
Online Self-Evaluation of Fetal Ultrasound Images for Medical Continuing Education: A Randomised Controlled Trial

Suha Jaudi^{1,2}, Louise Chevalier³, Nicolas Fries², Alain Daher¹, Sophie Tezenas du Montcel^{4,5}, Marc Dommergues^{1,2,*}

¹Department of Obstetrics and Gynecology, Hopital Universitaire Pitié Salpêtrière Charles Foix APHP and Sorbonne Université, Paris, France

²Collège Français d'Echographie Foetale (French College for Fetal Ultrasound, CFEF), Montpellier, France

³Department of Biostatistics, Hopital Universitaire Pitié Salpêtrière Charles Foix APHP and Sorbonne Université, Paris, France

⁴Department of Biostatistics and URC, Hopital Universitaire Pitié Salpêtrière Charles Foix APHP and Sorbonne Université, Paris, France

⁵ER4 Modélisation et Recherche Clinique, Sorbonne Université, Paris, France

Email address:

marc.dommergues@aphp.fr (M. Dommergues)

*Corresponding author

To cite this article:

Suha Jaudi, Louise Chevalier, Nicolas Fries, Alain Daher, Sophie Tezenas du Montcel, Marc Dommergues. Online Self-Evaluation of Fetal Ultrasound Images for Medical Continuing Education: A Randomised Controlled Trial. *Education Journal*. Vol. 8, No. 5, 2019, pp. 226-231. doi: 10.11648/j.edu.20190805.17

Received: May 30, 2019; **Accepted:** August 13, 2019; **Published:** September 5, 2019

Abstract: Continuing medical education in the field of fetal ultrasound imaging is based on expert audit of still images, a time consuming approach. Our objective was to determine if self-evaluation of the images a professional produced is as effective as audit and feedback by an expert as a method of continuous medical education. We designed a prospective blinded randomized controlled trial. 321 ultrasonographers uploaded on a continuous medical education website a first set of 30 biometry images (10 cephalic, 10 abdominal and 10 femoral) from 10 consecutive second or third trimester normal screening scans. In arm 1: participants (N = 151) assessed their own images online according to a standardized procedure, and received feedback with detailed recommendations for change, automatically generated based on their assessment. The images were also audited by an expert, but participants remained blinded to the expert's rating. In arm 2: participants (N = 177) had their images assessed by an expert and received a feedback, formatted as in arm 1, automatically generated based on the expert's assessment. Three to 6 months later, participants uploaded a second set of images, audited by an expert. A total of 19,680 images were audited. In the self-assessment group, the percentage of images meeting all criteria (IMAC) rose from 55 to 62.2 (p < 0.0001). In the expert-assessment and feedback group, the percentage of IMAC rose from 54.2 to 59.1 (p < 0.0001). Improvement in image quality was equivalent in both groups with a difference in IMAC increase of 2.3 percentage points (95%CI: -1.7 to +6.4). In conclusion, online training based on self-assessment of fetal ultrasound images was as effective as expert audit and feedback. NCT02074592.

Keywords: Continuing Medical Education, Online Audit and Feedback, Ultrasound Image Quality, Fetal Biometry, Self-training, Fetal Ultrasound Screening, Quality Assurance

1. Introduction

Audit and feedback of ultrasound images is widely used in continuous medical education as a strategy to improve practice [1]. Various programs of audit and feedback of

second trimester ultrasound images are available [2-5] Online audit and feedback programs have the advantage of being user friendly, and allow for blind evaluation of the trainee. However, they are time consuming for the experts in charge of evaluating the images submitted by the participants. So

far, the potential of self- evaluation of ultrasound images for continuing education in medicine has not been assessed by a randomized trial. Self-evaluation however, has been promoted for initial education as a process during which students collect information about their own performance or progress; compare it to explicitly stated criteria, and revise accordingly [6]. The fact that self-evaluation was not applied to continuing medical education for producing sonographic imaging is unexpected, since ultrasound basically consists in producing images meeting predefined criteria. Producing standard images is particularly meaningful in foetal ultrasound. Indeed, precise criteria were devised to define how each body part should be imaged, and on which particular still images measurements should be taken [7]. Guidelines stress the importance of quality criteria for still images because failing to produce an image meeting all criteria may reveal a structural anomaly in the fetus. Besides, foetal measurements are crucial, mainly because they allow for estimating the foetal weight, thus detecting growth abnormalities. Measurements made on a poor still image are flawed, underscoring the importance of appropriate images.

Our hypothesis was that self-assessment of images, mediated by an online process allowing for both a standardized assessment of images and an automated feedback generated based on this assessment, would be as effective as audit by an expert in order to improve image quality.

To test this hypothesis, we focussed on the head circumference, abdominal circumference, and femur length images, which are crucial to screen for foetal growth abnormalities. We designed a prospective randomized

controlled trial to compare the improvement of image quality according to whether feedback was provided following self-assessment or audit by an expert.

2. Methods

This study is a prospective blinded randomised controlled equivalence trial. We contacted by email 3510 French ultrasonographers, doctors or midwives affiliated to the French college for fetal ultrasound (CFEF). Those who volunteered to enrol were asked to upload a first set of 30 biometry images (10 cephalic, 10 abdominal and 10 femoral) obtained from 10 consecutive screening scans performed in the second or third trimester of pregnancy. Participants were instructed to upload images from examinations in which no fetal structural defect was found. Participants had not been involved in the study design.

The images were audited via the internet by 7 expert reviewers (5 doctors and 2 midwives) experienced in the field of screening ultrasound and trained to evaluate biometry images. Each image appeared on the left hand side of the reviewer's screen while the scoring criteria appeared on the right. Electronic processing of the image contour blinded the reviewer to the patient's and the ultrasonographer's identity. At any time, via a help button, the reviewer could access a tutorial explaining each criterion and providing examples of images meeting or missing each criterion. The reviewer assessed the images against the standard recommended by the French committee for prenatal ultrasound screening [8, 9], using a scoring grid published by Salomon et al [10] as detailed in table 1.

Table 1. Scoring grid.

Image type		
Cephalic	Abdominal	Femoral
Symmetrical plane	Symmetrical plane	Both ends of femur clearly visible
Image showing the thalami	Stomach image visible	<45° angle to the horizontal
Image showing the cavum septi pellucidi	Portal sinus visible	Femur image occupying more than half the total image size
Cerebellum not visible	Kidneys not visible	Calipers placed correctly
Head image occupying more than half the total image size	Abdominal image occupying more than half the total image size	
Calipers and ellipse placed correctly	Calipers and ellipse placed correctly	

For each image, the reviewer had to determine whether each criterion was met (criterion scored 1) or not (criterion scored 0). The score of a given image was the sum of the scores of its criteria. It ranged from 0 to 6 for cephalic and abdominal images and from 0 to 4 for femoral images. The score for a set of 3 images was the number of criteria present, divided by 16, and expressed as a mark over 20.⁸

For each set of images sent by a given ultrasonographer, we evaluated:

1. The percentage of images meeting all criteria (IMAC).
2. The mean of a score (expressed as a grade over 20) based on attributing one point per criterion present on a given image.

The following data were also collected: (i) gestational age, (ii) demographic characteristics of ultrasonographers

enrolled: age, gender, experience in fetal ultrasound (years), profession (doctor vs. midwife), type of fetal ultrasound practice (screening only, vs. screening plus referral ultrasound), medical practice other than fetal ultrasound and membership of the French college for fetal ultrasound (College Francais d'Echographie Foetale).

Ultrasonographers were randomized in two arms with a 1: 1 ratio, the expert reviewer being blinded to the allocation.

In arm 1, ultrasonographers were invited to assess their own images online according to the same standardized procedure that was used by the expert reviewers. Their self-assessment generated the report they received, with detailed recommendations for change. Their images were also audited by one of the experts, but the result of this audit remained concealed to the ultrasonographer.

In arm 2, ultrasonographers did not assess their own images. Their images were assessed by one of the experts according to the same standardized online procedure. They received an automatically generated report with detailed recommendations for change, based on the expert's assessment.

The report provided the number of images meeting all criteria (IMAC) and the mean score of the image set. Whenever one criterion was not met, the corresponding image was displayed on the report and recommendations for change corresponding to the criterion were provided. Recommendations consisted in a short tutorial illustrated by images showing the presence or absence of the criterion.

Three months after sending the first set, participants were asked to upload a second set of 30 biometry images, from ten consecutive routine ultrasound examinations. Images were audited by an expert reviewer using the same standardized online procedure as for the first set.

The main outcome criterion was the difference in the percentage of IMAC between the second and the first set of images, as assessed by expert ($\Delta\%$ IMAC).

The secondary outcome was the difference in the mean score for all images between the first and second set, as assessed by expert.

We also studied the agreement between self-assessment and expert-assessment for the first set of images in the self-assessment arm.

Statistical analysis

Ultrasonographers were randomly assigned, to one of the two groups, in a 1: 1 ratio. We used a blocking-scheme of 20 and a centralized procedure for randomization, without stratification. The distribution of subjects was balanced at the time of the randomization, with a maximum difference of 10 between the two groups.

We studied the difference in main outcome ($\Delta\%$ IMAC) between the two arms using an equivalence test for quantitative data. To test the equivalence in $\Delta\%$ IMAC between the two arms we chose the equivalence margins of $\pm 6.67\%$, i.e. a difference of 2 IMAC for each set of 30 images. We estimated the expected $\Delta\%$ IMAC based on a previous study [11], in which a similar online audit and feedback method was applied to the four chambers view. In this study, mean $\Delta\%$ IMAC was 13% with a standard deviation of 20. This would correspond approximately to a number of images meeting all criteria increasing by four out of 30 in the second set (after feedback) as compared to the baseline set of 30 images.

For a two sided alpha of 5% and a power of 80% (beta = 20%), 156 ultrasonographers were needed in each group. We thus expected to enrol 320 ultrasonographers in the study.

We studied the difference in percentage of IMAC by a paired t-test. We studied agreement between self-assessment and expert-assessment scores by intra class correlation coefficient. Confidence intervals were computed using a nonparametric bootstrap with bias-corrected and accelerated method (1000 replicates). For all tests, we considered a P value less than 0.05 statistically significant. We used the R 2.12.1 statistical package (R Foundation, Vienna, Austria).

We submitted the study to Institutional Review Board approval and it was exempt because patients were not involved, and the study had no impact on clinical management. The study was advertised through the French College of Fetal Ultrasound (<http://www.cfef.org/>), an association of 3510 professionals dedicated to post graduate training and quality assurance in fetal ultrasound. The web site was constructed by Horizon, Paris and Montpellier, France (<http://www.horizon-web-agency.com/>).

ClinicalTrials. gov Identifier: NCT02074592. Acronym: Self-Assessment Versus Expert assessment of fetal Ultra Sound (SAVEUS).

3. Results

Between July 2012 and September 2014, 416 ultrasonographers volunteered to enrol in the study. Two hundred and nine were randomised in arm 1 (self-assessment) and 207 in arm 2 (expert assessment). During the study period, 27 ultrasonographers (19 in arm 1 and 8 in arm 2) failed to send their first set of images and thus were excluded from the study. By March 2015, 61 ultrasonographers (39 in arm 1 and 22 in arm 2) failed to send their second set of images. They were also excluded from the analysis. Three hundred and twenty eight ultrasonographers (151 in arm 1 and 177 in arm 2) sent their two sets of images and their 19,680 images were analysed (Figure 1). The baseline ultrasonographers' characteristics are shown in Table 2. The distribution of gestational ages was similar. In both arms, the percentage of IMAC increased significantly between the first and the second set (table 3). In the self-assessment arm (arm 1), it rose from 55.0 to 62.3%, $\Delta\%$ IMAC = 7.3 ($p < .0001$). In the expert-assessment arm (arm 2), it rose from 54.2 to 59.1%, $\Delta\%$ IMAC = 4.9 ($p < .0001$). As judged by the increase in IMAC, improvement in image quality was equivalent in both groups. Between self-assessment and expert assessment, the difference in IMAC increase from the first to second set ($\Delta\%$ IMAC arm 1 - $\Delta\%$ IMAC arm 2) was 2.3 (95% CI = -1.7 to 6.4), which falls between the equivalence margins of $\pm 6.67\%$ for a two sided alpha of 5% and a power of 80%.

Table 2. Baseline population characteristics.

	Arm 1 (n= 151)	Arm 2 (n = 177)
Age: mean in years (SD*)	50 (9)	49 (10)
Gender: Female (n, %)	89 (59%)	115 (65%)
Experience in fetal ultrasound		
Mean duration of practice in years (SD*)	19 (12)	18 (14)
Profession		

	Arm 1 (n= 151)	Arm 2 (n = 177)
Midwife (n, %)	20 (14)	37 (21)
Doctor (n, %)	131 (86)	140 (79)
Type of fetal ultrasound practice		
Screening only (n, %)	129 (86)	165 (93)
Screening and referral (n, %)	22 (14)	12 (7)
Full time fetal ultrasonographer (n, %)	39 (26%)	58 (33%)
CFEF† membership (n, %)	80 (53%)	75 (42%)
Time between two sets:		
Mean in days (SD*)	138.9 (61.2)	147.3 (74.2)
Baseline % of IMAC [95%CI‡]	55 [52-58]	54 [51-57]
Baseline mean score (/20 [95%CI‡])	17.3 [17.0-17.6]	17.3 [17.1-17.6]

* SD: standard deviation. †: French College of Fetal Ultrasound (College Francais d'Echographie Foetale). ‡: 95%CI: 95% confidence interval. IMAC: images meeting all criteria.

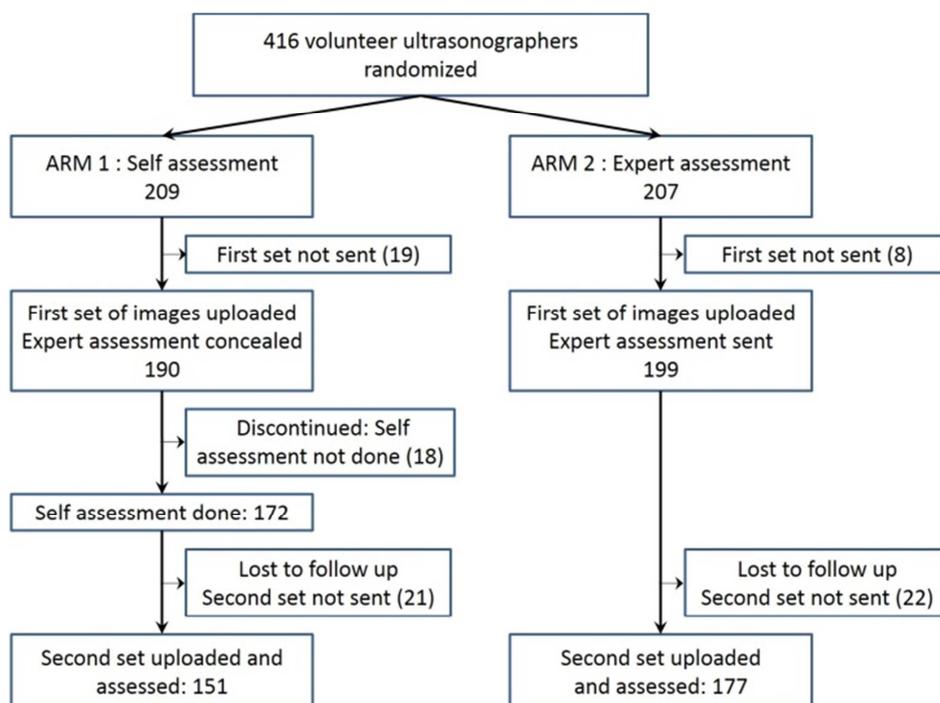


Figure 1. Flow Chart.

Table 3. Mean percentages of images meeting all criteria, mean score, and their differences between the first and second set of images.

	First set	Second set	Difference	p value†
% IMAC*				
Arm 1 mean (SD‡)	55.0 (20.5)	62.3 (19.1)	7.3 (18.0)	< 0.0001
Arm 2 mean (SD‡)	54.2 (20.9)	59.1 (20.9)	4.9 (18.9)	0.0007
Mean scores (/20)**				
Arm 1 mean (SD)	17.3 (2.0)	17.8 (1.6)	0.5 (1.8)	0.0006
Arm 2 mean (SD)	17.3 (1.7)	17.7 (1.5)	0.4 (1.4)	0.0002

*% IMAC: percentage of images meeting all criteria, calculated per set of 30 images uploaded by a given sonographer.

** Scores calculated on each set of 30 images uploaded by a given sonographer.

‡ SD: standard deviation †: Paired t test comparing the first to the second set.

Scores and % IMAC shown were calculated based on the assessment of an expert, who reviewed online the images uploaded by participants. In Arm 1 (N = 151), participants were blinded to the expert's evaluation of their first set of images, and received automated feedback based on their self-assessment of the images they had uploaded. In Arm 2 (N = 177), participants did not perform self-assessment of their images, and received automated feedback based on the expert's assessment of the images they had uploaded. In both arms, %IMAC and scores rose significantly in second set. The difference in IMAC improvements (Δ % IMAC) between arm 1 and 2 was 2.3 (95% CI = -1.7 to 6.4), which fell between the equivalence margins of \pm 6.67.

Mean scores also increased significantly between the first and second set (table 3). The difference in score increase between the self-assessment arm and the expert assessment arm was 0.1, with a 95% confidence interval of -2 to + 0.5,

suggesting the increase in scores was similar in both groups.

In arm 1, images from first set (n = 4530) had both expert and self-assessment, allowing us to analyse agreement between participants (n = 151) and experts. The agreement

between expert-assessment and self-assessment was moderate. The intraclass correlation coefficient (ICC) was 0.36 (CI: 0.21 – 0.48) for the percentage of IMAC and 0.52 (CI: 0.37 – 0.65) for mean scores. Overall, in comparison with experts, the ultrasonographers underrated their images significantly ($p < 0.001$).

4. Discussion

This is the first randomised controlled trial showing that online continuous medical education based on self-assessment of ultrasound images and automated feedback is as effective as expert audit and feedback, regarding the quality of fetal sonographic images.

Image audit was based on a national standard and used a previously published grid, with good intra and inter reviewer agreement [10]. Feedback was automated and included detailed recommendations for change. We elected to audit still images versus cine-loops to match recommendations for documenting ultrasound reports [4-9, 12-13]. We did not study the impact of equipment based on the assumption that over the study period, the impact of feedback on image quality would be mainly mediated by ultrasonographers improving their technical skills or knowledge. Yet, it is possible that image improvement may occasionally have resulted from upgrading equipment.

We decided to express the score derived from Salomon et al [11] as a fraction of twenty to facilitate understanding of French participants, who are familiar with marks over 20. We used as a primary criterion the percentage of images meeting all criteria (IMAC), because we believed this would reflect practice more intuitively than a score would do.

Although statistically significant, the increase in the number of images meeting all criteria was modest. On average, out of a set of 30, there were two additional images meeting all criteria after audit and feedback took place. This is in accordance with what we [11, 14] and others [15, 16] observed in previous studies.

It is surprising that only 62% of images met all criteria after audit and feedback. This could result from under-estimating image quality, or from the difficulty to provide perfect images in routine practice.

The effect of self-assessment mediated feedback may result in part from the fact that participants re-analysed their own images, even though participants and expert agreement on scoring was moderate [17]. Similar results concerning agreement between self-assessment and expert-assessment in medical education were reported previously [18, 19].

5. Conclusion

Our results suggest that image quality improves when sonographers undergo volunteer online training based on self-evaluation generating an automated report. We believe that such training kits are easy to implement due to the absence of extra workload from experts. They might be made widely available for the ultrasound community, to improve the quality of any

type of image that can potentially be analysed by self-scoring. More generally, our results support the concept that self-assessment is an effective training method.

Funding

This study was supported by a grant from the French ministry of health (Direction de la Recherche Clinique, PHRC regional AOR0802), and by a grant from the Agence de la Biomédecine, Paris, France.

Disclosures

The authors have no conflict of interest to declare.

Acknowledgements

We are grateful to the expert reviewers doctors and midwives who participated in this study: Daher A., Frayssé J. P., Le Gac M. P., Nguyen Herpin C., Pelissie M., Roueli A.

We thank all the members of the French College of Fetal Ultrasound (<http://www.cfef.org/>), who accepted to enrol in this trial.

References

- [1] Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, O'Brien MA, Johansen M, Grimshaw J, Oxman AD. (2012) Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev.* 2; 6. Art. No: CD000259.
- [2] AIUM. org: AIUM Ultrasound Practice Accreditation. <http://www.aium.org/accreditation/accreditation.aspx> [04/08/2019].
- [3] Fetalmedicine.com: The Fetal Medicine Foundation Training & Certification <http://www.fetalmedicine.com/fmf/training-certification/certificates-of-competence/> [05/05/2017].
- [4] CFEF. org: Collège Français d'Échographie Foetale F. M. C. E. P. P. <https://www.epp-echofoetale.fr/> [04/08/2019].
- [5] SFAPE.com: Société Française pour l'Amélioration des Pratiques Echographiques 2ème trimestre/3ème trimestre Analyse d'images. <http://www.sfape.com/> [04/08/2019].
- [6] Heidi Andrade & Anna Valtcheva (2009) Promoting Learning and Achievement Through Self-Assessment, *Theory Into Practice*, 48: 1, 12-19.
- [7] Salomon L. J., Alfirevic Z, Berghella V, Bilardo C., Hernandez-Andrade E, Johnsen S. L., Kalache K., Leung K.-Y., Malinger G., Munoz H., Prefumo F., Toi A and Lee W on behalf of the ISUOG. Clinical Standards Committee Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan *Ultrasound Obstet Gynecol.* 37 (1): 116-26.
- [8] Sureau C., Henrion R. Rapport du Comité National Technique de l'Echographie de Dépistage Prénatal 2005 <http://www.ladocumentationfrancaise.fr/rapports-publics/054000356/index.shtml> [05/05/2017].

- [9] Dommergues M, Bessis R, Henrion R. (2006) Rapport du Comité national technique de l'échographie de dépistage prénatal: quelles conséquences pour la pratique? *Gynécologie Obstétrique & Fertilité*, 34: 1090–1095.
- [10] Salomon LJ, Bernard JP, Duyme M, Doris B, Mas N, Ville Y (2006.) Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester *Ultrasound Obstet Gynecol.* 27 (1): 34-40.
- [11] Jaudi S, Granger B, Herpin CN, Fries N, Du Montcel ST, Dommergues M (2013). Online audit and feedback improve fetal second-trimester four-chamber view images: a randomised controlled trial. *Prenat Diagn.* 33 (10): 959-64.
- [12] International Society of Ultrasound in Obstetrics & Gynecology (2006). Cardiac screening examination of the fetus: guidelines for performing the 'basic' and 'extended basic' cardiac scan. *Ultrasound Obstet Gynecol*; 27: 107-13.
- [13] RCOG. org. uk: Logbook for fetal anatomy <https://www.rcog.org.uk/en/careers-training/resources--support-for-trainees/useful-resources-for-trainees/trog-news/ultrasound-updates/>.
- [14] Jaudi S, Du Montcel ST, Fries N, Nizard J, Desfontaines VH, Dommergues M (2011). Online evaluation of fetal second trimester four chamber view images: a comparison of six evaluation methods. *Ultrasound Obstet Gynecol*; 38 (2): 185-190.
- [15] Sairam S, Awadh A. M. A, Cook K, Papageorgiou A. T, Carvalho J. S (2009). Impact of audit of routine second-trimester cardiac images using a novel image scoring method. *Ultrasound Obstet Gynecol*; 33: 545-551.
- [16] Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD (2006). Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care*; 15 (6): 433-436.
- [17] Koch, Gary G. Intraclass correlation coefficient. In: Kotz S, Johnson N (1982). *Encyclopedia of Statistical Sciences*. New York: John Wiley & Sons; p. 213–217.
- [18] Gordon MJ (1991). A review of the validity and accuracy of self-assessments in health professions training. *Acad Med*; 66: 762-9.
- [19] Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*; 296: 1094-102.