

## NONINVASIVE OUTCOME MEASURES OF URINARY INCONTINENCE AND LOWER URINARY TRACT SYMPTOMS: A MULTICENTER STUDY OF MICTURITION DIARY AND PAD TESTS

ASNAT GROUTZ, JERRY G. BLAIVAS, DAVID C. CHAIKIN, NEIL M. RESNICK, KURT ENGLEMAN,\* DEBORAH ANZALONE,\* BRIAN BRYZINSKI\* AND ALAN J. WEIN

### ABSTRACT

**Purpose:** We assessed the test-retest reliability of a 24, 48 and 72-hour micturition diary and pad test in patients referred for the evaluation of urinary incontinence and lower urinary tract symptoms.

**Materials and Methods:** We prospectively enrolled 109 patients referred for the evaluation of lower urinary tract symptoms in our multicenter study. Patients were requested to complete a 72-hour micturition diary and pad test, and repeat each test during a 1-week interval. The test-retest reliability of various parameters of the 72-hour micturition diary and pad test was analyzed and compared. Further analysis was done to compare the test-retest reliability of 24, 48 and 72-hour studies performed on the same days after a 1-week interval. Reliability was assessed by Lin's concordance correlation coefficient (CCC) with a cutoff value of 0.7 indicating test-retest reliability.

**Results:** Of the 109 patients 106 (97%) with a median age of 64 years completed the study. The number of pads and total weight gain appeared to be reliable measures of the 24, 48 and 72-hour pad tests. For the 24-hour diary the total number of incontinence episodes was a reliable measure, while the total number of voiding episodes was marginally reliable (mean CCC 0.785 and 0.689, respectively). For the 48-hour diary the number of incontinence episodes and total number of voiding episodes were reliable measures (mean CCC 0.78 and 0.83, respectively), while for the 72-hour diary each parameter was highly reliable (CCC 0.86 and 0.826, respectively). However, an increased test period was associated with decreased patient compliance.

**Conclusions:** The 24-hour pad test and micturition diary are reliable instruments for assessing the degree of urinary loss and number of incontinent episodes, respectively. Increasing test duration to 48 and 72 hours increases reliability but is associated with decreased patient compliance.

**KEY WORDS:** urinary incontinence, reproducibility of results, incontinence pads, outcome assessment

Subjective symptoms of voiding dysfunction are common but measuring these symptoms objectively has always been a clinical challenge. Various clinical tools are available for assessing lower urinary tract symptoms, such as history, physical examination, symptom questionnaire, severity index, urine culture, micturition diary, pad test, uroflowmetry, urodynamic evaluation, electromyography, urethroscopy and radiographic or sonographic imaging of the urinary tract. Although invasive measures may provide accurate physiological information that may correlate with symptoms, such tests have several drawbacks. They are expensive, require a special setting and training, are inconvenient to patients and may be misleading due to the artificial testing environment. Conversely noninvasive measures are inexpensive, convenient, easy to perform and associated with relatively high patient compliance. However, to our knowledge noninvasive tools such as pad tests and micturition diaries have seldom been subjected to scientific testing for reproducibility.

In the research or clinical setting where the accurate evaluation of urinary symptoms and repeat measurements with time are required validity and reliability analyses of these measures are mandatory. A valid test measures what it intends to measure and a reliable test requires consistent results. Test-retest reliability is confirmed when there is a

high level of agreement among short-term replicate observations within a specific interval when a stable condition would be expected.<sup>1</sup> Data on the reliability of noninvasive, objective measures of lower urinary tract symptoms are sparse. We assessed the test-retest reliability of a 72-hour micturition diary and pad test in patients referred for the evaluation of lower urinary tract symptoms, and compared the test-retest reliability of 24, 48 and 72-hour studies.

### MATERIALS AND METHODS

We prospectively enrolled 109 patients referred for the evaluation of lower urinary tract symptoms in our multicenter study. The centers included 8 general urological practices at which clinical research studies are regularly performed. The protocol was approved by the institutional review board and all patients signed an informed consent agreement. Study inclusion criteria were age 18 years or older and the ability to communicate, understand and comply with study requirements. Exclusion criteria were a confusional state or depression, urinary tract infection, vaginitis, medication known to affect voiding, pregnancy, restricted mobility, severe grade 3 or 4 urogenital prolapse and excessive total daytime urinary output greater than 2,500 ml. or nocturnal output greater than 35% of total output due to disease, such as overt congestive heart failure, diabetes or alcohol abuse. In addition, care was taken to avoid the pad test during menstruation.

All patients underwent a detailed clinical evaluation com-

Accepted for publication April 14, 2000.

Supported by Zeneca Pharmaceuticals, and the Institute for Bladder and Prostate Research.

\* Financial interest and/or other relationship with AstraZeneca.

TABLE 1. Test-retest results of the 72-hour micturition diary

	No. Pts.	Mean $\pm$ SD Diary 1	No. Pts.	Mean $\pm$ SD Diary 2	CCC
Voiding (No. episodes):					
Diurnal	108	26.9 $\pm$ 10.3	106	26.2 $\pm$ 10.3	0.797
Nocturnal	108	4.6 $\pm$ 3.8	106	4.5 $\pm$ 4.3	0.605
Overall	108	31.5 $\pm$ 11.8	106	30.7 $\pm$ 10.9	0.826
No. incontinence episodes	108	7.6 $\pm$ 9.0	106	7.7 $\pm$ 9.8	0.860
No. urgency episodes	108	8.0 $\pm$ 11.0	106	6.6 $\pm$ 10.0	0.702
Av. daily voided vol. (ml.)	100	1,820 $\pm$ 1,013	101	1,849 $\pm$ 961	0.872

prising a complete history and physical examination, symptom questionnaire, laboratory assessment, post-void residual urine volume determination, renal and bladder ultrasound, urodynamics and urethrocytostomy. Patients were requested to complete a 72-hour micturition diary and pad test, and repeat each test after a 1-week interval, that is on the same days of the following week. All diaries and pad tests were completed before any further invasive diagnostic or therapeutic intervention was done.

The 72-hour micturition diary included certain data, such as the time of voiding, voided volume in ml., urgency and incontinence episodes, and urge, stress or unaware incontinence. Urgency was defined as a strong urge to urinate that made the patient feel that urination was about to occur and if he/she did not urinate immediately, he/she would have lost control. Patients were also requested to rate on a scale of 0—best day to 10—worst day the extent to which the diary represented symptoms. The diary was repeated after 1 week during the same 3 days of the following week.

A 72-hour pad test was performed at the time of each micturition diary. Patients were allowed to wear as many pads as they believed necessary. Used pads were then enclosed in a self-sealing plastic bag. Pad number and weight were recorded before being given to the patients and after they were returned to the research center.

Test-retest reliability was assessed using Lin's concordance correlation coefficient (CCC).<sup>2</sup> The test-retest reliability correlation coefficient reflects the degree of the correspondence of results at different time points. The most reliable measures are indicated by a reliability correlation coefficient closest to 1. Generally a minimum test-retest reliability of 0.7 is satisfactory for studies focusing on group level differences.<sup>3</sup> Corresponding to this cutoff point we considered a CCC of 0.7 or more as a statistically quantitative measure of test-retest reliability. Data are presented as the mean plus or minus standard deviation or percent according to the variables.

As the main outcome measures the test-retest CCCs of various parameters of the 72-hour micturition diary and pad test were analyzed and compared. Further analysis was done to compare the CCCs of 24 versus 48 versus 72-hour studies. The protocol specified that the micturition diary and pad test were maintained for 2 periods of 72 hours each. Therefore, we compared the test-retest reliability of the 24, 48 and 72-hour subsets. The 3 possible 24-hour periods to analyze were days 1, 2 and 3 of each 72-hour period. We compared the equivalent days of the 2 records. Similarly the 2 possible 48-hour periods to analyze were days 1 and 2 or the last 2 days of each 72-hour period.

## RESULTS

Of the 109 patients enrolled 106 (97%) 22 to 84 years old (median age 64) completed the study, including 92 women (84%) and 17 men (16%). Patient characteristics and diagnoses were similar at participating centers. Three patients withdrew from the trial, including 1 before completing any diary or pad test and 2 because of an adverse event (urinary tract infection) or failure to comply after trial period 1. The main lower urinary tract symptoms were urge, stress and mixed incontinence, and the frequency-urgency syndrome without incontinence in 34%, 22%, 26% and 16% of the study

population, respectively. The symptom questionnaire included 4 symptoms (frequency, urgency, nocturia and incontinence) that the patients were asked to rate on a scale of 0 to 10 how much each symptom bothered them. Mean bothersomeness scores were 6.51, 6.67, 5.46 and 7.40, respectively.

Table 1 shows the test-retest results of the 72-hour micturition diary. For all examined parameters except nocturnal voiding episodes the CCC was  $>0.7$ . Furthermore, the total number of voiding and incontinence episodes during 72 hours were highly reliable measures (CCC  $>0.8$ ). Generally patient compliance was 97% across all parameters. However, voided volume compliance decreased from 92% for the 24-hour to 76% for the 72-hour study. To maximize the number of patients used to determine the reproducibility of this parameter we calculated the daily average for each period rather than for a 72-hour total. Table 2 shows the results of the test-retest analysis of the 72-hour pad test. Total number and total weight gain of the pads were highly reliable (CCC 0.875 and 0.935, respectively). Overall compliance for the pad tests decreased from 96% for the 24-hour to 90% for the 72-hour study.

Table 3 shows comparisons of the 24, 48 and 72-hour studies. For the 24-hour diary the total number of incontinence episodes was a reliable measure, while the total number of voiding episodes reached marginal reliability (mean CCC 0.785 and 0.689, respectively). For the 48-hour diary the number of incontinence episodes and total number of voiding episodes were reliable measures, while for the 72-hour diary each was highly reliable (CCC  $>0.07$  and  $>0.8$ , respectively). The total number of urgency episodes reached marginal reliability only when measured for 72 hours. All pad test parameters, that is the number of pads and total weight gain, appeared to be reliable during the 24, 48 and 72-hour periods although longer test periods were associated with increased CCC values.

## DISCUSSION

Micturition diaries are widely used by clinicians interested in voiding dysfunction. Diaries make it possible to document voiding patterns in the patient environment and during various daily activities. Therefore, they are particularly important for providing objective evidence of changes in subjectively measured symptoms after therapy. In addition, close self-monitoring by micturition diary may allow the patient insights into behavioral alterations that may decrease urge incontinence episodes.<sup>4</sup> To our knowledge the formal standardization of structure, content and duration of voiding diaries is lacking.

Micturition diaries vary in duration from 24 hours to 14 days.<sup>4-8</sup> Although a 7-day chart usually includes a complete variety of social activities, most patients find it too protracted.<sup>7</sup> In our study we compared the test-retest reliability of 24, 48 and 72-hour diaries. Since daily activities may vary according to specific days of the week, patients were requested

TABLE 2. Test-retest results of the 72-hour pad test

	Mean $\pm$ SD Test 1	Mean $\pm$ SD Test 2	CCC
Total No. pads	6.86 $\pm$ 3.98	6.87 $\pm$ 4.11	0.875
Total wt. gain (gm.)	172.4 $\pm$ 317.0	159.7 $\pm$ 316.0	0.935

TABLE 3. CCC comparing the 24, 48 and 72-hour studies

	24-Hr. CCC			48-Hr. CCC		72-Hr. CCC
	Day 1	Day 2	Day 3	Days 1-2	Days 2-3	
Micturition diary (No. episodes):						
Voiding	0.673	0.704	0.690	0.781	0.770	0.826
Incontinence	0.810	0.758	0.787	0.839	0.827	0.860
Urgency	0.577	0.698	0.633	0.688	0.709	0.702
Pad test:						
No. pads	0.737	0.726	0.729	0.856	0.828	0.875
Total wt. gain	0.721	0.889	0.890	0.877	0.946	0.935

to complete the tests after a 1-week interval, that is on the same days of the following week. Results of this analysis demonstrate that the main parameters of the 24 or 48-hour diaries (total number of voiding and incontinence episodes) as well as all parameters of the pad tests may be considered reliable measures (CCC >0.7). In addition, there was a decrease in within-patient variability during the 72-hour period because extreme days were offset by typical days. For many parameters the longer test period was associated with little change in patient compliance. However, voided volume compliance decreased to about 76% when the test period was increased to 72 hours, implying that a longer test period impaired patient compliance.

The optimal structure and content of the diary are also controversial. Complex charts yield more data but are also associated with low compliance. Abrams and Klevmark suggested differentiating among frequency charts in which only the number of incontinence and voiding episodes, frequency-volume charts in which voided volume is also recorded and urinary diaries in which additional information is recorded, such as the number and type of drink, food and activities, are collected.<sup>7</sup> They recommended using frequency-volume charts in routine clinical practice and reserving the more complex charts for research projects. They also pointed out that the decreased compliance associated with complex charts may be counteracted by providing more detailed instructions and encouragement in the research environment.

However, the various components of urinary charts have seldom been tested for their test-retest reliability. Only Wyman et al previously examined the test-retest reliability of a frequency chart.<sup>8</sup> The highest test-retest correlation of a 1-week frequency chart was the number of diurnal voiding and incontinence episodes. A lower correlation was observed with the number of nocturnal voiding episodes. Inconsistency in reporting nocturnal voiding episodes may be due to several reasons, such as recall capability since nocturnal voiding is occasionally reported the following morning, sleep patterns and quality as well as hypnotic medications or the actual variability in nocturnal diuresis. The results of our study support these findings. Nocturnal voiding episodes had the lowest correlation value on test-retest analysis. We also observed that voided volume measurements during a 72-hour period were associated with lower patient compliance. Furthermore, the amount of urinary loss during incontinence episodes should also be considered when evaluating total voided volume. Therefore, we believe that one should be careful when interpreting total voided volume values unless total urinary loss is also analyzed.

More data are available on the reliability of pad tests but results are inconclusive. Victor et al examined the test-retest reliability of 2 consecutive 24-hour pad tests in 46 women, of whom 15 repeated the test after 6 to 28 days.<sup>9</sup> The correlation coefficient was 0.66 for 2 consecutive 24-hour periods and 0.90 for 2, 48-hour periods. Victor et al thought that this finding indicated that prolonging the test period may enhance test reliability. Versi et al stated that the 48-hour pad test was only marginally more reliable than the 24-hour test.<sup>10</sup> In contrast to these studies, Lose et al reported significant variation in the 24-hour pad test.<sup>11</sup> In most cases

this discrepancy would be explained by different activities during the 2 test periods. Therefore, when comparing 2, 24-hour pad tests one should confirm that the tests were performed on a representative day and during similar physical activities.<sup>12</sup> This factor should be considered, especially when the test is used for scientific purposes. We believe that the longer the duration of the pad test, the more accurate and reliable the information obtained. However, we are also aware that there is a limit to patient tolerance. Our results suggest that pad weight gain and number may be considered reliable parameters during the 24, 48 or 72-hour period. These data imply that the 24-hour pad test may be used as a reliable measure for monitoring urinary incontinence provided that it is performed during representative times and activities.

It is important to stress that using a strict cutoff point to outline reliable parameters is arbitrary unless the clinical relevance of test-retest data is independently assessed. For example, a difference of 100 ml. in urinary loss from 300 to 200 ml./24 hours may have no clinical relevance, whereas a difference of 100 to 0 ml. has obvious relevance. According to our cutoff point CCC >0.7 indicates test-retest reliability, although it is a mistake to infer that lower values are unreliable. Correlation coefficient values may have various strengths from little if any, low, moderate and high to very high. According to Aday<sup>3</sup> a minimum test-retest coefficient of 0.7 is satisfactory for studies focusing on group level differences<sup>3</sup> but a value of 0.9 or greater is preferred when the emphasis is on changes in individuals with time. Our results imply that test-retest reliability may be enhanced by extending the test period. However, prolonged test periods may also be associated with lower compliance. Therefore, we believe that in the daily clinical setting 24-hour studies are adequate. Furthermore, in 2 subsequent studies these instruments were valid outcome measures for assessing anti-incontinence therapy.<sup>13,14</sup> For research projects one may prefer to use 72-hour studies but care should be taken to ensure high compliance.

#### CONCLUSIONS

Data on the optimal structure, content and duration of noninvasive measures of voiding dysfunction are sparse and nonconclusive. Although test-retest reliability is clearly enhanced by prolonging the test period, complex and prolonged studies are also associated with impaired compliance. We believe that these data support the use of a 24-hour pad test and micturition diary for assessing the degree of urinary loss and number of incontinent episodes, respectively.

#### REFERENCES

- O'Leary, M. P., Barry, M. J. and Fowler, F. J., Jr.: Hard measures of subjective outcomes: validating symptom indexes in urology. *J Urol*, **148**: 1546, 1992
- Lin, L. I.: A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**: 255, 1989
- Aday, L. A.: *Designing and Conducting Health Surveys*, 2nd ed. San Francisco: Jossey-Bass Publishers, pp 53-54, 1996
- Burgio, K. L., Locher, J. L., Goode, P. S. et al: Behavioral vs. drug

- treatment for urge urinary incontinence in older women: a randomized controlled trial. *JAMA*, **280**: 1995, 1998
5. Larson, G. and Victor, A.: The frequency/volume chart in genuine stress incontinent women. *Neurourol Urodyn*, **11**: 23, 1992
  6. Rabin, J. M., McNett, J. and Badlani, G. H.: Computerized micturition diary. *Neurourol Urodyn*, **12**: 541, 1993
  7. Abrams, P. and Klevmark, B.: Frequency volume charts: an indispensable part of lower urinary tract assessment. *Scand J Urol Nephrol*, suppl., **179**: 47, 1996
  8. Wyman, J. F., Choi, S. C., Harkins, S. W. et al: The urinary diary in evaluation of incontinent women: a test-retest analysis. *Obstet Gynecol*, **71**: 812, 1988
  9. Victor, A., Larsson, G. and Asbrink, A. S.: A simple patient-administered test for objective quantitation of the symptom of urinary incontinence. *Scand J Urol Nephrol*, **21**: 277, 1987
  10. Versi, E., Orrego, G., Hardy, E. et al: Evaluation of the home pad test in the investigation of female urinary incontinence. *Br J Obstet Gynaecol*, **103**: 162, 1996
  11. Lose, G., Jorgensen, L. and Thunedborg, P.: 24-hour home pad weighing test versus 1-hour ward test in the assessment of mild stress incontinence. *Acta Obstet Gynecol Scand*, **68**: 211, 1989
  12. Mouritsen, L., Berild, G. and Hertz, J.: Comparison of different methods for quantification of urinary leakage in incontinent women. *Neurourol Urodyn*, **8**: 579, 1989
  13. Chaikin, D. C., Blaivas, J. G., Rosenthal, J. D. et al: Results of pubo-vaginal sling for stress incontinence: a prospective comparison of 4 different instruments for outcome analysis. *J Urol*, **162**: 1670, 1999
  14. Groutz, A., Blaivas, J. G. and Rosenthal, J. D.: A simplified urinary incontinence score for the evaluation of treatment outcomes. *Neurourol Urodyn*, **19**: 127, 2000