

URODYNAMIC INTER-RATER RELIABILITY BETWEEN LOCAL AND CENTRAL PHYSICIAN REVIEWERS FOR THE FILLING CYSTOMETROGRAM IN SISTER (STRESS INCONTINENCE SURGICAL TREATMENT EFFICACY TRIAL)**Hypothesis / aims of study**

Few urodynamic outcomes or inter-rater reliability reports exist from urodynamic studies (UDS) performed at multiple centers. Prior reports from our group demonstrated the need for both standardized testing and interpretation guidelines¹, and a recent report of an inter-rater reliability study using these interpretation guidelines established excellent intraclass correlation for cystometrogram (CMG) numerical values among four experienced central physician reviewers². As part of an on-going quality control (QC) effort to study the inter-rater reliability of UDS performed at multiple centers, we now report a study comparing agreement level between the local physician reviewers (LR) and the four central physician reviewers (CR) for the filling CMG.

Study design, materials and methods

A standardized testing protocol was implemented study-wide. Standardized urodynamic interpretation guidelines were internally developed, approved by outside expert review, tested for inter-rater reliability in a prior pilot study², and formally adopted (7/25/03). Subsequently, 36 urodynamic tracings from 9 centers and 13 different UITN-certified testers were randomly selected. These tracings were read and interpreted by 11 different LR. Each of the 4 CR reviewed 9 randomly assigned CMG tracings. Tracings were electronically transferred to CR from the UITN UDS central repository held at the UITN Coordinating Center. No CR reviewed tracings from his/her own center. LR and CR were instructed to categorize values as "invalid" if specified technical quality assurance standards were not met, or the signal pattern suggested implausible values because of technical deficiencies. Intraclass Correlation Coefficients (ICC) were calculated treating the data as if the LR's were one rater and the CR's another. Acceptable agreement was defined a priori as $ICC \geq 0.6$. p-values were obtained as a test of the hypothesis that CR mean and LR mean are equal.

Results

Filling CMG baseline pressure measurements, Valsalva leak point pressure (VLPP) measurements, and volume and pressure measurements at maximum cystometric capacity (MCC) had excellent ICC ranging from 0.74 to 0.99 (table). There were no significant differences between the LR and CR means, indicating excellent agreement.

Description	Intra-class correlation
Pves at CMG baseline	0.96
Pabd at CMG baseline	0.96
Pdet at CMG baseline	0.97
Mean VLPP	0.99
Volume at MCC	0.74
Pves at MCC	0.96
Pabd at MCC	0.96
Pdet at MCC	0.94

Although detrusor overactivity was also noted by both LR and CR, the relative few instances observed did not permit meaningful statistical calculations.

Interpretation of results

The urodynamic interpretation guidelines developed for the SISTEr protocol resulted in excellent agreement for critical filling CMG variables between 11 local and 4 central physician reviewers at 9 American medical centers.

Concluding message

This recent QC study demonstrates that with proper QC measures in place, i.e. well-trained, certified testers; a standard UDS testing protocol with clear annotation of all events occurring during UDS; and a set of standardized interpretative guidelines, excellent inter-rater reliability between LR and CR can be achieved for numerical filling CMG variables.

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).*

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