

Validation of the Urogenital Distress Inventory (UDI-6) and Incontinence Impact Questionnaire (IIQ-7) in a Dutch Population

Elaine Utomo, Ida J. Korfage, Mark F. Wildhagen,
Anneke B. Steensma, Chris H. Bangma and Bertil F.M. Blok

Neurourol Urodyn. 2015 Jan;34(1):24-31.

ABSTRACT

Aims

The Urogenital Distress Inventory (UDI-6) and Incontinence Impact Questionnaire (IIQ-7) assess symptom distress and the impact on daily life of urinary incontinence. The UDI-6 has not been validated before in males. Our aim was to validate the UDI-6 and IIQ-7 in Dutch men and women.

Methods

The translation to Dutch followed standardized procedures. We validated the IIQ-7 with and without an additional gender-neutral item (IIQ-SF). Adults with urinary incontinence for at least three months, completed the measures at inclusion; one week after inclusion to evaluate the test-retest reproducibility; and six months after inclusion with the addition of the RAND-36 health transition item to assess responsiveness and interpretability. To assess the discriminate ability, a reference population was enrolled. To assess construct validity, the urodynamic diagnosis was used.

Results

Questionnaire data of 160 patients were analysed. Patients reported more symptoms and bother than the reference population ($p < .001$). The internal consistency was good in the IIQ-SF baseline scores (Cronbach's alphas 0.86-0.92); though moderate in the UDI-6 (Cronbach's alphas 0.44-0.66). Both measures showed good reproducibility at the test-retest (Intraclass Correlations Coefficients 0.75-0.85). Construct was adequate with 75% confirmed hypotheses of urodynamic data with measure scores. The measures were responsive after treatment with smaller measurement errors than the minimal important change. No floor or ceiling effects were observed in baseline data.

Conclusion

The Dutch UDI-6 and IIQ-7 are reliable, valid, and responsive instruments for assessing symptom distress of urinary incontinence and its impact on daily life in both men and women.

INTRODUCTION

Urinary incontinence (UI) affects approximately 30% to 60% of women and 1% to 39% of men, depending on age, aetiology, and the definition of incontinence used.¹ Various measures are available to assess the impact of UI on health related quality of life (HRQOL). Generic HRQOL measures lack sensitivity to the unique aspects of a specific disease^{2,3}, therefore disease-specific HRQOL measures can be considered more applicable in capturing the impact of a specific disease such as UI. These measures may also serve to assess the impact of treatment on UI and to facilitate future research in incontinence treatment.

The urogenital distress inventory (UDI) and incontinence impact questionnaire (IIQ) were both developed to assess the impact of UI on HRQOL.^{4,5} Short versions of the UDI and IIQ, UDI-6 and IIQ-7, were developed to reduce the respondents' burden.⁶ The UDI-6 and IIQ-7 are both "A grade" recommended by the Fourth International Consultation on Incontinence.⁷ A measure that is valid and reliable for a particular language and culture may not prove so when used in a different population.⁸ The UDI and IIQ long forms were previously translated and revisited in Dutch⁹, and are nowadays widely used in the Netherlands in urogynecology clinical practice and research.¹⁰⁻¹² An unvalidated Dutch translation of the UDI-6 and IIQ-7 is also commonly used.^{13,14} However, by using a questionnaire without the measurement properties tested¹⁵, it remains unknown whether a reliable, valid, and responsive measure for the population of interest is used. Though not yet in Dutch, the UDI-6 and IIQ-7 have been translated and validated in other languages.¹⁶⁻²² Although both measures⁶ were developed for female patients, we aimed to validate these in both sexes. Unlike the UDI-6, the IIQ-7 has earlier been validated in men.²³ The objective of this study was to validate the UDI-6 and IIQ-7 in Dutch to provide a useful evaluation tool for the use in both men and women suffering from UI.

METHODS

This observational study was conducted at a tertiary Urology and Gynaecology centre as part of a larger validation study of HRQOL pelvic floor measures. The study was approved by the Institutional Ethical Committee (MEC-2008-376).

Questionnaires

The UDI-6 is a six item symptom inventory, specific to symptoms associated with lower urinary tract dysfunction, and combines information on irritative, stress and obstructive/discomfort symptoms.⁶ The IIQ-7 is a seven item life-impact assessment

instrument specific to UI, and covers separate domains of physical activity, travel, social/relationships, and emotional health.⁶ Both measures were developed for self-administration and are intended to be used in combination. Patients rated how much they experience impaired function of UI and the extent to which this UI affects their daily functioning with four response options per item ((0) “not at all”; (1) “slightly”; (2) “moderately”; (3) “greatly”). We maintained the scoring procedure as described by the original developer: if more than two items are missing, a total score is not to be calculated.⁶ The mean score of items is multiplied by 33 1/3 to convert to a 0–100 scale.⁶ Higher scores indicates more symptom distress (UDI-6), or more impact on daily life (IIQ-7).

In consultation with the original developer we concluded that the UDI-6 had good face validity concerning male patients, i.e. the questions seemed plausible, relevant, and to span the domain adequately. In the IIQ-7 however, the item addressing household chores (“Has urine leakage affected your ability to do household chores (cooking, housecleaning, laundry)?”) might be related more to females. Therefore, based on the factor loadings of the IIQ long form⁴ and as suggested by the original developer, we added a question about employment (“Has urine leakage affected your employment (work) outside the home?”). The addition of this “employment” item resulted in three versions of the IIQ-7 completed by both men and women: (1) *IIQ-7 original*⁶; (2) *IIQ-7 adjusted*; in which the item addressing “household chores” was substituted by the item “employment” (“Has urine leakage affected your employment (work) outside the home?”)⁴ (3) *IIQ-8*; in which the item “employment” was added to the original IIQ-7. We refer to these versions as the IIQ-short forms (IIQ-SF). Referring to the IIQ long form⁴, the “employment” item was part of the “travel” domain while the “household chose” item was part of the “physical activity” domain. However, unlike the IIQ long form, the IIQ-7 does not contain separate domain scores,⁶ and therefore only total scores of the IIQ-SF are reported.

Linguistic Validation

The translation of the measures to Dutch followed standardized forward-backward procedures⁸: three independent forward translations, and backward translation by a native speaker. The Dutch versions of the measures were tested in 10 patients with UI, where potential problems were explored and discussed guided by a checklist. UI was defined according to the International Continence Society, thus “the complaint of any involuntary leakage of urine”.²⁴ As a result, the lay-out was adapted to clarify the instructions on how to indicate the chosen answer. Several minor problems were identified which did not indicate a need to adapt the content. The Dutch versions of UDI-6 and IIQ-SF were then finalized (See ‘Vragenlijsten’).

Study population and study design

Inclusion criteria included males and females aged 18 years or older, symptoms of UI for at least three months, and fluent and literate in the Dutch language. Exclusion criteria were symptomatic urinary tract infections, neurologic diseases (except diabetic neuropathy), active malignant tumours, dementia, and mental retardation. At the initial physician office visit, the treating physician explained the study to all consecutive patients potentially eligible for inclusion. The physician logged the name and gender of patients who were interested in participation, and gave an information package including two sets of questionnaires. These patients were phoned by the principal investigator for further explanation, and were instructed to complete the first set of questionnaires (baseline) at home, and to return it by postal mail directly after completion, together with a signed informed consent form. The second set of questionnaires were completed one week after baseline, and returned by postal mail. Six months after baseline, patients received through postal mail a third set of questionnaires to be completed at home and returned by postal mail. Age and education were documented through the questionnaire. Educational level was classified as “low” (primary school), “intermediate” (high school) or “high” (college or university degree). The test-retest period of one week between the repeated administrations was long enough to prevent recall, though short enough to ensure that clinical change had not occurred. During this period no treatment was initiated or changed. Given the observational nature of the study, the treating physician was free to perform any investigation (e.g. urodynamic studies) and prescribe any treatment after completion of the second round of questionnaires, for individual patient care.

The third questionnaire included the health transition item of the RAND 36-Item Health Survey (RAND-36-HTI)²⁵, in which patients were asked to score the change in their general health compared to one year ago. The response options were: (1) ‘much better’; (2) ‘a little better’; (3) ‘same’; (4) ‘a little worse’; (5) ‘much worse’. This RAND-36-HTI was used as the anchor (external criterion)²⁶ for the evaluation of responsiveness and interpretability.

Urodynamic studies

Because of the observational nature of this study, multichannel urodynamic testing was only carried out if indicated according to the standard evaluation protocol of our clinic as recommended by the Fourth International Consultation of Incontinence Recommendations.²⁷ Urodynamic testing was performed according to the International Continence Society standards.²⁸ Using urodynamic studies, patients were categorized into groups: no urodynamic abnormality detected, stress urinary incontinence (SUI), detrusor overactivity (DO), and detrusor underactivity. When mixed SUI/DO incontinence was detected the type of treatment patient received during follow-up determined the classification for analysis.

Reference group

Baseline data from a reference group were collected using a random subsample of an ISO-certified (ISO 26362) Dutch panel²⁹. This subsample was stratified by gender, age, educational level and residential area, and therefore representative for the Dutch population above the age of 18. No medical data was available for this internet panel. A total of 450 panel participants were invited in order to meet our targeted sample size of 250 participants.

Statistical methods

Statistical analysis was conducted using IBM® SPSS-software 20.0. Statistical significance was defined as p -value $< .05$. The mean and standard deviation (SD) are reported for continuous data. For discrete data, the count and percentage are reported.

As described by the original developer⁶, if one or two items were missing in the measures, they were replaced by the average of the respondent's observed values (unconditional mean imputation³⁰). The UDI-6 or IIQ-7 scores were not calculated if more than two items were missing.⁶

To assess the differences between patient and reference group, the Student's T-test was used for continuous data, and the Chi-squared test for categorical variables. General linear models were used to compare measure scores, controlling for variables that differed significantly between the patient and reference group in univariate analysis.

The Welch's F statistic was used as a single global test to compare means between groups, since homogeneity was not assumed. In case of significant differences between means, pairwise comparisons using Games-Howell analysis was performed.

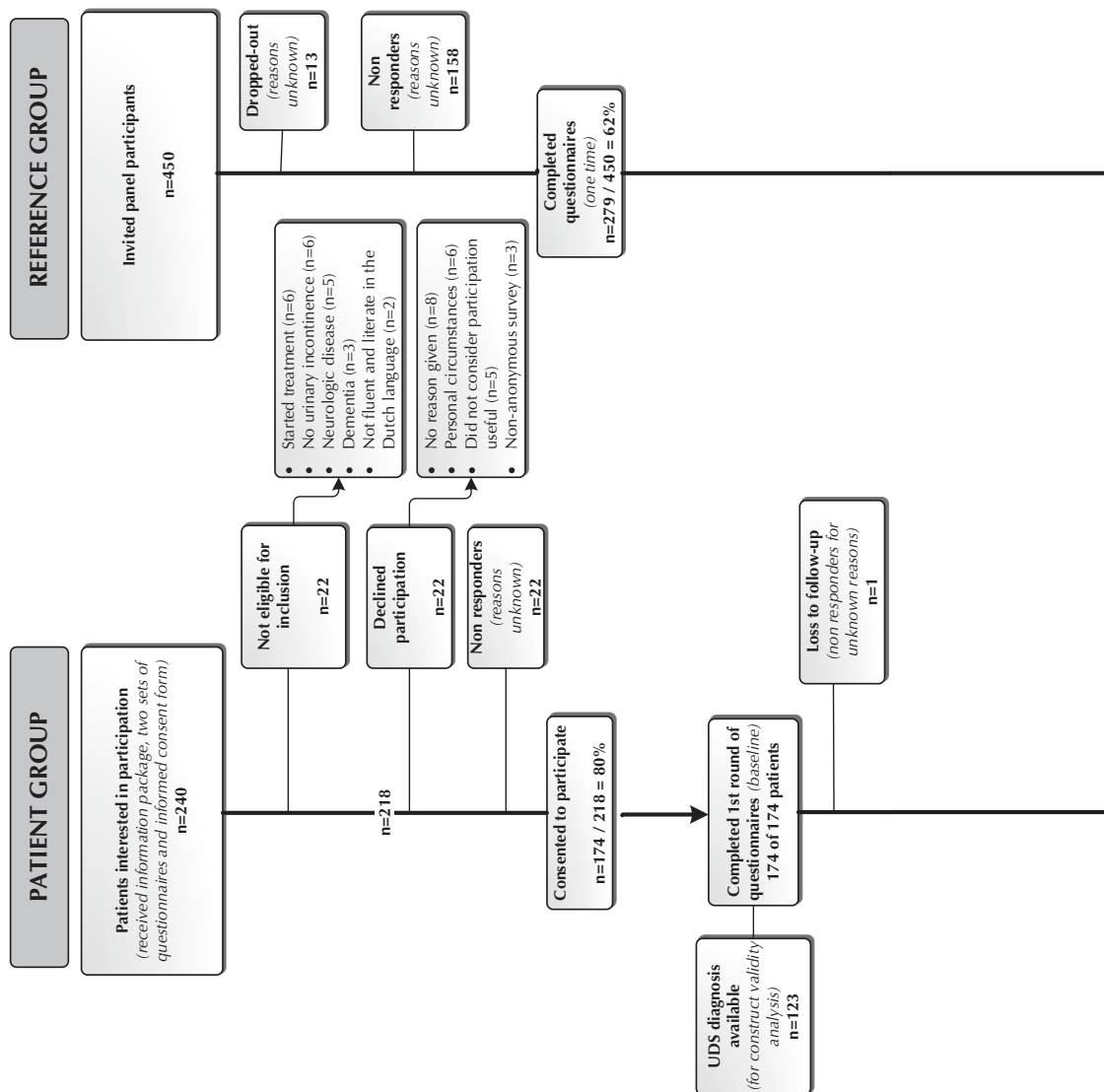
For receiver operating characteristics (ROC) analysis the anchor RAND-36-HTI was dichotomized: patients who reported that they were "a little better" and "much better" were classified as "improved", while "same", "a little worse", and "much worse" were classified as "not improved".

We tested the measurement properties of the measures³¹:

- The **internal consistency**, i.e. the extent to which the items are measuring the same underlying construct, was tested by calculating Cronbach's alpha using baseline scores. A Cronbach's alpha between 0.70 and 0.95 was considered to reflect good internal consistency.³¹
- The **reproducibility**, i.e. the degree to which repeated measurements in the test-retest period provide similar answers, concerns reliability and agreement. The **reliability**, i.e. the extent to which patients can be distinguished from each other despite measurement errors, was calculated with the Intraclass Correlation Coefficient (ICC) according to McGraw and Wong for agreement³², and was considered acceptable when the ICC was ≥ 0.70 .³¹ **Agreement**, i.e. the extent to

which the scores on repeated measures are close to each other, was quantified using the limits of agreement (LOA) as described by Bland and Altman.³³ The LOA were calculated as the absolute mean change in scores ($\text{mean}_{\text{change}}$) of repeated measurements during the test-retest period $\pm 1.96 * \text{SD} (\text{SD}_{\text{change}})$. The LOA estimates where 95% of individual differences fall, that is the absolute measurement error. Agreement is considered good if the LOA are smaller than the minimal important change (MIC; see interpretability).³¹

- The **construct validity**, i.e. the extent to which the scores relate to other measures, was verified by urodynamic diagnosis for discriminative validation. We formulated specific hypotheses and expected at least 75% of the results to be in accordance with these hypotheses³¹:
 1. We expect patients without an abnormality detected on urodynamic studies, to have significantly better IIQ-SF scores than patients with a urodynamic diagnosis;
 2. We expect patients with DO to score significantly worse in the irritative domain of the UDI-6, than patients with other urodynamic diagnoses;
 3. We expect patients with SUI to score significantly worse in the stress domain of the UDI-6, than patients with other urodynamic diagnoses;
 4. We expect patients with dysfunctional voiding or underactive detrusor, to score significantly worse in the obstructive domain of the UDI-6, than patients with other urodynamic diagnoses.
- To assess **responsiveness**, i.e. the ability of a measure to detect clinically important changes over time, in treated patients we firstly evaluated the linear relationship of the $\text{mean}_{\text{change}}$ in measure scores between baseline and 6 month follow-up with the RAND-36-HTI score, using the Pearson correlation coefficient. In addition, the area under the ROC curve (AUC) for the UDI-6 and IIQ-SF measures was determined. The AUC indicates the probability that a measure correctly classified patients as improved, using the RAND-36-HTI as an anchor. The AUC was considered adequate if ≥ 0.70 .¹⁵
- To assess **interpretability**, i.e. the degree to which one can assign qualitative meaning to quantitative scores, the anchor based ROC method was used to assess the MIC.¹⁵ The MIC was defined as the optimal ROC cut-off point, i.e. the value for which the sum of the proportions of misclassifications ($([1-\text{sensitivity}]) + [1-\text{specificity}]$) is smallest.¹⁵
- **Floor and ceiling effects** occur when high proportions of respondents report scores at the lower or upper end of the scale, indicating limited content validity, and thus may lead to a reduced reliability and limited responsiveness.¹⁵ Floor or ceiling effects are considered to be present if more than 15% of respondents achieve the lowest or highest possible score.³¹



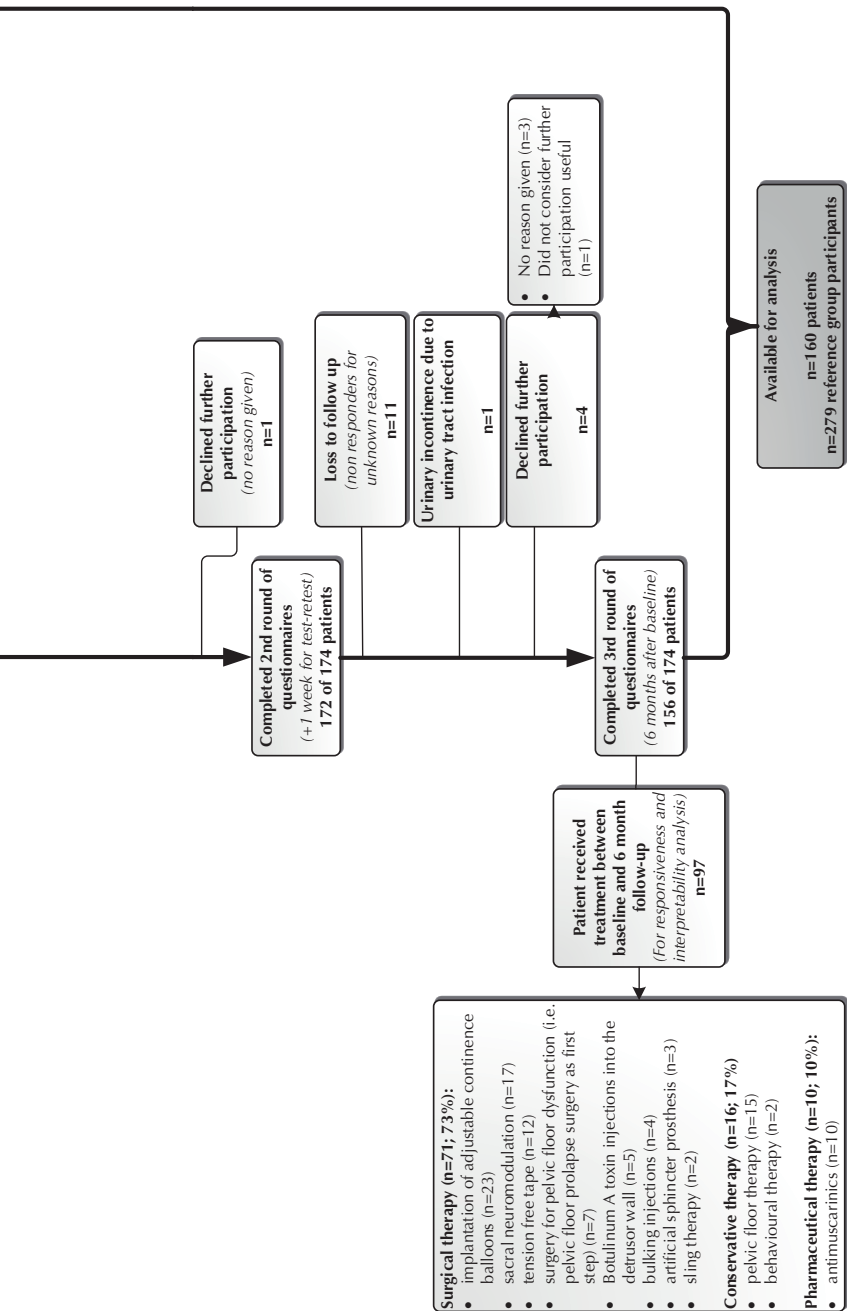


Figure 2.1 Study flowchart
Abbreviations: UDS urodynamic studies

RESULTS

Of 240 male and female consecutive patients, 218 were found to be eligible for inclusion (Figure 2.1). Of the 218 patients, 174 (80%) consented to participate. After three rounds of questionnaires data were available for 160 out of 174 patients (92%) to test at least one measurement property (i.e. patients completed the questionnaires at baseline and at least at one additional time-point) of at least one questionnaire (i.e. patients completed the UDI-6 and/or IIQ-7). Regarding the reference group, the measures were sent out to 450 panel participants of which 279 (62%) responded.

The study population had a mean age of 62±12 years, and 59% were female (Table 2.1). At baseline, patients reported significantly more symptoms of UI (UDI-6, $p<.001$) and more impact of these symptoms on daily life (IIQ-7 original, IIQ-7 adjusted and IIQ-8 all $p<.001$) than the reference group. After adjusting for age and educational level, these differences remained significant ($p<.001$).

Table 2.1 Characteristics of the respondents and questionnaires scores

	Patients (n = 160)	Reference group (n = 279)	p-value univariate	p-value GLM***
Age (years)	62 ± 12	49 ± 16	<.001	
Gender			.08	
Males	65 (41%)	138 (50%)		
Females	95 (59%)	141 (51%)		
Educational level*			<.001	
Low	59 (37%)	21 (8%)		
Intermediate	73 (46%)	172 (62%)		
High	28 (18%)	86 (31%)		
Scores at baseline (0-100)**				
UDI-6	46.4 ± 16.8	12.2 ± 12.7	<.001	<.001
IIQ-7 original	38.4 ± 24.6	4.0 ± 10.9	<.001	<.001
IIQ-7 adjusted	39.4 ± 25.4	4.2 ± 11.2	<.001	<.001
IIQ-8	37.3 ± 24.2	12.2 ± 12.7	<.001	<.001

Data in mean ± standard deviation or number (%).

*missing n=2

**higher scores indicate more symptom distress (UDI-6) or more impact on daily life (IIQ-SF)

***p-value: corrected for age and educational level with general linear model (GLM)

Internal consistency

In the patient group, Cronbach’s alpha was 0.49 for the UDI-6 total score and the domains ranged between 0.46–0.65, indicating moderate internal consistency (Table 2.2). In the reference group, Cronbach’s alpha was 0.66 for the UDI-6 total score, and

the domains ranged between 0.44–0.58, also indicating moderate internal consistency. Internal consistency remained moderate after stratification by gender (not in table).

The Cronbach's alpha of the IIQ-7 original, IIQ-7 adjusted, and IIQ-8 were respectively 0.87, 0.86, and 0.87 in the patient group, indicating good internal consistency. In the reference group, these were 0.91, 0.91, and 0.92 respectively, also indicating good internal consistency.

Table 2.2 Internal consistency and reproducibility

	Patients	Reference group (n = 279)	Patients Test-retest reliability		
	Internal consistency (Cronbach's alpha)		ICC _{agreement}	mean _{change} ± SD _{change}	Limits of agreement*
UDI-6	0.49 (n=149)	0.66	0.84 (n=152)	6.3 ± 7.0	-7.42 to 20.0
Irritative	0.65 (n=155)	0.53	0.83 (n=152)	8.9 ± 13.1	-16.8 to 34.6
Stress	0.46 (n=152)	0.58	0.83 (n=152)	8.4 ± 11.5	-14.1 to 30.9
Obstructive/discomfort	0.54 (n=154)	0.44	0.85 (n=152)	7.3 ± 11.8	-15.8 to 30.4
IIQ-7 original	0.87 (n=147)	0.91	0.76 (n=153)	9.9 ± 14.1	-17.7 to 37.5
IIQ-7 adjusted	0.86 (n=132)	0.91	0.77 (n=150)	10.2 ± 14.1	-17.4 to 37.8
IIQ-8	0.87 (n=131)	0.92	0.75 (n=149)	10.0 ± 14.1	-17.6 to 37.6

All scale scores are transformed to a 0 – 100 point scale.

Higher scores indicate more symptom distress (UDI-6) or more impact on daily life (IIQ-SF).

* Limits of agreement described by Bland and Altman³³ = mean_{change} ± 1.96*SD_{change}

Reproducibility (reliability and agreement)

The retest assessments were completed on average 10±25 days after baseline assessment. The ICC_{agreement} was 0.84 for the UDI-6, and the domains ranged between 0.83–0.85, indicating good reliability (Table 2.2). The ICC_{agreement} of the IIQ-7 original, IIQ-7 adjusted, and IIQ-8 were 0.76, 0.77, and 0.75 respectively, also indicating good reliability.

The mean_{change} and the LOA in the test-retest period are also presented in Table 2.2. The LOA are smaller than the MIC (see also results *Interpretability*), indicating good agreement.

Construct validity

In 124 patients urodynamic data were available and classified according to urodynamic diagnosis. Sixteen of 124 patients were diagnosed with mixed incontinence and were classified into either SUI (n=11) and DO (n=4), while one was treated for mixed incontinence and therefore not included in this analysis. Table 2.3 presents the measure scores in 123 patients stratified into urodynamic diagnosis.

Table 2.3 Construct validity

	No urodynamic abnormality (n = 9)	Detrusor overactivity (n = 34)	Stress Urinary Incontinence (n = 59)	Dysfunctional voiding (n = 9)	Underactive Detrusor (n = 12)	p-value [#]
UDI-6 (total score)	39.5 ± 12.5	48.1 ± 13.1	48.2 ± 18.9	37.6 ± 17.9	43.5 ± 12.0	.22
Irritative	37.0 ± 23.2*	68.6 ± 23.1	51.6 ± 26.0*	37.0 ± 23.2*	55.5 ± 27.8	.003
Stress	50.0 ± 23.5*	50.9 ± 24.2*	66.9 ± 22.8	50.9 ± 24.2	45.8 ± 21.4*	.006
Obstructive/discomfort	31.5 ± 24.2	23.7 ± 22.8	25.0 ± 25.8	31.5 ± 24.2	29.1 ± 33.4	.90
IIQ-7 original	14.8 ± 15.3	41.0 ± 22.4**	38.8 ± 25.0**	24.9 ± 19.3	33.7 ± 17.0	.004
IIQ-7 adjusted	15.3 ± 15.2	41.7 ± 21.9**	39.2 ± 25.8**	25.5 ± 18.9	35.3 ± 16.3	.003
IIQ-8	13.9 ± 13.3	39.9 ± 22.2**	37.5 ± 24.5**	23.0 ± 17.9	33.0 ± 16.0	.001

Measure scores according to urodynamic diagnosis at baseline in mean ± SD

All scores are transformed to a 0 – 100 point scale

Higher scores indicate more symptom distress (UDI-6) or more impact on daily life (IIQ-SF)

[#]Welch's F was used for comparing means

Data **in bold** were used for pair wise comparisons of hypotheses using post hoc Games-Howell with mean difference significant at level * $p < .05$; ** $p < .01$

The distribution of the mean scores of the UDI-6 irritative domain, UDI-6 stress domain, IIQ-7 original, IIQ-7 adjusted and IIQ-8, differed significantly between the urodynamic groups. Pairwise comparisons of predefined hypotheses on outcome resulted in 75% (3 out of 4) confirmed hypotheses indicating acceptable construct validity. Our confirmed hypotheses were:

- The mean score of the irritative domain of the UDI-6 was significantly higher in the DO group than in patients with SUI ($p = .02$), dysfunctional voiding ($p = .04$), or without urodynamic abnormality ($p = .02$).
- The mean score of the stress domain of the UDI-6 was significantly higher in the SUI group than in patients with DO ($p = .02$) or underactive detrusor ($p < .05$).
- The mean scores of the IIQ-7 original, IIQ-7 adjusted and IIQ-8 were significantly lower in the group of patients without urodynamic abnormality than in patients with SUI ($p < .01$) or DO ($p < .01$)

Responsiveness

The treatment work-up was individually determined for each patient by the treating physician after completion of the second round of questionnaires. Thus, 97 patients were treated and completed the third round of questionnaires on average 5.7 ± 1.3 months after baseline assessment. Treatment consisted of surgery ($n = 71$, 73%), conservative therapy ($n = 16$, 17%), or pharmaceuticals ($n = 10$, 10%).

Table 2.4 shows the mean_{change} between baseline and six month follow-up of each measure classified by the responses of the RAND-36-HTI. Since two patients (2%) reported their health was “much worse”, and 10 patients (11%) reported it was “a little worse” compared to one year ago, we combined those categories. The relationships between the mean_{change} of measure scores and the RAND-36-HTI as anchor were significantly linear (Table 2.4). The AUC for the UDI-6 was 0.81 ($p < .001$) indicating good responsiveness. The AUC for the IIQ-7 original was 0.65 ($p = .02$); for the IIQ-7 adjusted 0.70 ($p = .004$); and for the IIQ-8 0.66 ($p = .02$), indicating acceptable responsiveness.

Interpretability

The MIC was -16.7 for the UDI-6 with a sensitivity of 0.76 and specificity of 0.78. The MIC was -19.0 for the IIQ-7 original, -26.2 for the IIQ-7 adjusted, and -28.8 for the IIQ-8, with sensitivity ranging from 0.38–0.43 and specificity ranging from 0.84–0.90 (Table 2.4).

All MIC fell outside the LOA (Table 2.2), indicating acceptable measurement errors which are smaller than the values for MIC.¹⁵

Table 2.4 Responsiveness and interpretability

	Number (%) n = 97	UDI-6 n = 80	IIQ-7 original n = 81	IIQ-7 adjusted n = 80	IIQ-8 n = 80
Health transition item (RAND-36)*:					
A little worse / much worse	12 (12%)	5.8 ± 19.7	-2.8 ± 34.2	0.3 ± 30.5	-1.8 ± 31.8
Same	40 (41%)	-7.8 ± 17.5	-1.8 ± 22.2	-1.7 ± 21.3	-1.1 ± 21.8
A little better	13 (13%)	-19.3 ± 15.3	-18.3 ± 20.6	-20.1 ± 21.6	-18.9 ± 21.0
Much better	17 (18%)	-27.6 ± 14.8	-16.8 ± 28.6	-21.2 ± 28.1	-18.2 ± 27.0
Missing	15 (16%)				
Pearson correlation coefficient <i>r</i>		0.54	0.25	0.34	0.29
<i>p</i> -value		<.001	.03	.002	.01
Area under the ROC curve		0.81	0.65	0.70	0.66
<i>p</i> -value		<.001	.02	.004	.02
Minimal important change		-16.7	-19.0	-26.2	-28.8
Sensitivity; specificity		0.76; 0.78	0.43; 0.84	0.41; 0.88	0.38; 0.90

Data presented in number (%) or mean_{change} ± SD_{change} between baseline and six month follow-up
Negative scores indicate improvement of urinary incontinence symptoms (UDI-6) and impact on daily life (IIQ-SF)

* “Compared to one year ago, how would you rate your health in general now?”

Floor and ceiling effects

We observed no floor or ceiling effects in the baseline data of our patient group. The lowest score was reported by at most 5% of respondents (IIQ-SF), and the highest score by at most 0.6% of respondents (IIQ-7 adjusted).

As expected in the reference group, no ceiling effects were observed, though floor effects were present in the UDI-6 (n=76; 27%); IIQ-7 original (n=212; 76%); IIQ-7 adjusted (n=211; 76%); and IIQ-8 (n=211; 76%).

DISCUSSION

The objective of this study was to validate the UDI-6 and IIQ-7 in Dutch to provide a useful evaluation tool for the use in both men and women suffering from UI. In our population, the UDI-6 and IIQ-SF had good discriminant ability. The reproducibility (reliability and agreement) and construct validity of these measures were good, and both measures proved to be responsive and interpretable. By using the values of the MIC we determined, a decrease of ≥ 16.7 points in the UDI-6 score and a decrease of ≥ 19.0 points in the IIQ-7 original, ≥ 26.2 in the IIQ-7 adjusted, and ≥ 28.8 in the IIQ-8, can be considered as “clinically relevant” in patients undergoing any treatment for UI.

The internal consistency of the IIQ-SF in the patient group was good with Cronbach's alphas of 0.91 and 0.92. However, analysis of the UDI-6 showed moderate internal consistency of the total score and the domains, in both patient and reference group. Other validation studies found comparable Cronbach's alphas for the UDI-6.^{20,21} A possible explanation is the use of unconditional mean imputation³⁰ for measure scores, which is the scoring procedure as described by the developer.⁶ This procedure allows imputing a score of one UDI-6 domain into another UDI-6 domain, e.g. missing items of the irritative domain may be imputed by scores of the stress domain. As a consequence, domain scores might be ‘contaminated’. Another possible explanation of a moderate internal consistency of the UDI-6 was suggested by Franzén et al.²⁰ The UDI-6 contains items assessing symptoms of disease called “causal indicators”. These are called “causal indicators” since the occurrence of these symptoms can cause deterioration on HRQOL, but the impact of these symptoms on HRQOL may vary from patient to patient. Furthermore, the patient's perception of the severity of their symptoms may be influenced by other factors and concomitant symptoms.³⁴ The IIQ-SF contains items measuring effects of UI on HRQOL, which reflect the level of HRQOL, and are regarded as “effect indicators”.³⁴ Scales based on causal indicators are more heterogeneous than scales based on effect indicators. For example, a patient with family problems may perceive UI as having a greater effect upon their HRQOL than usual. Since psychometric analysis is based on the assumption that all items are

effect indicators reflecting the same latent HRQOL construct, the properties of causal items (UDI-6) affect reliability measures such as Cronbach's alpha leading to a lower internal consistency.³⁴

We recognize several limitations to this study. First, we recruited patient data from a tertiary urology and gynaecology centre. In this setting we expect more patients with severe symptoms of UI, and therefore the extent to which the results of this study can be generalized to other settings may be limited. Second, we did not validate the measures separately for men and women. A comparison between the two genders is desirable since health-seeking behaviour and bother from lower urinary tract symptoms (LUTS)³⁵ and UI^{36,37} differs significantly between men and women. Twenty-five percent of women were bothered by the presence of LUTS compared to 18% men, with equal prevalence of treatment seeking of 28% and 24% respectively, however in the absence of evaluation of the effect of incontinence.³⁵ In this hospital-based survey, men were more likely than women to seek treatment as age and severity of LUTS increased.³⁵ When urinary incontinence was taken into account in a community based survey, 45% of women versus 22% of men suffering from incontinence had sought medical care for their problem.³⁶ However in a different community based study, men with incontinence reported a higher rate of healthcare seeking of 56% compared to 46% of women.³⁷ Even though we did not intend to compare sexes and their urological functioning as mirrored by the questionnaires, we can highly recommend this for future research. This study was not designed for this purpose as we would have taken gender distribution into account in our sampling procedure.

The strength of this study is the study design, which enabled us to address almost all quality criteria for measurement properties of Terwee et al.³¹ An exception was the criterion validity, which requires the availability of a gold standard. Since there are no gold standards of symptoms and bother of UI, we were not able to test this property. Our study design also allowed us to test the UDI-6 and IIQ-SF in male patients. As suggested in an earlier study²³, we revised the gender-specific items of the IIQ-7. Hence, we studied the original IIQ-7 along with two variants: the IIQ-7 "adjusted" and the IIQ-8. The psychometric properties of these variants of IIQ-SF were equally good. Therefore we would like to recommend using the "original" IIQ-7 in women as well as in men, to allow better international comparison of research and treatment outcome data.

CONCLUSION

In conclusion, the UDI-6 and IIQ-7 in Dutch are reliable, valid and responsive instruments, and therefore suitable for assessing symptom distress and HRQOL of UI in both men and women, as well as evaluating outcome of treatment.

Acknowledgments

Thanks to John S. Uebersax for giving permission to validate and his advice how to validate these measures in male patients; to Pim J. van Leeuwen for his contribution to the design of this study; to Christa A. Lips and Ben Wasscher for their translations of the questionnaires; to Caroline B. Terwee for her methodology advice; to Gerard J.J.M. Borsboom for his statistical assistance; and especially to all the patients who gave their time and effort to participate.

This study was funded by the Stichting Urologisch Wetenschappelijk Onderzoek (SUWO, Rotterdam, the Netherlands) to make reference data collection possible.

SUWO was not involved in the design, conduct, interpretation, or analysis of the study or review or approval of the manuscript.

REFERENCES

1. Milsom I. Epidemiology of Urinary and Faecal Incontinence and Pelvic Organ Prolapse. In: Abrams P, Cardozo L, Khoury S, Wein A, eds. 4th International Consultation on Incontinence. 4th ed. Paris: Health Publication Ltd; 2009.
2. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27:S217-32.
3. Korfage IJ, Hak T, de Koning HJ, Essink-Bot ML. Patients' perceptions of the side-effects of prostate cancer treatment--a qualitative interview study. *Soc Sci Med* 2006;63:911-9.
4. Shumaker SA, Wyman JF, Uebersax JS, McClish D, Fantl JA. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. *Qual Life Res* 1994;3:291-306.
5. Wyman JF, Harkins SW, Choi SC, Taylor JR, Fantl JA. Psychosocial impact of urinary incontinence in women. *Obstet Gynecol* 1987;70:378-81.
6. Uebersax JS, Wyman JF, Shumaker SA, McClish DK, Fantl JA. Short forms to assess life quality and symptom distress for urinary incontinence in women: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program for Women Research Group. *Neurourol Urodyn* 1995;14:131-9.
7. Staskin D, Kelleher C. Incontinence, Patient-Reported Outcome Assessment. In: Abrams P, Cardozo L, Khoury S, Wein A, eds. 4th International Consultation on Incontinence. Paris 2009:363-412.
8. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417-32.
9. van der Vaart CH, de Leeuw JR, Roovers JP, Heintz AP. Measuring health-related quality of life in women with urogenital dysfunction: the urogenital distress inventory and incontinence impact questionnaire revisited. *Neurourol Urodyn* 2003;22:97-104.
10. van der Vaart CH, de Leeuw JR, Roovers JP, Heintz AP. [The influence of urinary incontinence on quality of life of community-dwelling, 45-70 year old Dutch women] De invloed van urine-incontinentie op de kwaliteit van leven bij thuiswonende Nederlandse vrouwen van 45-70 jaar. *Ned Tijdschr Geneesk* 2000;144:894-7.
11. Labrie J, van der Graaf Y, Buskens E, Tiersma SE, van der Vaart HC. Protocol for Physiotherapy Or TVT Randomised Efficacy Trial (PORTRET): a multicentre randomised controlled trial to assess the cost-effectiveness of the tension free vaginal tape versus pelvic floor muscle training in women with symptomatic moderate to severe stress urinary incontinence. *BMC Womens Health* 2009;9:24.
12. van Leijsen SA, Kluivers KB, Mol BW, et al. Protocol for the value of urodynamics prior to stress incontinence surgery (VUSIS) study: a multicenter randomized controlled trial to assess the cost effectiveness of urodynamics in women with symptoms of stress urinary incontinence in whom surgical treatment is considered. *BMC Womens Health* 2009;9:22.
13. Vervest HA, Bisseling TM, Heintz AP, Schraffordt Koops SE. The prevalence of voiding difficulty after TVT, its impact on quality of life, and related risk factors. *Int Urogynecol J Pelvic Floor Dysfunct* 2007;18:173-82.
14. van Leijsen SA, Kluivers KB, Mol BW, et al. Can preoperative urodynamic investigation be omitted in women with stress urinary incontinence? A non-inferiority randomized controlled trial. *Neurourol Urodyn* 2012;31:1118-23.

15. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. 1st ed. Cambridge: Cambridge University Press; 2011.
16. Altaweel W, Seyam R, Mokhtar A, Kumar P, Hanash K. Arabic validation of the short form of Urogenital Distress Inventory (UDI-6) questionnaire. *Neurourol Urodyn* 2009;28:330-4.
17. Cam C, Sakalli M, Ay P, Cam M, Karateke A. Validation of the short forms of the incontinence impact questionnaire (IIQ-7) and the urogenital distress inventory (UDI-6) in a Turkish population. *Neurourol Urodyn* 2007;26:129-33.
18. Chan SS, Choy KW, Lee BP, et al. Chinese validation of Urogenital Distress Inventory and Incontinence Impact Questionnaire short form. *Int Urogynecol J* 2010;21:807-12.
19. El-Azab AS, Mascha EJ. Arabic validation of the Urogenital Distress Inventory and Adapted Incontinence Impact Questionnaires--short forms. *Neurourology and urodynamics* 2009;28:33-9.
20. Franzen K, Johansson JE, Karlsson J, Nilsson K. Validation of the Swedish version of the incontinence impact questionnaire and the urogenital distress inventory. *Acta Obstet Gynecol Scand* 2013; 92 (5):555-61.
21. Ruiz de Vinaspre Hernandez R, Tomas Aznar C, Rubio Aranda E. [Validation of the Spanish version of the short forms of the Urogenital Distress Inventory (UDI-6) and the Incontinence Impact Questionnaire (IIQ-7) in pregnant women] Validacion de la version espanola de las formas cortas del Urogenital Distress Inventory (UDI-6) y del Incontinence Impact Questionnaire (IIQ-7) en mujeres embarazadas. *Gac Sanit* 2011;25:379-84.
22. Zhu L, Yu SJ, Lang JH, et al. [Validation of incontinence impact questionnaire short form in Chinese population]. *Zhonghua Fu Chan Ke Za Zhi* 2011;46:505-9.
23. Moore KN, Jensen L. Testing of the Incontinence Impact Questionnaire (IIQ-7) with men after radical prostatectomy. *J Wound Ostomy Continence Nurs* 2000;27:304-12.
24. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. *Neurourol Urodyn* 2002;21:167-78.
25. Zee vdK, Sanderma R. Het meten van de algemene gezondheidstoestand met de RAND-36. Een handleiding. Groningen.
26. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395-407.
27. Abrams P, Andersson KE, Birder L, et al. Fourth International Consultation on Incontinence Recommendations of the International Scientific Committee: Evaluation and treatment of urinary incontinence, pelvic organ prolapse, and fecal incontinence. *Neurourol Urodyn* 2010;29:213-40.
28. Schafer W, Abrams P, Liao L, et al. Good urodynamic practices: uroflowmetry, filling cystometry, and pressure-flow studies. *Neurourol Urodyn* 2002;21:261-74.
29. Flycatcher Academic Support Centre. 2012, at <http://www.flycatcher.eu/nld/academic-support/>.
30. Little RJA, Rubin DB. *Statistical analysis with missing data*. 2nd [rev] ed: Hoboken, N.J. : Wiley-Interscience; 2002.
31. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
32. McGraw K, Wong S. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996;1:30-46.

33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.
34. Fayers PM, Hand DJ, Bjordal K, Groenvold M. Causal indicators in quality of life research. *Qual Life Res* 1997;6:393-406.
35. Apostolidis A, Kirana PS, Chiu G, Link C, Tsiouprou M, Hatzichristou D. Gender and age differences in the perception of bother and health care seeking for lower urinary tract symptoms: results from the hospitalised and outpatients' profile and expectations study. *Eur Urol* 2009;56:937-47.
36. Harris SS, Link CL, Tennstedt SL, Kusek JW, McKinlay JB. Care seeking and treatment for urinary incontinence in a diverse population. *J Urol* 2007;177:680-4.
37. Li Y, Cai X, Glance LG, Mukamel DB. Gender differences in healthcare-seeking behavior for urinary incontinence and the impact of socioeconomic status: a study of the Medicare managed care population. *Med Care* 2007;45:1116-22.