Taylor & Francis
Taylor & Francis Group

The Journal of Maternal-Fetal & Neonatal Medicine

ISSN: 1476-7058 (Print) 1476-4954 (Online) Journal homepage: http://www.tandfonline.com/loi/ijmf20

An objective scoring method to evaluate image quality of middle cerebral artery Doppler

S. Ruiz-Martinez, G. Volpe, S. Vannuccini, A. Cavallaro, L. Impey & C. Ioannou

To cite this article: S. Ruiz-Martinez, G. Volpe, S. Vannuccini, A. Cavallaro, L. Impey & C. Ioannou (2018): An objective scoring method to evaluate image quality of middle cerebral artery Doppler, The Journal of Maternal-Fetal & Neonatal Medicine, DOI: 10.1080/14767058.2018.1494711

To link to this article: https://doi.org/10.1080/14767058.2018.1494711

Accepted author version posted online: 27 Jun 2018.

An objective scoring method to evaluate image quality of middle cerebral artery Doppler

Ruiz-Martinez S[1], Volpe G[2], Vannuccini S[2], Cavallaro A[2,3], Impey L[3], Ioannou C[2,3]

Short title: MCA image scoring

Word count: 1893

Table count: 5

Figure count: 2

Author's institutional affiliation:

1. Aragon Institute of Health Research (IIS Aragón), Obstetrics Department, Hospital Clínico Universitario Lozano Blesa Zaragoza, Spain.
2. Nuffield Department of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK
3. Fetal Medicine Unit, Department of Maternal and Fetal Medicine, Women's Center, John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, Oxford OX3 9DU, UK.

Corresponding author:
Dr Christos Ioannou
John Radcliffe Hospital, Women's Centre
Headley Way, Headington
Oxford, OX3 9DU
Email: christos.ioannou@ouh.nhs.uk

Conflict of interest: the authors report no conflict of interest

Funding source: none

Keywords: Image scoring, reliability, quality, agreement, reproducibility, middle cerebral artery, Doppler

Abstract

Objective: To validate an objective scoring system for middle cerebral artery (MCA) pulsed wave Doppler images.

Method: From an image database of routine 36 week scans, a random sample of MCA Doppler images was selected. Two reviewers rated the images subjectively as acceptable or unacceptable. Subsequently they used an objective 6-point image scoring system and awarded one point for each of: 1) anatomical site; 2) magnification; 3) angle of insonation; 4) image clarity; 5) sweep speed adjustment; and 6) velocity scale and baseline adjustment. Image scores 4-6 were defined as good quality whereas 0-3 as poor. The subjective and objective agreement between the two reviewers was compared using the adjusted Kappa statistic.

Results: A total of 124 images were assessed. Using objective scoring the agreement rate between reviewers increased to 91.9% (k = 0.839) compared to subjective agreement 75.8% (k = 0.516). The agreement for each criterion was: anatomical site 91.1% (k=0.823); magnification 95.2% (k=0.903); clarity 83.9% (k=0.677); angle 96.0% (k=0.919); sweep speed 98.4% (k=0.968); velocity scale and baseline 94.4% (k=0.887).

Conclusion: Objective assessment of MCA Doppler images using a 6-point scoring system has greater interobserver agreement than subjective assessment and could be used for MCA Doppler quality assurance.

Introduction

Doppler assessment of the middle cerebral artery (MCA) is receiving increasing attention in research and clinical practice. The cerebroplacental ratio (CPR) has become part of the assessment of small for gestational age and potentially compromised fetus[1,2]. It may also form part of the more difficult but potentially more useful task of identifying the 'growth restricted' but appropriate for gestational age fetus[3]. A number of studies have correlated the MCA and particularly the CPR with adverse perinatal outcomes, although its role in clinical decision making remains less clear. It is also an essential part of the assessment of the potentially anaemic fetus, most commonly in the evaluation of pregnancies exposed to red cell antibodies[4] or parvovirus[5]. In this role its usage can dictate the need and timing of intrauterine transfusion and can prevent unnecessary invasive fetal assessment.

Doppler ultrasonography is a safe and non-invasive technique commonly used in fetal monitoring[6]. Like other medical techniques, it requires learning and training. The measurement of the MCA must follow standards so that the values obtained are adequate and reproducible in order to maximise the potential of Doppler assessment in clinical practice. Objective scoring has been shown to be useful for fetal biometry[7] and crown-rump length[8]. For nuchal translucency for instance, it is known that appropriate feedback and intervention for individual sonographers may improve their performance[9]. Guidelines describe the correct Doppler assessment of the MCA[10] but there are no published studies objectively assessing its key criteria and no scoring method for assessing whether an MCA Doppler image has been recorded accurately.

The aim of this study is to evaluate an objective scoring system for MCA Doppler images and to compare this with subjective assessment.

Methods

This is an ultrasound image scoring reliability study. All pregnant women in the John Radcliffe Hospital, Oxford, UK are offered MCA Doppler assessment as part of routine growth scan at 36 weeks gestation. 1270 examinations were performed between December 2016 and January 2017 and a random 10% sample was selected. All images were taken by four different trained sonographers, following the same institutional protocol and using two different machines (Phillips Epiq 7 and GE Voluson E8). According to a previously published power calculation, a sample of 125 examinations is adequate to detect a 10% difference between two reviewers with 90% power, assuming a rate of inter-observer agreement of 80%[7]. The current study is covered by ethics reference REC 17/SC/0374 granted on 27/07/2017; patient informed consent is not required as this is a retrospective review of routinely collected data.

An objective scoring system was developed based on the ISUOG Practice Guidelines[10]. The following six criteria (Table 1) were defined: 1) anatomical site; 2) magnification; 3) image clarity; 4) angle of insonation; 5) sweep speed adjustment; and 6) velocity scale and baseline adjustment. Once these features are fulfilled, the pulsatility index (PI), resistance index (RI) and peak systolic velocity (PSV) is obtained automatically from at least three uniform waves.

Two assessors were blinded to each other's rating results and they rated all images subjectively as "acceptable" or "unacceptable". To assess the images objectively, the same two assessors used the 6-point image scoring system. One point was awarded for each criterion satisfied; and zero if the criterion was not satisfied (Figure 1). All criteria were accorded equal weight and the sum of points was the final score. We considered images scoring 4 or more points as good quality; and those scoring less than 4 points as poor quality.

Score distributions were compared between the two observers using the Wilcoxon test. Subjective and objective agreement between observers was assessed using the unadjusted (Cohen) and the prevalence-adjusted and bias-adjusted (PABAK) Kappa statistics[11]. Inter-item consistency of the six criteria of the scoring system was assessed for each observer using the Cronbach's alpha statistic[12]. Analyses were performed using IBM SPSS statistics version 23.

Results

A total of 124 middle cerebral artery Doppler images from 4 sonographers were used. Using subjective scoring, reviewers A and B judged 71 (57.3%) and 79 (63.7%) images respectively to be acceptable. 60 (48.4%) images were acceptable by both assessors whereas 34 (27.4%) were unacceptable by both, an agreement rate of 75.8%.

The distribution of objective scores amongst subjectively rated images is shown in Table 2. Images deemed subjectively acceptable would unsurprisingly have an objective score most often 4-6 and never 1-2. Conversely images deemed unacceptable would usually have an objective score 1-3, but it is interesting to note that 10-20% of those images could have an objective score of 4-6.

Using the objective scoring method the agreement rate between reviewers increased to 91.9%, adjusted k = 0.839 compared to subjective rating agreement 75.8%, adjusted k = 0.516 (Table 3). Both reviewers had a median score of 4 (range 1-6) and the score distributions are shown in Figure 2. Reviewer A had a mean score 4.27 whereas reviewer B a mean score 4.17 and this small difference was statistically significant (Wilcoxon signed rank P = 0.022).

Table 3 demonstrates that objective assessment of the image quality using the overall image score has the highest reliability between the two reviewers when compared to any other combination of assessment methods. Table 4 highlights that amongst the individual scoring criteria, highest reliability was noted for the sweep speed adjustment (adjusted k = 0.968); and lowest reliability for the criterion of image clarity (adjusted k = 0.677).

Criteria interdependency was almost non-existent as demonstrated by the low or negative Cronbach a values for each individual criterion (Table 5).

Discussion

This study demonstrates that there is better inter observer agreement using a 6-point objective scoring system of MCA Doppler images than there is for subjective evaluation.

It is widely accepted that quality assurance mechanisms should be in place when obstetric ultrasound is used in research or as part of an established screening programme[13]. Objective scoring systems have been validated for use in fetal biometry[7], crown rump length (CRL) assessment[8] and nuchal translucency (NT) measurement[9]. MCA Doppler has a proven role in the assessment of small for gestational age fetuses[1,2,14] and also in monitoring fetal anaemia[4,5,15]. The potential of MCA in reducing still birth in those contexts makes accurate assessment particularly important. Quality assurance using objective image scoring and targeted feedback has been shown to improve the consistency of ultrasound measurements even amongst trained sonographers[16].

We employed an established methodology and we devised a 6-point objective scoring method which incorporates technical recommendations from an international guideline for MCA Doppler assessment[10]. In addition to high inter observer agreement for the overall objective score we demonstrated excellent agreement for most of the individual criteria. Clearly defined and unequivocally quantifiable criteria - such as scale velocity adjustment – tend to have very high agreement. On the other hand image clarity – an essential but less quantifiable criterion – demonstrates comparatively lower agreement.

There are some interesting observations when comparing this Doppler based image quality score to a previously validated score for CRL measurement[8]. Very low Cronbach a-scores for our scoring method suggest that there is very little or no interdependence between the individual criteria. Higher Cronbach scores were observed for CRL scoring[8] and this suggests higher consistency and interdependence between the individual criteria for CRL i.e. higher likelihood that more than one criteria would score positive for the same underlying reason. As a result, CRL score distributions tend to be positively skewed with median score 6. In our study very low Cronbach a-scores suggest that there is no interdependence between the individual criteria. This can be explained because each individual criterion assesses fundamentally different quality properties of a Doppler image such the anatomic plane of acquisition or the angle of Doppler application or the features of waveform Doppler optimisation. As a result, our score distributions are less skewed and the median score is 4.

Amongst other strengths of this study are that the sample was randomly drawn from an unselected pregnant population undergoing routine examinations at 36 weeks i.e. the gestation where MCA - as part of the CPR - is most useful in assessing placental insufficiency. The sonographers who performed the measurements were appropriately trained sonographers or fellows performing ultrasound examinations within the standards of a tertiary UK NHS service and they were using regularly upgraded scan machines. The study was adequately powered and the use of PABAK is the recommended methodology for assessing image scoring agreement[8].

We also acknowledge some limitations. There was a statistically significant difference of mean score between the two reviewers but the magnitude of this difference was clinically insignificant and did not affect the overall reliability. The scoring cut-off used to define good and poor quality – though consistent with previous image scoring literature - is rather arbitrary. Given the low interdependence between criteria it could be argued that each criterion is absolutely essential for the image to be deemed satisfactory and that the quality threshold should be set higher than 4. However this does not invalidate the conclusions of this exercise which demonstrates that the scoring method is reliable. It is up to institutions to define the level of threshold in their clinical practice. Another limitation is that these MCA images were taken as part of a universal screening programme for fetal growth and cerebroplacental ratio but they were not performed specifically for peak systolic velocity measurement. Nevertheless the same quality principles underlie the latter measurement and it is reasonable to assume that the proposed scoring method is valid when MCA is used as a screening tool for fetal anaemia. This scoring method cannot assess the effect of pressure applied on the ultrasound probe which is known to alter MCA values[17]. Image analysis is time consuming and should not be the sole measure of quality control. However it is a useful tool as part of a quality control strategy which may also include quantitative analysis of measurement distributions of individual sonographers, similar to those strategies employed for quality assurance of nuchal translucency programmes[9].

We propose that this objective scoring system is used to assess MCA Doppler measurement quality. Providers should use this regularly to audit the performance of individual sonographers in their institutions; and should decide locally the satisfactory score threshold in order to determine the need for feedback and retraining.

Figures

Figure 1 Representative examples of MCA images (A) where all scoring criteria are met; (B) wrong anatomical site: circle of Willis and MCA poorly identified with gate too lateral and near the skull; (C) inadequate magnification; (D) suboptimal image clarity resulting in inaccurate tracing of the waveform; (E) no angle correction; (F) no sweep speed adjustment resulting in too many waves per image; and (G) no baseline and velocity scale adjustment so that waveform does not fill up the screen

Figure 2 Distribution of total image scores for both reviewers

References

1. Flood K, Unterscheider J, Daly S, et al. The role of brain sparing in the prediction of adverse outcomes in intrauterine growth restriction: results of the multicenter PORTO Study. *Am J Obstet Gynecol.* 2014;211(3):288.e281-285.

2. Odibo AO, Riddick C, Pare E, Stamilio DM, Macones GA. Cerebroplacental Doppler ratio and adverse perinatal outcomes in intrauterine growth restriction: evaluating the impact of using gestational age-specific reference values. *J Ultrasound Med.* 2005;24(9):1223-1228.

3. Bardien N, Whitehead CL, Tong S, Ugoni A, McDonald S, Walker SP. Placental Insufficiency in Fetuses That Slow in Growth but Are Born Appropriate for Gestational Age: A Prospective Longitudinal Study. *PLoS One.* 2016;11(1):e0142788.

4. Mari G, Deter RL, Carpenter RL, et al. Noninvasive diagnosis by Doppler ultrasonography of fetal anemia due to maternal red-cell alloimmunization. Collaborative Group for Doppler Assessment of the Blood Velocity in Anemic Fetuses. *N Engl J Med.* 2000;342(1):9-14.

5. Borna S, Mirzaie F, Hanthoush-Zadeh S, Khazardoost S, Rahimi-Sharbaf F. Middle cerebral artery peak systolic velocity and ductus venosus velocity in the investigation of nonimmune hydrops. *J Clin Ultrasound.* 2009;37(7):385-388.

6. Stampalija T, Gyte GM, Alfirevic Z. Utero-placental Doppler ultrasound for improving pregnancy outcome. *Cochrane Database Syst Rev.* 2010(9):CD008363.

7. Salomon LJ, Bernard JP, Duyme M, Doris B, Mas N, Ville Y. Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester. *Ultrasound Obstet Gynecol.* 2006;27(1):34-40.

8. Wanyonyi SZ, Napolitano R, Ohuma EO, Salomon LJ, Papageorghiou AT. Image-scoring system for crown-rump length measurement. *Ultrasound Obstet Gynecol.* 2014;44(6):649-654.

9. Snijders RJ, Thom EA, Zachary JM, et al. First-trimester trisomy screening: nuchal translucency measurement training and quality assurance to correct and unify technique. *Ultrasound Obstet Gynecol.* 2002;19(4):353-359.

10. Bhide A, Acharya G, Bilardo CM, et al. ISUOG practice guidelines: use of Doppler ultrasonography in obstetrics. *Ultrasound Obstet Gynecol.* 2013;41(2):233-239.

11. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46(5):423-429.

12. Bland JM, Altman DG. Cronbach's alpha. *BMJ.* 1997;314(7080):572.

13. Holt R, Abramowicz JS. Quality and Safety of Obstetric Practices Using New Modalities- Ultrasound, MR, and CT. *Clin Obstet Gynecol.* 2017;60(3):546-561.

14. Meher S, Hernandez-Andrade E, Basheer SN, Lees C. Impact of cerebral redistribution on neurodevelopmental outcome in small-for-gestational-age or growth-restricted babies: a systematic review. *Ultrasound Obstet Gynecol.* 2015;46(4):398-404.

15. Schenone MH, Mari G. The MCA Doppler and its role in the evaluation of fetal anemia and fetal growth restriction. *Clin Perinatol.* 2011;38(1):83-102, vi.

16. Sarris I, Ioannou C, Dighe M, et al. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. *Ultrasound Obstet Gynecol.* 2011;38(6):681-687.

17. Su YM, Lv GR, Chen XK, Li SH, Lin HT. Ultrasound probe pressure but not maternal Valsalva maneuver alters Doppler parameters during fetal middle cerebral artery Doppler ultrasonography. *Prenat Diagn.* 2010;30(12-13):1192-1197.

Table 1

Criterion	Description
Anatomical site	Axial brain section visualizing the thalami and sphenoid wings and identifying the circle of Willis by colour Doppler with the gate placed in the proximal third of the MCA
Magnification	MCA image occupies at least 50% of the screen
Image clarity	Waveform should be clear without artefacts and tracing should be accurate
Angle of insonation	Less than 15^{o} followed by angle correction as close as possible to 0^{o}
Sweep speed adjustment	3-10 uniform waveforms are visualised
Velocity scale and baseline adjustment	Waveforms occupy 75% of the screen

Table 1 Image scoring criteria for MCA Doppler image

	Objective image score					
Subjective assessment	**1**	**2**	**3**	**4**	**5**	**6**
Unacceptable A	1 (1.9%)	11 (20.8%)	15 (28.3%)	15 (28.3%)	10 (18.9%)	1 (1.9%)
Acceptable A	-	-	3 (4.2%)	26 (36.6%)	32 (45.1%)	10 (14.1%)
Unacceptable B	2 (4.4%)	7 (15.6%)	16 (35.6%)	13 (28.9%)	6 (13.3%)	1 (2.2%)
Acceptable B	-	-	5 (6.3%)	21 (26.6%)	37 (46.8%)	16 (20.3%)
Unacceptable by both A and B (Objective score A)	1 (2.9%)	11 (32.4%)	10 (29.4%)	10 (29.4%)	2 (5.9%)	-
Unacceptable by both A and B (Objective score B)	2 (5.9%)	7 (20.6%)	13 (38.2%)	8 (23.5%)	3 (8.8%)	1 (2.9%)
Acceptable by both A and B (Objective score A)	-	-	1 (1.7%)	20 (33.3%)	29 (48.3%)	10 (16.7%)
Acceptable by both A and B (Objective score B)	-	-	1 (1.7%)	17 (28.3%)	28 (46.7%)	14 (23.3%)

Table 2

Table 2 Comparison between subjective assessment and objective scoring for both observers

Table 3

Criterion	Agreement (%)	Kappa Cohen (95% CI)	Adjusted Kappa PABAK (95% CI)
Subjective A / Subjective B	75.8	0.496 (0.341 – 0.651)	0.516 (0.365 – 0.667)
Subjective A / Objective B	78.2	0.493 (0.334 – 0.652)	0.565 (0.419 – 0.710)
Objective A / Subjective B	75.0	0.460 (0.311 – 0.609)	0.500 (0.348 – 0.652)
Subjective A / Objective A	79.8	0.530 (0.375 – 0.685)	0.597 (0.456 – 0.738)
Subjective B / Objective B	76.6	0.494 (0.347 – 0.641)	0.532 (0.383 – 0.681)
Objective A / Objective B	91.9	0.780 (0.651 – 0.909)	0.839 (0.743 – 0.935)

Table 3 Agreement between subjective assessment and objective scoring for MCA Doppler image

Table 4

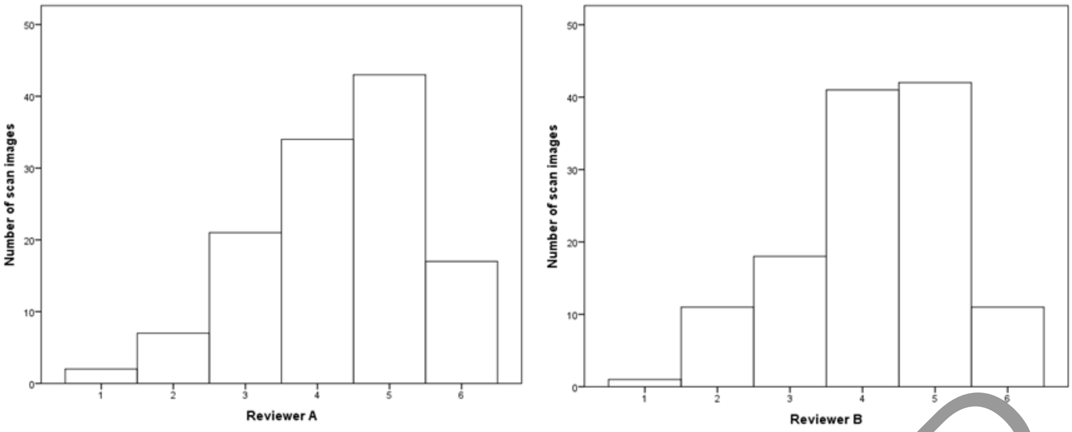
Criterion	Agreement (%)	Kappa Cohen (95% CI)	Adjusted Kappa PABAK (95% CI)
Anatomic site	91.1	0.783 (0.661 - 0.905)	0.823 (0.722 – 0.923)
Magnification	95.2	0.845 (0.725 - 0.965)	0.903 (0.828 – 0.979)
Image clarity	83.9	0.644 (0.503 - 0.785)	0.677 (0.548 – 0.807)
Angle	96.0	0.917 (0.846 - 0.988)	0.919 (0.850 – 0.989)
Sweep speed	98.4	0.849 (0.643 - 1.000)	0.968 (0.923 – 1.000)
Velocity and baseline	94.4	0.868 (0.774 - 0.962)	0.887 (0.806 – 0.968)

Table 4 Agreement between reviewers for each scoring criterion

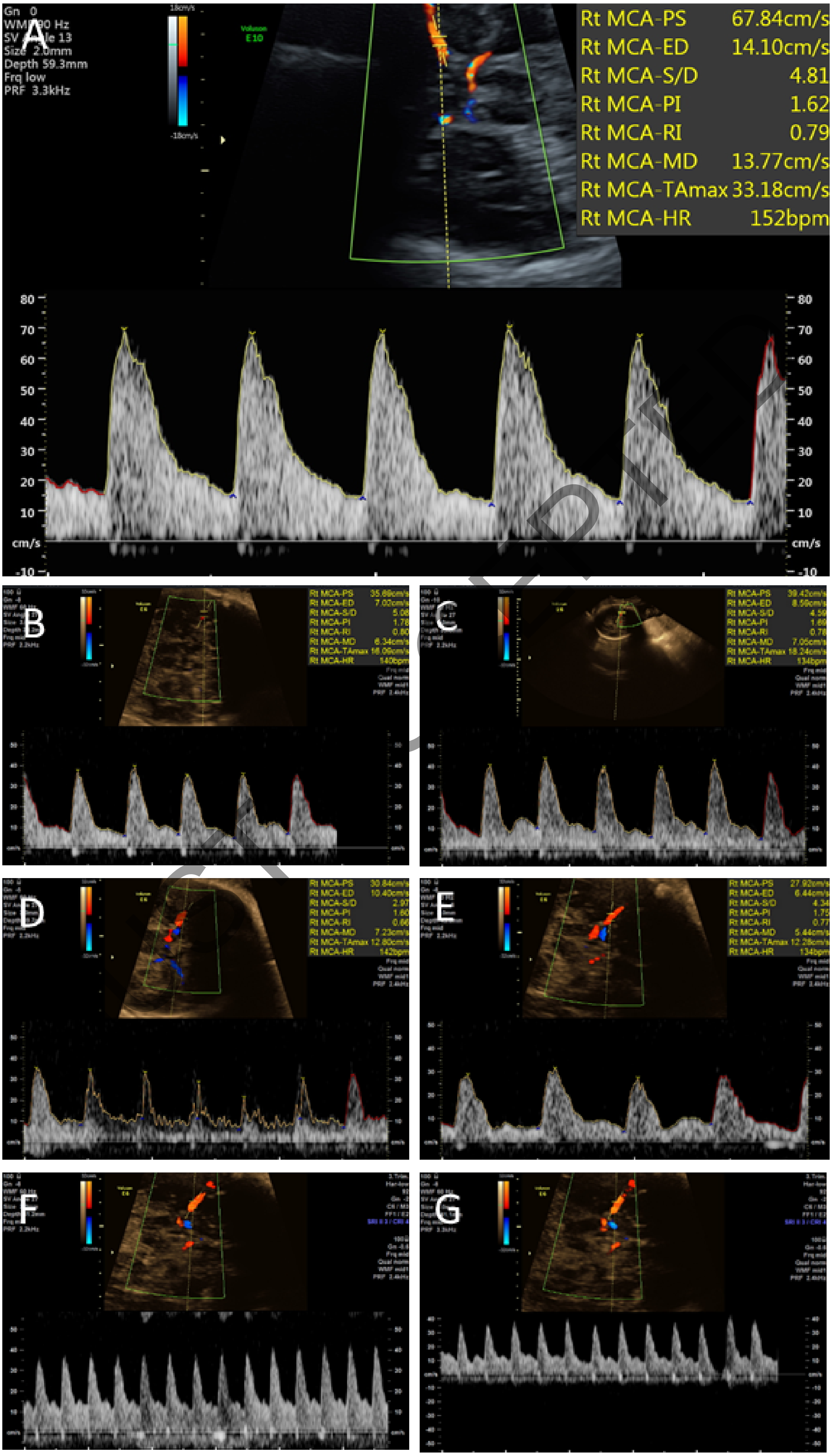
Table 5

	Cronbach's alpha for excluding each item individually	
	Reviewer A	**Reviewer B**
Anatomic site	0.123	-0.124
Magnification	0.341	0.323
Image clarity	0.118	0.031
Angle	0.269	0.162
Sweep speed	0.246	0.118
Velocity and baseline	0.107	-0.048
All six items	0.243	0.116

Table 5 Inter-item consistency amongst image scoring criteria for MCA Doppler image



Number of scan images
50
40
30
20
10
0
1
2
3
4
5
6
Reviewer A
Number of scan images
50
40
30
20
10
0
1
2
3
4
5
6
Reviewer B


A
Gn 0
Size 2.0mm
Depth 59.3mm
Frq low
PRF 3.3kHz
Rt MCA-PS 67.84cm/s
Rt MCA-ED 14.10cm/s
Rt MCA-S/D 4.81
Rt MCA-PI 1.62
Rt MCA-RI 0.79
Rt MCA-MD 13.77cm/s
Rt MCA-TAmax 33.18cm/s
Rt MCA-HR 152bpm
cm/s
B
C
D
E
F
G