Universiteit Leiden
The Netherlands

Quality assessment of ultrasonic foetal biometry during the IUGR Risk Selection (IRIS) trial: a cross sectional study

Verfaille, V.; Haak, M.C.; Pajkrt, E.; Jonge, A. de; Henrichs, J.; Franx, A.; ... ; IRIS Study Grp

Citation
Verfaille, V., Haak, M. C., Pajkrt, E., Jonge, A. de, Henrichs, J., Franx, A., & Jellema, P. (2020). Quality assessment of ultrasonic foetal biometry during the IUGR Risk Selection (IRIS) trial: a cross sectional study. *Midwifery, 91*. doi:10.1016/j.midw.2020.102842

Version:	Publisher's Version
License:	Creative Commons CC BY 4.0 license
Downloaded from:	https://hdl.handle.net/1887/3635436

Note: To cite this publication please use the final published version (if applicable).

Midwifery 91 (2020) 102842

Contents lists available at ScienceDirect

Midwifery

journal homepage: www.elsevier.com/locate/midw

Quality assessment of ultrasonic foetal biometry during the IUGR Risk Selection (IRIS) trial: A cross sectional study

Viki Verfaille[a,b,*], Monique C. Haak[c], Eva Pajkrt[d], Ank de Jonge[a], Jens Henrichs[a], Arie Franx[e,f], Petra Jellema[a,g], IRIS study group

[a] Amsterdam University Medical Centre, Vrije Universiteit Amsterdam, Department of Midwifery Science, AVAG/Amsterdam Public Health, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands
[b] Dutch Professional Organisation of Sonographers, Edisonweg 8, 3442 AC Woerden, the Netherlands
[c] Department of Obstetrics and Gynaecology, Leiden University Medical Centre, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[d] Department of Obstetrics and Gynaecology, Amsterdam University Medical Centre, AMC, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands
[e] University Medical Centre Utrecht, Lundlaan 6, 3584 EA Utrecht, the Netherlands
[f] Department of Obstetrics and gynaecology, Erasmus Medical University Centre, Dr. Molewaterplein 40, 3015 GD Rotterdam, the Netherlands
[g] Zorginstituut Nederland, Willem Dudokhof 1, 1112 ZA Diemen, the Netherlands

ARTICLE INFO

Article history:
Received 3 March 2020
Revised 30 June 2020
Accepted 18 September 2020

Keywords:
Prenatal ultrasonography
Audit
Quality assessment
Foetal biometry
Image scoring

ABSTRACT

Objective: Intrauterine growth restriction is a major risk factor for perinatal morbidity and mortality. Ultrasonic foetal biometry is an important tool to monitor foetal growth. Therefore, the quality of these biometry scans is vital to achieve good diagnostic accuracy. We assessed the quality of foetal biometry during a nationwide trial and explored its association with sonographer's characteristics.

Methods: Four scans from every sonographer ($n = 154$), performed at 29 and 35 weeks gestational age were collected. Two assessors scored these scans according to a national audit system. A quality score $\geq$ 65% was considered 'adequate'.

We compared the quality scores per scoring criterion (i.e. foetal head measurements, abdominal circumference and femur length with regard to magnification, correctness of the plane and calliper placement) and gestational age. We analysed the associations between characteristics of the sonographers and their scores. In a subsample of scans of 30 sonographers we determined the interrater agreement on the quality scores given by the two assessors independently.

Findings: The mean score was 81.3%. Thirteen sonographers (8.4%) failed to achieve 'adequate quality'. Scores for femur length (83.8%) were significantly higher than those for head (77.9%) and abdominal circumference (78.6%) (both $P < 0.05$). Scores for correctness of the plane (73.4%) were lower than those for magnification (81.2%) and calliper placement (85.7%) (both $P < 0.05$). Gestational age did not affect the quality scores. Only the number of scans performed in the previous year was positively associated with the scores ($\beta = 0.01$; $P < 0.05$). The mean interrater difference in quality scoring was 11.1%, with 77.6% agreement on scans of 'adequate quality', but with no agreement on scans with 'insufficient quality'.

Key conclusions and implications for practice: Most sonographers achieved an 'adequate quality' score. Highest quality scores were attained for femur length, lowest quality scores for the correct plane. The number of scans one performs is associated with the quality scores, yet the minimum number of scans to perform for guaranteed quality still needs to be determined. Further research is needed to develop a standardized method to assess and maintain good ultrasonic foetal biometry quality.

© 2020 Elsevier Ltd. All rights reserved.

Introduction

Ultrasonic foetal biometry is important for the estimation of foetal weight and monitoring of foetal growth during preg-

Abbreviations: GA, gestational age; IRIS-trial, IUGR Risk Selection trial; IUGR, intrauterine growth restriction; SAS, standard anomaly scan; SD, standard deviation.

* Corresponding author at: Viki Verfaille, Amsterdam UMC, Vrije Universiteit, Department of Midwifery Science, Amsterdam Public Health, Van der Boechorststraat 7, G-102A, 1081 BT Amsterdam, the Netherlands.

E-mail addresses: verfaillevumc@gmail.com (V. Verfaille), m.c.haak@lumc.nl (M.C. Haak), e.pajkrt@amsterdamumc.nl (E. Pajkrt), ank.dejonge@amsterdamumc.nl (A. de Jonge), j.henrichs@amsterdamumc.nl (J. Henrichs), a.franx@erasmusmc.nl (A. Franx), PJellema@zinl.nl (P. Jellema).

https://doi.org/10.1016/j.midw.2020.102842
0266-6138/© 2020 Elsevier Ltd. All rights reserved.

nancy. Intrauterine growth restriction (IUGR) is a major risk factor for perinatal morbidity and mortality (Flenady et al., 2011; Gardosi et al., 2013), therefore high quality foetal biometry is essential to achieve diagnostic accuracy (Dudley and Chapman, 2002; Dudley, 2005; Ville, 2008). This requires attention because foetal biometry is frequently performed by sonographers with a variety of backgrounds such as obstetricians and midwives, and the results of these scans, regardless of their quality, can directly influence management decisions such as induction of premature labour.

In the Netherlands, sonographers are trained in basic ultrasonography in gynaecology and obstetrics, during or after an initial (para-)medical training (NVOG, 2009); further requirements are not standardized. Therefore educational and professional backgrounds, experience, ultrasonography equipment and the settings sonographers work in, including hospitals, may vary.

Nuchal translucency measurements and standard anomaly scans (SAS) require additional courses, certification and are, in some countries, subjected to audits (FMF, 2020a; Fetal Medicine Foundation (FMF) 2020b; Rijksinstituut voor Volksgezondheid en Milieu (RIVM) 2019). Although IUGR has a higher incidence (NVOG, 2017) than congenital anomalies (Anderson et al., 1995; Dolk et al., 2010), relatively little effort has been made to ensure the quality of ultrasonic foetal biometry in daily practice. Organizations like the International Society of Ultrasound in Obstetrics and Gynaecology have defined requirements for sonographers who perform foetal biometry and stress that quality assurance should be conducted routinely (Salomon et al., 2011). However, there is no guideline to assess the quality of foetal biometry. To our knowledge, quality assessments of foetal biometry in a substantial group of sonographers working in different daily practice settings, have not been reported before. It is unknown whether characteristics of sonographers are associated with the quality of foetal biometry.

In the IUGR Risk Selection (IRIS) trial, the (cost-)effectiveness of routine third trimester ultrasonic foetal biometry is examined using a nationwide stepped wedge cluster randomized trial (Henrichs et al., 2016). During this trial we explored the quality of foetal biometry performed in daily practice. Additionally we explored which characteristics of the sonographers are related to foetal biometry quality.

Methods

Participants

All sonographers ($n = 154$) performing routine foetal biometry in the IRIS-trial participated in the quality assessment. They had to complete an e-learning module on foetal biometry (LUMC, 2014) before the start of the trial. This internet-based module contained questions concerning the requirements for a biometry plane to be correct and concerning calliper placement. The sonographers had to obtain a score of at least 80% for this e-learning module.

As sonographers without SAS-certification do not participate in a biannual audit, we performed a biometry quality assessment according to the SAS-requirements to ensure a comparable level in this group ($n = 42$) before entering the trial. The three most recent, complete biometry scans (including pictures of HC, BPD, AC and FL measurements) were collected by one of the research assistants during a site visit. The sonographers were not informed this was how the selection was executed. The scans were evaluated as described in one of the following paragraphs. In analogy to the Dutch SAS-audit in 2014 (de Groot et al., 2014), a quality score of at least 65% was a prerequisite for participation in the IRIS-trial. Ten sonographers had a score below 65%; they received written feedback on their biometry images; they were free to follow additional training. All ten succeeded during a voluntary second attempt, at which point they were allowed to perform biometry scans in the IRIS-trial.

Data collection

During the IRIS-trial, images were collected between August 2015 and March 2016 (Henrichs et al., 2019). The six most recent biometry scans were collected from each participating sonographer: three scans performed at a gestational age (GA) between 28 and 30 weeks and three between 34 and 36 weeks. Each scan was from a different pregnant woman. At the research centre the scans were checked for completeness: including (a) pictures of HC, BPD, AC and FL measurements; and including (b) at least two scans performed at each GA. If more than four scans remained available, the four scans (two scans performed at each GA period) subjectively evaluated to be of best quality were forwarded for assessment as described in one of the following paragraphs.

During a site visit before entering the trial, information about the sonography centre and the sonographers was collected: years of experience since graduation from basic sonography training, SAS-certification, the number of sonography training activities attended in the previous year, the number of biometry scans performed in the previous year and type of work setting (e.g. midwifery practice or sonography centre).

Quality assessment

The quality was assessed by two experienced sonographers, both board members of the Dutch Professional Association of Sonographers for Obstetrics and Gynaecology (BEN, 2019) and these assessors were blinded to the name of the sonographer. All scans of the same sonographer were scored by a single assessor. At receipt, the assessors divided the scans based on their time available in order to guarantee a swift processing time.

Both assessors were certified for the SAS and nuchal translucency measurements. Each assessor had experience working in both sonography centres and in hospitals. One assessor had also performed advanced ultrasound examinations in a university hospital. These assessors had standardized their scoring beforehand through a consensus meeting and by independently scoring images until satisfactory agreement was achieved. This was defined as a maximum of one point discrepancy in the score (i.e. for plane, or magnification or calliper placement) per foetal measure per round, which is a maximum difference of 7.5% (3/40 points). Discrepancies were discussed in order to improve standardized scoring. After three rounds, each consisting of four biometry scans, the disagreement remained within the predefined limits allowing the quality assessment to start.

Ultrasonic foetal biometry was performed as described by Verburg et al. (2008) because the Verburg charts (Verburg et al., 2008; de Graaff et al., 2009) are recommended by the Dutch Society of Obstetrics and Gynaecology (de Graaff et al., 2009; NVOG, 2017). The image scoring criteria in Table 1 form the basis of the scoring instrument. Combining the scores for cephalic (range 0–4), abdominal (range 0–3) and femoral measurements (range 0–3) adds up to a maximum sub score of 10 points per scan, and a maximum total score of 40 points for 4 scans per sonographer. All quality scores, total score per sonographer and sub scores were recalculated into percentages to enable comparison. The mean scores for biparietal diameter and head circumference were utilized to compare the sub score for calliper placement in the cephalic plane to the calliper placement in the other planes.

All sonographers were informed about their achieved quality scores during the IRIS-trial and received detailed written feedback on their score. If the total score was below 65% (26/40 points), they

Table 1
Image scoring criteria for assessing the quality of ultrasonic foetal biometry Point per scoring criterion: 0=insufficient; 1=adequate. US, ultrasound; HC, head circumference; BPD, biparietal diameter.

	Cephalic plane(maximum 4 points)	*Abdominal plane*(maximum 3 points)	*Femoral plane*(maximum 3 points)
Magnification (max. 1 point) *Plane* (max. 1 point)	Area of interest is visualized full screen		
	• Symmetrical plane, oval shaped • Horizontal mid-line • Cavum septi pellucidi in the anterior third of the fronto-occipital diameter • Posterior horn of lateral ventricle visible	• Transverse section with a vertebra and a rib visible • Umbilical vein/ portal sinus in the anterior third of the abdomen • Kidneys not visible • Stomach preferably visible	• Line of the skin is parallel to the femur • Both ends of the bone clearly visible • Femur preferably in a horizontal plane, measuring the femur closest to the probe if both are visible
Calliper placement (max. 1 point)	BPD is measured at the widest part of the skull, from the outer borders of the parietal bones, without skin	At the outer borders of the body outline, skin covering	At the outer borders of the edges of the femoral diaphysis, without the cartilage of the epiphysis
(max. 1 point)	HC is measured using the ellipse facility placed at the outer borders of the skull: with the same callipers for BPD measurement and callipers placed on the outer borders of the occipital and frontal bones, without skin		

had to repeat the e-learning on foetal biometry within two weeks and confirm this by submitting their new e-learning score.

At the end of the study, a subsample of scans of 30 sonographers was selected randomly in SPSS to evaluate interrater agreement. For reasons of feasibility this was performed on a smaller subsample of 30 instead of 154. SPSS was programmed to oversample sonographers with an 'insufficient quality' score: at least 25% of the sample had to have a score <65% to enable the calculation of specific agreement. Both assessors performed first (i.e. during the IRIS trial) and second assessments (i.e. at the end of the trial). The second assessor was blinded to the score given by the first assessor. The second score was not reported to the sonographers.

Statistical analyses

All statistical analyses were performed with IBM SPSS Statistics 22. For all tests, a *P*-value below 0.05 was considered statistically significant.

In order to analyse whether the quality scores were related to the GA at which the foetal biometry was performed, we compared the images from 29 weeks to those from 35 weeks performed by the same sonographers using a paired *t*-test. We also compared the scores per scoring criterion with paired t-tests.

Using univariable and multivariable linear regression analyses with backward selection we investigated the association of characteristics of the sonographers with the quality score, i.e. SAS-certification, years of experience, work setting, number of biometry scans performed and training activities in the previous year.

Interrater agreement was analysed using a Bland-Altman plot. A difference up to 10% of the total score was decided to be acceptable. The proportions of specific agreement on 'adequate' (i.e. ≥65%) or 'insufficient quality' scores were reported in a contingency table.

Results

In this study 154 sonographers participated, some of them also performed scans for obstetrician-led care in hospitals. Their characteristics are presented in Table 2. Twenty-five of the organisations they worked in also conduct SAS. Participating organisations had a median number of two ultrasonography machines (range 1–12), with a median age of three years (range 0–7). The median number of sonographers was five (range 1–25) per organisation.

For the quality assessment, images of 616 biometry scans were used. The mean quality score of all sonographers (n = 154) was 81.3% (32.5/40 points, SD 12.0%, range 40–100%). Thirteen (8.4%) sonographers that had fulfilled the entrance requirements failed to achieve an adequate score of ≥ 65% during the study.

Because 42 sonographers without SAS-certification had undergone a quality assessment before entering the trial we could calculate their quality score before and during the trial. Their mean score increased from to 71.2% (SD 12%, range 37%–93%) to 80.4% (SD 12%, range 50%–98%) during the IRIS-trial. The ten sonographers who initially failed the entry-assessment but achieved an 'adequate quality' score at the repeated evaluation, managed to maintain this during the IRIS-trial.

Table 3 shows the quality scores from all 154 sonographers specified per scoring criterion and by GA at which the scan was performed. Combined scores obtained for the foetal femur (83.8%) were significantly higher than those obtained for the head (77.9%) and abdomen (78.6%). The score obtained for the correct plane (73.4%) was the lowest. There were no significant differences in quality scores between the foetal biometry performed at 29 weeks compared to 35 weeks of gestation.

The univariable associations between the characteristics of the sonographers and their obtained quality scores are shown in Table 4. In multivariable regression analyses including all characteristics of sonographers as predictor variables, the number of biometry scans performed in the previous year was the only significant correlate of quality scores (β=0.01, 95%CI; 0.00 to 0.02, P<0.05). This latter association is illustrated in Fig. 1.

The mean interrater difference was 11.1% (4.4/40 points, SD 12.4%) (Fig. 2). The 95% limits of agreement show differences up to 35.4% (14.2/40 points). Fig. 2 shows that average scores below 75% were structurally rated lower by one specific assessor and showed larger differences in the quality scoring. Higher scores seemed

Table 2
Characteristics of the sonographers n, absolute number, numbers differ due to missing values. The percentage of missing values ranged from 1.9% for number of biometry scans and training activities previous year to 5.8% for years of experience.

	Median (range) or number (%)
Certified for the standard anomaly scan (n = 154)	104 (67.5%)
Years of experience (n = 145)	8 (1–26)
Number of biometry scans performed previous year (n = 151)	223 (0–1232)
Number of training activities previous year (n = 151)	3 (0–28)
Workplace (n = 154) sonography centre midwifery practice combination/multiple sites	75 (48,7%) 52 (33,8%) 27 (17,5%)

Table 3
Quality scores (%) per scoring criterion and per gestational age group GA, gestational age; SD, standard deviation; CI, confidence interval.

	Mean score% (SD)	*Compared to*	*Mean difference*	*95% CI*	*P-value*
Femur Head Abdomen	83.8 (17.6) 77.9 (14.6) 78.6 (16.9)	Head Abdomen Femur	5.9 −0.6 −5.2	**3.0; 8.8** −3.2; 1.9 **−8.4; −2.1**	**<0.01** 0.61 **<0.01**
Plane Magnification Calliper placement	73.4 (16.1) 81.2 (23.8) 85.7 (16.6)	Magnification Calliper placement Plane	−7.7 −4.5 12.3	**−11.9; −3.6 −8.9;−0. 2 15.5; 9.1**	**<0.01** **0.04** **<0.01**
29 weeks GA 35 weeks GA	82.0 (14.3) 80.6 (12.9)	35 weeks GA	1.4	−0.6; 3.5	0.17

Table 4
Univariable associations between characteristics of the sonographer and the quality score β, regression coefficient; CI, confidence interval; SAS, standard anomaly scan.

Characteristic	β	*95% CI*	*P-value*
SAS-certification	0.56	−3.54; 4.67	0.79
Years of experience	0.15	−0.33; 0.64	0.53
Number of training activities	0.15	−0.32; 0.61	0.54
Number of biometry scans	0.01	**0.00; 0.01**	**0.03**
Workplace Midwifery practice Sonography centre Combination/multiple	reference 2.45 2.97	−1.84; 6.75 −2.67; 8.62	0.26 0.30

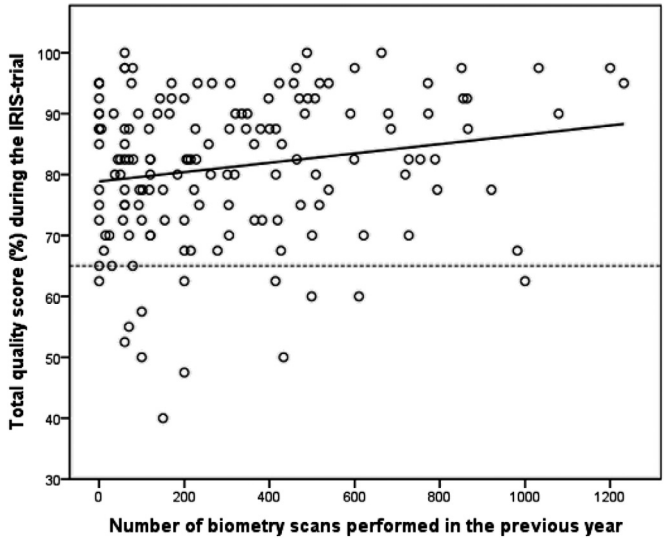

Fig. 1. Association between the number of biometry scans performed in the previous year and the quality score obtained during the IRIS-trial — 65% cut-off for an 'adequate quality' score.

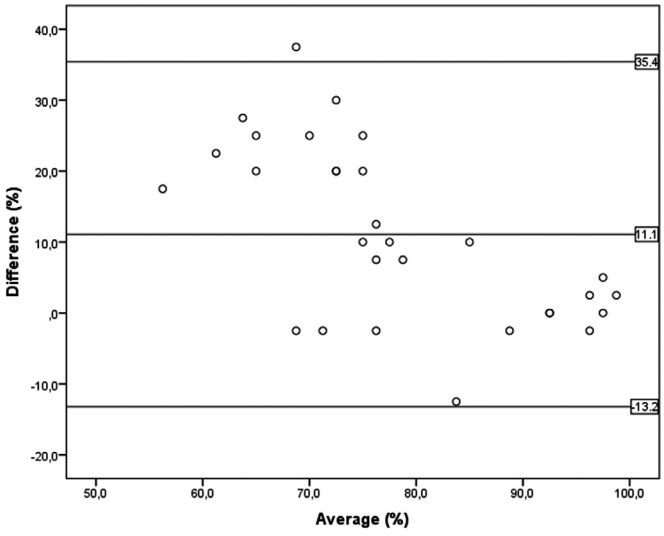

Fig. 2. Bland-Altman plot of the interrater agreement on biometry quality scores —– mean difference and 95% limits of agreement.

to be scored lower randomly by one assessor or the other, with smaller differences closer to the threshold of 10% (4/40 points).

Table 5 presents the agreement between the two assessors after dichotomizing the total quality scores ('adequate' vs 'insufficient quality'). In 77.6% of the cases the assessors agreed on 'adequate quality'. In other words, if one assessor rated the quality with a score $\geq$ 65%, there was a probability of 77.6% that the other assessor drew the same conclusion. There was no agreement on 'insufficient quality' because one of the assessors rated all scores as 'adequate'.

Discussion

In this study we assessed the quality of foetal biometry performed by sonographers in the IRIS-trial. The majority of sonographers achieved an 'adequate quality' score. Sonographers had most difficulties with obtaining the correct plane and those who performed more scans achieved higher quality. The interrater agreement on the quality scores was, however, limited.

This is one of the few studies that report quality scores for foetal biometry performed in daily practice on a large scale. In contrast to former studies performed in hospitals (Dudley and Potter, 1993; Dudley and Chapman, 2002) this study included a

Table 5
Agreement on the quality score by the two assessors Quality score ≥65%= 'adequate quality'; <65% ='insufficient quality'.

		Assessor A		
		≥65%	<65%	Total
Assessor B	≥65%	19	11	30
	<65%	0	0	0
	Total	19	11	30

large group of extramurally working sonographers, some of whom worked in hospitals as well. An average score of 81.3% is considered 'adequate quality', but leaves room for improvement. Thirteen (8%) sonographers failed to reach the minimum score of 65%.

The measurement of the foetal femur achieved the highest quality score. This is similar to earlier studies (Salomon et al., 2005a; Ursem et al., 2017). The femur is apparently easier to measure than the head and abdomen.

Of the scoring criteria, correct magnification was expected to be the simplest to achieve, yet it scored lower than calliper placement. However, the impact of magnification on accuracy is probably limited compared to not obtaining the correct plane or misplacing the callipers. It is important that sonographers are aware of these aspects of biometry quality and the consequences on diagnostic accuracy (Dudley and Chapman, 2002) as it can affect further management of pregnancy. Including self-assessment in quality assessments and access to reports as presented in Table 3 could encourage this awareness. Biometry training should focus on the criteria with the lowest quality scores.

Unexpectedly, the scores of the SAS-certified sonographers did not differ from the other sonographers, regardless of the additional training and requirements necessary to maintain their certification (FMF, 2020b; RIVM, 2016). Sonographers without SAS-certification may have benefited from the e-learning and feedback prior to entrance into the study. We found an increase of 9.2% (3.7/40 points) in quality score amongst the 42 sonographers without SAS-certification, scored before and during the study. A positive effect of audits and feedback has been reported on in other studies (Dudley and Potter, 1993; Dudley and Chapman, 2002; Sarris et al., 2011; Ursem et al., 2017), yet this subsample of 42 was too small to be statistically evaluated.

The number of biometry scans performed in the previous year was the only factor that was significantly associated with the quality score. We explored three different quantifications of experience: years of experience, number of training activities and biometry scans performed in the previous year. Generally, exposure improves performance. Our result is similar to findings in studies assessing the quality of SAS or nuchal translucency measurements (Hermann et al., 2013; Ursem et al., 2017). Therefore we recommend operationalizing experience in a required number of biometry scans per year rather than years of experience. Yet as Fig. 1 demonstrates, caution is needed in determining a threshold, as even some sonographers with a large number of 600 scans or more failed to acquire an 'adequate quality' score. Further research is needed to identify the minimum exposure required and whether a policy of ongoing certification increases diagnostic accuracy of foetal biometry.

Quality assessment of ultrasonic foetal biometry is complex. Quantitative approaches with Z-scores or cumulative summation techniques (Salomon et al., 2005b; Balsyte et al., 2010) are less prone to subjectivity or disagreement, but are labour intensive and require complex logistics that were not achievable in the IRIS-trial. Therefore, quality assessment was limited to the qualitative evaluation of biometry images, as performed in other studies (Dudley and Chapman, 2002; Sarris et al., 2013). The scoring instrument is readily available and is part of the SAS-audit in the Netherlands (Rijksinstituut voor Volksgezondheid en Milieu (RIVM) 2019; Ursem et al., 2017). It is similar to the image scoring method of Salomon et al. (2005a) There are few studies evaluating and reporting on these assessments. The 65% threshold for an 'adequate quality' of biometry scans was chosen because it was the same threshold used for the SAS-audit in the Netherlands in 2014 (de Groot et al., 2014). It therefore seemed a good starting point but an 'adequate' score of 65% still leaves room for all four scans to be partially 'insufficient' or one in three scans to be completely 'insufficient'. The Intergrowth 21st Project applied a comparable 67% threshold for certification (Sarris et al., 2013). An update in 2016 of the SAS-audit recommends a higher threshold of 75% (Rijksinstituut voor Volksgezondheid en Milieu (RIVM) 2019) which seems achievable in daily practice (Ursem et al., 2017). Dudley and Potter (1993) even reported proportions of more than 80% of scans meeting all pre-set quality criteria. Yet initially, sequential sampling and evaluation of substantial numbers of scans per sonographer is required for this scoring method to be applicable in daily care. Future research should explore the direct relationship between achieved quality scores and diagnostic accuracy in order to validate standards for pragmatic quality assessment.

The 10% (4/40 points) threshold for acceptability of differences in scoring between the assessors was arbitrary. Efforts to standardize the scores of the two assessors did not result in a continued high level of agreement. Dudley and Chapman (2002) also reported limited agreement between sonographers evaluating their own images and one auditor, especially for the measurement of the abdominal circumference. In contrast, (Salomon et al., 2005a) concluded that their method allowed for fair to good interrater reproducibility between three assessors. Yet they evaluated images obtained at 20–24 weeks GA by four sonographers, using the same probe and ultrasound machine. They advised repetition in a larger group, which has been done in the current study based on thirty sonographers of which 25% had an 'insufficient quality'. As the Bland-Altman plot shows, the disagreement occurs mainly in scores below 75%, so there may have been more agreement if it were determined on all 154 sonographers. Unfortunately, it was unfeasible to perform this on the complete group of 154 sonographers. Furthermore, although the assessors were explicitly informed about the threshold of <65% for an 'insufficient quality' score, the scoring instrument uses a continuous scale, that may have distracted the assessors from the dichotomous conclusion of 'adequate' versus 'insufficient quality'. The results of studies using image scoring to assess biometry quality, including ours, need to be considered with caution due to the lack of agreement between assessors. Further research is needed to develop or refine a reference standard for quality assessment that is generally accepted, affordable and applicable on a large scale and in daily practice.

We assessed the quality of ultrasonic foetal biometry in daily practice. Even though the majority of sonographers met the quality criteria, there is room for improvement. Training on the measurement of the foetal head and abdomen, and in particular on obtaining the right plane needs to be improved. The number of biometry scans performed per year is associated with better quality and the minimum number required should be determined. Further development and optimisation of the method to score ultrasonic foetal biometry quality in daily practice is warranted.

IRIS study group

Anneloes L. van Baar, Utrecht University, Utrecht, The Netherlands;

Joke M.J. Bais, Medical Centre Alkmaar (MCA), Alkmaar, The Netherlands;

Gouke J. Bonsel, University Medical Centre Utrecht, Utrecht, The Netherlands;
Judith E. Bosmans, VU University Amsterdam, Amsterdam, The Netherlands;
Jeroen van Dillen, Radboud University Medical Centre, Nijmegen, The Netherlands;
Noortje T.L. van Duijnhoven, Radboud University Medical Centre, Nijmegen, The Netherlands;
William A. Grobman, Northwestern University, Chicago, USA;
Henk Groen, University of Groningen, Groningen, The Netherlands;
Henriëtte van der Horst, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of General Practice and Elderly Care Medicine, Amsterdam Public Health, Amsterdam, Netherlands
Chantal W.P.M. Hukkelhoven, Wageningen University & Research, Wageningen, The Netherlands;
Trudy Klomp, Amsterdam UMC/AVAG, Amsterdam, The Netherlands;
Marjolein Kok, Amsterdam University Medical Centre, Amsterdam, The Netherlands;
Marlou L. de Kroon, University of Groningen, Groningen, The Netherlands;
Maya Kruijt, Dutch Society of Obstetrics and Gynecology NVOG, Utrecht, The Netherlands;
Anneke Kwee, University Medical Centre Utrecht, Utrecht, The Netherlands;
Sabina Ledda, Dutch Professional Association of Sonographers (BEN), Woerden, The Netherlands;
Harry N. Lafeber, Amsterdam University Medical Centre, The Netherlands;
Jan M. van Lith, Leiden University Medical Centre, Leiden, The Netherlands;
Ben Willem Mol, University of Adelaide, Adelaide, Australia;
Bert Molewijk, Amsterdam University Medical Centre, Amsterdam, The Netherlands;
Marianne Nieuwenhuijze, Academie Verloskunde Maastricht, Maastricht, The Netherlands;
Guid Oei, Maxima Medical Centre, Eindhoven, The Netherlands;
Cees Oudejans, Amsterdam University Medical Centre, Amsterdam, The Netherlands;
K. Marieke Paarlberg, Gelre Hospitals, Apeldoorn, The Netherlands;
Aris T. Papageorghiou, University of Oxford, UK;
Uma M. Reddy, Yale University, New Haven, USA;
Paul A.O.M. De Reu, Jiboro Health Care Centre/retired midwife, Eindhoven, The Netherlands;
Marlies Rijnders, TNO, Leiden, The Netherlands;
Alieke de Roon-Immerzeel, Royal Dutch Organisation of Midwives (KNOV), The Netherlands;
Connie Scheele, Dutch Professional Association of Sonographers (BEN), Woerden, The Netherlands;
Sicco A. Scherjon, University Medical Centre Groningen (UMCG), Groningen, The Netherlands;
Rosalinde Snijders, Amsterdam University Medical Centre, Amsterdam, The Netherlands;
Marc E. Spaanderman, Univarsity Medical Centre Maastricht, Maastricht, The Netherlands;
Pim W. Teunissen, Maastricht University, Maastricht, The Netherlands;
Hanneke W. Torij, Rotterdam University of Applied Sciences, Rotterdam, The Netherlands;
Jos Twisk, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam Public Health, Amsterdam, Netherlands;
Tanja G. Vrijkotte, Amsterdam University Medical Centre, Amsterdam, The Netherlands;
Kristel C. Zeeman, Royal Dutch Organisation of Midwives (KNOV), The Netherlands;
Jun Jim Zhang, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

Ethical approval

The Dutch Institutional Review Board of the VU Medical University Centre Amsterdam has approved the IRIS-trial (reference number: 2013.409). Women participating in the main study provided written informed consent about data collection, collection of ultrasonic foetal measurements included.

Funding sources

The Netherlands Organisation for Health Research and Development (ZonMW) funded the IRIS-study with grant no. 209030001, The Hague. The funder did not have any part in the study design, in the collection, analysis or interpretation of data, in the writing of the report, or in the decision to submit this article.

Clinical trial registry and registration number

The IRIS-trial is registered at the Netherlands Trial Registration, number NTR 4367.

Author contributions

Viki Verfaille, Petra Jellema, Monique Haak, Eva Pajkrt, Ank de Jonge and Arie Franx were involved in study conception and design.
Viki Verfaille, Ank de Jonge and Petra Jellema were involved in the development, implementation and/or data collection of the study.
Viki Verfaille and Jens Henrichs conducted data-analyses.
Viki Verfaille and Ank de Jonge drafted the manuscript and interpreted the results.
Viki Verfaille, Ank de Jonge and Jens Henrichs had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
All authors critically revised the manuscript for important intellectual content, approved its final version, and agree to be accountable for all aspects of the work presented in this manuscript.
The final version has also been revised and approved by the complete IRIS study group.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgments

We are grateful to all the sonographers who participated in this quality assessment on behalf of the IRIS-trial. We want to thank the two assessors who scored this impressive number of images alongside their daily obligations. We also want to thank professor Henrica C.W. de Vet for her expert advice in how to measure and finally analyse image scores. Finally we want to thank Jessica Hrudey for her language editing services.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.midw.2020.102842.

References

Anderson, N., Boswell, O., Duff, G., 1995. Prenatal sonography for the detection of fetal anomalies: results of a prospective study and comparison with prior series. Am. J. Roentgenol. 165, 943–950. https://doi.org/10.2214/ajr.165.4.7676997.

Balsyte, D., Schäffer, L., Burkhardt, T., Wisser, J., Zimmermann, R., Kurmanavicius, J., 2010. Continuous independent quality control for fetal ultrasound biometry provided by the cumulative summation technique. Ultrasound Obstet. Gynecol. 35, 449–455. https://doi.org/10.1002/uog.7545.

Beroepsvereniging Echoscopisten Nederland voor verloskunde en gynaecology (BEN) (Dutch Professional Association of Sonographers), 2019. Home page. Available from: https://www.echoscopisten.nl/en (Accessed 5 Feb 2020).

de Graaff, K., Laudy, J., Haak, M.C., 2009. Foetale echoscopie (Foetal Sonography) Versie 2.0. NVOG kwaliteitsnorm. Available from: https://www.nvog.nl/kwaliteitsdocumenten/kwaliteitsnormen/ (Accessed 5 Feb 2020).

de Groot, A., Lorijn-Peeters, E., Cohen-Overbeek, T., Go, A., Haak, M., 2014. Kwaliteitsbeoordeling SEO (Quality Assessment SAS). 7–8. Available from: http://www.rivm.nl/dsresource?objectid=rivmp:250228&type=org&disposition=inline&ns_nc=1. (Accessed 15 Jan 2018).

Dolk, H., Loane, M., Garne, E., 2010. The Prevalence of Congenital Anomalies in Europe. Springer, Netherlands, pp. 349–364.

Dudley, N.J., Potter, R., 1993. Quality assurance in obstetric ultrasound. Br. J. Radiol. 66, 865–870. https://doi.org/10.1259/0007-1285-66-790-865.

Dudley, N.J., Chapman, E., 2002. The importance of quality management in fetal measurement. Ultrasound Obstet. Gynecol. 19, 190–196. https://doi.org/10.1046/j.0960-7692.2001.00549.x.

Dudley, N.J., 2005. A systematic review of the ultrasound estimation of fetal weight. Ultrasound Obstet. Gynecol. 25, 80–89. https://doi.org/10.1002/uog.1751.

Flenady, V., Koopmans, L., Middleton, P., Frøen, J.F., Smith, G.C., Gibbons, K., Coory, M., Gordon, A., Ellwood, D., McIntyre, H.D., Fretts, R., Ezzati, M., 2011. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. Lancet 377, 1331–1340. https://doi.org/10.1016/S0140-6736(10)62233-7.

Fetal Medicine Foundation (FMF), 2020a. Nuchal translucency scan | Certificates of competence | Training & Certification | The Fetal Medicine Foundation. Available from: https://fetalmedicine.org/nuchal-translucency-scan (Accessed 5 Feb 2020).

Fetal Medicine Foundation (FMF), 2020b. The 18-23 weeks scan | Certificates of competence | Training & Certification | The Fetal Medicine Foundation. Available from: https://fetalmedicine.org/training-n-certification/certificates-of-competence/the-18-23-weeks-scan (Accessed 5 Feb 2020).

Gardosi, J., Madurasinghe, V., Williams, M., Malik, A., Francis, A., 2013. Maternal and fetal risk factors for stillbirth: population based study. Br. Med. J. 346, f108. https://doi.org/10.1136/bmj.f108.

Henrichs, J., Verfaille, Viester, V., Westerneng, L., Molewijk, M., Franx, B., van der Horst, A., Bosmans, H., de Jonge, J.E., Jellema, A., IRIS Study Group, P., 2016. Effectiveness and cost-effectiveness of routine third trimester ultrasound screening for intrauterine growth restriction : study protocol of a nationwide stepped wedge cluster-randomized trial in The Netherlands (The IRIS Study). BioMed Cent. Pregnancy Childbirth 16, 1–11. https://doi.org/10.1186/s12884-016-1104-8.

Henrichs, J., Verfaille, V., Jellema, P., Viester, L., Pajkrt, E., Wilschut, J., van der Horst, H., Franx, A., de Jonge, A.IRIS study group, 2019. Effectiveness of routine third trimester ultrasonography to reduce adverse perinatal outcomes in low risk pregnancy (the IRIS study): nationwide, pragmatic, multicentre, stepped wedge cluster randomised trial. Br. Med. J. 367, l5517. http://dx.doi.org/10.1136/bmj.l5517.

Hermann, M., Fries, N., Mangione, R., Boukobza, P., Ville, Y., Salomon, L.J., 2013. Nuchal translucency measurement: are qualitative and quantitative quality control processes related? Prenat. Diagn. 33 (8), 770–774. https://doi.org/10.1002/pd.4121.

Leids Universitair Medisch Centrum (LUMC), Haak, M.C., 2014. Medical Education. Basic Fetal Biometry. Available from: www.medicaleducation.nl//LRS.Net/Student/SearchResult.aspx?Search=ultrasoundbiometry (Accessed 5 Feb 2020).

Nederlandse Vereniging voor Obstetrie & Gynaecologie (NVOG) (Netherlands Society of Obstetrics and Gynaecology), 2009. Kwaliteitsnorm Foetale echoscopie (Quality Norm Foetal Sonography) Versie 2.0. Available from: https://www.nvog.nl/kwaliteitsdocumenten/kwaliteitsnormen/ (Accessed 5 Feb 2020).

Nederlandse Vereniging voor Obstetrie & Gynaecologie (NVOG) (Netherlands Society of Obstetrics and Gynaecology), 2017. NVOG-richtlijn Foetale groeirestrictie (FGR) (Guideline Foetal Growth Restriction). Available from: https://www.nvog.nl/kwaliteitsdocumenten/richtlijnen/perinatologie/ (Accessed 5 Feb 2020).

Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2019. Kwaliteitsbeoordeling Foetale Nekplooimeting (NT-meting) (Quality Assessment Fetal Nuchal Translucency Measurement) Versie 2. Available from: https://www.rivm.nl/documenten/kwaliteitsbeoordeling-foetale-nekplooimeting-nt (Accessed 5 Feb 2020).

Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2019. Kwaliteitsbeoordeling SEO (Quality Assessment Structural Anomaly Scan) Versie 3. Available from: https://www.rivm.nl/documenten/kwaliteitsbeoordeling-seo (Accessed 5 Feb 2020).

Salomon, L.J., Bernard, J.P., Duyme, M., Doris, B., Mas, N., Ville, Y., 2005a. Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester. Ultrasound Obstet. Gynecol. 27, 34–40. https://doi.org/10.1002/uog.2665.

Salomon, L.J., Bernard, J.P., Ville, Y., 2005b. Analysis of Z-score distribution for the quality control of fetal ultrasound measurements at 20-24 weeks. Ultrasound Obstet. Gynecol. 26, 750–754. https://doi.org/10.1002/uog.2640.

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., Lee, W., 2011. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet. Gynecol. 37, 116–126. https://doi.org/10.1002/uog.8831.

Sarris, I., Ioannou, C., Dighe, M., Mitidieri, A., Oberto, M., Qingqing, W., Shah, J., Sohoni, S., Al Zidjali, W., Hoch, L., Altman, D.G., Papageorghiou, A.T., 2011. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. Ultrasound Obstet. Gynecol. 38, 681–687. https://doi.org/10.1002/uog.8997.

Sarris, I., Ioannou, C., Ohuma, E.O., Altman, D.G., Hoch, L., Cosgrove, C., Fathima, S., Salomon, L.J., Papageorghiou, A.T., 2013. Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project. Br J Obstet Gynaecol 120, 33–37. https://doi.org/10.1111/1471-0528.12315.

Ursem, N., Peters, I.A., Kraan, -., van der Est, M.N., Knapen, M., Reijerink, J.C., Cohen-Overbeek, T.E., 2017. An audit of second-trimester fetal anomaly scans based on a novel image-scoring method in the Southwest region of the Netherlands. J. Ultrasound Med. 36, 1171–1179. doi:10.7863/ultra.16.06055.

Verburg, B.O., Steegers, E.A.P., De Ridder, M., Snijders, R.J.M., Smith, E., Hofman, A., Moll, H.A., Jaddoe, V.W., Witteman, J.C.M., 2008. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. Ultrasound Obstet. Gynecol. 31, 388–396. https://doi.org/10.1002/uog.5225.

Ville, Y., 2008. “Ceci n’est pas une échographie”: a plea for quality assessment in prenatal ultrasound. Ultrasound Obstet. Gynecol. 31, 1–5. https://doi.org/10.1002/uog.5248.