

## DOES THE USE OF 3D ENDOANAL ULTRASOUND IMPROVE INTER-OBSERVER AGREEMENT COMPARED WITH 2D ULTRASOUND IN PATIENTS WITH FAECAL INCONTINENCE?

HIMANSHU WADHAWAN, MRCC<sup>1,2</sup>, LYNNE F. SMITH, MSc<sup>2</sup>, STEVEN R. BROWN, FRCS<sup>2</sup>

School of Medicine and Biomedical Sciences University of Sheffield, Beech Hill Road, Sheffield, S20 2RX United Kingdom<sup>1</sup>

Sheffield Teaching Hospitals NHS Foundation Trust Northern General Hospital Herries Road, Sheffield, S7 5AU, United Kingdom<sup>2</sup>

Endoanal ultrasound (EAUS) is used in the assessment of the anal sphincter in patients with faecal incontinence. However, interpretation is very operator dependent. 3D technology allows capture of the image of the whole anal canal in three dimensions and manipulation of the image by others not carrying out the scan reducing operator dependence.

**The aim of the study** was to determine whether inter-observer agreement is better using 3D technology compared with 2D images.

**Material and methods.** For the first part of the study inter-observer variability was compared using a small number of patients and a large number of interpreters. Study images of ten randomly selected patients undergoing endoanal ultrasound for faecal incontinence were obtained in 2D format and using 3D technology. Images were interpreted by 4 specialists (defined as personnel who regularly reported scans) and 9 non-specialists with an interest in coloproctology (1 radiologist and 8 colorectal surgeons). For the second part of the study images of forty patients were randomly selected in both formats and interpreted by only 2 specialists. Each image was graded as normal, showing internal sphincter injury, external sphincter injury or a combination.

**Results.** There appeared to be minimal to no advantage for the 3D format over the 2D format for any of the groups in terms of inter-observer variability. For interpretation of the 10 images as expected the inter-observer agreement was low for the non-specialist group ( $k = 0.11$  for 2D and  $k = 0.16$  for 3D) but was surprisingly only moderate for the specialists alone ( $k = 0.42$  for 2D and  $k = 0.44$  for 3D). In the second part of the study there was a higher overall agreement and a slight improvement in inter-observer variability with the 3D format. Agreement was moderate for 2D and substantial for 3D ( $k = 0.60$  and  $k = 0.67$  respectively).

**Conclusions.** Despite the ability to view the whole anal canal in different planes, the 3D technology appears to only slightly improve inter-observer agreement and only in expert hands.

**Key words:** imaging, endoanal ultrasound, three-dimensional, anorectal

Imaging the anal sphincters with anal endosonography allows interpretation of the structure and integrity of the sphincter complex and is the test most likely to alter management of patients with faecal incontinence (1). Unfortunately, interpretation can be difficult and is very operator dependent. The external sphincter has an echogenicity similar to that of the longitudinal muscle medially and the ischioanal fat laterally making reproducible assessments of

thickness a problem. In addition, anatomical variations in the structure of the sphincter complex may result in discrepancies in interpretation particularly as conventional ultrasound only allows image visualization in two planes. Differentiating abnormalities from normal anatomy depends on the skill of the operator and their ability to interpret the sphincter configuration in all planes. This can only really be done at the time of the test. Al-

though photographic images are usually taken at standard sites within the anal canal, subsequent interpretation is even more difficult.

The recent development of 3D anal endosonography allows reconstruction of a multiplanar image of the anal canal. These images can be stored on a hard disc and used subsequently by reviewers. This allows more accurate delineation of sphincter complex integrity despite not being present at the time of the test.

We hypothesised that the use of a 3D ultrasound machine was less subjective and improved the inter-observer variability of interpretation of images for those undergoing anal endosonographic assessment for faecal incontinence.

## MATERIAL AND METHODS

The study was carried out in two parts. Firstly an independent researcher who had no knowledge of the ultimate scan diagnosis randomly selected ten patients with faecal incontinence who had undergone ultrasound assessment of their anal sphincters at a teaching hospital during a one-month period.

Each group included scans considered as normal and those that had previously been interpreted as showing internal and/or external sphincter defects. The procedure was performed with the patient in lithotomy position, using Pro Focus ultrasound scanner with a 13 MHz rotating 2050 probe (B-K Medical, Gentofte, Denmark). The pictures were stored on the hard disc of the computer. The 3D reconstruction of the images was carried out using BK-3D viewer for Windows software version 7.0 (B-K Medical). The observer was able to manipulate the 3D images using the software. For each patient standard photographic prints of the proximal, middle and distal anal canal were obtained along with the full 3D multiplanar image.

Each of these ten cases were shown to 8 colorectal surgeons and one radiologist, all with an interest in pelvic floor disorders. All of these observers had minimal experience in interpreting ultrasound images. Four specialists with previous theoretical and practical experience of carrying out and reporting endoanal ultrasound images (2 gastrointestinal physiologists and 2 colorectal surgeons) were also shown the scans. Each set of images was given one of four possible diagnoses; normal sphincter, internal

sphincter damage, external sphincter damage or combined sphincter damage.

In order to increase the statistical strength of the study, we considered an alternative study design for the second part of the study. An independent observer randomly selected scans of forty patients with faecal incontinence carried out over a one month period. 2D photographic prints and 3D images of these scans were obtained as in first part of the study.

Each of the forty scans was shown to only two of the specialist endoanal ultrasound interpreters (one gastrointestinal physiologist and one colorectal surgeon).

## Statistical analysis

Statistical advice was sought from the University of Sheffield statistical department regarding a power calculation and subsequent statistical analysis. As this was a pilot study with semi-quantitative analysis, it was not possible to specify a sample size a priori. Our hypothesis was to simply evaluate if better inter-observer agreement resulted from the use of a 3D scanning technique and it was not to evaluate if 3D scanning is better than the 2D scanning.

Interpretation of the images was compared for inter-observer variability using the Fleiss' Kappa (2). This is a statistical measure which allows assessment of the reliability of agreement among number of observers rating a number of items. The measure calculates the degree of agreement amongst the observers over that which would be expected by chance and is scored as a number between 0 and 1.

Strength of agreement was judged according to Landis and Koch criteria (3): kappa <0 poor: 0-0.20 slight, 0.21-0.40 fair; 0.41-0.6 moderate; 0.61-0.80 substantial; 0.81-1 almost perfect.

Data was analysed using an online calculator for Fleiss' Kappa (4) and Chi square test was performed for comparing the interpretation outcome of 2D and 3D images in each group, using using GraphPad Prism version 5.00 for Mac, GraphPad Software, San Diego California USA, [www.graphpad.com](http://www.graphpad.com).

## RESULTS

In first part of the study for the group of 13 observers as a whole the overall agreement in

interpretation of the 2D images was 37% and 3D was 42%. The kappa coefficient for variability was slight for the ten 2D images, kappa =0.16 and fair for the images seen in 3D, kappa =0.22. There was no statistically significant advantage in interpretation with either modality.

If only those with the greatest experience in interpreting the images (defined as those who report on anal ultrasound) were considered as a sub-group, the overall agreement increased to 56% for 2D images and 58% for 3D with a corresponding improvement in the kappa coefficient which became moderate for the 2D, kappa = 0.42 and 3D, k= 0.44 image interpretation. Again when the findings of 2D and 3D images in this group were compared against each other, there was no significant advantage with either modality.

There was also no interpretation advantage for the 3D device with the subgroup of 8 surgeons and one radiologist who do not routinely report scans. The inter-observer agreement was slight for both the 2D, kappa = 0.11 and 3D, kappa = 0.16. Interpretation and percentage of agreement was 33% for 2D and 37% for the 3D scans for this subgroup.

For the second part of the study, when looking at 2D scans the two observers agreed that 16 patients had no sphincter defect, 4 had EAS defect, 1 had isolated IAS defect and 7 had combined injury (tab. 1). The overall agreement in the interpretation of 2D images was 70% and kappa coefficient for variability was moderate (kappa =0.60). When interpreting the 3D scans the two observers agreed that 24 patients had no sphincter defect, 1 has EAS defect, 1 had isolated IAS defect and 3 had combined injury (tab. 2). The overall agreement in the interpretation of 3D images was 75% and kappa coefficient for variability was substantial (kappa =0.67). Despite the apparent improvement in kappa coefficient using the 3D modality the difference was not statistically significant (p=0.18).

## DISCUSSION

Anal endosonography has been validated histologically (5), intraoperatively (6) and physiologically (7) as the gold standard for assessment of sphincter integrity. It is better than MRI particularly at visualising the internal sphincter with a better inter-observer

Table 1. Agreement between two observers for the presence of external and internal anal sphincter defect with 2D scans

Observer 1	Observer 2				
	normal	EAS defect	IAS defect	combined defect	total
Normal	16	7	2	2	27
EAS defect	0	4	0	0	4
IAS defect	0	0	1	1	2
Combined defect	0	0	0	7	7
Total	16	11	3	10	40

Table 2 Agreement between two observers for the presence of external and internal anal sphincter defect with 3D scans

Observer 1	Observer 2				
	normal	EAS defect	IAS defect	combined defect	total
Normal	24	6	0	0	30
EAS defect	0	1	0	0	1
IAS defect	0	1	1	1	3
Combined defect	1	0	2	3	6
Total	25	8	3	4	40

variability (8). A detailed assessment of the integrity of the sphincter complex is important for selection of patients who may be suitable for surgical repair although some of these patients may respond to sacral neuromodulation (9, 10, 11).

Unfortunately, conventional endoanal ultrasound is a very subjective test. It depends on the operator being able to interpret the images at the time of the test. Although traditionally 'hard copies' of images are taken at three levels (proximal, middle and distal anal canal) subsequent interpretation and confirmation of the findings by another observer can be difficult. The 3D ultrasound device overcomes this problem by providing a multiplanar image that can be manipulated by subsequent observers. We felt this should improve the agreement in interpreting the scans.

In the first part of our study despite our hypothesis we failed to show no significant improvement in inter-observer variability using the 3D device when compared with the 2D device for the group as a whole. The level of agreement was very low for both groups. The variability improved if only those who report images regularly were analysed, the level of agreement was moderate for the 2D and the 3D device but there was no advantage of one technique over the other.

In the second part of our study with only two specialists looking at forty scans of 2D and 3D each we showed that the level of agreement improved further. It was moderate for the 2D and substantial for the 3D. But again there was no advantage of one technique over the other even in expert hands.

Several studies have examined the inter-observer variability or reliability for interpretation of 2D ultrasound. Faltin et al. (12) interpreted ultrasounds of 54 primiparous women who had undergone vaginal delivery. Using 4 observers who had formal theoretical and practical training, a similar variability in interpretation to our 'reporting' observers was found (kappa 0.34, fair for prints and kappa 0.42, moderate for video). They suggested the low reliability may have related to the acute setting with local oedema and changes associated with pregnancy modifying the echogenicity of the tissues. Perhaps more relevant to the reliability was the experience of the observers. Norderval et al. (13) showed that an experienced sonologist achieved a higher degree of

intra-observer agreement than an inexperienced sonologist. We showed a significantly improved reliability in the subset of observers who regularly report scans.

A subgroup of surgeons and a radiologist with an interest in pelvic floor problems was chosen despite the fact that they did not regularly report scan results as it was felt that this group would have the most benefit from the multiplanar imaging opportunity offered with the 3D device. Gold et al. (14) argued that surgeons may be more appropriate to interpret scans as they are more familiar with anal canal anatomy. Unfortunately this was not the case with the surgical 'non-reporters' having only a slight level of agreement regardless of the technique.

We found only one previous study that has compared diagnostic confidence with 2D and 3D endosonography. Inter-observer agreement was shown to approach 98.2% using the 3D method and was better than 2D. The reasons for the difference in accuracy between this study and our own is difficult to explain. It is not clear whether these two observers were experts although they were identified as the study investigators (1). A possible explanation for the lack of advantage with the 3D images in our study includes the quality of the 2D images. Using the 2050 transducer in our study for the 2D 'snapshots' probably results in more accurate acquisition, sharper images and less chance of operator movement/error than using the standard 1850 with a water filled cone. However, we would have expected the level of agreement to be higher with higher quality images. This was not the case.

Gold et al. (14) previously showed that agreement in interpretation of internal anal sphincter defects was usually very good as the internal sphincter is often clearly seen. Conversely isolated external anal sphincter tears are more difficult to see and disagreement in interpretation can occur. Internal sphincter injury is almost always seen in combination with external sphincter injury and the presence of the internal sphincter injury therefore directs the sonographer towards the accompanying external sphincter defect, facilitating diagnosis. Our study included all combinations of sphincter injury as well as scans showing intact sphincter complexes. Unfortunately, due to the design of the study it was not possible to analyse each of these parameters individually.

In summary, in the first part of our study using large numbers of interpreters with a pelvic floor interest we failed to show any improvement in inter-observer variability using the 3D ultrasound device compared with 2D snapshots. The level of agreement was only slight for both techniques. In the

second part of the study we were able to demonstrate that greater experience in interpretation improves agreement to a moderate level for the 2D scans and to a substantial extent for the 3D technique but we still failed to show any advantage in using the 3D multiplanar imaging.

## REFERENCES

1. Christensen AF, Nyhuus B, Nielsen MB et al.: Three-dimensional anal endosonography may improve diagnostic confidence of detecting damage to the anal sphincter complex. *Br J Radiol* 2005; 78: 308-11.
2. Fleiss JL: Measuring nominal scale agreement among many raters. *Psychological Bulletin* 1971; 76: 378-82.
3. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
4. Randolph, J. J 2008. Online Kappa Calculator. From <http://justus.randolph.name/kappa>
5. Sultan AH, Kamm MA, Talbot IC et al.: Anal endosonography for identifying external sphincter-defects confirmed histologically. *Br J Surg* 1994; 81: 463-65.
6. Sultan AH, Nicholls RJ, Kamm MA et al.: Anal endosonography and correlation with in vitro and in vivo anatomy. *Br J Surg* 1993; 80: 508-11.
7. Sultan AH, Kamm MA, Hudson CN et al.: Anal-sphincter disruption during vaginal delivery. *N Engl J Med* 1993; 329: 1905-11.
8. Malouf AJ, Williams AB, Halligan S et al.: Prospective assessment of accuracy of endoanal MR imaging and endosonography in patients with fecal incontinence. *Am J Roentgenol* 2000; 175: 741-45.
9. de Leeuw JW, Vierhout ME, Struijk PC et al.: Anal sphincter damage after vaginal delivery: relationship of anal endosonography and manometry to anorectal complaints. *Dis Colon Rectum* 2002; 45: 1004-10.
10. Bartram C: Anal endosonography after sphincter repair. *Abdom Imaging* 1999; 24: 574-75.
11. Boyle DJ, Knowles CH, Lunniss PJ et al.: Efficacy of sacral nerve stimulation for fecal incontinence in patients with anal sphincter defects. *Dis Colon Rectum* 2009; 52(7): 1234-39.
12. Faltin DL, Boulvain M, Stan C et al.: Intraobserver and interobserver agreement in the diagnosis of anal sphincter tears by postpartum endosonography. *Ultrasound Obstet Gynaecol* 2003; 21: 375-77.
13. Norderval S, Dehli T, Vonen B: Three-dimensional endoanal ultrasonography: intraobserver and interobserver agreement using scoring systems for classification of anal sphincter defects. *Ultrasound Obstet Gynaecol* 2009; 33: 337-43.
14. Gold DM, Halligan S, Kmiot WA et al.: Intraobserver and interobserver agreement in anal endosonography. *Br J Surg* 1999; 86: 371-75.

Received: 4.07.2010 r.

Address correspondence: 12 Woburn Close, Doncaster DN4 9ER, United Kingdom