Test-retest reliability and discriminant validity for the Brazilian version of "The Interstitial Cystitis Symptom Index and Problem Index" and "Pelvic Pain and Urgency/Frequency (PUF) Patient Symptom Scale" instruments

Marcella Lima Victal[1], Carlos Arturo Levi D'Ancona[2], Roberto Gomes Junqueira[3], Daniel Carlos da Silva[2], Henrique Ceretta Oliveira[1], Maria Helena Baena de Moraes Lopes[1]

[1]School of Nursing, State University of Campinas, Campinas, Brazil; [2]Department of Urology, Clinical Hospital, State University of Campinas, Campinas, Brazil; [3]Departament of Urology, University of the Joinville Area, Joinville, Brazil

Contributions: (I) Conception and design: ML Victal, MH de Moraes Lopes; (II) Administrative support: ML Victal, MH de Moraes Lopes, CA D'Ancona, RG Junqueira; (II) Provision of study materials or patients: ML Victal, CA D'Ancona, RG Junqueira, D Carlos da Silva; (IV) Collection and assembly of data: ML Victal; (V) Data analysis and interpretation: ML Victal, MH de Moraes Lopes, HC Oliveira; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Marcella Lima Victal. School of Nursing, State University of Campinas (UNICAMP), Street Tessália Vieira de Camargo, 126 - Cidade Universitária "Zeferino Vaz" Campinas, SP, Brazil. Email: mavictal@hotmail.com.

Background: The purpose is to evaluate the psychometric properties of reliability and discriminant validity of the Brazilian Portuguese versions of two instruments used in the diagnosis of interstitial cystitis (IC): "The Interstitial Cystitis Symptom Index and Problem Index" (The O'Leary-Sant), and "Pelvic Pain and Urgency/ Frequency (PUF) Patient Symptom Scale".

Methods: Three groups of patients were examined: a study group (subjects with IC), control group 1 (individuals with at least one IC symptom), and control group 2 (subjects without IC symptoms). Test-retest stability was evaluated at intervals of 3 to 7 days in the study group. Discriminant validity was examined in all three groups.

Results: The intraclass correlation coefficient (ICC) [95% confidence interval (CI)] results were 0.56 (range, 0.21-0.78) for The O'Leary-Sant Symptom Index, 0.48 (range, 0.10-0.73) for The O'Leary-Sant Problem Index, and 0.49 (range, 0.12-0.74) for the PUF. To analyze discriminant validity between groups, we used Fisher's exact test and odd ratio (OR) to identify differences. We obtained a P value<0.0001, which indicated that the null hypothesis was rejected; in other words, there was evidence that at least two different groups were compared to the proportion of patients with IC.

Conclusions: The analyzed instruments did not reach appropriate values for reliability. Future studies are needed to analyze the psychometric measures of these instruments on a larger sample of patients with IC.

Keywords: Diagnosis; interstitial cystitis (IC); questionnaires; validation study

Introduction

Interstitial cystitis (IC) is characterized by painful urinary symptoms in the absence of a bacterial infection. It is diagnosed through clinical signs and the exclusion of other diseases (1). Symptoms of IC may include discomfort, pain, or pressure on the bladder extending to the pelvic area, associated with urgency and/or polyuria (2). Some authors include nocturia and pain during sexual intercourse as symptoms (2-8). Because of the difficulty of diagnosing IC,

its exact prevalence remains unclear. In the United States, the reported prevalence ranges from 10 to 67 in 100,000 inhabitants (4). To our knowledge, no prevalence studies have been performed in Brazil.

In 1987, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) developed criteria for the diagnosis of IC. These criteria were designed to identify a homogenous subgroup of patients for epidemiological research and treatment protocols. The criteria required patients to have agglomerations and/or Hunner's ulcers on the cystoscopic exam, as well as pain in the bladder or urinary urgency (5).

In 1999, the Interstitial Cystitis Database (ICDB) study group, financed by the National Institutes of Health (NIH), studied whether the NIDDK criteria were required to diagnose IC. The only difference between the ICDB and the NIDDK criteria was that the former did not necessarily require cystoscopic results for diagnosis. Even with rigorous application of the NIDDK criteria, two-thirds of patients who had strong indicators for IC by the ICDB criteria would still be excluded. According to the latter, IC may be characterized by painful vascular symptoms in the absence of infections or other identifiable clinical conditions (2).

Diagnosing IC is a long and complex process that begins with a urologic or urogynecologic exam. Symptoms are evaluated from several perspectives, including whether they last for more than 3 months. The next step is to determine whether infections or other diseases are present that might cause the same symptoms. Results of cystoscopy, urodynamic exams, biopsy, and questionnaires support the exclusion other diseases and aid in determining the diagnosis (8-13).

Questionnaires can be used to investigate the urinary, emotional, physical, and sexual aspects of the disease, as well as the patient's menstrual cycle and quality of life, leading to a precise diagnosis (9-13). However, no questionnaire related to the diagnosis of IC is available in Brazilian Portuguese.

We translated "The Interstitial Cystitis Symptom Index and Problem Index" (The O'Leary-Sant) and the "Pelvic Pain and Urgency/Frequency (PUF) Patient Symptom Scale" instruments into Brazilian Portuguese and adapted them to Brazilian society. In this process, we followed all of the steps for cultural adaptation (i.e., translation, synthesis of the translations, back-translation, review by a committee of specialists, and pre-test) developed by the American Academy of Orthopedic Surgeons.

We adapted the content, obtaining Portuguese versions that were faithful to the original English versions (14). To use the instruments in research and clinical practice,

we must first evaluate their psychometric properties to determine their validity and reliability. Thus, the goal of this study was to determine the psychometric properties of the Brazilian versions of these instruments, to judge their test-retest reliability (stability) and their discriminant (divergent) validity.

Methods

Authorization and ethical considerations

This was a methodological study that aimed to evaluate the reliability and discriminant validity of "The O'Leary-Sant" and "PUF" instruments.

To perform the study, we made prior contact with the authors of the instruments and obtained formal authorization to translate and perform the cultural adaptation of the instruments. We observed all of the ethical principles involved in research with human subjects, with a positive report from the Research Ethics Committee of the Faculty of Medical Sciences at UNICAMP (case No. 545/2010). All participants read and signed an informed consent form.

Instruments for data collection

Demographic information
To define the population profile of the sample studied, we collected sociodemographic information, including age, income (value in Brazilian reais, R$), work activity, educational level ("no education" to "graduate studies"), and the results of previous exams performed to determine a diagnosis of IC.

Verification list
This instrument was used to evaluate exclusion criteria for IC (15).

The O'Leary-Sant
The goal of this instrument is to evaluate and diagnose patients with IC. The O'Leary-Sant instrument is comprised of a Symptom Index (score range: 0-20 points) and a Problem Index (score range: 0-16 points), each of which contains four questions related to urinary and pain symptoms.

For each index, the score is calculated by summing the points for each item. On either index, a score ≥ 6 points indicates IC.

The Symptom Index covers various areas, including: whether the patient feels the need to urinate with little or no warning, has to urinate more frequently than every

2 hours, needs to get up during the night to urinate, and has pain in the bladder. The Problem Index evaluates other aspects, such as: urinary frequency during the day, urinary frequency at night, the need to urinate with little or no warning, and burning, pain, discomfort, or pressure on the bladder. Both indices evaluate the situation over the previous month.

Pelvic pain and urgency/frequency (PUF)

This instrument is another method for diagnosing IC. It consists of eight items that cover areas of pain, urgency, urinary frequency, and symptoms associated with sexual intercourse. Items 1, 2a, 4a, 5, 6, 7a, and 8a measure symptoms of IC. These items are related to urinary frequency during the day and night, as well as symptoms of pain during sexual intercourse or associated with the bladder or pelvis.

Items 2b, 4b, 7b, and 8b relate to discomfort caused by IC. These items ask about the discomfort of nocturia and pain, as well as how often urinary urgency and dyspareunia negatively affect the respondent's life. The score ranges from 0 to 35 points. A score ≥5 points is considered to indicate IC.

Study groups

Three groups of patients participated in the study. The study group consisted of 30 patients who had a diagnosis of IC confirmed by clinical signs and biopsy. A verification list containing the exclusion criteria for IC was applied. Although it was not possible to evaluate these criteria, the clinical symptoms together with a positive biopsy exam for chronic or unspecific cystitis were sufficient to include the person in this group. Control group 1 consisted of 29 patients who had at least one symptom suggestive of IC (pelvic pain, urgency, polyuria, or nocturia). Control group 2 consisted of 14 patients who did not have any symptom suggestive of IC. Patients for the study group came from a private urology clinic located in Joinville, Santa Catarina, and from private clinics and a public hospital in the city of Campinas, São Paulo. Patients for control groups 1 and 2 came from a public hospital in the city of Campinas, São Paulo, and from a specialized medical walk-in clinic in the city of Limeira, São Paulo.

Reliability

The reliability of a research instrument is defined as the degree to which the instrument produces the same results in repeated measurements. It concerns coherence, precision, stability, equivalence, and homogeneity. A reliable measure will produce the same results if the behavior is measured again using the same scale (16). The coefficient of reliability ranges from 0 to 1, expressing the relationship between error variance, true variance, and the score observed. The closer the coefficient is to 1, the more reliable the instrument is (16).

In this study, we used the test-retest reliability, which involves the administration of the same instrument to the same research subjects under similar conditions on two or more occasions. To evaluate this measurement, we used the patients from the study group. The questionnaires were applied at two different times at an interval of 3 to 7 days, before any factor (e.g., treatment) was applied that could influence the responses.

Validity

When a measurement instrument exactly measures what it should measure (i.e., truly reflects the concept that it is supposed to measure), then it is considered valid (16). This study evaluated the discriminant validity, also called the divergent validity, in which measurement approaches are used to differentiate one construct from others that may be similar to it (16). To evaluate this validity, we used the study group, control group 1, and control group 2.

Treatment and data analysis

The intraclass correlation coefficient (ICC) was determined with the Statistical Package for the Social Sciences (SPSS) program. An ICC ≥0.70 is considered adequate for indicating an instrument's reliability (16). Discriminant validity was evaluated by using the Chi-square test, as measured by the SAS 9.2 program.

Results

Analyses of test-retest reliability and discriminant validity were performed on the basis of the instruments' calculated scores. Patients with a score ≥6 points on either index of The O'Leary-Sant instrument or a score ≥5 points on the PUF instrument were considered to have a diagnosis of IC. In total, 73 patients (67 women, 6 men) participated in the study, with a mean overall age of 48 years. The mean age in each group was 45.2±11.9 years in the study group,

Table 1 ICCs for The O'Leary-Sant Index and the PUF scale

Questionnaire	ICC	95% CI	P value
The O'Leary-Sant Symptom Index	0.56	0.21-0.78	0.002
The O'Leary-Sant Problem Index	0.48	0.10-0.73	0.007
PUF	0.49	0.12-0.74	0.006

ICC, intraclass correlation coefficient; PUF, pelvic pain and urgency/frequency; 95% CI, 95% confidence interval.

Table 2 ICC by city for the tested instruments

Variable	City	ICC	95% CI
The O'Leary-Sant Symptom Index	Joinville	0.68	0.44-0.93
	Campinas	0.20	−0.34-0.76
The O'Leary-Sant Problem Index	Joinville	0.64	0.36-0.91
	Campinas	0.09	−0.47-0.66
PUF	Joinville	0.49	0.14-0.85
	Campinas	0.26	−0.27-0.80

ICC, intraclass correlation coefficient; 95% CI, 95% confidence interval; PUF, pelvic pain and urgency/frequency.

50.9±12.6 years in control group 1, and 48.2±14.6 years in control group 2 (P=0.180 by Kruskal-Wallis test).

The average monthly income for the three groups was R$1,908.18. The average monthly income was R$1,059.03±797.59 for control group 1 and R$1,528.57±1,078.77 for control group 2 (P=0.001 by Kruskal-Wallis test). The income in the study group was significantly higher than that of the control groups, but there was no difference between the control groups.

There were no differences in terms of educational levels between the groups (P=0.1111 by Fisher's exact test). In the study group, we observed a higher proportion of people with a middle school (36.67%) and high school education (33.33%); in the control groups 1 and 2, there was a greater concentration of elementary (57.14% and 35.71%, respectively) and middle school education (25% and 50%, respectively).

Of the 30 patients who participated in the test-retest, only 24 returned for the second application of the instrument. The main reason that subjects did not come back for the retest was that they could not miss work. The ICC was used to measure the test-retest reliability of The O'Leary-Sant and PUF instruments (*Table 1*).

The ICC for test-retest agreement for the O'Leary-Sant and PUF instruments was compared between the cities of Joinville and Campinas (*Table 2*).

The Chi-square test was used to evaluate the discriminant validity between the three groups (*Tables 3-5*).

We analyzed the validity of both instruments, and obtained a P value of <0.0001. Using a 5% significance value, we rejected the null hypothesis. Thus, the results indicated that there was evidence that at least two groups differed from each other with respect to the proportion of cases with IC.

Discussion

Differences in income and educational levels between the study group and the control groups might have occurred because the study group came from an area of Brazil with a higher per capita income than the other areas. The lower educational levels of patients in control groups 1 and 2 might have compromised the ability to obtain correct responses to the questions on the instruments, which, in turn, may have influenced the scores obtained.

In this study, The O'Leary Symptom Index and Problem Index were analyzed separately. The ICC of 0.56 for the Symptom Index and 0.48 for the Problem Index between the test and retest scores did not reach a value of 0.70, indicating insufficient reliability (15). We also calculated the ICC by city, because the interval of time between the test and retest differed between the cities of Joinville (3 days) and Campinas (7 days). This fact might have affected the agreement between the tests. The low values for the ICC in the city of Campinas for both instruments showed that the agreement between the test and retest was lower than

598

Victal et al. Instruments for diagnosis of interstitial cystitis

Table 3 Discriminant validity for The O'Leary-Sant Symptom Index between groups

Group	No IC, n (%)	IC, n (%)	OR (95% CI)
Study group	3 (10.0)	27 (90.0)	2.864 (0.662-12.392)
Control group 1	7 (24.1)	22 (75.9)	1.00
Control group 2	20 (80.0)	5 (20.0)	0.080 (0.020-0.291)

$P < 0.0001$ by Fisher's exact test. IC, interstitial cystitis; OR, odd ratio; 95% CI, 95% confidence interval.

Table 4 Discriminant validity for The O'Leary Problem Index between groups

Group	No IC, n (%)	IC, n (%)	OR (95% CI)
Study group	5 (16.7)	25 (83.3)	3.529 (1.051-11.855)
Control group 1	12 (41.4)	17 (58.6)	1.00
Control group 2	22 (88.0)	3 (12.0)	0.096 (0.023-0.396)

$P < 0.0001$ by Fisher's exact test. IC, interstitial cystitis; OR, odd ratio; 95% CI, 95% confidence interval.

Table 5 Discriminant validity for the PUF scale between groups

Group	No IC, n (%)	IC, n (%)
Study group	0 (0.0)	30 (100.0)
Control group 1	1 (3.5)	28 (96.5)
Control group 2	13 (52.0)	12 (48.0)

$P < 0.0001$ by Fisher's exact test. PUF, pelvic pain and urgency/frequency; IC, interstitial cystitis.

in the city of Joinville. This difference in the test and retest scores may stem from the fact that IC is an unstable disease, with continuously changing symptoms, which could have influenced the patients' responses.

Nevertheless, previous studies have found different results. In the study for the development and validation of The O'Leary-Sant instrument (11), the test-retest reliability, as measured by the ICC, was higher than 0.90 for both indices, in a total of 45 patients. Subsequently, another study, evaluated only the Symptom Index in 67 people with IC. The test-retest reliability, as measured by the ICC, was 0.80, indicating an excellent level of reproducibility (12). We obtained a value of 0.49 for the ICC of the PUF instrument, which is below the value of 0.70 proposed by another author (17). The lack of other studies of the test-retest reliability of this instrument makes it impossible to compare these findings.

The analysis of discriminant validity of the Symptom and Problem Index of The O'Leary-Sant instrument identified that the study group and control group 1 were very different from control group 2 in terms of the proportion of cases classified as IC; however, when we compared between them, this difference did not prove relevant. This result may stem from the fact that there were many people with urinary problems and symptoms characteristic of IC.

However, because of the difficulty in diagnosing IC, these people might not have been treated properly. During the interview for inclusion in control group 2, we asked about symptoms of IC, so that only patients who denied having these symptoms were included.

As shown in *Table 4*, the PUF instrument confirmed the presence of IC in 100% of patients classified with IC in the study group. However, in control group 1, the PUF instrument was not capable of discriminating between patients who only have some symptoms of the disease, but do not have a diagnosis of IC. A study concluded that The O'Leary-Sant and PUF instruments are important ways to help diagnose IC, but neither alone is sufficient to guarantee the diagnosis (13). Moreover, a study stated that the PUF is not a rubric for diagnosis, but that it is essential to measure the disease's evolution and, consequently, develop proper treatment (18).

The fact that the ICC values in both instruments did not reach a sufficient value for reliability probably reflects low case numbers. Other factors that might have interfered are inadequate understanding of the instruments or a change in the patient's clinical picture. These factors reinforce the importance of performing further validation studies on a larger sample of patients with IC to confirm or negate these findings.

Conclusions

Analyzing the test-retest stability for The O'Leary-Sant and PUF instruments revealed below adequate levels for reliability. Analysis of the discriminant validity showed that there were differences in both instruments with respect to the proportion of subjects classified as having IC.

Acknowledgements

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The study was approved by Research Ethics Committee of the Faculty of Medical Sciences at UNICAMP (case No. 545/2010). Written informed consent was obtained from the patients for publication of this article. A copy of the written consent is available for review by the editor-in-chief of this journal.

References

1. Kim SH, Oh SJ. Comparison of voiding questionnaires between female interstitial cystitis and female idiopathic overactive bladder. Int Neurourol J 2010;14:86-92.
2. Nickel JC. Diagnosis of interstitial cystitis: another look. Rev Urol 2000;2:167.
3. National Kidney and Urologic Diseases Information Clearinghouse. Interstitial Cystitis/Painful Bladder Syndrome. Available online: http://www.niddk.nih.gov/health-information/health-topics/urologic-disease/interstitial-cystitis-painful-bladder-syndrome/Pages/facts.aspx
4. Warren JW, Brown J, Tracy JK, et al. Evidence-based criteria for pain of interstitial cystitis/painful bladder syndrome in women. Urology 2008;71:444-8.
5. Parsons CL, Dell J, Stanford EJ, et al. Increased prevalence of interstitial cystitis: previously unrecognized urologic and gynecologic cases identified using a new symptom questionnaire and intravesical potassium sensitivity. Urology 2002;60:573-8.
6. Clemons JL, Arya LA, Myers DL. Diagnosing interstitial cystitis in women with chronic pelvic pain. Obstet Gynecol 2002;100:337-41.
7. Macdiarmid SA, Sand PK. Diagnosis of interstitial cystitis/painful bladder syndrome in patients with overactive bladder symptoms. Rev Urol 2007;9:9-16.
8. Meijlink JM. Interstitial Cystitis/Bladder Pain Syndrome: Diagnosis & Treatment. Cited Mar 12, 2012. Available online: http://www.painfulbladder.org/pdf/Diagnosis&Treatment_IPBF.pdf
9. Evans RJ, Sant GR. Current diagnosis of interstitial cystitis: an evolving paradigm. Urology 2007;69:64-72.
10. Sirinian E, Azevedo K, Payne CK. Correlation between 2 interstitial cystitis symptom instruments. J Urol 2005;173:835-40.
11. O'Leary MP, Sant GR, Fowler FJ Jr, et al. The interstitial cystitis symptom index and problem index. Urology 1997;49:58-63.
12. Lubeck DP, Whitmore K, Sant GR, et al. Psychometric validation of the O'leary-Sant interstitial cystitis symptom index in a clinical trial of pentosan polysulfate sodium. Urology 2001;57:62-6.
13. Kushner L, Moldwin RM. Efficiency of questionnaires used to screen for interstitial cystitis. J Urol 2006;176:587-92.
14. Victal ML, Lopes MH, D'Ancona CA. Adaptation of the O'Leary-Sant and the PUF for the diagnosis of interstitial cystitis for the Brazilian culture. Rev Esc Enferm USP 2013;47:312-9.
15. Rebola J, Coelho MF. Cistite Intersticial: Etiopatogenia e atitudes terapêuticas. Acta Urológica 2003;20:19-24.
16. LoBiondo-Wood G, Haber J. Pesquisa em enfermagem. Métodos, avaliação crítica e utilização. 4th ed. Rio de Janeiro: Guanabara Koogan, 2001:187-98.
17. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 3rd ed. New York: Oxford University Press, 2003:64-5.
18. Brewer ME, White WM, Klein FA, et al. Validity of Pelvic Pain, Urgency, and Frequency questionnaire in patients with interstitial cystitis/painful bladder syndrome. Urology 2007;70:646-9.