Research paper

Reliability and validity of a pattern identification questionnaire for liver-qi stagnation infertility

Ji-Yeon Lee[a], Soo-Yeon Ahn[a], Da-Hee Jun[b], Ki-Tae Ha[c], Jeong-Eun Yoo[a],*

[a] Department of Obstetrics and Gynecology, College of Korean Medicine, Daejeon University, Daejeon, South Korea
[b] College of Korean Medicine, Daejeon University, Daejeon, South Korea
[c] School of Korean Medicine and Healthy Aging Korean Medical Research Center, Pusan National University, Yangsan, South Korea

ABSTRACT

Introduction: Traditional Korean medicine recognizes six types of female infertility. One of these types is liver-qi stagnation infertility. A pattern identification questionnaire for liver-qi stagnation infertility was first developed to objectively assess the diagnosis and treatment of liver-qi stagnation according to traditional Korean medicine. This study examined the reliability and validity of that questionnaire.
Methods: Forty-two infertile women identically diagnosed by two liver-qi stagnation diagnostic experts were studied. The reliability of the questionnaire was measured using Cronbach's α, and validity was examined using the McNemar test and kappa statistic to assess the agreement between the diagnoses by the two experts. Construct validity was tested using factor analysis. The cutoff value for the diagnosis of liver-qi stagnation was determined by the Youden index and concordance probability.
Results: The questionnaire had relatively high reliability, with a Cronbach's α of 0.858. The kappa statistic for validity was 0.6383, indicating significant agreement between the two experts. Factor analysis to test construct validity revealed four major factors. The cutoff score for the diagnosis of liver-qi stagnation was 37, with a sensitivity of 82.35% and specificity of 92.00%.
Conclusions: The questionnaire for liver-qi stagnation infertility may be appropriate for identifying and assessing liver-qi stagnation and has statistical reliability and validity. Follow-up studies with large samples should be performed to review potential problems and confirm statistical significance.

1. Introduction

Infertility or subfertility can be defined as the decreased ability or inability to conceive because of a disease or unknown cause. Generally, 85–90% of women who attempt to conceive will do so within 1 year; however, the remaining 10–15% who fail to conceive within 1 year are usually diagnosed with infertility [1]. The incidence of infertility is increasing as a result of many factors, including older age at the time of marriage, prolonged use of contraception during marriage, artificial abortion, stress, and environmental pollution [2]. Female-related causes of infertility (45–55%) include ovulation disorders (15–25%), fallopian tube and peritoneal factors (25–35%), and infertility of unexplained cause (10–20%) [3].

Traditional Korean medicine (TKM) and combined Western

medicine and TKM treatments have been determined to be effective for treating infertility [4,5]. Recently, TKM practice guidelines for infertility that encompass study findings and expert recommendations have been developed to promote evidence-based practice [6,7]. These practice guidelines outline the traditional Korean diagnostic approach of pattern differentiation to identify the imbalance among the visceral organs, qi, and blood, which affect infertility [8]. Pattern differentiation is a unique diagnostic approach used in TKM; symptoms and signs manifested in patients are examined as a whole to determine the diagnosis [9]. In TKM, pattern identification is an important index that assists in medical decision-making [10].

Causes of female infertility according to TKM include kidney deficiency, liver-qi stagnation (LQS), blood stasis, phlegm-dampness, qi and blood deficiency, and damp heat [11]. LQS infertility is described as a
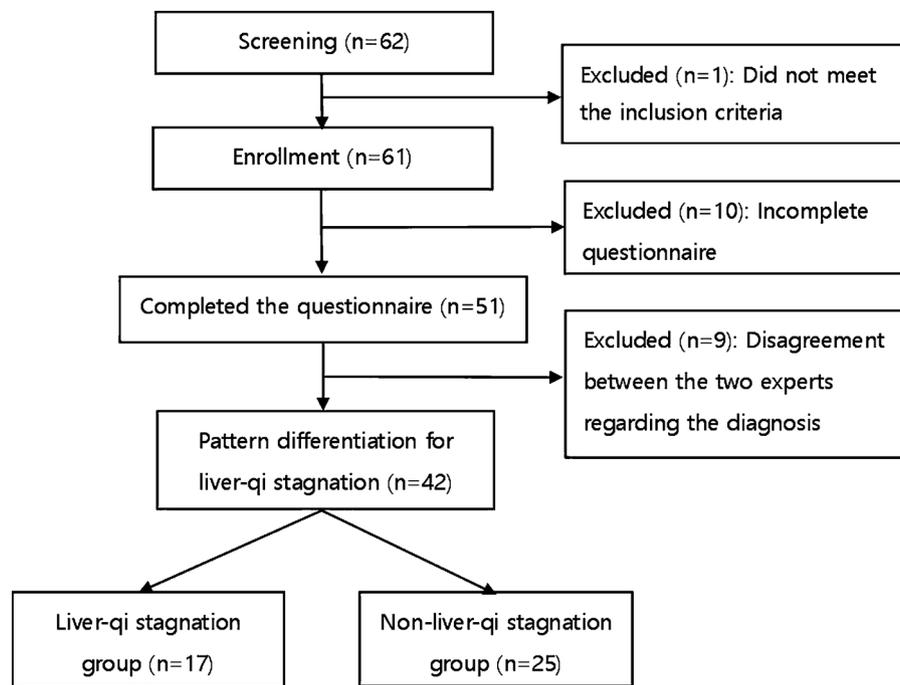
Fig. 1. Flowchart of the study.

condition in which LQS delays circulation to the uterus, which stagnates qi and blood in the uterus and induces obstruction of the Chong meridian and Ren meridian, thereby hindering pregnancy [12]. Attention focused on the identification of the LQS infertility pattern has increased because *Paeonia lactiflora*, an ingredient in an LQS standard prescription known as *jokyungjongoktang*, has been found to promote implantation by increasing endometrial receptivity [6,13]. Furthermore, LQS infertility has been speculated to be related to repeated implantation failure, ovulation disorder, and infertility due to unexplained causes; therefore, its clinical importance is emphasized [14].

Pattern identification in TKM involves the differentiation of conditions by collecting information about symptoms and signs using four traditional diagnostic methods: inspection, listening and smelling, inquiring, and palpation [15]. These four TKM diagnostic methods are effective, but the results may differ according to the physician's level of knowledge and proficiency in diagnostic methods [16,17]. To overcome these shortcomings, multiple studies have attempted to develop pattern identification tools with enhanced objectivity and reproducibility [10,18,19]. Similarly, a pattern identification questionnaire for LQS infertility (QLQSI) was first developed to increase the objectivity and usefulness of the TKM diagnostic method for infertility. Symptoms of LQS infertility were identified based on reviews of Korean and Chinese literature to develop a 16-item draft of the questionnaire. The draft underwent three rounds of review by 20 Delphi experts who were traditional Korean gynecologists at TKM schools nationwide. Two items with low agreement were excluded, resulting in a 14-item QLQSI [20].

The present study evaluated the reliability and validity of the QLQSI for diagnosing women with infertility. To standardize the diagnosis of LQS, the cutoff score of the QLQSI was computed.

2. Methods

2.1. Participants

Participants were recruited for 6 months, from March 1, 2018, to August 31, 2018, at the Women's Medical Center at Cheonan Korean Medicine Hospital of Daejeon University. Eligible participants (per the inclusion and exclusion criteria) were enrolled in the study.

2.2. Inclusion and exclusion criteria

The inclusion criteria were as follows: women who attempted to conceive without using contraceptives but who did not conceive within 1 year; ovarian insufficiency; reduced egg quality; repeated implantation failure; repeated miscarriage; and infertility of unexplained cause. All participants signed and submitted an informed written consent form for this study.

The exclusion criteria were as follows: women diagnosed with infertility due to their spouse's aspermia; women diagnosed with infertility due to other spouse-related factors; and women who were determined by the investigator to be unfit for the clinical trial.

2.3. Ethical consideration

This study was approved by the Institutional Review Board of the Cheonan Korean Medicine Hospital of Daejeon University (DJUMC-2018-BM-02). Participants were provided a thorough explanation of the study purpose and methods before they voluntarily signed the study consent form. This study complied with the Declaration of Helsinki.

2.4. Study design and procedure

This study was designed as a single-group observational study. After signing the informed consent form, the participants were given a screening number based on when they signed the form. Those who met the inclusion criteria completed the pattern identification QLQSI and provided their demographic information (Fig. 1).

2.4.1. QLQSI completion and pattern identification

The QLQSI (Appendix A) is a 14-item questionnaire; each item is rated using a 6-point Likert scale ranging from 0 (not at all) to 5 (extremely severe). Items 1 to 10 are related to LQS symptoms and are completed by the participants. Items 11 to 14 are related to tongue diagnosis and pulse diagnosis for LQS and are completed by physicians; in this study, items 11 to 14 were completed by two TKM physicians who had at least 2 years of clinical experience and who specialized in traditional Korean gynecology.

After the participants completed items 1–10, the two TKM doctors

independently examined the participants in private rooms and completed items 11–14. After comprehensively reviewing all responses provided in the questionnaire, the two TKM physicians indicated whether the diagnosis was LQS or non-LQS (NLQS). A diagnosis of NLQS involves five of the six causes of female infertility except LQS, namely, kidney deficiency, blood stasis, phlegm-dampness, qi and blood deficiency, and damp heat.

In addition, the total score for items 1–10 (participant score) and that for items 11–14 (expert score) were separately recorded. Then, the total score for items 1–14 (total score) was computed.

2.4.2. Demographic and clinical characteristics

Participants' age, specific diagnosis of infertility, childbirth history, and miscarriage history were examined using the questionnaire.

2.5. Sample size

The effective sample size to ensure representativity of the sample was 42–70, which was approximately three- to five-times the number of questionnaire items [21]. We recruited 62 participants, and 61 were enrolled. Fifty-one participants completed the QLQSI, but the diagnosis determined by two TKM physicians was identical for only 42 of them; therefore, these 42 participants were included in the final analysis. However, during the diagnostic agreement analysis, 51 participants who completed the questionnaire were analyzed (Fig. 1).

2.6. Statistical analysis

Collected data are presented as mean ± standard deviation or number (%). To examine demographic and clinical characteristics, continuous variables were analyzed using the independent *t*-test and categorical variables were analyzed using Fisher's exact test. To analyze the reliability of the QLQSI, internal consistency was assessed using Cronbach's α. For validity testing, the McNemar test and kappa statistic were used to assess the agreement regarding the diagnosis between the two experts on LQS patterns. Construct validity was tested using the principal component method from factor analysis and varimax rotation. To determine the cutoff score for diagnosing LQS, the Youden index [22] and concordance probability [23] were calculated. All data were analyzed using SAS 9.4 for Windows (SAS Inc., Cary, NC, USA).

3. Results

3.1. Demographic and clinical characteristics

The mean ages of participants in the LQS and NLQS groups were 35.18 ± 3.71 years and 36.08 ± 3.90 years, respectively. Regarding the specific cause of infertility, seven had infertility of unexplained cause (53.85%), five had ovulation-related infertility (38.46%), and one had uterus-related infertility (7.69%) in the LQS group. In the NLQS group, 12 had infertility of unexplained cause (60%), six had ovulation-related infertility (30%), and two had uterus-related infertility (10%). Childbirth history referred to the number of times the woman gave birth after 20 weeks of gestation; these values were 0.15 ± 0.38 in the LQS group and 0.10 ± 0.31 in the NLQS group. Miscarriage history referred to spontaneous termination of pregnancy before 20 weeks of gestation; these values were 0.85 ± 1.34 in the LQS group and 0.65 ± 1.18 in the NLQS group. There were no significant differences between the two groups in terms of all these characteristics. Questions regarding age were completed by 42 participants, and questions regarding the specific infertility diagnosis, childbirth history, and miscarriage history were completed by 33 participants (Table 1).

3.2. Reliability of the QLQSI

To test the internal consistency of the QLQSI, Cronbach's α was

Table 1
Demographic and clinical characteristics.

	LQS		NLQS		p-value
	n	Mean ± SD or n (%)	n	Mean ± SD or n (%)	
Age	17	35.18 ± 3.71	25	36.08 ± 3.90	0.4572
Specific infertility diagnosis	13		20		
Ovulation factor		5(38.5)		6(30.00)	0.3283F
Uterine factor		1(7.69)		2(10.00)	
Unexplained		7(53.85)		12(60.00)	
Childbirth history	13	0.15 ± 0.38	20	0.10 ± 0.31	0.6556
Miscarriage history	13	0.85 ± 1.34	20	0.65 ± 1.18	0.6620

Analyzed by *t*-test, F: Fisher`s Exact test.

computed for 42 participants. Cronbach's α for this questionnaire was 0.858, which is relatively high. The values computed after removing each item ranged from 0.8337 to 0.8774, similar to the total Cronbach's α. The total correlation for all items was higher than 0.3, with the exception of item 10 ("I have irregular menstrual periods") and item 13 ("tongue coating is thin and whitish"). In particular, item 13 had a negative correlation (Table 2).

3.3. Validity of the QLQSI

3.3.1. Diagnostic agreement

The diagnostic agreement between expert 1 and expert 2 was analyzed. For the 51 participants who completed the questionnaire, both experts diagnosed 17 women with LQS and 25 women with NLQS; however, the two experts disagreed about the diagnosis for the remaining nine participants. Therefore, the sensitivity and specificity of the diagnostic agreement between the two experts were 77.27% and 86.21%, respectively. The significance probability of the McNemar test (0.7389) confirmed that there were no significant differences between the two experts, and the kappa statistic (0.6383) confirmed that the two experts showed significant diagnostic agreement (Table 3).

3.3.2. Construct validity

Factor analysis was performed for 14 items of the QLQSI. A principal component analysis was performed, and the number of factors was determined using the eigenvalue. Varimax rotation led to four factors with an eigenvalue of 1 or more. Exploratory factor analysis showed that the sum of eigenvalues of the four factors was 9.9472, and that

Table 2
Internal consistency of the QLQSI items.

	Items-Total correlation	Cronbach`s α if items deleted
Q1. I get annoyed frequently	0.7434	0.8352
Q2. I am a little sensitive	0.4829	0.8504
Q3. I get angry easily	0.7134	0.8370
Q4. I tend to get nervous often	0.6866	0.8386
Q5. I feel like my chest is tightened	0.7680	0.8337
Q6. My flank sides are full	0.6037	0.8434
Q7. I have breast pain before my menstrual period	0.4177	0.8540
Q8. I have indigestion when I am bothered about something	0.5586	0.8461
Q9. I sigh often	0.5813	0.8447
Q10. I have irregular menstrual periods	0.1486	0.8685
Q11. The pulse is wiry or string-like	0.6642	0.8399
Q12. Tip of the tongue is red	0.4291	0.8534
Q13. Tongue coating is thin and whitish	−0.0256	0.8774
Q14. Tongue body is red	0.4031	0.8548
Cronbach`s α = 0.858 (n = 42)		

Table 3
Diagnostic concordance among experts regarding the QLQSI.

		Expert 2 diagnosis			p-value
	n(%)	NLQS	LQS	Sum(%)	
Expert 1 diagnosis	NLQS	25(86.21)	5(22.73)	30 (58.82)	0.7389
	LQS	4(13.79)	17(77.27)	21 (41.18)	
	Sum(%)	29 (56.86)	22 (43.14)	51 (100)	
Kappa value = 0.6383 95% Confidence Interval = (0.4244, 0.8522)					

P-value was analyzed by McNemar test.

these values explained 99.47% of the total variance. The rotated component matrix was examined to select items with factor loading of 0.5 or more. Factor loading values for factors 1 and 2 were more than 0.5 for item 11 ("the pulse is wiry or string-like"), but it was classified as 1 because the loading was slightly greater. The common factor analysis showed that all items had a value of 0.5 or greater, so no item was removed. However, Cronbach's α for factor 4 had a negative value, as did the factor loading for item 13 ("tongue coating is thin and whitish"); therefore, construct validity should be tested again after removing or modifying item 13 (Table 4).

Factor 1 consisted of eight items: item 3 ("I get angry easily"), item 1 ("I get annoyed frequently"), item 5 ("I feel like my chest is tightened"), item 2 ("I am a little sensitive"), item 4 ("I tend to get nervous often"), item 9 ("I sigh often"), item 11 ("the pulse is wiry or string-like"), and item 8 ("I have indigestion when I am bothered about something"); therefore, factor 1 was referred to as the "stress response factor." Factor 2 consisted of two items: item 12 ("tip of the tongue is red") and item 14 ("tongue body is red"); therefore, factor 2 was referred to as the "tongue factor." Factor 3 consisted of two items: item 7 ("I have breast pain before my menstrual period") and item 6 ("my flank sides are full"); therefore, factor 3 was referred to as the "physical symptom factor." Factor 4 consisted of item 10 ("I have irregular menstrual periods") and item 13 ("tongue coating is thin and whitish"); therefore, factor 4 was referred to as the "other factor" (Table 4).

3.3.3. Cutoff score for diagnosing LQS

To calculate the optimal cutoff score for diagnosing LQS, the 20th percentiles for participant scores, expert scores, and total scores were computed, resulting in a participant score of 23, an expert score of 11, and a total score of 37 (Table 5). We computed cutoff scores for QLQSI with a sensitivity of at least 80%.

Table 5
Cutoff scores of the questionnaire scores for a sensitivity of at least 80%.

Expert diagnosis	n	Variable	Mean ± SD	20th percentile
LQS	17	Participant score	29.18 ± 7.8	23
		Expert score	12.74 ± 1.71	11.00
		Total score	42.41 ± 8.18	37.00

Table 6
The Youden index and concordance probability for optimal cutoff values of the questionnaire scores.

	n	Variables	Cutoff value	Sensitivity(%)	Specificity(%)	YI	CP
LQS	17		22	82.35	68.00	0.504	0.560
		Participant score	**23**	**82.35**	**72.00**	**0.544**	**0.593**
			24	64.71	76.00	0.407	0.492
			10	94.12	56.00	0.501	0.527
		Expert score	11	82.35	72.00	0.544	0.593
			12	**76.47**	**88.00**	**0.645**	**0.673**
			36	82.35	80.00	0.624	0.659
		Total score	**37**	**82.35**	**92.00**	**0.744**	**0.758**
			38	64.71	92.00	0.567	0.595

YI (Youden Index) = sensitivity + specificity–1, CP (Concordance Probability) = sensitivity × specificity.

To confirm the goodness of fit for the cutoff scores, the Youden index and concordance probability were calculated for the varying scores. Their corresponding optimal cutoff scores were a participant score of 23, an expert score of 12, and a total score of 37. The total score of 37 showed higher values than the participant score of 23 and expert score of 12 in the Youden index and concordance probability (Table 6). Therefore, it was most appropriate to set the cutoff score as a total score of 37 and to diagnose LQS when the total score was 37 or more.

4. Discussion

Reliability refers to the variance of repeated measures obtained under the same conditions, and it encompasses the consistency, stability, predictability, and accuracy of a measure. Reliability is tested using the test–retest method, split-half method, multiple forms technique, and internal consistency [24]. Cronbach's α is used to evaluate internal

Table 4
Factor analysis to test the construct validity of the QLQSI.

Questionnaire items	Total (n = 42)				Common factor analysis
	Factor1	Factor2	Factor3	Factor4	
Q3. I get angry easily	**0.8434**	0.0326	0.1622	0.0767	0.7446
Q1. I get annoyed frequently	**0.8035**	0.2179	0.1595	−0.0664	0.7260
Q5. I feel like my chest is tightened	**0.7979**	0.2184	0.2089	0.2106	0.7723
Q2. I am a little sensitive	**0.7967**	0.0357	−0.2322	−0.0189	0.6903
Q4. I tend to get nervous often	**0.7199**	0.1725	0.2039	−0.0228	0.5901
Q9. I sigh often	**0.6464**	0.0236	0.2931	−0.2549	0.5693
Q11. The pulse is wiry or string-like	**0.5490**	0.5431	0.1840	0.0306	0.6311
Q8. I have indigestion when I am bothered about something	**0.5321**	0.2792	0.2374	−0.0266	0.4182
Q12. Tip of the tongue is red	0.1081	**0.9331**	0.1482	0.0530	0.9072
Q14. Tongue body is red	0.1965	**0.8753**	−0.0292	−0.2477	0.8669
Q7. I have breast pain before my menstrual period	0.1198	0.1714	**0.8479**	−0.0695	0.7675
Q6. My flank sides are full	0.4784	0.0398	**0.6687**	0.1363	0.6961
Q10. I have irregular menstrual periods	0.0677	−0.2114	0.2293	**0.8552**	0.7376
Q13. Tongue coating is thin and whitish	0.0768	−0.0851	0.4647	**−0.7131**	0.8332
Eigenvalue	4.4529	2.2204	1.8244	1.4495	
% Variance	44.53	22.20	18.24	14.50	
% Accumulated Variance	44.53	66.73	84.97	99.47	
Cronbach`s α	0.892	0.883	0.694	−0.933	

Factor extraction method: Principal axis analysis, Rotation method: Varimax, Shaded section indicates factor loading > 0.500.

consistency and is a statistic that measures the degree of agreement among responses to questionnaire items designed for the same purpose. Internal consistency is considered high when the pattern of change is nearly identical, and a value of 0.7 or more is generally considered to satisfy reliability [25]. Cronbach's α for the QLQSI was 0.858, which is relatively high. In general, Cronbach's α and the test–retest method are used when testing the reliability of the TKM pattern identification questionnaire [10,26]. In this study, we could not use reliability tests other than Cronbach's α; therefore, additional studies are needed to test the stability and reliability using the test–retest method.

The kappa coefficient, which was found to be 0.6383, was used to assess diagnostic agreement between the two experts. The kappa coefficient is a statistic used to understand the inter-rater agreement for categorical items. A kappa statistic between 0 and 0.2 indicates slight agreement, a value between 0.21 and 0.4 indicates fair agreement, a value between 0.41 and 0.6 indicates moderate agreement, and a value between 0.61 and 0.80 indicates substantial agreement [27]. In the present study, there was substantial diagnostic agreement between the two experts using the pattern identification questionnaire. Because the experts referred to the questionnaire items and made a diagnosis, the objectivity and reproducibility of the diagnosis could be improved [28].

Validity refers to the degree to which a test appropriately measures the variable of interest. It is generally classified as content validity, criterion validity, and construct validity. In a previous study involving the development of a QLQSI, content validity of the questionnaire items was established based on the results of surveys by Delphi experts. Criterion validity is a measure of comparing the tool of interest with an existing validity-established instrument [29]. In the present study, we could not find a valid instrument that could be used as a reference to compare with our QLQSI; therefore, we used construct validity instead. Factor analysis, which is most frequently used to test construct validity, clusters highly correlated items into a single factor and maintains independence among factors [30]. To identify the factors, we used principal component analysis, which is a technique used to identify a smaller number of factors while minimizing data loss. Among factor rotation techniques, varimax rotation is useful when evaluation of the features of each factor rather than the variance structure of each variable is desirable; therefore, we chose varimax rotation in this study. When the eigenvalue (sum of the squares of factor loadings) is more than 1, the variables are clustered into a factor because the explanatory power of the factor is greater than that of each variable [31]. In our study, four factors ("stress response factor," "tongue factor," "physical symptoms factor," and "other factor") were found to have an eigenvalue more than 1, which explained 99.47% of the variance.

The reliability test indicated that the total correlation coefficients of the items were less than 0.3 for item 10 ("I have irregular menstrual periods") and item 13 ("tongue coating is thin and whitish"), suggesting that these two items had lower agreement than other items in the questionnaire. Factor analysis also showed that the reliability coefficient for factor 4, which included items 10 and 13, was negative, and that the factor loading of item 13 was negative.

Item 13 ("tongue coating is thin and whitish") had a negative correlation, suggesting that this item was in opposition to the diagnosis of LQS. In other words, a higher score for item 13 was associated with a lower chance of the existence of LQS. The phrasing "tongue coating is thin and whitish" implies both thin and whitish; this needs to be analyzed to determine whether a higher score for this phrasing using a 6-point scale signified an increase in both "thin" and "whitish." In particular, the wording "thin" connotes a decrease in thickness; therefore, it is possible that an increased score for the item can reverse the meaning. Therefore, the rating scale or item phrasing should be modified.

In addition, item 10 ("I have irregular menstrual periods") had a low correlation; therefore, it requires modification. For this item, it was difficult for participants to assess themselves objectively; therefore, we asked the participants to write the dates of the first day of their last six menstrual periods. A normal menstrual cycle is defined as ranging from 21 to 35 days; the periods that exceeded this range by 7 days or more and the number of periods with more than a 7-day difference in starting dates were considered abnormal. A score of 0 was given for no such irregularities, and a score of 5 was given for five such irregularities. However, because the item had low correlation in our study, it is necessary to clarify the meaning of "irregular menstrual periods" by specifically classifying "irregular periods" as "delayed menstrual period," "early menstrual period," and "irregular menstruation."

In general, the optimal cutoff score has a sensitivity of approximately 80% [32,33]; therefore, we computed the participant, expert, and total scores of the 20th percentile for the LQS group. Additionally, we computed the Youden index and concordance probability of each cutoff score to analyze their goodness of fit.

The Youden index and concordance probability methods are often used to obtain the optimal cutoff value. The Youden index [22] was first introduced by Youden and consists of a function of sensitivity and specificity. The score with the greatest value among the values obtained by subtracting 1 from the sum of the sensitivity and specificity for all values of the data is set as the optimal cutoff value. That is, a score with the greatest sum of sensitivity and specificity is used as the optimal cutoff score. The concordance probability method [23] was first introduced by Liu. The optimal cutoff score is defined as the score that maximized the product of sensitivity and specificity. This method has a value between 0 and 1.

The results showed that a participant score of 23 and a total score of 37 were cutoff values of satisfying 80% sensitivity, the Youden index, and concordance probability in common. However, cutoff value of the expert score was 11 points based on 80% sensitivity and 12 points based on the Youden index and concordance probability. When the total score was 37, the Youden index and concordance probability scores were higher than the participant score of 23 and expert score of 12. Therefore, determining LQS based on a total score of 37 was the most appropriate diagnostic method, with a sensitivity of 82.35% and specificity of 92.00%.

One of the limitations of this study was that there were many dropouts during the study. Of 61 enrolled participants, 10 participants were withdrawn due to omissions on the questionnaire. Most of these omissions were in the expert section due to a scheduling conflict for the two experts. Therefore, the final sample size was smaller than the initial sample size. Furthermore, of 42 patients who received the same diagnosis by both experts, nine participants did not complete the demographic information items, resulting in missing answers regarding the specific infertility diagnosis, childbirth history, and miscarriage history.

We analyzed the reliability and validity of the QLQSI and identified the cutoff score for diagnosing LQS. Subsequent studies should be performed to attempt to modify the items based on the issues found in this study. Additionally, follow-up regarding the questionnaire should be performed with a larger sample so that it could be used in clinical practice and for infertility research as a useful tool for objectively appraising TKM diagnoses and treatments.

5. Conclusion

We tested the reliability and validity of the QLQSI for women with infertility. Cronbach's α for the entire questionnaire was 0.858, indicating relatively high reliability. Validity testing showed that the two experts exhibited significant diagnostic agreement, with a kappa statistic of 0.6383. Construct validity was tested using factor analysis; our four factors explained 99.47% of the total variance. The cutoff score for LQS diagnosis was a total score of 37, which had sensitivity and specificity of 82.35% and 92.00%, respectively. The QLQSI may be an appropriate tool for identifying and assessing LQS with statistical reliability and validity. Potential problems determined by our research should be studied further to confirm the statistical significance of our results.

Author's contribution

All research was performed by the authors.

Funding statement

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Appendix A. Pattern identification questionnaire for liver qi stagnation infertility (QLQSI)

	Items	0	1	2	3	4	5
		Not at all (Severity) Extremely severe					
1	I get annoyed frequently.	0	1	2	3	4	5
2	I am a little sensitive.	0	1	2	3	4	5
3	I get angry easily.	0	1	2	3	4	5
4	I tend to get nervous often.	0	1	2	3	4	5
5	I feel like my chest is tightened.	0	1	2	3	4	5
6	My flank sides are full.	0	1	2	3	4	5
7	I have breast pain before my menstrual period.	0	1	2	3	4	5
8	I have indigestion when I am bothered about something.	0	1	2	3	4	5
9	I sigh often.	0	1	2	3	4	5
10	I have irregular menstrual periods.	0	1	2	3	4	5
	* Please write down the dates of the first day of your last 6 menstrual periods.						
	MM/DD/YYYY						
	MM/DD/YYYY						
	MM/DD/YYYY						
	MM/DD/YYYY						
	MM/DD/YYYY						
	MM/DD/YYYY						
Participant score:							
The following items will be rated by an expert.							
11	The pulse is wiry or string-like.	0	1	2	3	4	5
12	Tip of the tongue is red.	0	1	2	3	4	5
13	Tongue coating is thin and whitish.	0	1	2	3	4	5
14	Tongue body is red.	0	1	2	3	4	5
Expert's diagnosis (liver qi stagnation):							
Expert score:		Total score:					

References

[1] L. Aghajanova, J. Hoffman, E.M. Lin, C.N. Herndon, Obstetrics and gynecology residency and fertility needs: national survey results, Reprod. Sci. 24 (2017) 428–434.

[2] N.M. Hwang, The status of infertility and policy direction in Korea, Health Welf. Policy Forum 82 (2003) 88–97.

[3] R.S. Makar, T.L. Toth, The evaluation of infertility, Am. J. Clin. Pathol. 117 (suppl_1) (2002) S95–S103.

[4] K. Ried, K. Stuart, Efficacy of traditional Chinese herbal medicine in the management of female infertility: a systematic review, Complement. Ther. Med. 19 (2011) 319–331.

[5] J.Y. Jo, T.H. Kim, M.K. Hyun, H.H. Kim, D.I. Kim, Traditional Korean medicine for female infertility: a review of results from infertility support programs in Korea, Eur. J. Integr. Med. 8 (2016) 847–853.

[6] D.I. Kim, Korean Medicine Clinical Practice Guideline of Infertility, the Association of Korean Medicine, Seoul, 2010, pp. 35–53.

[7] T.Y. Choi, J. Choi, J.A. Lee, J.H. Jun, B.K. Park, M.S. Lee, The quality of clinical practice guidelines in traditional medicine in Korea: appraisal using the AGREE II instrument, Implement. Sci. 10 (2015) 104.

[8] G. Maciocia, The Foundations of Chinese Medicine E-Book: A Comprehensive Text, third ed., Elsevier Health Sciences, China, 2015, pp. 43–89.

[9] WHO, WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region, WHO Regional Office for the Western Pacific, Manila, Philippines, 2007, p. 80.

[10] S.J. Lee, J.B. Park, H.S. Lee, K.H. Kim, Development and validation of yin-deficiency questionnaire, Am. J. Chin. Med. 35 (2007) 11–20.

[11] G. Maciocia, Obstetrics and Gynecology in Chinese Medicine E-Book, second ed., Elsevier Health Sciences, China, 2011, pp. 683–695.

[12] C. Kovalik, Regulating Menstruation and Optimizing Female Fertility With TCM, Naturopathic Doctor News & Review, (2006) https://ndnr.com/womens-health/regulating-menstruation-and-optimizing-female-fertility-with-tcm/.

[13] H.J. Choi, T.W. Chung, M.J. Park, K.S. Lee, Y.J. Yoon, H.S. Kim, J.H. Lee, S.M. Kwon, S.O. Lee, K.J. Kim, Paeonia lactiflora enhances the adhesion of trophoblast to the endometrium via induction of leukemia inhibitory factor expression, PLoS One 11 (2016) e0148232.

[14] T.A. Maughan, X.P. Zhai, The acupuncture treatment of female infertility-with particular reference to egg quality and endometrial receptiveness, J.C.M. (98) (2012) 5–13.

[15] J. Yang, H. Huang, L. Zhu, Introduction to Formulae of Traditional Chinese Medicine, World Scientific Publishing Co. Pte. Ltd., Singapore, 2013, pp. 25–28.

[16] G.G. Zhang, B. Bausell, L. Lao, B. Handwerger, B.M. Berman, Assessing the consistency of traditional Chinese medical diagnosis: an intergrative approach, Altern. Ther. Health Med. 9 (2003) 66–71.

[17] C.J. Hogeboom, K.J. Sherman, D.C. Cherkin, Variation in diagnosis and treatment of chronic low back pain by traditional Chinese medicine acupuncturists, Complement. Ther. Med. 9 (2001) 154–166.

[18] B.K. Kang, T.Y. Park, J.A. Lee, T.W. Moon, M.M. Ko, J. Choi, M.S. Lee, Reliability and validity of the Korean standard pattern identification for stroke (K-SPI-Stroke) questionnaire, BMC Complement. Altern. Med. 12 (2012) 55.

[19] J.H. Lee, J.W. Park, S.J. Ko, J.S. Kim, Development and validation of a new pattern identification scale for stomach qi deficiency, Eur. J. Integr. Med. 17 (2018) 56–63.

[20] J.Y. Lee, S.E. Baek, E.J. Park, S.Y. Ahn, D.H. Lee, K.T. Ha, J.E. Yoo, A study on the development of the questionnaire for pattern identification of liver-qi stagnation infertility, J. Korean Obstet. Gynecol. 31 (2018) 68–79.

[21] T.A. Cheng, J.T. Wu, M.Y. Chong, P. Williams, Internal consistency and factor structure of the Chinese Health Questionnaire, Acta Psychiatr. Scand. 82 (1990) 304–308.

[22] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1950) 32–35.

[23] X. Liu, Classification accuracy and cut point selection, Stat. Med. 31 (2012) 2676–2686.

[24] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.

[25] L.J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika 16 (1951) 297–334.

[26] H.H. Ryu, H.J. Lee, H.G. Kim, J.Y. Kim, Reliability and validity of a cold–heat pattern questionnaire for Traditional Chinese Medicine, J. Altern. Complement. Med. 16 (2010) 663–667.

[27] M.S. Kramer, A.R. Feinstein, Clinical biostatistics: LIV. The biostatistics of concordance, Clin. Pharmacol. Ther. 29 (1981) 111–123.

[28] S. Mist, C. Ritenbaugh, M. Aickin, Effects of questionnaire-based diagnosis and training on inter-rater reliability among practitioners of traditional Chinese medicine, J. Altern. Complement. Med. 15 (2009) 703–709.

[29] C.L. Kimberlin, A.G. Winterstein, Validity and reliability of measurement instruments used in research, Am. J. Health. Syst. Pharm. 65 (2008) 2276–2284.

[30] B. Thompson, L.G. Daniel, Factor Analytic Evidence for the Construct Validity of Scores: a Historical Overview and Some Guidelines, SAGE Publications, Thousand Oaks, CA, USA, 1996, pp. 197–208.

[31] H. Abdi, Encyclopedia for Research Methods for the Social Sciences, SAGE Publications, Thousand Oaks, CA, USA, 2003, pp. 792–795.

[32] B.K. Kang, T.Y. Park, J.Y. Jung, M.M. Ko, M.S. Lee, J.A. Lee, The optimal cut-off value of blood stasis syndrome score in BSS diagnosis in Korea, Evid. Complement. Altern. Med. 2017 (2017) 8049481.

[33] E. Rojo, O. Pino, G. Guilera, J. Gómez-Benito, S.E. Purdon, B. Crespo-Facorro, M.J. Cuesta, M. Franco, A. Martínez-Arán, N. Segarra, Neurocognitive diagnosis and cut-off scores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S), Schizophr. Res. 116 (2010) 243–251.