317

Zimmern P[1], Nager C[1], Albo M[1], Fitzgerald M[1], Mohr B[1], McDermott S[1]
1. Urinary Incontinence Treatment Network (UITN)

INTERPRETATION OF PRESSURE-FLOW STUDIES: INTER-RATER RELIABILITY BETWEEN CENTRAL REVIEWERS IN SISTER (STRESS INCONTINENCE SURGICAL TREATMENT EFFICACY TRIAL), A MULTI-CENTER STRESS INCONTINENCE SURGERY TRIAL

Hypothesis / aims of study

As part of an on-going effort to study the inter-rater reliability (IRR) of urodynamic studies (UDS) performed at multiple centers, we established standard testing procedures and interpretative guidelines that resulted in excellent IRR among local reviewers (LR) and central reviewers (CR) for non-invasive flow and CMG quantitative measurements (1,2,3). However, the pressure-flow study (PFS) remains a challenge (2). Despite use of standardized interpretation guidelines that were internally developed, approved by an outside expert, and tested for IRR in a prior pilot study (2), there was still a need to refine the guidelines for PFS. We report here on our most recent study of IRR among four experienced central physician reviewers for PFS parameters using refined guidelines for PFS interpretation.

Study design, materials and methods

To enhance IRR for PFS among CR, refined PFS interpretive guidelines including more precise definitions of measurements for PFS baseline, delta Pdet, maximum flow and 5 "unambiguous" voiding patterns were added to prior interpretive guidelines (2). The refined guidelines were then used in an IRR study of CR interpreting 36 recent urodynamic tracings from 13 UITN-certified testers from 9 centers. Each CR reviewed all 36 tracings. CR categorized variables as "invalid" if specified technical standards were not met, or the signal pattern suggested implausible values because of technical deficiencies. Cases were classified as plausible only if all of the following criteria were met for all 4 CR: Pdet at both CMG and PFS baseline between –5 and +10, and pre-void cough pressure transmission ratio (Pabd/Pves) $\geq$70%. Intraclass Correlation Coefficients (ICC) were calculated for numerical variables. The percentage of cases where all CR agreed and a kappa statistic (k) were calculated for the voiding pattern. Acceptable agreement was defined as ICC or $k \geq 0.6$. P-values were obtained as a test of the hypothesis that reviewer means are equal.

Results

1. The 4 CR achieved acceptable agreement on whether the cases were plausible (kappa=0.76). For 78% of the 36 cases, all 4 CR agreed on plausibility, and in 100% of the cases, at least 3 of 4 CR agreed.
2. All 4 CR rated 18 of 36 cases (50%) plausible and 10 of 36 (28%) implausible. Nine of the 36 cases (25%) were judged implausible due solely to Pdet at PFS baseline.
3. Using data from the 18 plausible cases only, the CR achieved excellent IRR for the quantitative PFS measures (see table). There were no significant differences between the CR means indicating consistency across CR.
4. Among the plausible cases, agreement for voiding pattern remained poor (k=0.43). The 4 CR agreed on voiding pattern for only 39% of the cases; for 61% of the cases at least 3 out of 4 CR's agreed.

Summary of IRR among 4 Central Reviewers for selected quantitative PFS variables.

Description	**ICC**	**"Acceptable" Agreement[1]**
Pves at PFS baseline	0.97	Y
Pabd at PFS baseline	0.94	Y
Qmax	0.93	Y
Pves at Qmax	0.99	Y
Pabd at Qmax	0.99	Y
Pdet at Qmax	0.998	Y
Delta Pdet (Pdet at Qmax – Pdet at PFS baseline)	0.997	Y

[1] "Acceptable" agreement is ICC>=0.60 for continuous variables and kappa>=0.60 for categorical variables.

Interpretation of results

The plausibility criteria resulted in a large percentage of cases being considered implausible. Among plausible cases, defined by strict criterion that resulted in a large number of cases being removed, the refined guidelines led to excellent inter-rater reliability for PFS quantitative variables. Agreement for voiding pattern did not achieve acceptable IRR.

Concluding message

Reliable interpretation of qualitative PFS variables remains a challenge. Further refinements in both testing procedures and interpretive guidelines are currently under consideration.

References

1 SUFU 2003: *Urodynamic challenges posed by a multicenter study of surgery for female stress urinary incontinence*.

2 SUFU 2004: *Urodynamic inter-rater reliability in SISTEr (Stress Incontinence Surgical Treatment-Efficacy Trial)*. (abstract #7)

3 *Urodynamic inter-rater reliability between local and central physician reviewers for the filling cystometrogram in SISTEr*. (submitted ICS 2004)

FUNDING: NIH/NIDDK UO1 DK: 58225, 58229, 58231, 58234, 60397, 60380, 60393, 60395, 60397, 60401