대한경락경혈학회지 Vol.22, No.4, pp.1~8, 2005
The Korean Journal of Meridian & Acupoint

A Pilot Study on Reliability of Pulse Diagnosis in Eight-Constitution Medicine

Hyang-Sook Lee[1] · Yong-Bum Lee[2,*] · Yong-Sup Shin[3] · Hee-Ju Kim[4] · Jung-Chul Seo[5] · Joon-Moo Lee[1] · Hye-Jung Lee[6] · Sun-Mi Choi[7]

[1]Dept of Meridian & Acupoint and [2]Medical Classics & History, College of Korean Medicine, SangJi University, Wonju ; [3]Shin Korean Medicine Clinic, Seoul ; [4]Ganghwa County Health Centre, Incheon ; [5]Dept of Acupuncture & Moxibustion, College of Korean Medicine, Daegu Haany University, Gumi ; [6]Acupuncture and Meridian Science Research Center, Kyung Hee University, Yongin ; [7]Dept of Medical Research, Korea Institute of Oriental Medicine, Daejeon

팔체질의학 맥진의 신뢰성 연구

이향숙[1], 이용범[2,*], 신용섭[3], 김희주[4], 서정철[5], 이준무[1], 이혜정[6], 최선미[7]

[1]상지대학교 한의과대학 경혈학교실 ; [2]상지대학교 한의과대학 원전의사학교실 ; [3]서울 신한의원 ; [4]강화보건소 ; [5]대구한의대학교 한의과대학 침구학교실 ; [6]경희대학교 침구경락과학연구센터 ; [7]한국한의학연구원 의료연구부

한글초록

배경 : 체질의학에서 진단의 중요성에도 불구하고 팔체질의학의 맥진의 신뢰성에 대한 과거 연구가 거의 없었다.

목적 : 본 예비연구는 팔체질의학의 맥진에 대한 진단자 간의 신뢰성 및 진단자 내의 신뢰성을 검사하기 위함이다.

재료 및 방법 : 팔체질의학을 이용한 진료 경력이 3년 내지 5년 되는 한의사 두 명이 실험에 참여하였다. 31명의 건강한 대학생들의 팔체질을 맥진을 통해 진단하고 그 결과를 확신도와 함께 각각 기록하도록 하였으며 이 중 5명은 각각 진단자로부터 3회씩 진단을 받도록 하였다. 그 결과를 분석하여 진단자 간의 신뢰성과 진단자 내의 신뢰성을 검사하였다. 아울러 진단에 대한 확신도와 결과와의 상관성이 있는지도 조사하였다. 대상자와 진단자는 서로를 알 수 없도록 가리개를 하고 실험을 진행하였다.

결과 : 진단자 간의 일치도는 35.7%였으며 카파 계수는 0.232로 보통 이하 수준인 것으로 나타났다. 진단자 내 일치도는 진단자 별로 각각 89%, 66.83%로 나타났다. 다만 두 진단자 사이의 결과가 일치하는 경우 불일치하는 경우보다 진단자의 확신도가 높은 것으로 드러났다.

결론 : 본 연구에서는 팔체질의학의 맥진의 진단자 간, 진단자 내의 신뢰성이 비교적 낮은 것으로 나타났다. 이는 진단의 변수가 많고 진단자의 숙련도나 대표성의 문제 등 여러 가지 요소들로 인한 결과로 파악된다. 팔체질의학의 진단자 간, 진단자 내의 신뢰성에 대한 향후 연구가 절실히 요구된다.

핵심어 : 팔체질의학, 맥진, 신뢰성, 일치도

* Correspondence to : Yongbum Lee, associate professor, KMD, PhD
Department of Medical Classics & History, College of Korean Medicine, SangJi University, Woosan-dong 660, Wonju, Kangwon-do, 220-702, Tel. +82-33-730-0668, Fax. +82-33-730-0653,
E-mail : erc633@sangji.ac.kr
·투고 : 2005/11/01 ·수정 : 2005/12/13 ·채택 : 2005/12/19

Ⅰ. Introduction

It is possible to conduct research on the accuracy of diagnostic tests or techniques. For example, a series of specimen might be shown to two different doctors, each asked to say whether the specimen indicated tumour, and the results compared. Such a study measures the 'reliability' of a test, that is, whether it gives similar results under different conditions. Usually, the results of a test are compared with a 'gold standard' and such a study is called 'validity' study. For the accuracy of diagnostic tests or techniques, establishing validity and reliability may be one of the most important prerequisites. If a test is not reliable, it cannot be valid; it must be both reliable and valid to be of benefit[1].

Eight-Constitution Medicine (ECM) was first introduced by Dr. Dowon Kuon in 1965 [2]. According to the ECM theory, the inherent imbalance of the organ functions results in specific characteristics in physiology, appearance, personality and diathesis, but is regarded as a natural and healthy status. This healthy imbalance, or individual difference, is classified into 8 types and called 8 constitutions. Under continuous unfavourable mental or physical conditions, however, the mutual controls among various organs are gradually deteriorated until the strongest organ of each constitution becomes hyperactive and the weakest organ relatively less active, thus creating an overbalanced status. ECM acupuncture treatments seek to identify constitution and restore overbalanced state to its inherent, healthy imbalanced state. As with other treatments in traditional Korean Medicine, treatment within ECM focuses not on the symptoms but rather on the causes of disease[3].

To evaluate the effect of ECM acupuncture treatment, it would be the very first step to establish reliability of its diagnosis, i.e. whether ECM practitioners determine the constitution consistently. There arise two research questions; (1) does an ECM practitioner have the same diagnostic result on a single patient when he/she repeats diagnostic procedures? and (2) do the ECM practitioners agree with each other on the diagnostic results when they see the same patient? The first question is about intra-rater (intraobserver) reliability or test-retest reliability, i.e. the same observer rates the same patients multiple times and the results are correlated, in this case, the same. The second refers to inter-rater (interobserver) reliability. These are common issues in clinical assessment and ECM diagnosis should be examined on them[4].

Despite the significance of reliability of diagnosis, there has been little research on ECM pulse diagnosis. Although the role of pulse diagnosis in determining ECM constitutions is pivotal, we do not have sufficient evidence to support its reliability and thus validity. Therefore, we aimed at testing inter- and intra-rater reliability of pulse diagnosis in ECM.

Ⅱ. Materials and Methods

A. Inter-rater reliability test

The study was conducted at one Korean Medicine Clinic in Seoul. Thirty-one healthy college students (M:F=21:17, mean age 23.8, range 19-33 yr) participated in the study and written informed consents were obtained. Subjects were told not to smoke or drink a day before the examination. Two Korean Medicine Doctors (diagnostician A and B) practising ECM for over three years determined participants' constitutional types based on pulse diagnoses. We designed two rows of random numbers so that the participants given his/her own number would be randomly allocated to each diagnostician without overlapping. All subjects were diagnosed by each diagnostician at least once. Diagnostician A and B made pulse diagnoses for participants who were randomly chosen for each diagnostician using random number table, and reported how much they were sure about their diagnosis, on a numeric rating scale (NRS; 0, 0-25% of confidence; 1, 25-50% of confidence; 2, 50-75% of confidence; 3, 75-100% of confidence). To ensure blinding, diagnostician A and B, with a patch over the eyes, waited for a participant to come in each room. The participant comes in and lies down in bed, an assistant covers the participant with white cotton sheet from head to toe but the arms outside for pulse to be taken. When the participant is ready, the assistant allows the diagnostician to take off the eye patch and start pulse diagnosis. After making pulse diagnosis and writing down the result, the diagnostician again wears the eye patch and waits till the next participant is ready to be examined.

B. Intra-rater reliability test

Five participants who were randomly chosen from 31 volunteers by selecting the first 5 numbers in a random number table were examined three times again in random order by each diagnostician. One participant was examined twice by technical error. All procedures were the same as those described in the inter-rater reliability test.

C. Statistical analysis

For inter-rater reliability, agreement and kappa (κ) value were determined. Agreement is the proportion of times that two diagnosticians give the same constitutional type. Kappa is the proportion of agreement achieved that is greater than that expected by chance. It varies between 0 and 1, where 0 represents chance agreement and 1 represents perfect reliability[5].

To determine association between diagnostician's confidence and inter-rater agreement, χ^2 test was used; when the diagnostician marked 2 or 3 on an NRS, it was considered 'confident' and if marked 0 or 1, 'not confident'. The results were divided into 'agreed' when the two diagnosticians had the same constitution for one participant or 'disagreed' when they had a different constitutional type for the same participant.

Table 1. Inter-rater Agreement

A \ B	Chol	Hep	Gas	Pan	Col	Pul	Ves	Ren	Radial artery malformation	Total
Chol	1			1	3					5
Hep		4			4	2				10
Gas										
Pan		2		1		1	1			5
Col				2	4	3	1	1		11
Pul		2					1	2		5
Ves		1					1			2
Ren								1		1
Radial artery malformation									3	3
Total	1	9		4	11	6	4	4	3	42

A, diagnostician A; B, diagnostician B; Chol, Cholecystonia (Mokeum constitution); Hep, Hepatonia (Mokyang constitution); Gas, Gastrotonia (Toeum constitution); Pan, Pancreotonia (Toeyang constitution); Col, Colonotonia (Geumeum constitution); Pul, Pulmotonia (Geumyang constitution); Ves, Vesicotonia (Sooeum constitution); Ren, Renotonia (Sooyang constitution). Agreement was 35.7% and κ value was 0.232.

For intra-rater reliability, within-rater agreement was calculated.

Ⅲ. Results

A. Inter-rater reliability

Each diagnostician determined constitutional type of 31 participants; once for 25 participants, twice for one participant, and three times for 5 participants, hence 42 times in total. There was a poor inter-rater agreement between the two study diagnosticians, 15 times agreed out of 42 times (Table 1). Agreement was 35.7% and κ value was 0.232.

For the two diagnosticians, no one had a significant correlation between their confidence about the result and agreement with the other diagnostician; P=0.488 for diagnostician A, P=0.177 for diagnostician B, χ^2 test. We then analysed the 'agreed' and 'disagreed' results separately with a cut-off point of 50% for confidence. When the results between the two diagnosticians were the same (n=12), both of them were more than 50% sure about their diagnosis (n=11); diagnostician A was more than 50% sure about his results for all cases (n=12) and diagnostician B was so for 11 cases. As diagnostician A missed confidence data for one case, we had a total 26 cases of disagreement. When the results disagreed (n=26), on the contrary, both of them were more than 50% sure about their diagnosis in only 16 cases. For 2 cases, they were less than 50% confident about their diagnosis; while diagnostician A was more than 50% sure about his results for 22 cases and less

Table 2. Intra-rater Agreement

Participant No.	Diagnostician A	Diagnostician B
	Times for the same result / No. of diagnoses made (agreement, %)	Times for the same result / No. of diagnoses made (agreement, %)
1	3/3 (100%)	3/3 (100%)
2	2/2 (100%)	0/2 (0%)
3	3/3 (100%)	3/3 (100%)
4	3/3 (100%)	2/3 (67%)
5	2/3 (67%)	2/3 (67%)
6	2/3 (67%)	2/3 (67%)

Mean agreement was 89% for diagnostician A and 66.83% for diagnostician B.

than 50% so for 4 cases, diagnostician B had 18 cases with over 50% of confidence and 8 cases with less than 50% confidence. For none of 27 cases were the two diagnosticians sure about their decision with more than 75% of confidence.

B. Intra-rater reliability

Diagnostician A and B made diagnoses more than one time for 6 participants, three times for 5 participants each and twice for one participant. Diagnostician A had a complete agreement (100%) for 4 participants and 67% agreement for two participants. Diagnostician B had a complete agreement for two participants, 67% agreement for three participants. For one participant who was diagnosed twice, two different constitutional type diagnoses were made, resulting no agreement (0%). Mean agreement was 89% for diagnostician A and 66.83% for diagnostician B (Table 2). Compared with diagnostician B, diagnostician A had a better intra-rater reliability and seemingly more confidence for his diagnostic results. No formal statistics were done due to a small number of cases and missing data on confidence.

Ⅳ. Discussion

The findings of this study suggest that there exist poor inter-rater reliability in ECM pulse diagnosis. Although intra-rater reliability was good, small sample size limits its meaningful interpretation.

The results, however, should be analysed in consideration of several aspects. First of all, pulse diagnosis may require practitioners' high concentration and changed settings and other confounders such as presence of assistant and stress from being tested could have hampered their capability to perform as usual. Although pulse diagnosis plays a crucial role in ECM diagnosis, other information from questioning the patients and even response from ECM acupuncture should be integrated to finally determine the constitutional types. Therefore, determining the ECM constitutions on-

ly by pulse diagnosis and comparing the results to assess the reliability of ECM diagnosis can be open to criticism. As it is a pilot and probably the first study to assess the reliability of pulse diagnosis of ECM in an objective, blinded manner, further studies assessing the whole diagnostic method or system of ECM are warranted to truly determine the reliability of ECM diagnosis.

Secondly, regarding the inter-rater reliability or agreement of ECM pulse diagnosis, we applied κ value. The simplest approach to assessing agreement is simply to see how many exact agreements are observed, which here is 15 times out of 42 times. There is thus agreement for 15/42 = 0.357 (35.7%) of the ECM constitutions. There exist two weaknesses of this simple calculation; firstly, it takes no account of where in the table the agreement was, and secondly, we would expect some agreement between the diagnosticians by chance even if they were guessing. We can get a more reasonable answer by considering the agreement in excess of the amount of agreement that we would expect by chance, that is, κ value[6]. It has a maximum of 1.00 when agreement is perfect, a value of zero indicates no agreement better than chance, and negative values show worse than chance agreement, which is unlikely in this context. Here in this study, the κ value was 0.232 which can be interpreted as fair according to Landis et al.[7] (Table 3). In practice, any value of κ much below 0.5 will indicate poor agreement[6]. However, we contend that the degree of acceptable agreement must depend on circumstances. There are at least 8 variables, i.e. 8 constitutional types, and in this study 9 variables including radial artery malformation were analysed. Hence, the κ value should be lower than other usual diagnostic tests and head-to-head comparison of κ value might be misleading. Kappa value of 0.232 which was rated as fair in this study may be more than fair if the number of variables analysed, which is relatively large in this case, are appropriately weighted.

Table 3. Interpretation of κ Value (Landis et al., 1977)

Value of κ	Strength of agreement
< 0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very good

Thirdly, one can argue that the participating diagnosticians may not represent ECM practitioners. Although intra-rater agreement was good reaching 67% to 100% for diagnostician A, diagnostician B had a wider range of agreement, from 0% to 100%. Variability between diagnosticians should be due to multiple factors, e.g. duration of ECM training and practice. Some constitutional types may be easier to detect by pulse diagnosis than the others, or there may exist a specific constitutional type which is hard to discriminate from another specific type. One diagnostician would find one constitution easier to determine while

another diagnostician finds it difficult. Lack of 'gold standard' also limits assessment of validity of this diagnostic technique.

Regarding the association between degree of confidence and inter-rater agreement, it was not possible to find any association between degree of confidence and inter-rater agreement, probably owing to the small sample size. However, when the agreements and disagreements are analysed separately, it takes on a rather different picture. The more confident the diagnostician is about the result, the more inter-rater agreements he seems to reach. This seems to be in accordance with a homeopathy trial where homeopaths were confident about the chosen similimum a significant difference was observed in favour of the homeopathic treated group over placebo[8]. More research on the association between degree of confidence and inter-rater agreement is needed.

Future studies on this topic may extend investigations to associations between inter-rater reliability and other sociodemographic and health-related variables such as occupation, smoking, drinking, body mass index, and blood pressure. It should be born in mind, however, that establishing inter- and intra-rater reliability of ECM diagnosis is of the utmost importance.

V. Conclusion

This pilot study suggests that there is a poor inter-rater agreement in ECM pulse diagnosis and intra-rater reliability varies among diagnosticians. It seems that degree of confidence is related with inter-rater agreement. Due to a small sample size and other limitations, more appropriately designed studies on this topic are warranted.

1. Vickers AJ, van Haselen RA, Peng L, Berkovitz S. Inter-rater reliability of symptom repertorisation: a pragmatic empirical study. Br Homeopath J. 2000 ; 89 : 188-91.
2. Kuon D. A study of constitution-acupuncture. Journal of the International Congress of Acupuncture and Moxibustion. Tokyo, Japan Acupuncture and Moxibustion Society. 1965 ; 10: 149-67.
3. Kuon D. Eight-Constitution Medicine: An overview. Seoul : Institute for Modern Korean Studies, Yonsei University Press. 2003.
4. Polgar S. Measurement and instrumentation. In: Polgar S, Thomas SA. Introduction to Research in the Health Sciences. 3rd ed. Melbourne : Churchill Livingstone. 1998 : 161-8.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology. Boston : Little, Brown. 1991.
6. Altman DG. Some common problems in medical research. In: Altman DG. Practical statistics for medical research. 1st ed. London : Chapman & Hall/CRC. 1999 : 396-439.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 ; 33 : 159-74.

8. Jacobs J, Malthouse S, Chapman E, Jiménez LM, Crothers D. Childhood diarrhea; results from Nepal and combined analysis. Proceedings of the Congress of the Liga Medicorum Homoeopathica Internationalis. 1997 ; 52 : 81-3.