14

Clarkson B[1], Drinnan M[1], Caffarel J[1], Robson W[1], Pickard R[1], Griffiths C[1], McArdle F[1]
1. Newcastle upon Tyne Hospitals NHS Trust

EVALUATION OF NON-INVASIVE BLADDER PRESSURE AND URINE FLOW MEASUREMENT SYSTEMS FOR ROUTINE CLINICAL USE: ASSESSMENT OF INTERPRETER AND TEST-RETEST VARIABILITY

Hypothesis / aims of study

The development of the penile cuff test has allowed non-invasive measurement of bladder pressure. This can be combined with urinary flow rate on a pressure-flow (cuff-test) nomogram to categorise bladder outlet obstruction [1]. To be diagnostically useful however, test results must be reproducible within an individual (test-retest reliability) and between different operators (interpreter variability). Sites using the system were asked to recruit study participants to undertake two pressure-flow studies and analyse the results according to the guidelines given in the manual issued with the systems.The study aims to compare analysis from 'novice' sites using the system, with the 'expert' analysis of our centre. This will quantify variability between interpreters and may help to improve on the guidelines and tuition provided with the measurement system. The amount of test-retest data available for analysis has, to date, been limited, and the study will therefore evaluate the agreements of tests and retests for a large cohort of non-invasive studies in multiple sites. This will allow comparison with the test-retest statistics for conventional urodynamics. An evaluation can also be made of change in nomogram category caused by differing observers and multiple tests using the cuff-test nomogram [1].

Study design, materials and methods

Penile cuff devices were installed in six urodynamic laboratories without previous experience of non-invasive bladder pressure measurement. The urodynamic technicians were given 1 day of training in the setting up, recording and clinical use of the device, and an interpretation guidance manual. More than 70 men with LUTS were recruited per centre to take part in two pressure-flow studies of approximately 4 weeks apart. Staff at the 'novice' sites measured cuff interruption pressure and maximum flow rate, then 'expert' staff in our centre interpreted the same data blindly. Bland-Altman [2] analysis was used to quantify agreement between 'observers' - the remote 'novice' sites and our 'expert' centre. The test-retest results were analysed similarly using the same observer, and compared with those for conventional urodynamics as the gold standard. As outcome is based on the results from the cuff-test Nomogram, both sets of data were plotted on this to show how well the data agrees with respect to nomogram category. Any change in category due to either inter-observer or test-retest variability was assessed graphically.

Results

The results discussed here are those of a single remote site and our centre with 24 test-retest and 44 inter-observer subjects. 8 subjects were discarded due to unusable traces and the remainder were unavailable for analysis.

	Flow (ml/s)		**Pressure (cmH_20)**	
	Mean (μ)	**2SD**	**Mean (μ)**	**2SD**
Inter-observer	-0.3	3.0	-3.3	24.4
Test-retest	0.2	7.4	5.9	44.0

Table 1, Bland-Altman analysis for a single site: Mean shows bias of measurements and 2SD shows the closeness of agreement. An inter-observer negative mean represents a lower estimate on our centre's part. A test-retest positive mean represents a higher second test value.

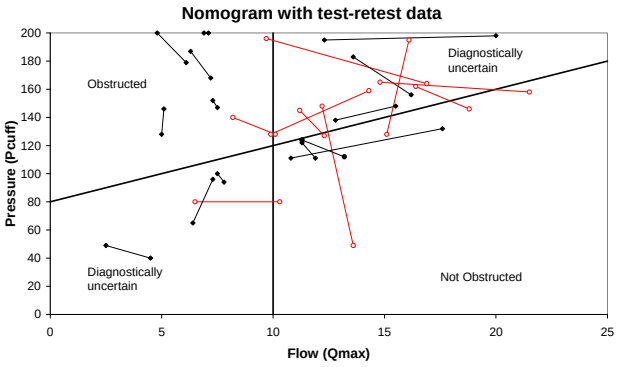

Fig 1, Test-retest data on the cuff-test nomogram, category change in red (open circles).

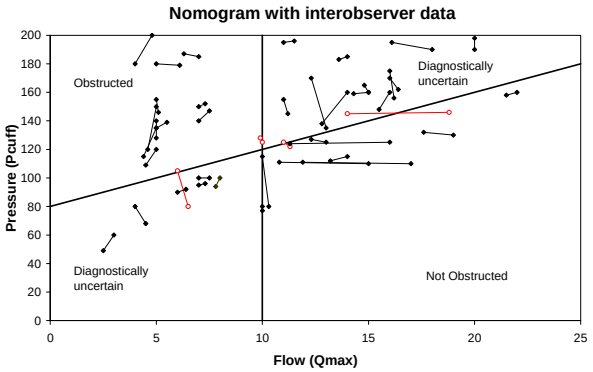

Fig 2, Inter-observer changes on the nomogram - category change in red (open circles).

2nd test/ Remote site's interpretation	**Test-Retest**			**Inter-observer**		
	1st test			**Our centre's interpretation**		
	Obstructed	Uncertain	Not Obstructed	Obstructed	Uncertain	Not Obstructed
Obstructed	5	3	0	12	1	0
Uncertain	0	6	2	1	18	0
Not Obstructed	0	4	4	0	2	10

Table 2, Contingency table showing change in nomogram category between tests and observers.

Test-retest results showed agreement of pressure within 44cmH_2O and flow within 7.4ml/s (Table 1, SD=2.5ml/s, within patients as calculated in [3]), which is comparable to the standard deviation quoted in [3] of 2ml/s. Figs 1 & 2 show the test-retest and inter-observer data plotted in pairs for each patient, the red (open circles) showing where category has changed. Table 2 shows the changes in category by number of test population. From this 38% of subjects changed category (from or to 'diagnostically uncertain') between the test and the retest. A much smaller percentage (9%) of inter-observer data changed nomogram category.

Interpretation of results

Bland-Altman analysis (Table 1) clearly shows much better agreement between observers than test-retest results, which is to be expected. Flow rates agreed very well (within 3ml/s) between observers, and obvious outliers were attributed to misinterpretation of spikes (flow surges) on cuff deflation. Many of the inter-observer pressure differences were again attributable to interpretation technique. Test-retest variability was much larger than the inter-observer, illustrating that exact agreement of such a test on different days is difficult, as is suggested by [3]. This also suggests that inter-observer variability is not the main source of error between tests. Agreement of test-retest results was analysed in a more qualitative way by using the cuff-test nomogram and comparing change in diagnostic outcome. Change in nomogram category was large for test-retest, but of these more than half had one test very close to a boundary, and none crossed from 'not obstructed' to 'obstructed' (see fig 1). This is comparable to the change of 39% in nomogram category quoted in [3] for the AG nomogram due to test-retest variability.

Concluding message

On revision of the Bland Altman data, agreement between observers could be improved by providing more comprehensive tuition in the analysis guidelines accompanying the machines. This is unlikely to improve test-retest results significantly, but this is a limitation of urodynamics, as illustrated by the comparison to test-retest studies on conventional urodynamics [3].

References

[1] Journal of Urology, Vol. 174, 1323–1326, October 2005
[2] Lancet, 1986; i: 307-310
[3] Neurourology and Urodynamics 19:637–656 (2000)

FUNDING: None

HUMAN SUBJECTS: This study was approved by the Newcastle and North Tyneside Ethics Committee and followed the Declaration of Helsinki Informed consent was obtained from the patients.