Quality of Life Research (2005) 14: 1901–1913
DOI 10.1007/s11136-005-5266-5

Quality of life of women with urinary incontinence: Cross-cultural performance of 15 language versions of the I-QOL

Donald M. Bushnell, Mona L. Martin, Kent H. Summers, Jan Svihra, Christos Lionis & Donald L. Patrick
Health Research Associates, Inc, 6505 216th st. SW Suite 105, Mountlake Terrace, WA 98403, USA (E-mail: bushnell@hrainc.net)

Accepted in revised form 29 March 2005

Abstract

Urinary incontinence (UI) has substantial and important impacts on health-related quality of life. The purpose of this research is to report the psychometric performance of 15 different language versions of the Incontinence-specific Quality of Life (I-QOL) measure, a patient-reported outcome measure specific to stress, urge and mixed urinary incontinence. The multi-national dataset consisted of data from four clinical trials for stress incontinent females and from two additional population studies, enrolling women with stress, urge and mixed UI. All enrolled patients completed the I-QOL and comparative measures at baseline. The clinical trial populations had multiple administrations up to 12 weeks, and the two population studies included a shorter retest. Country-specific psychometric testing for validity, reliability, and responsiveness followed standardized procedures. Confirmatory factor analyses were performed to assess the I-QOL subscales. The I-QOL measurement model was confirmed as three subscales. Summary and subscale scores for the 15 versions were internally consistent (alpha values = 0.91–0.96) and reproducible (ICC = 0.72–0.97). Using changes in the independent measures of incontinence episode frequency standardized response means were predominantly strong (ranged 0.71–1.05) across 13 versions (out of 15) in association with these measures and effect sizes. These additional language versions of the I-QOL instrument demonstrate psychometric properties similar to the original version. The I-QOL has shown good results in both community studies and clinical trials with varying types and severity of urinary incontinence. It is a reliable and valid measure of HRQOL, suitable for use in a variety of international settings.

Key words: Quality of life, Reliability, Responsiveness, Patient reported outcomes, Stress urinary incontinence, Urinary incontinence, Validity

Introduction

In the United States, urinary incontinence (UI) has been defined as 'involuntary loss of urine sufficient to be a problem' [1]. More recently, the International Continence Society (ICS) has defined UI as 'the complaint of any involuntary leakage of urine' [2]. In either case, the definition indicates urinary incontinence as being more than just a physiologic condition, but one that should be further described by specifying relevant factors such as type, frequency, severity, precipitating factors, social impact, effect on hygiene, and impacts on one's quality of life including physical, emotional, and social well-being [2]. While prevalence rates for UI vary due to differences in definitions, study characteristics, and target populations [3–6], it is estimated that 8–58% of the general population of adult women (3–11% of adult men) have symptoms of incontinence [6–9].

Conservative estimates show that UI of all types disrupts the lives of approximately 10–20% of women in the general population [10, 11].

Strategies for assessing outcomes of urinary incontinence and its treatment should incorporate the patient's perspective alongside the more proximal measures of bladder function [12]. Patient-reported outcomes (PROs) (including symptoms, functional status, and perceived quality of life) are increasingly used alongside clinical measures to monitor the course of UI and its treatment. Treatment outcomes, as perceived and reported by patients, compliment clinical evidence and judgment of efficacy and effectiveness.

A number of measures have been developed or used to assess the perceived impact of UI [13–23], particularly for women [13, 14, 18–21]. Many outcome measures assess UI specific symptoms, functional status, and restrictions of usual activities, but tend to not fully capture the subjective evaluation of life quality associated with UI and its treatment.

The initial development of the Incontinence-specific Quality of Life Measure (I-QOL) [24] was approached using a needs based model of health-related quality of life [25, 26]. This was accomplished using a research team made up of urologists and QOL experts who conducted interviews with different panels of UI patients (men and women) to evaluate and refine relevant items for inclusion. The I-QOL was then validated, independent of medical intervention. Further developmental research, using data from a clinical trial of incontinent females, included: a refinement of the I-QOL's measurement model; evaluation of its relationship with clinical measures, and its ability to detect change over time in response to treatment [27]. Development for use in multiple countries was based on standardized published methods for cross-cultural adaptation [28]. These methods included two forward translations, an evaluated backward translation, cognitive debriefing interviews with patients having UI in each country, and a between-country harmonization process to develop the final language versions [29]. The purpose of this research is to report results of the assessment of the psychometric properties of I-QOL translations, used in patients with UI from 15 different countries.

Methods

The data for this psychometric assessment came from two main sources: 4 clinical trial settings and 2 clinic (population) settings.

Clinical Trial Cohort

Incontinence female outpatients aged 18 years and above (with one trial having an age limit of 65) with a clinical diagnosis of stress urinary incontinence of at least 3 months duration were invited to participate in one of four multi-center, double-blind, placebo-controlled, randomized clinical trials to assess the efficacy of duloxetine in the treatment of stress urinary incontinence. In each trial, the case definition included a predominant symptom of stress UI with a weekly incontinent episode frequency meeting a certain criteria (at least four in one trial and at least 7 in the remaining); the lack of predominant symptoms of enuresis or urge incontinence; diurnal and nocturnal frequencies less than eight and less than three, respectively, on screening history; negative funnel infusion cystometry with a first sensation greater then 100 ml, a bladder capacity of at least 400 ml; and a positive fixed volume cough stress test and stress pad test (greater than 2 g). Subjects were excluded if they had prolapse stage II or greater [30]; had a post void residual volume of 50 ml or more; were using any pharmacologic agent or device for UI; had adopted or changed behavioral management for UI within 3 months; or had a history of prior continence surgery. The trials were conducted in Argentina, Australia, Belgium, Brazil, Canada, Denmark, Finland, France, Germany, The Netherlands, Poland, Spain, Sweden, South Africa, the United Kingdom, and the United States. Even though Argentina, Finland, France and Germany were countries included in the clinical trials, they are omitted from these analyses due to an insufficient number of patients to support the psychometric validation analyses. However, it should be noted that validation results on both the French and German versions have been previously published [29].

Additional Non-clinical Trial Studies

SLOVAKIA: A total of 61 adult females presenting with symptoms of urinary incontinence were

enrolled at a community outpatient urodynamic clinic in Martin between August and December, 2002. Patients were excluded if they had poor cognitive function, were in recovery from a stroke, had a neurological condition, or had poor communication skills. At the first visit patients were given the I-QOL and SF-36 questionnaires and were interviewed by the clinician. Patients were scheduled for a second clinic visit 2 weeks later, at which time the I-QOL was re-administered just prior to undergoing urodynamic testing. Completed questionnaires were returned by 52 of the 61 patients approached.

GREECE: In May and June, 2002, 52 women from the Neopolis Health Center and surrounding rural areas, were recruited for the validation study based on their responses to screening questions regarding involuntary loss of urine. Each participant completed a first administration of the I-QOL measure, followed by a second administration of the I-QOL 3 weeks later. Incontinence was typed into stress, urge or mixed UI by eliciting patient descriptions of symptoms and circumstances of episodes. Participants were asked to self report the severity of their incontinence (mild, moderate, severe), the frequency of incontinent episodes in the last month, the length of time they had experienced this problem, and the number of medical appointments they have had in the past year to treat their incontinence.

Measures

Patient HRQoL was measured using the I-QOL measure (see appendix). It contains 22 negatively framed items, each with a five-point Likert-type response scale (1 = extremely, 2 = quite a bit, 3 = moderately, 4 = a little, 5 = not at all). The I-QOL was scored according to the developers' instructions [27]. In addition to yielding a total score, the I-QOL consists of the following three domain scores: avoidance and limiting behaviors (8 items), psychosocial impacts (9 items) and social embarrassment (5 items). The I-QOL total and subscale sum scores are transformed onto a 0–100 scale for greater interpretability, with the higher scores representing greater quality of life. The I-QOL was self-administered in all studies.

In the clinical trials, incontinence episode frequency (IEF) was collected to measure efficacy. A self-report severity item (the Patient Global Impression of Severity (PGI-S)) was given to patients with response options of mild, moderate, and severe. Validity of the PGI-S has been demonstrated by Yalcin and Bump [31].

Measurement model

In previous validation work, subscale definition was determined via both *a priori* designation and subsequent exploratory factor analyses. Since I-QOL subscales have already been determined (avoidance or limiting behaviors, psychological impacts, and social embarrassment), a principal factor analysis (or principal axis factoring with oblique rotation) in each language version was applied. The rotated factor structures were examined to see if the items substantially loaded on the theoretically expected factors (or subscales) found in the original validation. An overall incontinence-related quality-of-life score (using all I-QOL items) was computed and reported in this paper.

Evaluation of I-QoL psychometric properties

Psychometric testing of the I-QOL was conducted using standardized procedures [32] and instrument review criteria developed by the Scientific Advisory Committee of the Medical Outcomes Trust, including missing data (we used the criteria of greater than 5%), ceiling effects (>50% indicating 'not at all' on any item), and floor effects (>50% indicating 'extremely' on any item) [28]. It is necessary to evaluate the psychometric properties of the instrument prior to pooling data from individual countries to ensure that all translations performed appropriately in each language group. All psychometric analyses were based on data collected at baseline visit, with the exception of assessments of responsiveness in the clinical trials that also used data from follow-up. Psychometric assessments were made using SPSS for Windows version 10.1.

Reliability

Cronbach's alpha was calculated using the baseline data to assess internal consistency, or the degree of association between the item and scale scores [33]. A minimum value of 0.70 for group and 0.90 for individual comparisons is desirable [28, 33]. Reproducibility (test/retest reliability) was assessed across the 2-week interval between the randomization visit and the baseline from the tri-

als; baseline to 2-week in Slovakia; and baseline to 3-week in Greece using the intraclass correlation coefficient (ICC) [34]. This is a preferred measure of strength of association for determining stability of scores over time because it measures the homogeneity within groups relative to the total variation. The recommended level for group comparisons is 0.70 [28].

Validity

Convergent validity, a type of construct validity, involves comparing a PRO measure of one concept to another logically related measure with the same concept. If previous predictions of association are accurate, then convergent validity is achieved. Convergent validity for the I-QOL has already been established using the Medical Outcomes Study Short Form-36 Health Survey (SF-36) and Psychological General Well-Being Schedule (PGWB) measures [24]. These measures were not included in the clinical trials, but the SF-36 was used in the Slovakian community study. To assess convergent validity, Pearson's correlation was computed to measure the association between the total and subscale scores of the I-QOL measure and the SF-36. Specifically, as in previous I-QOL validation studies, it is hypothesized, using a two-tailed test at a $p < 0.05$ level, that I-QOL scores will be more highly correlated with physical and mental well-being than with bodily pain.

Discriminant (known groups) validity involves assessing whether or not a PRO is able to distinguish between two or more recognized groups with theoretically different levels of the outcome to be measured. In this analysis, the I-QOL was assessed using definitions of incontinence-related severity (mild, moderate, severe). ANOVA statistics will be used with expectations that the overall I-QOL scores would be significantly lower ($p < 0.05$) for women indicating greater self-reported severity. Eta squared values, the proportion of variance in the I-QOL that is explained by differences among severity groups, will be examined.

Responsiveness

Responsiveness is the ability of an instrument to detect small but important changes [35, 36]. To interpret importance of changes in the I-QOL, its scores were anchored to the frequency of incontinence episodes. Using data from the clinical trials, change scores were calculated using the difference between the baseline visit and the final follow-up visit using a last observation carried forward approach. Improvements were defined at two different levels: (1) at least a 25% decrease with no change defined as between 0 and 24% as was evaluated in the I-QOL's original validation [27], and (2) at least a 50% decrease with no change defined as between 0 and 49% determined by recent studies as a better threshold [37]. The standardized response mean (SRM) was used to evaluate responsiveness (mean change score, divided by the standard deviation of change score) [38]. Higher values for the SRM indicate a greater sensitivity to change [35, 36, 38, 39].

Results

Sample characteristics

Completed I-QOL questionnaires were available for a total of 1815 patients enrolled in the clinical trials with an additional 104 patients in community studies in Slovakia and Greece. The total female population ranged in age from 22 to 83, with mean ages across the languages ranging from 48.9 ± 8.7 (in Slovakia) to 56.9 ± 12.1 (in Belgium). With the exception of Greece, patients predominantly self-reported the severity of their condition as being moderate to severe (percents range from 58.3% in the US to 76.4% in French-speaking Canada) (Table 1).

I-QOL item characteristics

In comparing the variation of individual items across countries, we found that with the exception of 2 items in Greece, no languages exhibited problems with missing data. In the Greek dataset, item 19 (*Feel no control over bladder*) had 9.1% missing, and item 11 (*Important to plan ahead*) had 6.1% missing. An analysis of ceiling effects showed all languages to have consistent results. Ceiling effects were detected in items such as #13 (*Hard getting a good nights sleep*) and #22 (*Worry about having sex*) indicating greater percentages of patients affected very little by these issues. Potential floor effects were noted in items #2 (*I worry*

Table 1. Language, sociodemographic, and clinical characteristics

Country (Language)	Patients		Age		Severity
	n	Percent	mean	StDev	% Moderate/Severe
Clinical Trial Cohort					
Australia (English)	77	4.0	53.2	12.1	71.4
Belgium (Dutch)	63	3.3	56.9	12.1	63.5
Brazil (Portuguese)	40	2.1	51.1	8.9	75.0
Canada (English)	65	3.4	54.2	11.4	70.8
Canada (French)	55	2.9	53.4	11.0	76.4
Denmark (Danish)	51	2.7	52.0	11.5	74.5
Spain (Spanish)	45	2.3	51.0	11.3	73.3
United Kingdom (English)	82	4.3	50.3	9.5	73.2
The Netherlands (Dutch)	102	5.3	52.1	8.8	59.8
Poland (Polish)	140	7.3	53.1	9.3	72.9
South African (English)[a]	117	6.1	53.5	10.9	70.1
Sweden (Swedish)	120	6.3	51.8	9.9	67.5
USA (English)[a]	858	44.7	51.7	10.4	58.3
Non-clinical Trial Studies					
Greece (Greek)	52	2.7	61.0	11.5	40.4
Slovakia (Slovak)	52	2.7	48.9	8.7	65.4
Total	1919	100	52.2	10.4	64.2

[a] There was a small amount of another language used but, based on local investigator's responses, proportions were unknown.

about coughing and sneezing), #18 (*I worry about wetting myself*), and #12 (*I worry about my incontinence getting worse as I grow older*) indicating a great deal of impact in responders across the various countries.

Domain structure of the I-QOL

A confirmatory factor model was performed to test the strength of each item to the subscales found by the developers. In all language versions, three factors were indicated (eigenvalues greater than 1) with the cumulative percent of variation explained ranging from 54% (in UK sample) to 69% (in the Slovakian sample). In assessing the item-loadings, there were only 8 instances where items had a better relationship with a different subscale. In Brazil, item #9 (*Incontinence is always on my mind*) had a slightly better fit in the social embarrassment vs. the psychosocial impacts subscale and #19 (*I feel like I have no control over my bladder*) fit better in the psychosocial impacts vs. social embarrassment subscale. In Canada (English and French) and Denmark, item #2 (*I worry about coughing and sneezing*) had higher associations with the social embarrassment vs. the avoidance and limiting behavior subscale. In the UK and South Africa, item #6 (*I don't feel free to leave my home for long periods of time*) fit better in the avoidance and limiting behavior vs. the psychosocial impacts subscale. Finally, in Greece, item #3 (*I have to be careful about standing up after sitting down*) fit slightly better in the psychosocial impacts vs. the avoidance and limiting behavior subscale.

Scoring

The I-QOL was then scored as previously suggested by the developers with the three subscales and an overall incontinence-specific quality of life score. Figure 1 shows the I-QOL scores plotted for each language version (line graph used for visual simplicity, no link between domains is implied). While there is some variation in scores by subscale, all versions show similar patterns with more positive scores (higher QOL) for the psychosocial subscale and lower scores in social embarrassment. Outlying versions include Greece with highest levels of QOL (as expected due to having predominantly self-reported UI as mild). The samples from Slovakia (clinic-based) and Poland report the lowest UI-related QOL. Also included in the graph are the means from the original validation study, which have comparatively greater means (higher QOL) across all scores.

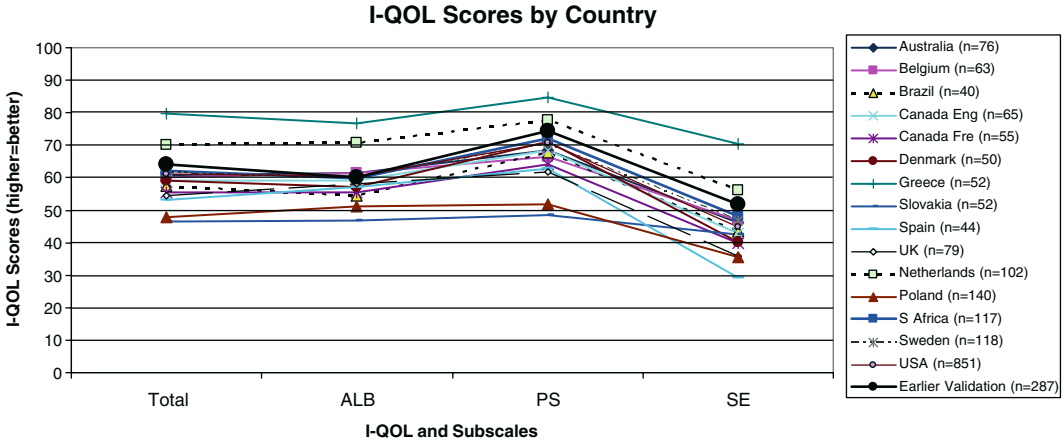

Figure 1. I-QOL scores by country. Legend: ALB = Avoidance and limiting behavior; PS = Psychosocial; SE = Social Embarrassment.

Reliability

The overall I-QOL summary score showed high internal consistency (*alpha* ranges between 0.91 and 0.94), indicating that the 22-items could be summed to form a composite score. Subscales also showed acceptable *alpha* values with values exceeding 0.74 for behaviors; 0.81 for psychosocial and 0.72 for social embarrassment subscales. The intraclass correlation coefficient assessing reproducibility at 2 weeks ranged from 0.72 (Denmark) to 0.91 (Sweden) for the total score, demonstrating stability of I-QOL scores over time (Table 2).

Validity

Table 3 presents I-QOL scores according to the categories of patient perceived severity (mild, moderate, severe). As expected, I-QOL scores were significantly different for women who reported mild incontinence compared with those who reported moderate or severe incontinence. These differences were significant at the 0.05 level for all languages except Brazilian Portuguese (most likely due to the lower numbers in this language). Eta squared values are also included in Table 3 showing the measure of association between severity and quality of life (11 of the 15 versions having greater than 0.25). I-QOL subscale results had similar results (not shown).

Convergent validity was assessed using the SF-36 domains for the data collected in Slovakia. Pearson's correlations were predominantly moderate (mostly between 0.30 and 0.60). *A priori* expectations were met in that I-QOL scores were more highly associated with physical functioning (PF = 0.48, RP = 0.50) and mental well being (MH = 0.43, RE = 0.44, SF = 0.57) than with bodily pain (0.34). General health perceptions and vitality were among the other SF-36 domains that had lower correlations to the I-QOL scores (Table 4).

Responsiveness and effect size interpretation

Standardized response means (SRM) are shown in Table 5 for changes in the number of incontinent episodes. Important changes were defined as the percent change in I-QOL score for the improved group using the number of incontinent episodes ($\geq$25% and $\geq$50%). In all language versions, improvements in I-QOL scores were larger for the groups exhibiting at least a 25% decrease in the number of incontinent episodes over 12 weeks (vs. 0–24% decrease), and at least twofold differences between those with a $\geq$50% decrease in episodes (vs. a 0–49% decrease). These improvements were associated with SRM values all greater than 0.50 (indicating moderate to high effect sizes) [39] demonstrating that the I-QOL is able to detect change across the reported language versions. I-QOL subscale results had similar results (not shown).

Table 2. Reliability statistics for the I-QOL total summary score

	Internal consistency (Cronbach's alpha)			
	I-QOL total score	I-QOL avoidance & limiting behaviors	I-QOL psychosocial impacts	I-QOL social embarrassment
Clinical Trial Cohort				
Australia (n = 76)	0.94	0.83	0.90	0.85
Belgium (n = 62)	0.94	0.84	0.91	0.86
Brazil (n = 40)	0.92	0.82	0.88	0.72
Canada (English) (n = 65)	0.91	0.82	0.85	0.78
Canada (French) (n = 53)	0.95	0.87	0.91	0.85
Denmark (n = 50)	0.93	0.88	0.81	0.78
Spain (n = 44)	0.92	0.74	0.89	0.77
United Kingdom (n = 79)	0.91	0.80	0.85	0.79
The Netherlands (n = 102)	0.94	0.85	0.89	0.90
Poland (n = 140)	0.94	0.82	0.92	0.87
South Africa (n = 117)	0.93	0.82	0.89	0.83
Sweden (n = 118)	0.94	0.86	0.91	0.84
USA (n = 851)	0.94	0.84	0.90	0.86
Non-clinical Trial Studies				
Greece (n = 52)	0.93	0.80	0.92	0.84
Slovakia (n = 52)	0.96	0.88	0.94	0.86
	Test-retest reproducibility (Intraclass Correlation Coefficient)			
	I-QOL total score	I-QOL avoidance & limiting behaviors	I-QOL psychosocial impacts	I-QOL social embarrassment
Clinical Trial Cohort				
Australia (n = 76)	0.94	0.88	0.93	0.92
Belgium (n = 62)	0.94	0.88	0.94	0.92
Brazil (n = 40)	0.92	0.88	0.90	0.89
Canada (English) (n = 65)	0.91	0.86	0.84	0.83
Canada (French) (n = 53)	0.95	0.93	0.94	0.92
Denmark (n = 50)	0.93	0.88	0.82	0.79
Spain (n = 44)	0.92	0.89	0.93	0.88
United Kingdom (n = 79)	0.91	0.90	0.95	0.91
The Netherlands (n = 102)	0.94	0.92	0.96	0.93
Poland (n = 140)	0.94	0.90	0.93	0.88
South Africa (n = 117)	0.93	0.90	0.92	0.90
Sweden (n = 118)	0.94	0.93	0.95	0.93
USA (n = 851)	0.94	0.93	0.89	0.88
Non-clinical Trial Studies				
Greece (n = 52)	0.93	0.91	0.94	0.95
Slovakia (n = 52)	0.96	0.93	0.91	0.91

Discussion

While the I-QOL has been widely translated using techniques that consider the cultural aspects of health, it is important to psychometrically validate these types of measures for each language adaptation. This analysis was conducted by individual language and then compared to the original validation results done in the US [27], allowing for patient responses to be examined for any cultural differences that may exist, either in that particular language or in the translation process.

The first step involved a close investigation of the items in the I-QOL and how each item fit into the domains defined in the original design. Evaluating the amount of missing data, 14 of the 15 language versions did not have any items with greater than 5% left unanswered. In the Greek

Table 3. Discriminant validity of the I-QOL total summary score (Self-reported Severity)

	Eta-squared[a]	Mild		Moderate		Severe		*F*-stat
		mean (sd)	n	mean (sd)	n	mean (sd)	n	
Clinical Trial Cohort								
Australia	0.241	73.8 (15.6)	22	59.0 (19.9)	43	42.0 (15.9)	11	11.6 ***
Belgium	0.442	73.3 (15.2)	16	62.5 (14.8)	27	34.9 (20.3)	13	21.0 ***
Brazil	0.085	66.1 (16.6)	10	55.6 (17.4)	24	50.4 (24.0)	6	1.7
Canada (Eng)	0.361	73.0 (12.0)	19	56.0 (13.4)	37	42.9 (16.6)	9	17.5 ***
Canada (Fre)	0.413	66.0 (19.9)	11	60.6 (15.1)	32	27.3 (18.8)	10	17.6 ***
Denmark	0.146	69.7 (20.0)	11	58.1 (18.2)	28	47.8 (14.2)	10	3.9 *
Spain	0.154	67.0 (19.4)	10	51.0 (20.4)	20	45.1 (19.4)	13	3.6 *
UK	0.312	69.1 (12.4)	17	53.1 (16.4)	46	37.1 (9.9)	11	16.1 ***
The Netherlands	0.457	81.0 (9.8)	40	69.6 (13.5)	44	45.9 (19.3)	17	41.2 ***
Poland	0.370	65.2 (19.4)	38	48.3 (17.3)	68	28.0 (15.8)	34	40.2 ***
South Africa	0.266	74.3 (16.6)	34	60.5 (16.4)	64	43.9 (17.1)	18	20.5 ***
Sweden	0.293	75.3 (16.0)	37	56.9 (17.2)	73	37.8 (12.9)	8	23.8 ***
USA	0.274	70.1 (15.7)	349	58.3 (16.3)	424	35.2 (15.8)	75	159.4 ***
Non-clinical Trial Studies								
Greece	0.509	89.6 (9.3)	27	68.3 (17.5)	11	52.0 (21.7)	5	20.8 ***
Slovakia	0.251	60.2 (18.9)	18	44.6 (23.7)	22	29.3 (16.8)	12	8.2 **

[a] Eta squared is measure of association from ANOVA.
Note: Discrepancies in numbers are a result of subjects missing self-reported severity data.
*** Significant at the 0.001 level; ** Significant at the 0.01 level; * Significant at the 0.05 level.

sample, 2 items had higher amounts of missing responses (6.1 and 9.1%). It was noted that patients had difficulty understanding the meaning of some items, indicating a possible need to revisit the Greek translation. While ceiling effects were noted in this analysis, they were consistent across all language versions. Only two items registered ceiling effects of greater than 50% in all languages indicating that over half of the population is not affected by these issues: *I worry about having sex* and *I have a hard time getting a good nights sleep*. While these items are conceptually important, they may be good candidates for removal if the development of a shorter version of the I-QOL is investigated as they can both be affected by a great number of conditions besides incontinence.

Using confirmatory factor analyses, the I-QOL measurement model was validated in all language versions. Even though a few items in select language versions did show a better fit in a different subscale, they were also highly correlated with the subscale they actually loaded into (from the original factoring). There was an overwhelming agreement of the three subscales appropriately referred to as social embarrassment, avoidance and limiting behavior, and psychosocial impacts.

Once the I-QOL was scored and plotted, variations in mean scores were noted across the language versions. While the quality of life scores differed in some of the countries (i.e., lower scores in Poland and Slovakia, and higher scores in Greece and the Netherlands), the trends of the I-QOL score components remained very similar. The differences are more than likely related to differences in population characteristics rather than measurement errors. For example, it was expected that the average I-QOL scores in the Greek sample would be higher, as approximately 60% of that population had self-reported the severity of their UI as 'fair.' Interestingly, it has also been demonstrated in previous studies that women in Greece tend to not report their symptoms of UI to their doctor because they consider it only a minor problem [9].

Reliability, using Cronbach's alpha, was confirmed in all language versions, where high internal consistency was demonstrated by alpha values all above 0.70. Similarly, test-retest reliability showed

Table 4. Convergent validity of the I-QOL (using data from Slovakia)

SF-36 domain	I-QOL total score (22-items)	I-QOL avoidance & limiting behaviors (8 items)	I-QOL psychosocial impacts (9 items)	I-QOL social embarrassment (5 items)
Physical Function (PF)	0.48***	0.42**	0.47***	0.51***
Role-Physical (RP)	0.50***	0.44**	0.48***	0.51***
Bodily Pain (BP)	0.34*	0.37**	0.29*	0.34*
General Health Perceptions (GH)	0.06	0.03	0.06	0.08
Vitality (VT)	0.33*	0.28*	0.34*	0.31*
Social Function (SF)	0.57***	0.53***	0.56***	0.53***
Role-Emotional (RE)	0.44**	0.37**	0.41**	0.52***
Mental Health (MH)	0.43**	0.32*	0.51***	0.37**

N = 52 for all Pearson correlations.
* Correlation is significant at the 0.05 level (2-tailed); ** Correlation is significant at the 0.01 level (2-tailed); *** Correlation is significant at the 0.001 level (2-tailed).

the I-QOL to be stable over time with ICC values all greater than the recommended 0.70.

Because of the limited number of data elements used across these different studies, validity of the I-QOL was demonstrated using known-groups classifications based on the UI severity of the patients. With the exception of Brazil, I-QOL scores in all language versions were able to discriminate between levels of severity (showing quality of life to be significantly higher in those patients in the mild category than for those who were much more severe). Data from Slovakia was used in making comparisons to hypotheses generated from the initial validation study regarding the SF-36. It was confirmed that the I-QOL measure was more closely associated to the SF-36's physical and mental functioning domains than with bodily pain. The highest relationships were actually seen with the social functioning domain, which is not surprising in that UI certainly impacts patients' social aspects of life such as going to social events. The lowest associations were noted with the general health perceptions indicating that the I-QOL does indeed measure disease-specific aspects of QOL.

An important characteristic of a measure is its ability to measure change over time. Since a majority of the data came from clinical trials, the use of an efficacy measure was used to assess responsiveness. Incontinence episode frequency (IEF) is the current 'gold standard' assessment for this and is used frequently in studies that evaluate incontinence treatments. In this analysis, the performance of the I-QOL was compared to $\geq$25% and a $\geq$50% reduction in the number of incontinent episodes. Statistics summarizing responsiveness across the language versions were generally high for the IEF indicated that the I-QOL moved in the expected direction with clinical change and was responsive to detected changes. These results provide the means for estimating sample sizes for subsequent clinical trials of behavioral or pharmacologic interventions and for interpreting observed effect sizes in terms of a frequently used clinical measure (number of reported incontinent episodes).

The consistent strength in the I-QOL's psychometric performance mirrors that of its earlier publications [27, 29], and confirms the robust contribution that development by the needs-based

Table 5. Responsiveness for the I-QOL total summary score by change in episode frequency (IEF)

	IEF % decrease	n	I-QOL Change	SRM*	IEF % decrease	n	I-QOL Change	SRM*
Australia	0–24%	7	6.8	0.52	0–49%	20	5.7	0.43
	≥25%	55	16.3	1.05	≥50%	42	19.7	1.38
Belgium	0–24%	9	2.7	0.26	0–49%	14	6.3	0.50
	≥25%	39	12.0	0.79	≥50%	34	11.9	0.76
Brazil	0–24%	4	−2.8	−0.08	0–49%	5	0.0	0.00
	≥25%	28	13.9	0.75	≥50%	27	14.0	0.74
Canada (English)	0–24%	6	15.7	1.45	0–49%	21	9.7	1.18
	≥25%	46	19.2	1.05	≥50%	31	25.0	1.29
Canada (French)	0–24%	7	6.8	0.63	0–49%	20	8.9	0.75
	≥25%	32	12.8	1.05	≥50%	19	14.8	1.25
Denmark	0–24%	8	1.7	0.15	0–49%	18	9.5	0.53
	≥25%	34	14.1	0.94	≥50%	24	13.5	1.06
Spain	0–24%	5	5.3	0.29	0–49%	14	1.9	0.12
	≥25%	27	9.1	0.81	≥50%	18	8.9	0.50
United Kingdom	0–24%	17	−0.4	−0.02	0–49%	33	1.6	0.08
	≥25%	47	15.5	0.73	≥50%	31	21.6	1.04
The Netherlands	0–24%	9	3.2	0.43	0–49%	27	5.0	0.39
	≥25%	67	9.4	0.71	≥50%	49	10.7	0.84
Poland	0–24%	24	−4.3	−0.24	0–49%	46	2.8	0.13
	≥25%	96	17.8	0.80	≥50%	74	20.0	0.91
South Africa	0–24%	16	1.5	0.11	0–49%	33	2.1	0.15
	≥25%	93	14.6	0.80	≥50%	76	17.4	0.95
Sweden	0–24%	14	8.4	0.53	0–49%	49	9.0	0.72
	≥25%	86	12.3	0.98	≥50%	51	14.5	1.10
USA	0–24%	89	9.2	0.66	0–49%	241	8.7	0.60
	≥25%	627	15.2	0.93	≥50%	475	17.3	1.07

* Standardized response mean of those with an improvement.
Note: Greece and Slovakia were cross-sectional, non-clinical studies only, therefore not included in this table.

model provides for the development of HRQoL measures [26]. The needs-based model offers an advantage of eliciting item content that is more universal and cross-culturally relevant, implicating high cross-cultural acceptance. Finally, one concern in most clinical trials is the ability to aggregate, or pool, the cross-cultural datasets that have been collected via multiple translations of a single measure. The data reported in this paper demonstrates strongly similar trends in both the summary and subscale scores of the I-QOL, which would suggest similar performance across cultures. However, the differences in ranges between mean scores across cultures would suggest the need to control for case mix variables such as UI severity before interpreting results from pooled analyses, particularly since the literature suggests a stronger impact by patient-perceived severity and patient age on incontinence-specific quality of life than the particular type of UI a patient is diagnosed with (except when symptoms include pain).

Limitations

While some of the countries represented in these analyses contributed larger sample sizes to the dataset (N = 55–858), there were others (such as Brazil, Denmark, Spain, Greece and Slovakia) with relatively small datasets (N = 40–52). Even though the majority of the analyses all demonstrated acceptable psychometric performance, it is likely that the picture would have been different (especially for the areas where there was greater variation in statistics or fit) had these language versions been tested on a larger population.

In the clinical trial dataset for this analysis, it was not possible to identify the data for two of the countries (US and South Africa) that were actually a different language (US Spanish instead of US English and South African English instead of Afrikaans). Even though there is a known mixture of some small degree of language difference in these two datasets, the psychometric performance

remained stable enough within those populations to produce acceptable statistics. It is not known how the data for those two countries would change were it possible to identify and extract the mixed language data.

Other limitations that could be noted in this study would be the lack of universal convergent validity measures. It would have been ideal to be able to test the convergent statistics across all language versions and compare them to each other and to the original data published on the US English version and first four language versions of the I-QOL. However, the SF-36 was only available in one of the community studies, so only those data could be reported.

Conclusion

The I-QOL is a disease-specific patient-reported outcome measure designed to evaluate health-related quality of life concerns that affect patients with urinary incontinence. It is easily self-administered and takes approximately 5 min for the average patient reading at a 5th grade level to complete. Using data from both randomized clinical trials and clinical community studies in Slovakia and Greece, the results support the reliability, validity and responsiveness to change of the I-QOL providing a sensitive measure of HRQoL. With consideration to properly prepared translations (following acceptable methodological criteria), and study methods and populations, cross-cultural comparisons of the I-QOL can be made.

The performance of the I-QOL in these analyses suggests that it can be an important addition to the compendium of outcome measures used to assess urinary incontinence and its treatment. Given the need for patient-reported outcome data as a complement to clinical outcomes in the assessment of UI treatments, we have demonstrated the psychometric properties of an instrument to collect such data in international studies.

Appendix A: Incontinence Quality of Life (I-QOL) Instrument

Q1, I worry about not being able to get to the toilet on time; Q2, I worry about coughing or sneezing because of my urinary problems or incontinence; Q3, I have to be careful standing up after I've been sitting down because of my urinary problems or incontinence; Q4, I worry about where toilets are in new places; Q5, I feel depressed because of my urinary problems or incontinence; Q6, Because of my urinary problems or incontinence, I don't feel free to leave my home for long periods of time; Q7, I feel frustrated because my urinary problems or incontinence prevents me from doing what I want; Q8, I worry about others smelling urine on me; Q9, My urinary problems or incontinence is always on my mind; Q10, It's important for me to make frequent trips to the toilet; Q11, Because of my urinary problems or incontinence, it's important to plan every detail in advance; Q12, I worry about my urinary problems or incontinence getting worse as I grow older; Q13, I have a hard time getting a good night of sleep because of my urinary problems or incontinence; Q14, I worry about being embarrassed or humiliated because of my urinary problems or incontinence; Q15, My urinary problems or incontinence makes me feel like I'm not a healthy person; Q16, My urinary problems or incontinence makes me feel helpless; Q17, I get less enjoyment out of life because of my urinary problems or incontinence; Q18, I worry about wetting myself; Q19, I feel like I have no control over my bladder; Q20, I have to watch what or how much I drink because of my urinary problems or incontinence; Q21, My urinary problems or incontinence limit my choice of clothing; Q22, I worry about having sex because of my urinary problems or incontinence.

I worry about not being able to get to the toilet on time.

All items use the following response scale:

1 = EXTREMELY
2 = QUITE A BIT
3 = MODERATELY
4 = A LITTLE
5 = NOT AT ALL

Subscale Structure:

Avoidance and Limiting Behavior: Items 1, 2, 3, 4, 10, 11, 13, and 20

Psychosocial Impacts: Items 5, 6, 7, 9, 15, 16, 17, 21, and 22

Social Embarrassment: Items 8, 12, 14, 18 and 19

FOR PERMISSION TO USE THE I-QOL:

For a copy of the I-QOL and permission to use it, please contact the Medical Outcomes Trust, Health Assessment Lab, 235 Wyman Street, Suite 130, Waltham, MA 02451, 781-890-5544 info@hal-health.org.

References

1. Fantl JA, Bump RC, Robinson D, McClish DK, Wyman JF. Efficacy of estrogen supplementation in the treatment of urinary incontinence. The Continence Program for Women Research Group. Obstet Gynecol 1996; 88(5): 745–749.
2. Abrams P, Cardozo L, Fall M, et al. The standardization of terminology of lower urinary tract function: Report from the standardization sub-committee of the International Continence Society. Neurol Urodyn 2002; 21: 167–178.

3. Svihra J, Baska T, Martin M, et al. Prevalence of female overactive bladder with urinary incontinence in Slovakia. European Urology Supplements 2002; 1(1): 85.
4. Moller LA, Lose G, Jorgensen T. The prevalence and bothersomeness of lower urinary tract symptoms in women 40–60 years of age. Acta Obstet Gynecol Scand 2000; 79(4): 298–305.
5. Thom D. Variation in estimates of urinary incontinence prevalence in the community: Effects of differences in definition, population characteristics, and study type. J Am Geriatr Soc 1998; 46(4): 473–480.
6. Hampel C, Wienhold D, Benken N, Eggersmann C, Thuroff JW. Prevalence and natural history of female incontinence. Eur Urol 1997; 32(Suppl 2): 3–12.
7. Hunskaar S, Burgio K, Diokno A, Herzog AR, Hjalmas K, Lapitan MC. Epidemiology and natural history of urinary incontinence in women. Urology 2003; 62(4 Suppl 1): 16–23.
8. Herzog AR, Fultz NH. Prevalence and incidence of urinary incontinence in community-dwelling populations. J Am Geriatr Soc 1990; 38(3): 273–281.
9. Lionis C, Vlachonikolis I, Bathianaki M, Daskalopoulos G, Anifantaki S, Cranidis A. Urinary Incontinence, the Hidden Health Problem of Cretan Women: Report from a Primary Care Survey in Greece. Women and Health 2000; 31.
10. Milsom I. The prevalence of urinary incontinence. Acta Obstet Gynecol Scand 2000; 79(12): 1056–1059.
11. Burgio KL, Matthews KA, Engel BT. Prevalence, incidence and correlates of urinary incontinence in healthy, middle-aged women. J Urol 1991; 146: 1255–1259.
12. Blaivas JG, Appell RA, Fantl JA et al. Standards of efficacy for evaluation of treatment outcomes in urinary incontinence: Recommendations of the Urodynamic society. Neurourol Urodynam 1997; 16: 145–147.
13. 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. Quality of Life Res 1994; 3: 291–306.
14. Jackson S, Donovan J, Brookes S, Eckford S, Swithinbank L, Abrams P. The Bristol Female Lower Urinary Tract Symptoms questionnaire development and psychometric testing. Br J Urol 1996; 77: 805–812.
15. Black N, Griffiths J, Pope C. Development of a symptom severity index and a symptom impact index for stress incontinence in women. Neurol Urodynam 1996; 15: 630–640.
16. Lee PS, Reid DW, Saltmarche A, Linton L. Measuring the psychosocial impact of urinary incontinence. The York Incontinence Perceptions Scale (YIPS). JAGS 1995; 43: 1275–1278.
17. Yu LC. Incontinence stress index. Measuring psychological impact. J Gerontol Nurs 1987; 13: 18–25.
18. Cella D. Functional Assessment of Incontinence Therapy, version 2. Chicago: Rush Division of Psychosocial Oncology, 1994.
19. Hunskaar S, Vinsnes V. The quality of life in women with urinary incontinence as measured by the sickness impact profile. J Am Geriatr Soc 1991; 40: 976–977.
20. Hunter DJW, McKee M, Black NA, Sanderson CFB. Health status and quality of life of British men with lower urinary tract symptoms: results from the SF-36. Urology 1995; 45(6): 962–971.
21. Renck-Hooper U, McKenna SP, Whalley D. Measuring quality of life in female urinary urge incontinence: development and psychometric properties of the IQoLI. Outcomes Res 1997; 1: 1–8.
22. Nochajski TH, Burns PA, Pranikoff K, Dittmar SS. Dimensions of urine loss among older women with genuine stress incontinence. Neurol Urodynam 1993; 12: 223–233.
23. Seim A, Hermstad R, Hunskaar S. Management in general practice significantly reduces psychosocial consequences of female urinary incontinence. Qual Life Res 1997; 6: 257–264.
24. Wagner TH, Patrick DL, Bavendam TG, Martin ML, Buesching DP. Quality of life in persons with urinary incontinence: Development of a new measure. Urology 1996; 47(1): 67–72.
25. Patrick DL, Erickson P. Health Status and Health Policy: Allocating Resources to Health Care. New York: Oxford University Press, 1993.
26. Hunt SM, McKenna SP. The QLDS: a scale for the measurement of quality of life in depression. Health Policy 1992; 11: 307–319.
27. Patrick DL, Martin ML, Bushnell DM, Yalcin I, Wagner TH, Buesching DP. Quality of Life of Women with Urinary Incontinence: Further Development of the Incontinence Quality of Life Instrument (I-QOL). Urology 1999a; 53: 71–76.
28. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002; 11: 193–205.
29. Patrick DL, Martin ML, Bushnell DM, Marquis P, Andrejasich CM, Buesching DP. Cultural adaptation of a quality-of-life measure for urinary incontinence. Eur Urol 1999b; 36: 427–435.
30. Bump RC, Mattiasson A, Bø K et al. The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. Am J Obstet Gyynecol 1996; 175: 10–17.
31. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. Am J Obstet Gynecol 2003; 189: 98–101.
32. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993; 118: 622–629.
33. Cronbach LF. Coefficient alpha and the internal structure of tests. Psychometricka 1951; 16: 297–334.
34. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Controlled Clin Trials 1991; 12: 142s–158s.
35. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. J Chronic Dis 1987; 40: 171–178.
36. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. J Chronic Dis 1986; 39: 897–906.

37. Yalcin I, Bump RC. The effect of previous treatment experience and incontinence severity on the placebo response of stress urinary incontinence. Am J Obstet Gynecol 2004; 191(1): 194–197.
38. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. Med Care 1992; 30(10): 917–925.
39. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed., Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.

Address for correspondence: Donald Bushnell, Health Research Associates, Inc., 6505 216th st. SW Suite 105, Mountlake Terrace, WA 98403, USA
E-mail: bushnell@hrainc.net

Table 2: Reliability Statistics for the I-QOL Total Summary Score

	Internal Consistency (Cronbach's alpha)			
	I-QOL Total Score	I-QOL Avoidance & Limiting Behaviors	I-QOL Psychosocial Impacts	I-QOL Social Embarrassment
Clinical Trial Cohort				
Australia (n=76)	0.94	0.83	0.90	0.85
Belgium (n=62)	0.94	0.84	0.91	0.86
Brazil (n=40)	0.92	0.82	0.88	0.72
Canada (English) (n=65)	0.91	0.82	0.85	0.78
Canada (French) (n=53)	0.95	0.87	0.91	0.85
Denmark (n=50)	0.93	0.88	0.81	0.78
Spain (n=44)	0.92	0.74	0.89	0.77
United Kingdom (n=79)	0.91	0.80	0.85	0.79
The Netherlands (n=102)	0.94	0.85	0.89	0.90
Poland (n=140)	0.94	0.82	0.92	0.87
South Africa (n=117)	0.93	0.82	0.89	0.83
Sweden (n=118)	0.94	0.86	0.91	0.84
USA (n=851)	0.94	0.84	0.90	0.86
Non-clinical Trial Studies				
Greece (n=52)	0.93	0.80	0.92	0.84
Slovakia (n=52)	0.96	0.88	0.94	0.86

	Test-Retest Reproducibility (Intraclass Correlation Coefficient)			
	I-QOL Total Score	I-QOL Avoidance & Limiting Behaviors	I-QOL Psychosocial Impacts	I-QOL Social Embarrassment
Clinical Trial Cohort				
Australia (n=76)	0.87	0.88	0.93	0.92
Belgium (n=62)	0.87	0.88	0.94	0.92
Brazil (n=40)	0.83	0.88	0.90	0.89
Canada (English) (n=65)	0.74	0.86	0.84	0.83
Canada (French) (n=53)	0.90	0.93	0.94	0.92
Denmark (n=50)	0.72	0.88	0.82	0.79
Spain (n=44)	0.86	0.89	0.93	0.88
United Kingdom (n=79)	0.87	0.90	0.95	0.91
The Netherlands (n=102)	0.90	0.92	0.96	0.93
Poland (n=140)	0.86	0.90	0.93	0.88
South Africa (n=117)	0.86	0.90	0.92	0.90
Sweden (n=118)	0.91	0.93	0.95	0.93
USA (n=851)	0.81	0.93	0.89	0.88
Non-clinical Trial Studies				
Greece (n=52)	0.90	0.91	0.94	0.95
Slovakia (n=52)	0.87	0.93	0.91	0.91