

Construct Validity of the Incontinence Severity Index

Miles Murphy,* Patrick J. Culligan, Cristina M. Arce, Carol A. Graham,
Linda Blackwell, and Michael H. Heit

University of Louisville HSC, Louisville, Kentucky

Aims: To assess the construct validity of the incontinence severity index (ISI) by testing its correlation with two health-related quality of life measures, the short forms of the Urogenital Distress Inventory (UDI-6) and the Incontinence Impact Questionnaire (IIQ-7), in women with urodynamic stress incontinence. **Materials and Methods:** A cohort of 170 women with the urodynamic stress incontinence who underwent corrective surgery completed the ISI, IIQ-7, and the UDI-6 both pre- and post-treatment. We correlated the pre- and post-treatment responses between the ISI, the IIQ-7, the UDI-6 and their subscales. We also assessed the sensitivity of the ISI to change by correlating the percent change in score between the three instruments. The results were analyzed using a non-parametric test of correlation, the Spearman's rho. **Results:** The ISI scores were generally not well correlated with the pre-treatment IIQ-7 and UDI-6 scores ($r < 0.40$). The post-treatment scores and percent change from pre- to post-treatment of the ISI, however, were highly correlated with that of the IIQ-7 and UDI-6 ($r^2 > 0.70$, $P < 0.001$). The ISI was most highly correlated with the UDI-6 stress symptoms subscale and most poorly correlated with the UDI-6 obstructive/discomfort subscale. **Conclusions:** This study provides valuable insight into the construct validity of the ISI. Evidence of its convergent validity is found in the high correlation with the stress symptoms subscales of the UDI-6, while the poorer correlation with the obstructive subscale provides evidence of its divergent validity. These data also suggest that the ISI is highly sensitive to change seen with treatment. *Neurourol. Urodynam.* 25:418–423, 2006. © 2006 Wiley-Liss, Inc.

Key words: construct validity; incontinence severity; questionnaire

INTRODUCTION

While urinary incontinence is clearly a very common problem in women, this condition suffers from a lack of advanced epidemiologic analyses [Hunskar et al., 2003]. Therefore, in the past few years, a great deal of attention has been paid to studying the epidemiology of incontinence and other lower urinary tract symptoms (LUTS). A severity measure that has been recommended by the World Health Organization (WHO) for use in epidemiologic studies [Brown et al., 2003] is the incontinence severity index (ISI) created by Sandvik et al. [2000]. This index was used in an epidemiologic study of nearly 28,000 women in a county in Norway from 1995 to 1997 [Hannestad et al., 2000]. It is derived from the product of the answers to two questions. The first question assesses the frequency of incontinence on a scale of 1–4 (0–4 in follow-up studies). The second question assesses the amount of leakage on a scale of 1–3. The first description of this instrument in the United States was published by Sandvik et al. [2000]. Until that time most LUTS questionnaires used in the United States focused on health-related quality of life.

Severity of incontinence has been defined in various ways in the literature. Traditional urodynamic testing with measures of leak point pressures and/or urethral closure pressures and direct measures of urine loss such as pad tests allow “objective” measurement of continuous variables related to

incontinence severity. Some studies have shown that the subjective bother associated with incontinence correlates with these objective measures of incontinence severity [Shumaker et al., 1994; Stach-Lempinen et al., 2004]. Some studies concentrate on physician assessment of incontinence severity, while others look more at the patient's “subjective” report. One study seeking to understand the relationship between physician and patient assessment of incontinence severity, found that agreement between these two groups decreases as the severity of incontinence increased [Melville et al., 2003].

The ISI has been shown to correlate well with pad weighing tests in incontinent women [Sandvik et al., 2000]. This correlation represents evidence for the criterion validity of the ISI. However, incontinence-related quality of life may be associated with more than a simple index of frequency and amount of urinary leakage. Incontinence is a chronic condition, and the impact this condition has on a woman's life is valuable

No conflict of interest reported by the author(s).

*Correspondence to: Miles Murphy, MD, MSPH, The Institute for Female Pelvic Medicine, 1010 Horsham Road, Suite 19454, North Wales, PA 19454.

E-mail: milesmurphy@comcast.net

Received 13 September 2005; Accepted 13 March 2006

Published online 1 May 2006 in Wiley InterScience

(www.interscience.wiley.com)

DOI 10.1002/nau.20246

information. The bother from these symptoms and how they impact a patient's psycho-social health is an important component of the severity of their incontinence. To date, there is no evidence of the ability of the ISI to measure incontinence-related quality of life. Therefore, we wish to quantify the construct validity of the ISI by correlating it with the IIQ-7, the UDI-6, and its subscales.

MATERIALS AND METHODS

Population

We identified a cohort of 170 women with urodynamic stress incontinence that underwent TVT as their sole operative procedure at our institution from April 1999 to January 2002. Two fellowship-trained attending urogynecologists (PJC, MHH) performed all the cases. Placement of the TVT was performed as previously described [Ulmsten et al., 1996]. The only exceptions to this were cases in which regional or general anesthesia was used. Written or verbal informed consent was obtained from each subject, and the University of Louisville Human Studies Committee approved the study protocol. Methods, definitions, and units conform to the standards recommended by the International Continence Society, except where specified [Abrams et al., 2002].

Data Collection

A database was developed to collect pre-operative data on these 170 subjects as part of a previous investigation [Murphy et al., 2003]. These data included a general medical history, a detailed urogynecologic history, prior surgical history, pelvic organ prolapse quantification [Bump et al., 1996], and multi-channel urodynamic testing. For the current study, we added to this database the pre-operative responses to three validated questionnaires.

We describe the first questionnaire, the ISI, in the Introduction. This questionnaire is composed of two questions. (1) How often do you experience urinary leakage? (2) How much urine do you lose each time? The two other questionnaires, the Urogenital Distress Inventory (UDI) and the Incontinence Impact Questionnaire (IIQ), were two of the first condition-specific quality of life measures focusing on LUTS to be validated [Shumaker et al., 1994]. The construct validity of these instruments was investigated by comparing their scores with those of other quality of life instruments such as the RAND 36-Item Health Survey, the Centers for Epidemiologic Studies—Depression Scale, and the Medical Outcomes Study measure of Social support 1992. The creators of these instruments also assessed the criterion validity (how an instrument compares to a “gold standard” measure [McDowell and Newell, 1996]) of them by correlating them with attending physician diagnosis and pad-weight tests.

The IIQ can be divided into four subscales: travel, social, emotional, and physical activity. The UDI can be divided into

three subscales: obstructive/discomfort, irritative symptoms, and stress symptoms. Subset regression analysis was then used to create short forms of these instruments, the IIQ-7 and UDI-6, which were found to maintain high psychometric quality [Uebersax et al., 1995]. We used these short forms in our study. The score limits of these instruments and their subscales are 0–100. The score limits of the ISI are 0–12. Our practice routinely mails these questionnaires to patients prior to their first office visit, and these data were collected retrospectively, thus explaining the lower response rate in the pre- versus post-operative data pool.

We then prospectively measured subjective surgical outcomes after tension-free vaginal taping by mailing these same three questionnaires (the ISI, IIQ-7, and UDI-6) to our subjects. If we received no response, an investigator (CMA) attempted to contact the subjects by phone. Subject data were only used if the subject completed both instruments under question for each individual analysis. Likewise, only subjects who completed both pre-and post-operative questionnaires were included in the analysis of percent change in score. Percent change in score $([\text{pre-operative score} - \text{post-operative score}] \div \text{pre-operative score})$ was used as our unit of change because it controls for pre-operative differences in baseline incontinence severity.

Statistical Analysis

Univariate analysis comparing responders (those who completed the post-operative ISI questionnaire) to non-responders was conducted using the Pearson chi-square statistic for categorical data and the independent-samples *t*-test for continuous data. Histograms were created to determine the normality of the questionnaire response data. Pre-operative, post-operative, and percent change in score values all had non-normal distribution across the three instruments as expected in this “diseased” population. Therefore, Spearman's rho, a non-parametric test of correlation, was used to compare the instrument responses. Statistical analysis was performed using SPSS 10.0 for Windows (SPSS Inc., Chicago, IL).

RESULTS

Recruitment

Three members of our cohort suffered non-surgically related deaths by the time of follow-up. We contacted 150 (89.8%) of the remaining 167 living subjects. Of these, 137 (82.0%) consented to be in the study. The median follow-up period was 32 months (18–51).

Sample Characteristics

The demographics of the study population can be found in Table I.

TABLE I. Demographic Analysis Comparing ISI Responders and Non-Responders

	Responders (n = 135)	Non-Responders (n = 35)	P
Age	58.4 (11.3)	52.8 (11.8)	0.01
BMI	27.8 (5.1)	31.0 (6.1)	0.002
Parity	2.5 (1.6)	2.6 (1.3)	0.77
Follow-up interval (months)	31.2 (8.4)	33.8 (10.3)	0.12
Race			
White	130 (96.3)	33 (94.3)	0.65
Non-white	5 (3.7)	2 (5.7)	
Tobacco use	15 (11.7)	11 (32.4)	0.004
Prior surgery for:			
Prolapse	49 (36.6)	7 (20.6)	0.08
Incontinence	45 (33.3)	3 (8.6)	0.004
Hysterectomy	86 (63.7)	20 (58.8)	0.60
Pre-operative POP-Q stage			
Zero	64 (48.1)	9 (26.5)	0.07
One	52 (39.1)	18 (52.9)	
Two	17 (12.8)	7 (20.6)	
Pre-operative urodynamics			
Intrinsic sphincter deficiency	34 (25.8)	0 (35.7)	0.99
Detrusor overactivity	38 (28.6)	13 (38.2)	0.28

Data presented as mean ($\pm$ standard deviation) for continuous data and number of positive responses (percentage) for categorical data. BMI = body mass index. POP-Q = pelvic organ prolapse quantification.

Intrinsic sphincter deficiency defined as a leak point pressure <60 cm H₂O or a maximal urethral closure pressure ≤ 20 cm H₂O.

Subjects who completed the post-operative ISI questionnaire (responders, $n = 135$) were comparable to non-responders ($n = 35$) in most regards. However, non-responders were on average younger, had a greater BMI, were more likely to use tobacco, and were less likely to have undergone previous surgery for incontinence. Histograms of the pre-operative data (Fig. 1) reveal a skewed distribution of the ISI scores, while the IIQ-7 and UDI-6 demonstrate a more normal distribution.

Outcomes

Mean pre-operative and post-operative scores for the three instruments are as follows: IIQ-7 scores went from 47.9 to 17.8, UDI-6 from 59.8 to 26.5, and ISI from 9.8 to 3.7. When defining "failure" as a post-operative ISI $\geq$ pre-operative ISI, "cure" as a post-operative ISI = 0, and "improved" as a post-operative ISI $<$ pre-operative ISI; the "failure" rate was 15.8% and the "cure/improved" rate was 84.2%. The absolute "cure" rate was 46.1%.

Correlation With the IIQ-7

The correlation between the ISI and the IIQ-7 and UDI-6 can be found in Table II. The pre-operative ISI was not significantly correlated with the pre-operative IIQ-7, though the

P value did approach statistical significance ($P = 0.08$). While the Social and Emotional Health subscales reached statistical significance, these coefficients only represented a weak correlation. On the other hand, the post-operative values of the ISI were strongly correlated with the IIQ-7 and its subscales, and all demonstrated statistical significance ($P < 0.001$).

Correlation With the UDI-6

The ISI showed an even greater association with the UDI-6. The ISI was significantly correlated with the pre-operative UDI-6, but as with the pre-operative IIQ-7, the correlation was weak ($r < 0.50$). However, the ISI was highly correlated (>0.70) with the post-operative and percent change values of the UDI-6. This level of correlation persisted in two of the UDI-6 subscales, the irritative and the stress symptoms subscales. The ISI showed a much weaker correlation with the obstructive/discomfort subscale.

Sensitivity to Change

The strongest correlations seen in this study were found when comparing the percent change in test score seen after treatment (Fig 2). The IIQ and UDI have been shown to be highly sensitive to change after intervention [Hagen et al., 2002], and the percent change in ISI score was strongly correlated with the IIQ-7, the UDI-6, and all their subscales. As would be expected in this cohort of women with stress urinary incontinence, the greatest correlation was seen in the percent change in score after surgical treatment between the ISI and the stress symptoms subscale of the UDI-6. This distribution is shown in Figure 1.

DISCUSSION

This study provides valuable insights into the construct validity of the ISI. Urinary incontinence can be objectively measured in a number of ways: standing stress test, pad-weight test, and urodynamic testing. Measuring the effect that the symptoms of incontinence have on patient quality of life, however, is more complex. Tests that attempt to measure this aspect of incontinence should be compared to multiple indicators of validity in a process known as construct validation [McDowell and Newell, 1996]. Construct validity can be divided into two basic types of evidence, convergent, and divergent validity. Convergent validity refers to how well a test correlates with other methods that measure the same concept, while divergent validity is confirmed by a test not correlating with other tests that measure different themes.

Evidence of the convergent validity of the ISI is provided by its strong correlation with the IIQ-7. The IIQ-7 measures the degree to which incontinence impacts a patient's activities, relationships, and feelings. While the ISI does not directly ask questions regarding these aspects of a patient's life, it does correlate strongly with the IIQ-7. It also correlates well with the

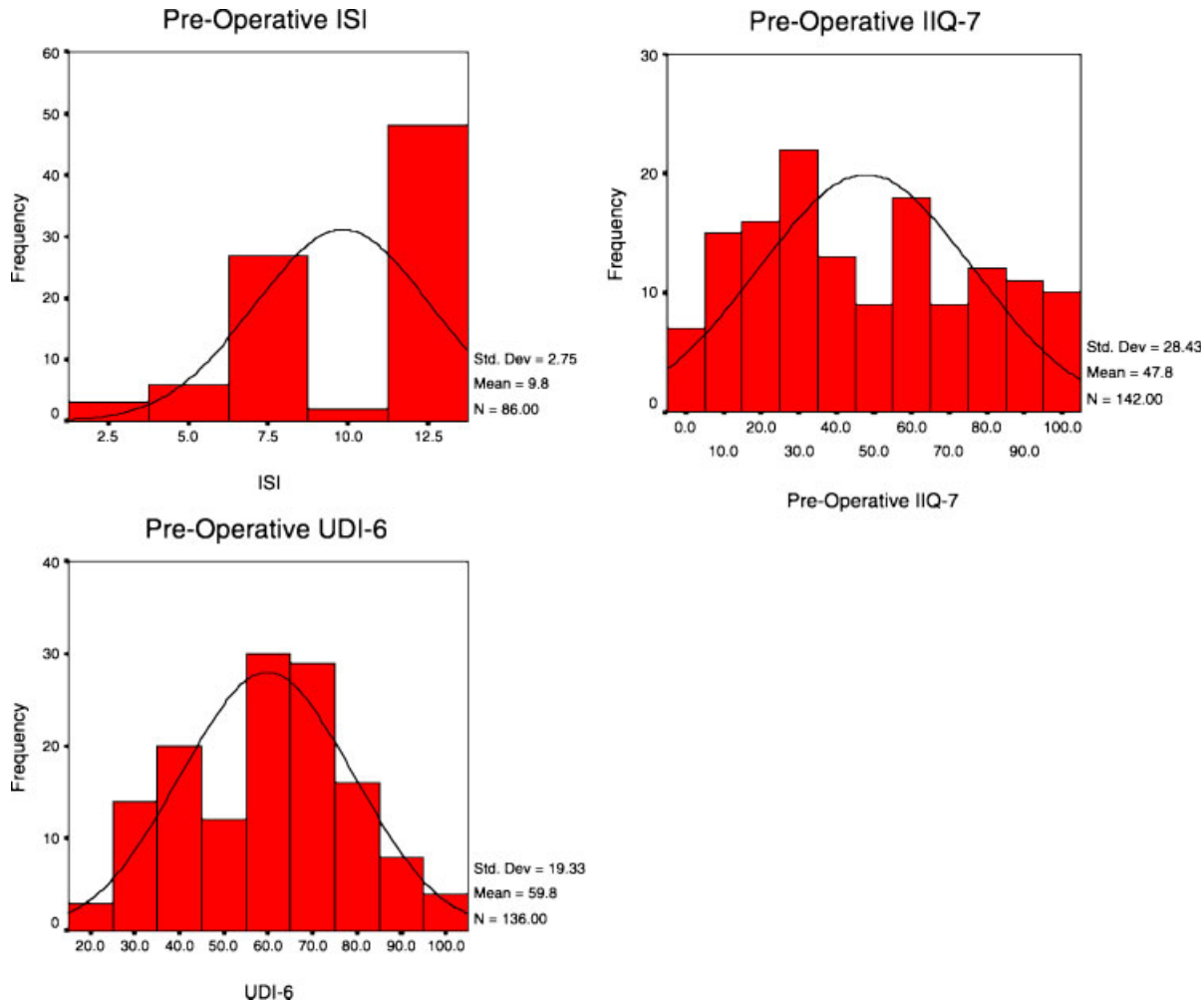


Fig. 1. Histograms showing the distribution of pre-operative scores of the three instruments.

TABLE II. Correlation of ISI With IIQ-7, UDI-6, and Subscales

	Pre-operative score—r (n)	Post-operative score—r (n)	Percent change in score—r (n)
IIQ-7 total	0.27 (79) ^{ns†}	0.70 (135) ^{***}	0.72 (68) ^{***}
Physical activity	0.13 (78) ^{ns}	0.71 (134) ^{***}	0.79 (57) ^{***}
Travel	0.23 (72) ^{ns†}	0.60 (135) ^{***}	0.58 (64) ^{***}
Social/relationships	0.25 (80) [*]	0.61 (135) ^{***}	0.65 (42) ^{***}
Emotional health	0.30 (73) ^{**}	0.64 (135) ^{***}	0.59 (60) ^{***}
UDI-6 total	0.36 (78) ^{***}	0.78 (135) ^{***}	0.76 (69) ^{***}
Irritative symptoms	0.43 (77) ^{***}	0.73 (134) ^{***}	0.68 (65) ^{***}
Obstructive/discomfort	0.14 (77) ^{ns}	0.25 (135) ^{**}	0.46 (34) ^{**}
Stress symptoms	0.23 (72) ^{ns†}	0.84 (135) ^{***}	0.84 (64) ^{***}

r = Spearman's rho correlation coefficient, (n) = number in study population.

ns = non-significant ($P > 0.10$).

ns[†] = non-significant but trending towards significance ($P = 0.05-0.10$).

* $P < 0.05$ (2-tailed).

** $P < 0.01$ (2-tailed).

*** $P < 0.001$ (2-tailed).

Percent change in score = (pre-operative—post-operative) ÷ pre-operative score.

UDI-6, which is a measure of how much a patient is bothered by symptoms of incontinence and/or pelvic organ prolapse.

The subscales of the UDI-6 are particularly useful for assessing the divergent validity of the ISI. Like the ISI, the stress symptoms subscale of the UDI-6 is designed to measure urinary leakage. The UDI-6 obstructive/discomfort subscale, on the other hand is designed to measure difficulty in bladder emptying and pain or discomfort in the pelvic area, symptoms more consistent with prolapse than incontinence per se. Of the three subscales, the obstructive/discomfort subscale is by far the most poorly correlated with the ISI. This discrepancy speaks to the divergent validity of the ISI.

In general, however, these conclusions are not as strongly supported when comparing the pre-operative responses to these instruments. One explanation for this finding is that the IIQ-7 and the UDI-6 measure subjective responses, and the impact and bother that women experience from a given severity of incontinence may vary more than their responses to the absence or resolution of incontinence. This explanation is supported by our finding that the pre-operative distribution of ISI scores is much more skewed than those of the IIQ-7 and

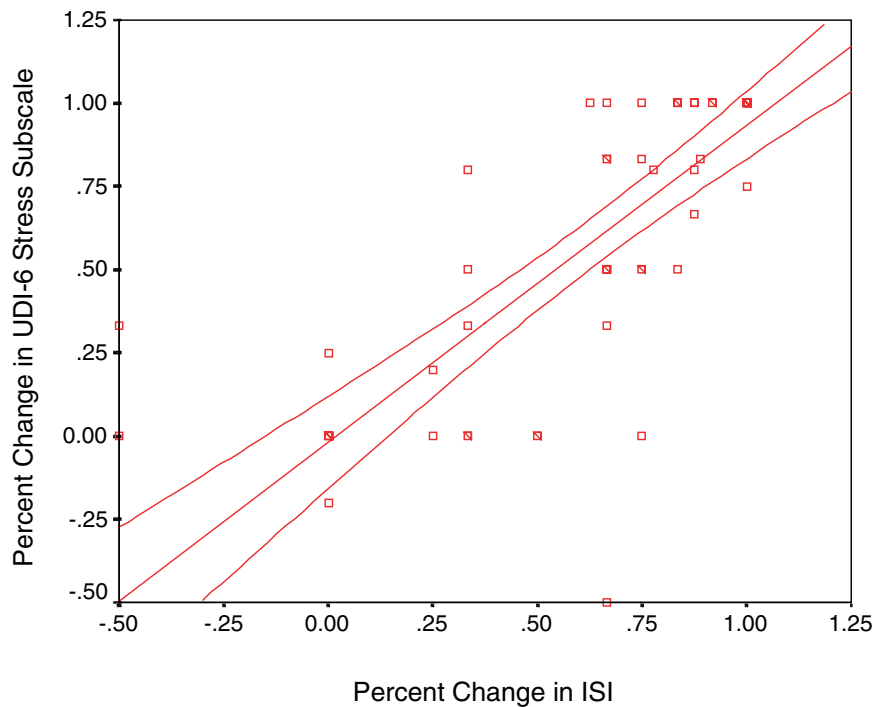


Fig. 2. Correlation between the percent change in score in UDI-6 stress symptoms subscale and the ISI seen after surgical treatment of urodynamic stress incontinence, with mean regression prediction line and 95% confidence intervals.

UDI-6. While the pre-operative ISI scores imply a high degree of incontinence severity one would expect in a population of women planning to undergo surgical correction, the pre-operative scores of the other two instruments reflect the more varied bother and impact on quality of life caused by these symptoms.

The external validity of this investigation is limited by the fact that this population suffered primarily from urodynamic stress incontinence. While about one third of the group also suffered from detrusor overactivity, findings in a population with this as their primary diagnosis might differ. There were also some small demographic differences between those who did and did not respond to the post-operative questionnaires. Non-responders were on average 5 years younger and more likely to be smokers, but less likely to have undergone a previous surgery for incontinence. Given these small statistical differences, we conclude that if any bias exists it is that our results may be applicable to more severe cases of incontinence. Lastly, the percent change in score results is only based on a partial sampling of our total population and should therefore be interpreted with caution.

CONCLUSIONS

This study provides valuable insight into the construct validity of the ISI. Our data only provides information relating to this instrument's correlation with incontinence-related

quality of life measures. While we feel quality of life is an important component of incontinence severity, the ISI was not designed to measure this aspect of incontinence severity. This investigation would have benefited from a comparison between the ISI and other instruments designed to measure the quantity of urine loss in incontinent subjects. Such comparison will certainly provide the validity of the ISI.

REFERENCES

- Abrams P, Cardozo L, Fall M, et al. 2002. The standardization of terminology of lower urinary tract function: Report from the standardization subcommittee of the International Continence Society. *Neurourol Urodyn* 21:167–78.
- Brown JS, Nyberg LM, Kuseck JW, et al. 2003. Proceedings of the National Institute of Diabetes and Kidney Diseases international symposium on epidemiologic issues in urinary incontinence in women. *Am J Obstet Gynecol* 188:S77–88.
- Bump RC, Mattiasson a, Bo K, et al. 1996. The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. *Am J Obstet Gynecol* 175:10–7.
- Hagen S, Hanley J, Capewell A. 2002. Test-retest reliability, validity, and sensitivity to change of the urogenital distress inventory and the incontinence impact questionnaire. *Neurourol Urodyn* 21:534–9.
- Hannestad YS, Rortveit G, Sandvik H, et al. 2000. A community-based epidemiological survey of female urinary incontinence: The Norwegian EPINCONT study. *J Clin Epidemiol* 53:1150–7.
- Hunskaar S, Burgio K, Diokno A, et al. 2003. Epidemiology and natural history of urinary incontinence in women. *Urology* 62:16–23.

- McDowell I, Newell C. 1996. *Measuring Health*, 2nd ed. New York: Oxford University Press. p. 31.
- Melville JL, Miller EA, Fialkow MF, et al. 2003. Relationship between patient report and physician assessment of urinary incontinence severity. *Am J Obstet Gynecol* 189:76–80.
- Murphy M, Heit MH, Fouts L, et al. 2003. Effect of anesthesia on voiding function after tension-free vaginal tape procedure. *Obstet Gynecol* 101:666–70.
- Sandvik H, Seim A, Vanvik A, et al. 2000. A severity index for epidemiological surveys of female urinary incontinence: Comparison with 48-hour pad-weighing tests. *Neurourol Urodyn* 19:137–45.
- Shumaker SA, Wyman JF, Uebersax JS, et al. 1994. Health-related quality of life measures for women with urinary incontinence: The incontinence impact questionnaire and the urogenital distress inventory. *Qual Life Res* 3:291–306.
- Stach-Lempinen B, Kirkinen P, Laippala P, et al. 2004. Do objective urodynamic or clinical findings determine impact of urinary incontinence or its treatment on quality of life? *Urology* 63:67–71.
- Uebersax JS, Wyman JF, Shumaker SA, et al. 1995. Short forms to assess life quality and symptom distress for urinary incontinence in women: The incontinence impact questionnaire and the urogenital distress inventory. *Neurourol Urodyn* 14:131–9.
- Ulmsten U, Hendriksson L, Johnson P, et al. 1996. An ambulatory surgical procedure under local anesthesia for treatment of female urinary incontinence. *Int Urogynecol J* 7:81–6.