238

Norderval S[1], Dehli T[1], Vonen B[2]

1. Dep. of Digestive Surgery, University Hospital of Northern Norway, Tromsø, Norway, 2. Dep. of Digestive surgery, University Hospital of Northern Norway and Dep. of Surgery, Institute for Clinical Medicine, Medical Faculty, University of Tromsø, Tromsø, Norway

SCORING SYSTEMS FOR CLASSIFICATION OF ANAL SPHINCTER DEFECTS USING THREE-DIMENSIONAL ENDOANAL ULTRASONOGRAPHY: INTRA- AND INTEROBSERVER AGREEMENT AND CORRELATION WITH ANAL INCONTINENCE

Hypothesis / aims of study

Three-dimensional endoanal ultrasonography (3D EAUS) has opened for assessment of the anal canal structures also in the longitudinal plane, and scoring systems for classification of anal sphincter defects based on three-dimensional assessment have been developed. The aim of this study was to determine the degree of intra- and interobserver agreement for an experienced and inexperienced ultrasonographer using two scoring systems for 3D EAUS assessment of anal sphincter defects, and to relate the extent of defects to the degree of anal incontinence.

Study design, materials and methods

All women aged between 20 and 40 years investigated in our outpatient clinic from January 2003 to December 2005 with a complete 3D EAUS using a 10 MHz rotating probe comprise the sample of this study. The EAUS datasets were assessed independently by two ultrasonographers; one experienced who had assessed more than 400 3D EAUS, and one inexperienced who had performed approximately 50 assessments prior to the present study. The EAUS datasets were retrieved from the hard disc of the scanner and selected by both ultrasonographers more than two months before the first formal assessment in the present study was undertaken. Assessments were performed blinded with respect to all other data. The EAUS datasets were assessed twice by both ultrasonographers, with an interval of two months between first and second assessment. Cases with intraobserver disagreement were resolved by a third (final) assessment one month after second assessment. The final assessment from each ultrasonographer was used to determine the degree of interobserver agreement and correlation with incontinence.

Sphincter defects were classified according to our EAUS defect score [1] and the Starck score [2]. Both scoring systems classify the extent of internal and external anal sphincter defects in three dimensions. Our EAUS defect score is a continuous score from 0 to 7, where score 0 is no defect and score 7 represents a maximal defect. The Starck score runs from score 0 to 16, with score 0 being no defect, score 3 the smallest possible defect and score 16 the maximal defect.

The continence status at the time of the EAUS recording was graded according to the
St. Mark's score and information was retrieved from the patient's medical record after the EAUS assessments were completed. Agreement was expressed with weighted *kappa* statistics, and rated as very good (*kappa* 0.81-1.0), good (*kappa* 0.61-0.80) moderate (*kappa* 0.41-060), fair (*kappa* 0.21-0.40) slight (*kappa* 0-0.20) and poor (*kappa* <0)

Results

EAUS datasets of 55 women were included. The EAUS was performed as a routine follow-up after primary repair of obstetric sphincter tears in 30 women and after secondary sphincter repair in 2 women. The remaining 23 women were investigated due to anal incontinence after primary repair of obstetric sphincter tears (13 women), after vaginal delivery without known sphincter injuries (7 women) and after secondary sphincter repair (3 women).

Intraobserver agreement for experienced and inexperienced ultrasonographer is shown in table 1 (95% confidence interval in parenthesis and identical scores between the various assessments expressed with per cents).

Table 1   Intraobserver agreement for our EAUS defect score and the Starck score

	Experienced ultrasonographer		Inexperienced ultrasonographer	
Our EAUS defect score	Weighted *Kappa*	Identical score	Weighted *Kappa*	Identical score
Assessment 1 vs assessment 2	0.75 (0.65-0.83)	56%	0.58 (0.47-0.70)	35%|
Assessment 1 vs final assessment	0.86 (0.80-0.92)	71%	0.72 (0.63-0.82)	49%
Assessment 2 vs final assessment	0.87 (0.79-0.95)	80%	0.82 (0.73-0.91)	65%
Starck score				
Assessment 1 vs assessment 2	0.73 (0.64-0.81)	35%	0.62 (0.51-0.73)	31%
Assessment 1 vs final assessment	0.84 (0.77-0.90)	55%	0.74 (0.64-0.84)	40%
Assessment 2 vs final assessment	0.84 (0.78-0.91)	55%	0.84 (0.77-0.91)	58%

Interobserver agreement between the two ultrasonographers regarding our EAUS defect score was obtained in 42% with a weighted *kappa* of 0.65 (95% confidence interval 0.55-0.76). Interobserver agreement regarding the Starck score was present in 25% with a weighted *kappa* of 0.74 (95% confidence interval 0.67-0.80).

Incontinence was present in 36 of the 55 women (65%). Median St. Mark's score was 4.0 (range 0 – 21), and 18 women reported a St. Mark's score ≥ 8.

The degree of incontinence correlated with our EAUS defect score (p=0.003 for the experienced and p=0.018 for the inexperienced ultrasonographer) and the Starck score (p=0.001 for the experienced and p=0.015 for the inexperienced ultrasonographer).

Interpretation of results

Intraobserver agreement for the experienced ultrasonographer was good for our EAUS defect score and the Starck score (weighted *kappa* 0.75 and 0.73 respectively). Intraobserver agreement for the inexperienced ultrasonographer was by definition moderate for our EAUS defect score and good for the Starck score, although the results were close (weighted *kappa* 0.58 and 0.62 respectively).

The intraobserver agreement for the inexperienced ultrasonographer was better between the second and the final assessment than between the first and the final assessment for both the Starck score and for our EAUS defect score (table 1). A similar improvement was not seen for the experienced ultrasonographer, and may reflect a learning curve for the inexperienced ultrasonographer. Since the inexperienced ultrasonographer had performed approximately 50 three-dimensional EAUS assessments prior to the study, our results may indicate that an examiner unaccustomed to ultrasonography has to perform at least 100 three-dimensional EAUS-assessments in order to achieve a high degree of repeatability when using the scoring systems described.

Interobserver agreement was good for both scoring systems, being somewhat better for the Starck score than our EAUS defect score. However, the frequency of identical scores between the two ultrasonographers was higher for our EAUS score than for the Starck score.

The extent of sphincter defects expressed by both scoring systems was correlated to the severity of incontinence symptoms.

Concluding message
Our EAUS defect score and the Starck score are reproducible scoring systems for ultrasonographic classification of anal sphincter defects. A possible learning curve for 3D EAUS assessment has been described. The extent of sphincter defects correlates with the degree of incontinence in our study group.

References
[1]   Ultrasound Obstet Gynecol (2008) 31; 78-84.
[2]   Ultrasound Obstet Gynecol (2003) 22; 609-15.

Specify source of funding or grant	**The study has solely been financed by the Dep. of Digestive Surgery, University Hospital of Northern Norway, Tromsø, Norway**
Is this a clinical trial?	**No**
What were the subjects in the study?	**HUMAN**
Was this study approved by an ethics committee?	**Yes**
Specify Name of Ethics Committee	**Regional Commite for Medical Research Ethics, Northern Norway**
Was the Declaration of Helsinki followed?	**Yes**
Was informed consent obtained from the patients?	**No**