Original Article

Developing an optimized cold/heat questionnaire

CrossMark

Youngheum Yoon, Hoseok Kim, Youngseop Lee, Jonghyang Yoo,
Siwoo Lee**

Mibyeong Research Center, Korea Institute of Oriental Medicine, Daejeon, Korea

ABSTRACT

Background: The cold/heat questionnaire is one of the most actively developed patient diagnostic tools in traditional Korean medicine (hereafter abbreviated as TKM) because of its objectivity. Unfortunately, the existing questionnaires contain too many items to hold the respondent's attention. In the current study, we aimed to develop an optimized cold/heat questionnaire to be used as a complement to the existing questionnaires.

Methods: We developed a new cold/heat questionnaire based on a domain analysis of the existing questionnaires. The questionnaire's reliability was examined via two test–retest reliability analyses involving 1890 individuals in November 2013 and February 2014. Its validity was examined using a professional cold/heat diagnosis kappa value.

Results: The new cold/heat questionnaire consisted of a total of seven items, which were created based on an analysis of the existing questionnaires. A reliability analysis performed using the study participants revealed a correlation coefficient of 0.609, 74.5% agreement with professional cold/heat diagnoses by TKM practitioners, and a kappa value of 0.487.

Conclusion: In the current study, we developed an optimized cold/heat questionnaire. The level of agreement between the questionnaire and professional cold/heat diagnoses by TKM practitioners was significant, which indicates great potential for its widespread use as a diagnostic tool in TKM.

1. Introduction

Pattern identification refers to a diagnostic system in which the patient's clinical data are analyzed based on four examination methods and are subsequently synthesized in order to determine the nature and location of the ailment and treatment options.[1] Pattern identification has been studied and developed throughout the long history of traditional Korean medicine (hereafter abbreviated as TKM). However, because the diagnosis is mainly based on the patient's subjective expression of symptoms and the practitioner's intuitive knowledge, an objective evaluation can be difficult.[2] Currently, evidence-based research and objective evaluation are more important than ever in medicine. In keeping with this trend, pattern identification in TKM must evolve from the individuals' subjective evaluation to objective and quantifiable evaluation methods.[3] The pattern identification

questionnaire is regarded as the primary answer to this demand.

Questionnaires are widely used across academic fields because their logical and systematic construction helps objectively identify the respondents' characteristics and current health status.[4] In addition, they have a substantial influence on the field of medicine, as evidenced by their wide adoption as tools for identifying the nature of various communicable disease outbreaks.[5]

Among the pattern identification types, cold/heat is regarded as the most important scale used to survey a patient's health status in Oriental or constitutional medicine.[6] Cold/heat is also a quintessential phenomenon studied in clinical TKM. Therefore, cold/heat questionnaires comprise an important part of TKM questionnaire development.

In Korea, questionnaires that effectively embody the characteristics of TKM have been developed and published since the late 1980s.[7] They have been used as a tool to determine an individual's constitution[8] and evaluate certain ailments such as back pain[9] and static blood.[10] Cold/heat questionnaire testing the agreement with Korean medicine doctor's diagnosis by using experts' surveys and the Delphi method is available.[7,11] Another questionnaire has been designed to evaluate deficiency of heat, true cold with false heat, and cold-heat complex.[2] In addition, cold/heat and deficiency/excess questionnaire has been developed to survey both of them.[12] However these questionnaires have some limitations such as having too many items and including similar contents, and these limitations caused difficulty in holding respondents' attention.

2. Methods

First, a new set of subdomains was established after analyzing the existing questionnaires, and a new set of optimized domains and items was created via a discussion between the authors. Second, the reliability and validity of the questionnaire were tested on survey participants who had been recruited in 2014 for the "health impact survey for the residents of communities that house disposal facilities" (Fig. 1).
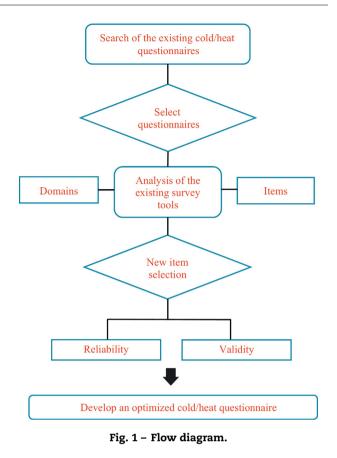
2.1. Selecting questionnaire items

2.1.1. Search for existing cold/heat questionnaire items
Using well-established nationwide academic paper search engines such as Oasis, DBPIA, and KISS, we searched for available academic papers using "cold/heat" and "questionnaire" as keywords. Of the retrieved questionnaires, those currently in wide usage were selected for the analysis. After a review of the number and nature of the items in the selected questionnaires, we established new domains and examined the reliability and validity of the items.

2.1.2. Restructuring the domains and items of the new questionnaire
Three authors performed a domain review, and eliminated items that focused on patients with diseases and domains that contained similar items. The new domains were created by extrapolating the 10 traditionally used questions (used in TKM

Fig. 1 – Flow diagram.

for the identification of the cold/heat pattern in the human body).[13]

2.2. Reliability and validity analysis

2.2.1. Study participants
The study participants consisted of 1890 individuals who had been recruited on two separate occasions (November 2013 and February 2014 in Gyeong-Ju, Korea) for the "Health impact survey for the residents of communities that house disposal facilities." Upon completion of the first and second surveys, the reliability of the questionnaire was tested on 1759 participants with no missing values. All 1759 participants were examined for cold/heat. The questionnaire's validity was tested on 1527 participants after eliminating 232 participants who did not belong to either the "cold" or the "heat" group.

2.2.2. Professional cold/heat diagnosis
TKM practitioners with a minimum clinical experience of 5 years performed the professional cold/heat diagnosis. The examination consisted of personal interviews of the participants regarding general health status and cold/heat diagnosis.

2.3. Statistical analysis

2.3.1. Reliability
A test–retest reliability examination that utilized the first and second surveys was performed to assess the reliability. Individual items, as well as the combined cold/heat scores of the items in the first and second surveys, were used for the kappa

Table 1 – Analysis of the existing survey tools				
Authors	Items	Domain	Reliability	Validity
Choi et al[14]	20	5—whole body, body parts, fluid intake, secretion/urine and stool, sensation	No	No
Kwon et al[12]	15	4—whole body, body parts, secretion/urine and stool	Cronbach α 0.721–0.849	Factor analysis
Kim & Park[15]	15	5—whole body, body parts, fluid intake, secretion/urine and stool, emotion	Cronbach α 0.5934–0.7526	Factor analysis
Ryu et al[7]	26	5—whole body, body parts, fluid intake, secretion/urine and stool, sensation	Cronbach α 0.605–0.722	Professional diagnoses agreement 94%

statistic, weighted kappa statistic, and Pearson's correlation coefficient analyses. Reliability was examined using the agreement and correlation obtained from each analysis.

2.3.2. Validity

To determine the participants' cold/heat status obtained through the survey, the cutoff values for the questionnaire were obtained via the receiver operator characteristic curve analysis. To examine the agreement between the cold/heat data based on the cutoff values and the professional diagnoses, the validity of the cold/heat questionnaire was examined using the kappa value.

3. Results

3.1. Analysis of the existing survey tools

The online search performed with "cold/heat" as a keyword yielded a total of 64 academic papers after the elimination of duplicates. The search performed with "questionnaire" as a keyword yielded a total of 81 academic papers. A total of 21 papers were confirmed to be relevant for the current study after eliminating the questionnaires and classic texts on bodily constitution, and the papers that did not include questionnaire use. The majority of the 21 papers addressed the relationship between symptoms and device measurements in the existing questionnaires, with the exclusion of several papers pertaining to the questionnaire design. Ultimately, five questionnaires were selected for the analysis. The questionnaires were composed of an average of 20 (19.8) items distributed across six domains consisting of "whole

body," "body parts," "fluid intake," "secretion/urine and stool," "sensation" (pain), and "emotion." In terms of reliability, a Cronbach α value was provided for all but one questionnaire. Factor analysis, agreement with professional diagnoses, and kappa values were used for validity analysis (Table 1).

3.2. Selecting new questionnaire items

For the healthy participants with no particular ailments, the subdomains of whole body, body parts, and fluid intake were selected upon elimination of the remaining subdomains of "urine and feces," "sensation" (pain), and "emotion" because they are highly specific. In addition, redundant items within the same domain were eliminated, and a few items were consolidated into one by extrapolating the semantic differential scale. Scores of the seven items were combined to derive the total questionnaire score, which ranged from a minimum of seven points to a maximum of 49 points. Lower scores indicated "cold," while higher scores indicated "heat," and the total combined score was defined as the "cold/heat" score (Table 2).

3.3. General characteristics of the participants

Of a total of 1898 participants, 1759 [607 (34.5%) males and 1152 (65.5%) females] participated in the reliability test and 1527 [530 (34.7%) males and 997 (65.3%) females] in the validity test. The body mass index, height, and weight of the male participants were significantly higher than those of their female counterparts. The participants who were aged o50 years accounted for 88.8% of the total participants (Table 3).

Table 2 – New cold–heat questionnaire								
A		←----------------→						B
(1) Hands usually cold	1	2	3	4	5	6	7	Hands usually warm
(2) Feet usually cold	1	2	3	4	5	6	7	Feet usually warm
(3) Abdomen usually cold	1	2	3	4	5	6	7	Abdomen usually warm
(4) Prefers to drink cool water	1	2	3	4	5	6	7	Prefers to drink warm water
(5) Often does not drink water	1	2	3	4	5	6	7	Prefers to drink a lot of water
Not at all	←----------------→							A lot
(6) Usually dislikes cold	1	2	3	4	5	6	7	
A lot	←----------------→							Not at all
(7) Usually dislikes warm	1	2	3	4	5	6	7	

Table 3 – General characteristics of the participants

	Reliability (N = 1759)	Validity (N = 1527)
Sex		
Male	607 (34.5)	530 (34.7)
Female	1152 (65.5)	997 (65.3)
Age (y)	64.12 ± 11.31	64.24 ± 11.23
≤19	7 (0.4)	5 (0.3)
20–39	80 (4.5)	66 (4.3)
40–49	112 (6.4)	100 (6.5)
50–59	343 (19.5)	296 (19.4)
60–69	591 (33.6)	507 (33.2)
≥70	626 (35.6)	553 (36.2)
Height	157.54 ± 8.96	157.55 ± 8.97
Weight (kg)	62.33 ± 10.33	62.27 ± 10.47
BMI (kg/m^2)	25.06 ± 3.22	25.03 ± 3.25

Data are presented as n (%) or mean ± standard deviation.
BMI, body mass index.

3.4. Reliability test results

To determine the questionnaire's reliability, a test–retest reliability examination was performed, which utilized the first and second surveys. The agreement analysis performed with the first and second cold/heat questionnaires indicated a relatively low-weighted kappa coefficient of 0.273–0.551 for individual item agreement. However, a correlation analysis performed with the overall cold/heat scores revealed a correlation coefficient of 0.609 (Table 4).

3.5. Validity test results

The overall cold/heat scores were grouped according to the participants' sex, age, and "sex + age" using the receiver operator characteristic curve analysis method in order to examine the cutoff values. The scores ranged from 26.5 points to 35.5 points. The grouping accuracy was determined by the area-under-the-receiver-operator-characteristic-curve values, which ranged from 0.703 to 0.817. An area-under-the-receiver-operator-characteristic-curve value close to 1 indicates a more accurate grouping. Using the cutoff values, the agreement between the professional diagnoses and the cold/heat grouping scores was examined based on the cold/heat scores. The agreement level ranged from a minimum of 68% to a maximum of 80.5%, and the kappa coefficient values ranged from 0.335 to 0.551 (Table 5).

The four cutoff values obtained from the sex + age group were compared with the professional cold/heat diagnoses; the agreement level was 74.5% and the kappa coefficient value was 0.487 (Table 6).

Table 4 – Reliability test results

	Primary cold/heat score	Secondary cold/heat score	Correlation coefficient
Test–retest	28.7 ± 6.6	27.4 ± 6.5	0.609[*]

Data are presented as mean ± standard deviation.
[*] p < 0.01.

Table 5 – Validity test results

	N	AUC (95% CI)	Cutoff value	Agreement (%)	Kappa
Total	1527	0.802 (0.780–0.824)	26.5	73.8	0.477
Sex					
Male	530	0.749 (0.705–0.792)	28.5	70.4	0.397
Female	997	0.812 (0.785–0.839)	27.5	75.6	0.498
Age (y)					
≤49	171	0.802 (0.735–0.870)	30.5	74.9	0.482
≥50	1356	0.806 (0.783–0.829)	26.5	74.3	0.488
Sex, age (y)					
Male, ≤49	53	0.703 (0.565–0.842)	35.5	64.8	0.335
Male, ≥50	477	0.761 (0.716–0.807)	28.5	71.2	0.419
Female, ≤49	118	0.817 (0.732–0.902)	30.5	80.5	0.551
Female, ≥50	879	0.813 (0.785–0.842)	27.5	76.0	0.506

AUC, area under the receiver operator characteristic curve; CI, confidence interval; Kappa, kappa statistics.

Table 6 – Agreement level and kappa value

		Cold/heat questionnaire		Total	Agreement (%)	Kappa
		Cold	Heat			
Professional diagnoses	Cold	635	156	791	74.5	0.487
	Heat	234	502	736		
Total		869	658	1527		

4. Discussion

In TKM, patients are examined using the following four basic categories of diagnostic observations: inspection, listening and smelling examination, inquiry, and palpation. The aim of pattern identification is to determine a treatment strategy upon identification of the root cause, location, and characteristics of the ailment through observation of external manifestations using the four diagnostic categories.[16] A variety of pattern identification methods have been developed in TKM, and the most widely used are the eight-principle, six-meridian, visceral and bowel, meridian and collateral, qi–blood–fluid–humor, defense–qi–nutrient–blood, triple energizer, and bodily constitution pattern diagnoses. The eight principles include yin/yang, exterior/interior, cold/heat, and deficiency/excess. These patterns were established based on the long history of observation in TKM, during which it was discovered that all ailments are rooted in imbalance in these patterns. Although the official nomenclature was established in modern times, the fundamental theory in TKM was formed long before.[16] Of these, cold and heat are the two fundamental patterns used to determine the nature of an ailment, and they serve as the basis for prescribing medicine or acupuncture therapy. The degree of cold and heat are judged by four diagnostic criteria; among them inquiry is a simple way to determine the patient's symptoms. Zhang Jie-bin, of the Ming Dynasty, established 10 principles for inquiry. These principles, which involve Q cold/heat, the whole body, body parts, sweat, urine and feces, food intake, fluid intake, the presence of other diseases, and the presence and nature of the patient's specific symptoms, can be quite useful. However, this type of pattern identification heavily relies on the practitioner's subjective evaluation and the patient's individual expression of the symptoms. Therefore, the diagnosis for the same patient can vary from one practitioner to another, which consequently affects the accuracy of the treatments prescribed.

With the current rising popularity of evidence-based medicine, the demand is increasing for objective and quantifiable diagnostic methods in TKM, and the TKM questionnaires are compatible with this trend in medical science. Through the scientific expression of quantifiable measurements, questionnaires are effective tools for evaluating the health status of a patient and the nature and severity of an ailment. The following cold/heat questionnaires are popular and currently used in the related field: a 24-item questionnaire,[14] a 15-item questionnaire that was developed based on the 30-item Dongui Medical Center questionnaire after validity verification,[12] a 15-item questionnaire developed after a four-step preliminary survey,[15] a 20-item questionnaire developed after validity verification with professional diagnoses by TKM practitioners,[7] and a 23-item questionnaire developed by the Korea Institute of Oriental Medicine for a project. The items in these questionnaires were selected using an expert Delphi method based on the classic TKM literature. The six domains consist of "whole body," "body parts," "fluid consumed," "bodily fluids/urine and feces," "pain," and "emotion," according to the symptoms expressed by the patients. This suggests that each domain consists of a minimum of two and a maximum of eight similar items. A well-designed questionnaire is concise; uses plain, everyday language; and avoids redundant items while maintaining the necessary items.[17] In particular, because the quality of the response decreases as the length of the item increases,[18] the total number of questionnaire items should not exceed 15 in order to keep the respondents attention on the subject.[19]

To this end, the current study restructured the domains of the existing questionnaires based on an internal discussion and reduced the number of items by half using the Likert scale. Thus, the semantic differentials were applied to dual items for the same body parts, such as hands and feet, and the domains in the existing version of the questionnaire were rated on the Likert scale. The last two items ("I generally dislike being cold/hot") and the corresponding Likert scales were retained because some respondents reported disliking both (i.e., summer and winter).

Items evaluating the cold/heat pattern in TKM tend to vary in the literature according to the presence/absence of pathological conditions. For instance, in the case of common cold, wind-cold and wind-heat are determined based on the color of phlegm and nasal discharge. However, it is inappropriate to include these as cold/heat pattern identification items that aim to identify the respondents' general state because the degree of nasal discharge or phlegm varies among healthy individuals. In addition, although the degree of pain or sweating is not effective for determining an individual's general cold/heat tendency, the determination of pain or sweating or the absence of both should be prioritized. A questionnaire completed in accordance with such standards is expected to be highly useful as a tool for evaluating the respondents' health status. However, in order for these diagnostic tools to be widely adopted, item reliability needs to be verified. Further, validity of the items needs to be verified by testing their agreement with professional diagnoses by TKM practitioners. The reliability and validity of the questionnaire developed in the current study were verified using individuals collected from clinical cases.

The results of the test–retest reliability examination based on the first and second survey data indicated low interitem reliability ratios (0.273–0.551). However, in terms of the overall cold/heat scores, a correlation coefficient of 0.609 was obtained. The validity was tested on 1572 participants selected after the elimination of 232 participants who did not belong to either cold or heat group. The test results indicated an agreement range of 64.8–80.5%, with a kappa coefficient value of 0.335–0.551. The four cutoff values, based on sex and age, were compared with professional diagnosis values; a 74.5% agreement level and a statistically significant kappa coefficient value of 0.487 were obtained.

In contrast, Ryu et al[7] reported a value of 0.605 for "cold" items and 0.722 for "heat" items, which had a relatively robust internal agreement. However, because these figures were obtained from data of patients with diseases, using the same items on healthy individuals would be inappropriate. Although the interitem reliability of the questionnaire developed in the current study was found to be low, the validity test involving healthy individuals revealed fairly high agreement with the professional diagnoses, which is relatively significant.

This suggests that the cold/heat questionnaire developed in the current study has great potential as a diagnostic tool

in clinical settings. However, because 88.8% of the study participants were individuals over the age of 50 years and the male-to-female ratio was 1:2, follow-up studies involving a wider range of age groups with a more balanced sex ratio are needed.

In conclusion, in the current study, we aimed to develop a concise cold/heat questionnaire designed to evaluate healthy individuals (as opposed to patients with diseases). Our analysis of the existing questionnaires revealed a pattern of six domains, including whole body, body parts, fluid intake, bodily fluids/urine and feces, sensation, and emotion. Following an internal discussion, domains consisting of items evaluating patients with diseases were eliminated, and, ultimately, a questionnaire consisting of seven items was developed. The questionnaire was applied in a study involving healthy individuals, and the results were compared with professional diagnoses. This comparison revealed a 74.5% agreement level, with a kappa coefficient value of 0.487, which was statistically significant. We hope that this questionnaire will be widely adopted as an assessment tool in the related field.

Conflicts of interest

The authors declare that they have no competing interests.

Acknowledgments

R E F E R E N C E S

1. Kim JH. A Comparative research of eight principle pattern identification—based on Zhang Jie-Bin, Cheng Guo-Peng, and Jiang Han-Tun. *J Orient Med Class* 2013;26:47–59.
2. Bae NS, Park YJ, Oh HS, Park YB. Preceding studies for questionnaires on Han-Yol patternization. *J Korea Inst Orient Med Diagn* 2005;9:98–111.
3. Cho HS, Bae GM. Research in developing of diagnosis questionnaires on cold, heat, deficiency and excess. *Korean J Orient Physiol Pathol* 2009;23:288–93.
4. Kim KT, Eom HS, Lee IS, Kim JW, Chi GY. Study on standard symptoms and gender differences of Qi.Xue.Yin.Yang asthenia syndromes based on questionnaire analysis. *Korean J Orient Physiol Pathol* 2006;20:1742–8.
5. Lee DW, Byun DG, Kim JW. Diagnostic significance of clinical history of atopic dermatitis in Koreans by questionnaire. *Korean J Dermatol* 1997;35:443–9.
6. Lee SH, Han SS, Jang ES, Kim JY. Clinical study on the characteristics of heat and cold according to Sasang Constitutions. *Korean J Orient Physiol Pathol* 2005;19:811–4.
7. Ryu HH, Lee JJ, Jang ES, Choi SM, Lee SG, Lee SW. Study on development of cold-heat pattern questionnaire. *Korean J Orient Physiol Pathol* 2008;22:1410–5.
8. Hwang DS, Cho JH, Lee CH, Jang JB, Lee KS. A study on reproducibility of responses to the Questionnaire for Sasang Constitution Classification 2 (QSCC2). *J Korean Orient Med* 2006;27:145–50.
9. Ahn MY, Song YK, Ko SG, Lim HH. Development of questionnaire for pattern identification of chronic low back pain by Delphi method. *J Korean Med Rehabil* 2013;23:143–58.
10. Yang DH, Park YJ, Park YB, Lee SC. Development of questionnaires for blood stasis pattern. *J Korea Inst Orient Med Diagn* 2004;10:141–52.
11. Kim SK, Nam TH, Park YB. Validity of questionnaires for cold-heat patternization (1). *J Korea Inst Orient Med Diagn* 2002;6:141–56.
12. Kwon OS, Kim JE, Lee JW, Seo CW, Han HY, Hong SH. Analytic study of diagnostic validity by the measure of cold-heat & deficiency-excess for Oriental medical examination. *Korean J Orient Physiol Pathol* 2009;23:180–5.
13. Liu GP, Wang YQ. Origin and development of inquiring diagnosis theory of TCM. *Acta Univ Trad Med Sinensis Pharmacol Shanghai* 2008;23:21–3.
14. Choi SH, Rhim HM, Oh JG, Rhim JY, Kang HW, Kim YS, et al. Standard pattern identifications for post stroke depression by Delphi method. *Korean J Orient Physiol Pathol* 2012;26:367–75.
15. Kim SK, Park YB. Development of questionnaire for cold-heat patternization. *J Korea Inst Orient Med Diagn* 2003;7:64–75.
16. Zhu WF. *Diagnostics of traditional Chinese medicine*. Beijing: China Press of Traditional Chinese Medicine; 2002.
17. Park JY. Basic principles for developing questionnaires. *Korean J Community Nutr* 2001;6:419–23.
18. Kim KH, Yoo DJ, Kim HJ, Kim CT. The influence of the questionnaire length on response quality. *Korean Assoc Surv Res* 2015;16:1–48.
19. Lee KS, Moon YS. A study on verbalization and wording in questionnaires and interviews: centered on the Korean Public Administration Review and the Korean Journal of Local Government Studies. *Korean J Local Gov Stud* 2014;17:165–91.