220

Richter F[1], Stær-Jensen J[1], Hoff Brækken I[2], Bø K[2], Ellstrøm Engh M[1]
1. *Akershus University Hospital, Department of Obstetrics and Gynaecology, Lørenskog, Norway,* ***2.*** *Norwegian School of Sport Sciences, Department of Sports Medicine, Oslo, Norway*

LEARNING AND INTER-RATER RELIABILITY STUDY OF THREE AND FOUR DIMENSIONAL TRANSPERINEAL ULTRASOUND IMAGING

Hypothesis / aims of study

To include three and four dimensional (3D/4D) transperineal ultrasound in clinical routine, the technique must be easy to perform, and the imaging reliable and easy to interpret. Both test-retest, intra- and inter-rater repeatability for the analyses of recorded 3D/4D ultrasound volumes have been found acceptable (1,2), but as far as we have ascertained there are no studies monitoring the learning process, including investigating the reliability of the whole procedure with instructions to the patient, recording and analysing of the ultrasound volumes. The primary aim of the present study was to study the learning process of performing a 3D/4D transperineal ultrasound examination and analysing the imaging. A second aim was to perform an inter-rater reliability study between two independent ultrasound examiners.

Study design, materials and methods

This prospective study was conducted from June to September 2010. Twenty women (two women at 22, eight women at 37 weeks of gestation and ten women 6 weeks after delivery) were included from a current running study designed to investigate the role of the pelvic floor muscles in women during pregnancy and after childbirth.

A physician (P) with no experience of 3D/4D ultrasound examinations, was introduced into an existing research team, on which an experienced gynaecologist (G), had performed more than 500 ultrasound examinations and analyses. The learning process for P was monitored. When P achieved acceptable agreement with G, an inter-rater study was performed between the two examiners. To estimate the size of the levator hiatus (LH) a 3D/4D transperineal ultrasound examination was performed using a GE Voluson E_8 system (GE Medical Systems) with 4-8 MHz curved array 3D/4D ultrasound transducer (RAB4-8l/obstetric). The ultrasound volumes were analysed in the axial plane at level of minimal hiatal dimension using the 4D View (v10.0). The measurements assessed were: area (LHarea), anterior-posterior (LHap) and transverse diameter (LHrl) and circumference (LHcirc), the bony part of the hiatus (Pubic arc) and levator urethral gap (LUG).

During three one-hour teaching sessions P was taught how to give instructions and record ultrasound examinations consisting of three contractions and three Valsalva manoeuvres. Thereafter P was instructed in off-line analysis for four five-hours sessions. Subsequently P and G each performed 3D/4D ultrasound examinations on 20 women in alternating order. To evaluate the learning process for the off-line analyses, the volumes recorded by G were used to compare measurements of LH dimensions between the two examiners. Agreement was estimated using intra-class correlation coefficient (ICC) in blocks of 10 volumes. When good to very good agreement was achieved for the majority of measurements, P analysed the ultrasound volumes recorded by herself and G, thus assessing, the learning process for the ultrasound examination. Finally, the inter-rater reliability of the technique was tested by having the two examiners each analyse the results of their own 20 examinations separately. During the study the examiners were blinded to previously collected data and to each others' results. The inter-rater reliability was analysed using ICC. The Bland-Altman-Scale was used to classify the reliability of the values. ICC values under 0.2 were considered poor, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 good and 0.81-1.00 very good. To quantify the inter-rater agreement, the differences between average values (bias) and SD were calculated.

Results

Learning how to perform ultrasound examinations: After having examined the first 10 women P achieved good to very good agreement with G in all LH measurements. In the next ten ultrasound examinations agreement was very good in all measured LH dimensions (Fig 1).

Learning how to analyse recorded volumes: After having analysed the first ten ultrasound volumes good to very good agreement between P and G was achieved for all LH measurements except for LUGright (ICC=0.47; SD=0.28), LUGleft (ICC=0.49; SD=0.21) and Pubic arch (ICC=0.13; SD=0.68). After having analysed the next ten volumes good to very good agreement was achieved for all LH measurements, except for Pubic arch (ICC=0.39; SD=0.76) (Fig 2).

Fig 1

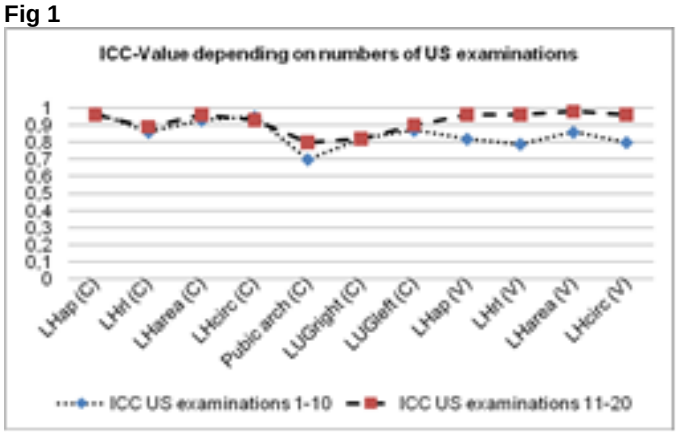
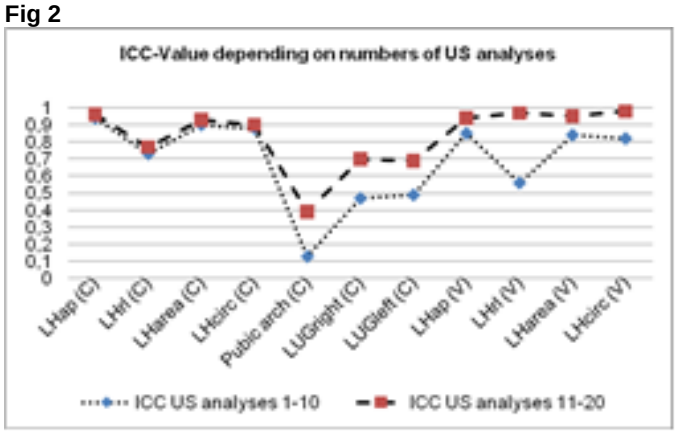


Fig 2



Inter-rater reliability study: One dataset showed insufficient imaging quality on Valsalva and was therefore excluded from further analysis. All measurements of the levator hiatus dimensions execpt Pubic arch showed very good reliability at contraction and Valsalva (Table 1). The measurements of Pubic arch showed good reliability. There was a significant systematic observer bias in the measurement of LHcirc during contraction ($p < 0.05$).

Table 1: Inter-rater differences in measurement of the levator hiatus (LH) dimensions at contraction (C) and Valsalva (V), *significant bias ($p = 0.01$)

Parameter	Mean P (95% CI)	Mean G (95% CI)	N	ICC	Bias	SD
LHap (C)	4.4 (4.1; 4.6)	4.3 (4.1; 4.6)	20	0.95	-0,04	0,19
LHrl (C)	3.6 (3.3; 3.9)	3.7 (3.4; 3.9)	20	0.89	-0.08	0.27
LHarea (C)	11.1 (10.0; 12.1)	11.5 (10.5; 12.4)	20	0.88	0.41	1.02
LHcirc* (C)	12.3 (11.7; 12.9)	12.7 (12.1; 13.2)	20	0.89	0.39	0.47
Pubic arch (C)	4.75 (4.5; 5.0)	4.6 (4.2; 5.0)	20	0.67	-0.14	0.56
LUGright (C)	1.9 (1.8; 2.1)	1.9 (1.8; 2.0)	20	0.81	0.01	0.18
LUGleft (C)	1.9 (1.7; 2.1)	1.9 (1.7; 2.1)	20	0.82	0.01	0.22
LHap (V)	6.2 (5.6; 6.8)	6.2(5.5; 6.8)	19	0.92	-0.01	0.51
LHrl (V)	4.5 (4.1; 4.9)	4.5 (4.2; 4.9)	19	0.94	-0.01	0.28
LHarea (V)	20.9 (16.6; 25.2)	21.9 (17.0; 26.9)	19	0.97	0.76	2.42
LHcirc (V)	17.0 (15.6; 18.5)	17.3 (15.6; 18.9)	19	0.94	0.29	1.11

Interpretation of results

Our data suggest that the 3D/4D ultrasound examination is easy to learn and can be an accessible tool for the clinician. The examination is easier to learn than the analysis process. Particularly defining the insertion of the puboviscerali muscle on the pubic ramus, which is required to determine LUG and Pubic arch seems to be demanding. One limitation of the learning study was that only one inexperienced examiner was included. As far as we have ascertained the inter-rater reliability study is the first study using two independent examiners and the agreement was excellent. Together with data from previous studies we suggest that this method can be considered a reliable research tool.

Concluding message

Three and four dimensional transperineal ultrasound is a reliable technique that can be learned in a relatively short period of time.

References

1. Brækken, I.H., Majida, M., Ellstrøm-Engh,M., Dietz, H.-P., Umek, W., Bø, K., (2008) Test-retest and intra-observer repeatability of two-, three- and four-dimensional ultrasound of pelvic floor muscle anatomy and function; Int Urogynecol J Pelvic Floor Dysfunct, 19: 227- 235
2. Majida, M., Brækken, I.H., Umek, W., Bø, K., Saltyte, B.J., Ellstrøm-Engh, M., (2009) Interobserver repeatability of three- and four-dimensional transperineal ultrasound assesment of pelvic floor muscle anatomy and function; Ultrasound Obstet Gynecol, 33:567- 573

Specify source of funding or grant	**none**
Is this a clinical trial?	**Yes**
Is this study registered in a public clinical trials registry?	**No**
Is this a Randomised Controlled Trial (RCT)?	**No**
What were the subjects in the study?	**HUMAN**
Was this study approved by an ethics committee?	**Yes**
Specify Name of Ethics Committee	**Ethic committee approval had been obtained for the main study by the Norwegian Regional Medical Ethics Committee REK Sør-Øst D 2009/170 REK Sør-Øst A 2009/289a**
Was the Declaration of Helsinki followed?	**Yes**
Was informed consent obtained from the patients?	**Yes**