depression. Journal of oriental neuropsychiatry. 2011; 12(2):1-15.
10. Lee SH, Kim BY, Kee JI, Kwon SJ, Chung DK. A Case of Mixed Anxiety and Depressive Disorder. Journal of Oriental Neuropsychiatry. 2003;14(2);127-34.
11. The Textbook Compilation Committee of Neuropsychiatry of Oriental Schools in Nation. The Neuropsychiatry of Oriental Medicine. Seoul: Jipmoon-Dang. 2007;299-300.
12. The Textbook Compilation Committee of Oriental Pathology. Oriental Pathology. Seoul: Hanimunhwasa. 2008; 172,286,288,579.
13. Park K. Jindanhaggwon-yeo. Daesung Publishing Company. 2000;491-2.
14. Kim BK. Chinese Medicine Diagnostics. Seoul: Seong-

- bosa. 1986;25,42.
15. Kim KK, Kang CW. Weighted Principal Component Analysis of the Oriental Gynaecology Experiments. *The Korean Journal of Applied Statistics*. 1999;12(1):17-28.
 16. Song NK, Kim JK, Shin SH, Kim YH, Hwang GD, Ahn KS, et al. A Study for the Objective Diagnosis by Statistical Analysis to the Bian Zheng Questionnaire. *Korean Journal of Oriental Medicine*. 2005;11(1):127-38.
 17. Kim NK, Lee DH, Choi IH, Ko SG. An expert survey for developing pattern diagnosis instrument of persistent allergic rhinitis. *The Journal of Korean Medical Ophthalmology & Otolaryngology & Dermatology*. 2013;26(2):1-9.
 18. Kim JB, Kim JH, Son CG, Kang WC, Cho JH. Development of Instrument of Pattern Identification for Functional Dyspepsia. *Korean J. Oriental Physiology & Pathology*. 2010;24(6):1094-8.
 19. Kim JM, Jeon HJ, Park SJ, Kim SY, Cho JK, Jung IC, Yoo HS. Preliminary Study to Develop the Instrument on Pattern Identification for Lung Cancer. 2014;28(6):585-92.
 20. Han GJ, Leem JT, Lee NL, Kim JS, Park JW, Lee JH. Development of a Standard Tool for Pattern Identification of Gastroesophageal Reflux Disease (GERD). *J. Int. Korean Med*. 2015;36(2):122-52.
 21. Kim KI, Shin SW, Lee NL, Lee BJ, Jung HJ, Jung SK, Lee JH. Preliminary Study for Development of Pattern Identification Tool of Chronic Cough. *J. Int. Korean Med.* 2015;36(1):22-39.
 22. Yim HJ, Kim SH, Lee SR, Jung IC. Study to Develop the Instrument of Pattern Identification for Hwa-byung. *Korean J. Oriental Physiology & Pathology*. 2008;22(5):1071-7.
 23. Kim KK, Seo BN, Kang WC, Jung IC. Guidelines for the Reliability and Validity of the Instrument on Pattern Identifications for Hwa-byung. *J of Oriental Neuropsychiatry*. 2013;24(4):331-42.
 24. Park DM, Lee SR, Kang WC, Jung IC. Preliminary Study to Develop the Instrument of Pattern Identification for Jing Ji and Zheng Chong. *J. of Oriental Neuropsychiatry*. 2010; 21(2):1-15.
 25. Choi WC, Lee HS, Seo BN, Kang WC, Jung IC. A Study to Assess the Reliability and Improvement of the Instrument of Pattern Identification for Jing Ji and Zheng Chong. *J of Oriental Neuropsychiatry*. 2014;25(4):371-82.
 26. Hwang JY, Jung IC, Lee SR. The Review for Studies of Aromatherapy, Relaxation Therapy and Exercise Therapy on Depression. *Journal of Oriental Neuropsychiatry*. 2011;22(4):1-10.
 27. Lee JE, Kwon YJ, Cho SH. A Review of Clinical Studies with Herbal Medicine for Depression - Based on Randomized Controlled Clinical Trial -. *Journal of Oriental Neuropsychiatry*. 2011;22(4):31-9.
 28. Park SW, Song CH. The Study of Literature on Moxa Treatment of Melancholia. *Korean Journal of Acupuncture*. 2002;19(1):131-44.
 29. Kim YJ, Park DS, Lee YH. A Study on Depression with Acupuncture & Moxibustion Treatment in Chinese Medical Literature. *The Journal of Korean Acupuncture & Moxibustion Society*. 2005;22(1):223-34.
 30. Park HS, Seo WH, Moon IR, Kim JK, Bae KY, Heo YK. A Case of Patient who have Depressive Episode Mixed with Organic Depressive Disorder. *Journal of Oriental Neuropsychiatry*. 2003;14(2):199-206.
 31. Shin HK, Kim JW, Kong HW, Kock KN, Kim HJ, Park CJ, et al. A Clinical Study on 2 Cases of Depression Patients with Autogenes Training. *Journal of Oriental Neuropsychiatry*. 2006;17(3):165-74.
 32. Kang HS, Kim JY, Lee JH, Sung WY. Report of 2 Depression Patients Accompanying Somatic-symptom Treated with Meditation of the Retaining Needle. *Journal of Oriental Neuropsychiatry*. 2008;19(2):241-50.
 33. Seok SH, Kim JH, Kim GW, Koo BS. A Case of Insomnia Incurable by Neuropsychiatric Medication Alone, but Made Possible with Sarn-Hwang-Sa-Sim-Tang. *Journal of Oriental Neuropsychiatry*. 2006;17(3):117-29.
 34. Lee SJ, Lee SH, Kim BY, Hwang SM, Chung DK. A Case about the Patient who has Deteriorated by Symptoms being R/O because of Side Effects by Western Medicine after having Diagnosis of Melancholia. *Journal of Oriental Neuropsychiatry*. 2002;13(2):213-23.
 35. Kang HS, Lee JH, Kim JY, Sung WY. A Clinical Report of Two Patients with Depression and Menopausal Symptoms Improved by Korean Traditional Treatment and Psychotherapy. *Journal of Oriental Neuropsychiatry*. 2009;20(2):177-186.
 36. Koo BS, Lee SJ, Han CH, Kim HJ, Park SH. The Basic Study for Building the Depression Prescription Guideline of Gamiguibi-Tang - The Evaluation of Reliability and Validity of the Depression Pattern-Identification Questionnaire -. *Journal of Oriental Neuropsychiatry*. 2009;20(4):1-13.
 37. Kim TH, Lee YK, Lyu YS. The Clinical Study of Digital Infrared Thermographic Imaging on Depressed Patients. *Journal of Oriental Neuropsychiatry*. 2000;11(2):87-102.
 38. Kim TH, Lee YK, Ryu YS. The clinical study of Digital Infrared Thermographic Imaging on Depressed patients. *Journal of Oriental Neuropsychiatry*. 2000;11(1):59-73.
 39. Lee JA, Kim JW, Whang WW, Kwak SY, Kim MJ, Park EH. The Effects of Quibitang on an Animal Model of Depression Induced by Chronic Mild Stress. *Journal of Oriental Neuropsychiatry*. 2001;12(1):123-35.
 40. Lee JH, Kim JW, Whang WW, Kim HT. The Effects of Koamsimsinwhan (without Cinnabarite) on Depression in Chronic Mild Stress (CMS)-treated Rats. *Journal of Oriental Neuropsychiatry*. 2002;13(2):107-19.
 41. Je JT, Lee SK. The Clinical Pilot Study of Hwangryunhaedok-tang Pharmacopuncture Therapy on Post-stroke Depression. *J of Pharmacopuncture*. 2010;13(2):67-73.

42. Cha YJ, Kim JW, Whang WW, Kim HT, Kwak SY, Kim MJ. The Effect of Guibiondamtang in an Animal Model of Depression using Chronic Mild Stress. *Journal of Oriental Neuropsychiatry*. 2001;12(2):53-68.
43. Lee JH, Kim GW, Koo BS. Experimental Study on the Anti-depressive Effect of Bee Venom Injection. *Journal of Oriental Neuropsychiatry*. 2010;21(2):61-73.
44. Chang MK, KimGW, Koo BS. Anti-depressant Effect of the Extracts of *Aconitum Carmichaeli*. *Journal of Oriental Neuropsychiatry*. 2009;20(3):49-64.
45. Lee ST, Kim GW, Koo BS. Experimental Study on the Antidepressant Effect of Sam-Jeong-Hwan. *Journal of Oriental Neuropsychiatry*. 2008;19(3):101-15.
46. Seol SH, Lee SR, Jung IC. The Effects of OnDam-tang-Kami-bang (ODK) in Antioxidant and Serotonin Metabolism Testing on P815 Cell. *Journal of Oriental Neuropsychiatry*. 2013;24(2):189-200.
47. Lee JW, Hong MC, Shin MK, Bae HS. Comparison of Nelumbinis Semen Extract with *Hypericum Perforatum* and Fluoxetine in Animal Model of Depression. *Journal of Oriental Neuropsychiatry*. 2006;20(4):830-43.
48. Ryu JM, Hwang UW, Kim JW, Chi SE, Kim EJ, Park EH. The effect of Jowiseungchungtang versus fluoxetine in the chronic mild stress model of depression in rats. *Journal of Oriental Neuropsychiatry*. 2004;15(1):27-41.
49. Sim TK, Jung IC, Lee SR. The Effect of Gamisoyo-san (Jiaweixiaoyaosan) on Serotonin Metabolism. *Journal of Oriental Neuropsychiatry*. 2011;22(1):37-51.
50. Eun Hee Lee, Woo Chang Choi, In Chul Jung. Preliminary Study to Develop the Instrument on Pattern Identifications for Depression. 2013;24(4):435-50.
51. Byun JS, Yang SY, Kang WC, Park YC. Primary Study to Develop the Instrument of Pattern Identification for Common Cold. *Korean J. Oriental Physiology & Pathology*. 2009;23(6):1226-31.
52. Portney LG & Watkins MP (2000) *Foundations of clinical research Applications to practice*. Prentice Hall Inc. New Jersey ISBN 0-8385-2695-0 p.560-7.