114

Sherburn M[1], Bo K[2], Galea M[1]
1. *The University of Melbourne,* ***2.*** *Norwegian School of Sports Science*

EVALUATION OF OUTCOME MEASURES FOR STRESS URINARY INCONTINENCE IN OLDER WOMEN

Hypothesis / aims of study

There is currently no 'gold standard' measure or set of measures to assess treatment outcomes for stress urinary incontinence in older women. The aims of this study were to determine outcome measures that are valid and responsive in an elderly female sample by evaluating concurrent and predictive validity, responsiveness, and minimal clinically important difference needed to detect a treatment effect, for a suite of measures.

Study design, materials and methods

The analyses for this study were performed using data generated in a previously reported RCT [1]. Outcomes assessment was performed by an investigator blinded to group allocation. Outcome measures evaluated were a cough stress test with and without a pre-contraction of the pelvic floor muscles, the ICIQ-UI SF (International Consultation on Incontinence Questionnaire - Urinary Incontinence Short Form), the ISI (Incontinence Severity Index), leakage episodes recorded in a seven-day Accident Diary, Bother 10cm VAS (visual analogue scale), and participant perceived global rating of change of symptoms (global change) at the primary end-point (20 weeks of intervention).

a) **Concurrent validity** was evaluated by correlating global change with mean change from baseline to the primary end-point of the study for each outcome measure using Spearman's Correlation Coefficient. In addition global change was correlated with baseline scores of all outcome measures, and baseline scores correlated with primary end-point scores, to ascertain the **predictive validity** of all outcome measures. Correlations were defined a priori, r=0.25-0.5 as fair to moderate correlation, r=0.5-0.75, moderate to good correlation and r=>0.75, good to excellent correlation [2].

b) **Responsiveness** was determined by:

i. Comparison of median/mean change scores between the two intervention groups
ii. Comparison of effect size
iii. Relative efficiency to detect a difference between groups

using global change and ICIQ-UI SF as the reference measures for comparison.

c) The **minimal clinically important difference** (MCID) for each outcome measure was calculated by two methods:

i. Independent t-tests tests to determine if mean change in the outcome measures discriminated between improvement and no improvement in global change.

ii. Receiver operating characteristics (ROC) analyses were used to select a score that discriminated between improved and not improved participants. Sensitivity and specificity of change scores for each outcome measure were calculated and ROC curves generated to determine the cut-off point between improvement and no improvement.

Results

Seventy-six participants who enrolled in the RCT [1], (41 pelvic floor muscle training and 35 bladder training) completed the primary end-point assessment at the end of the 20 week intervention. Data from these participants was used in the analyses.

Validity: Concurrent validity correlations are presented in Table 1. In addition, there were no correlations between global change and baseline scores on any outcome measure, thus demonstrating poor predictive validity (r<.1 for all measures) for the global change measure. However, baseline scores predicted end-point scores on most measures with moderate positive correlations (Table 1).

Responsiveness: For all outcome measures other than the two cough stress tests, the mean global change scores were significantly higher in the group of participants who improved than in the group who were worse or remained the same (Table 1).

Effect size: Results showed the ICIQ-UI SF and Accident Diary had the largest effect sizes (.51 and.39). The Accident Diary showed 59% efficiency compared to the ICIQ-UI SF (the 'gold standard') in detecting a treatment effect (Table 2).

The scores needed to detect a **MCID** are presented in Table 1 as values calculated by the two methods, and the percentage change each value represents for that outcome measure.

Table 1 Summary of the testing of outcome measures

			Responsiveness	
Outcome measure(values)	**Concurrent validity#**	**Predictive validity##**	**Mean change t score**	**MCID 2 methods (change)**
ICIQ-UI SF (score 0-21)	*r*=-.44** moderate	*r* =.50** moderate	3.38**	2 2.5 (9.5% 12%)
Accident Diary (leaks per week)	*r*=-.39** moderate	*r* =.67** good	2.12*	4.5 6 (-0.7 -0.9/day)
Bother VAS (score 0-10)	*r*=-.43** moderate	*r* =.59** moderate	3.24**	1 2 (10% 20%)
ISI (score 1-8)	*r*=-.36** fair	*r* =.40** fair	2.24*	1 (12.5%)
Cough stress test (gm loss)	*r*=-.12 nil	*r* =.51** moderate	-0.31	n/a
Brace/cough stress test (gm)	*r*=-.18 nil	*r* =.33** fair	-0.77	n/a

* p<0.5 ** p<0.001

mean change scores of outcome measures correlated with global change

mean baseline scores correlated with primary end-point scores of that measure

Table 2 Effect sizes and relative efficiency of outcome measures to detect a treatment effect

Outcome measure change scores baseline to end-point	SES	Relative treatment effect	RE
Cough stress test	.16	.43	.09
Brace/cough stress test	.15	.32	.10
ICIQ-UI SF	-.51	-1.00	1
Accident diary	-.39	-3.35	.59
Bother VAS	-.27	-.37	.28
ISI	-.22	-.42	.19

SES = standardised effect size. Ratio of the treatment effect to the pooled standard deviation of these differences
Relative Treatment effect = relative magnitude of the treatment effect across different measures
RE = relative efficiency. Relative efficiency to detect a treatment effect with respect to the 'gold standard' ICIQ-UI SF, where this measure is standardised to 1

Interpretation of results

The results of this analysis indicate that the ICIQ-UI SF and Accident Diary are the most valid and responsive outcome measures for stress urinary incontinence in older women. The two cough stress tests were the least responsive measures, demonstrating a small effect size and no concurrent validity. However, these cough tests have face validity as they target the activity enshrined in the definition of stress urinary incontinence - leakage with a cough. It may be that the cough test with a pre-contraction was learned by participants in both groups from participating in the outcome assessment. It has been shown that participants can learn 'the knack', or pre-contraction, within one week [3].
Since global rating of change incorporates many constructs in one measure, strong correlations should not be expected with more homogeneous outcome measures

Concluding message

Until a 'gold standard' is established, a comprehensive suite of results should incorporate participant perceived global response to treatment. Use of the ICIQ-UI SF and Accident Diary are recommended, however the two cough stress tests performed poorly and cannot be recommended. This information can be used for planning future trials in older populations.

References

1. Neurourol Urodyn (2007) 26; 665-666
2. Foundations of Clinical Research; Norwalk, Appleton and Lange,1993 (442)
3. Obstet & Gynecol (1998) 91; 705-709

Specify source of funding or grant	**National Health and Medical Research Council of Australia, Grant number 251632**
Is this a clinical trial?	**No**
What were the subjects in the study?	**NONE**