Ultrasound Obstet Gynecol 2023; **61**: 481–487
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.26051.

Standardization and quality control of Doppler and fetal biometric ultrasound measurements in low-income setting

S. ALI[1,2,3], J. BYAMUGISHA[2], M. G. KAWOOYA[3], I. M. KAKIBOGO[4], I. AINEMBABAZI[4], E. A. BIIRA[5], A. N. KAGIMU[3], A. MIGISA[3], M. MUNYAKAZI[3], S. KUNIHA[6], C. SCHEELE[7], A. T. PAPAGEORGHIOU[8], K. KLIPSTEIN-GROBUSCH[1,9] and M. J. RIJKEN[1,7]

[1]*Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands;* [2]*School of Medicine, Makerere University College of Health Sciences, Kampala, Uganda;* [3]*Ernest Cook Ultrasound Research and Education Institute (ECUREI), Mengo Hospital, Kampala, Uganda;* [4]*Antenatal and Maternity Unit, Kagadi Hospital, Kagadi District, Uganda;* [5]*The Woman's Place, Kampala, Uganda;* [6]*Department of Radiology, Mulago Hospital Complex, Kampala, Uganda;* [7]*Division of Woman and Baby, Department of Obstetrics, Birth Center Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands;* [8]*Nuffield Department of Women's and Reproductive Health, John Radcliffe Hospital, University of Oxford, Oxford, UK;* [9]*Division of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa*

KEYWORDS: antenatal; developing country; Doppler ultrasound; quality assurance; reliability; training

CONTRIBUTION

What are the novel findings of this work?
This is the first study in a low-resource setting to demonstrate that the quality of pulsed-wave Doppler and fetal biometry ultrasound measurements can be assessed reliably using freely available objective evaluation tools.

What are the clinical implications of this work?
Well-trained healthcare providers in underserved regions can perform Doppler and fetal biometry scans with consistency. The impact of in-service retraining and clinical audits on the quality of scans should be evaluated in future studies.

ABSTRACT

Objective *The aim of this study was to determine the quality of fetal biometry and pulsed-wave Doppler ultrasound measurements in a prospective cohort study in Uganda.*

Methods *This was an ancillary study of the Ending Preventable Stillbirths by Improving Diagnosis of Babies at Risk (EPID) project, in which women enroled in early pregnancy underwent Doppler and fetal biometric assessment at 32–40 weeks of gestation. Sonographers undertook 6 weeks of training followed by onsite refresher training and audit exercises. A total of 125 images for each of the umbilical artery (UA), fetal middle cerebral artery (MCA), left and right uterine arteries (UtA), head circumference (HC), abdominal circumference (AC) and femur length (FL) were selected randomly from the EPID study database and evaluated independently by two experts in a blinded fashion using objective scoring criteria. Inter-rater agreement was assessed using modified Fleiss' kappa for nominal variables and systematic errors were explored using quantile–quantile (Q–Q) plots.*

Results *For Doppler measurements, 96.8% of the UA images, 84.8% of the MCA images and 93.6% of the right UtA images were classified as of acceptable quality by both reviewers. For fetal biometry, 96.0% of the HC images, 96.0% of the AC images and 88.0% of the FL images were considered acceptable by both reviewers. The kappa values for inter-rater reliability of quality assessment were 0.94 (95% CI, 0.87–0.99) for the UA, 0.71 (95% CI, 0.58–0.82) for the MCA, 0.87 (95% CI, 0.78–0.95) for the right UtA, 0.94 (95% CI, 0.87–0.98) for the HC, 0.93 (95% CI, 0.87–0.98) for the AC and 0.78 (95% CI, 0.66–0.88) for the FL measurements. The Q–Q plots indicated no influence of systematic bias in the measurements.*

Conclusions *Training local healthcare providers to perform Doppler ultrasound, and implementing quality control systems and audits using objective scoring tools in clinical and research settings, is feasible in low- and*

Correspondence to: Dr S. Ali, School of Medicine, Makerere University College of Health Sciences, P.O. Box 7072, Kampala, Uganda (e-mail: alisambecker@gmail.com; s.ali-2@umcutrecht.nl)

Accepted: 1 August 2022

middle-income countries. Although we did not assess the impact of in-service retraining offered to practitioners deviating from prescribed standards, such interventions should enhance the quality of ultrasound measurements and should be investigated in future studies.

INTRODUCTION

Stillbirth and its associated psychosocial and economic costs are critical global health problems affecting disproportionately parents, healthcare providers and communities in low- and middle-income countries (LMICs)[1–3]. Over 50% of stillbirths in sub-Saharan Africa are antepartum and the majority are linked to placental failure[4]. Placental dysfunction leads to impaired exchange of oxygen and nutrients at the maternal–fetal interface and manifestation of acute or chronic fetal hypoxia, which is known to be associated with increased impedance to flow in the umbilical (UA) and uterine (UtA) arteries[5]. This stresses the importance of examining the placental and fetal circulation (commonly using Doppler ultrasound), in addition to fetal biometry, in order to identify at-risk pregnancies.

Given the essential role of Doppler ultrasonography and fetal biometry in clinical practice, these measurements should be accurate and reproducible. Slight variation in measurements may be unavoidable, but significant systematic errors can lead to misinterpretation, inappropriate intervention and harmful effects on pregnant women[6]. In the context of research, erroneous findings could prompt misguided public health policies. Thus, measurements should be performed and interpreted by well-trained healthcare providers, using adequate equipment and following standardized procedures.

In addition, regular departmental audits are necessary to identify ultrasonographers who may require tailored feedback and retraining to ensure that desired examination standards are upheld[7,8]. Audit tools for two-dimensional fetal biometry[9], crown–rump length (CRL)[10] and pulsed-wave Doppler measurements[7,8] have been published. Such tools are based on objective equally weighted scoring criteria and have been found to be more reproducible compared with subjective evaluation in high-income countries (HICs)[7–9]. However, studies documenting the use and applicability of obstetric ultrasound quality assurance procedures and tools in routine clinical and research settings in LMICs are limited. Thus, it remains fertile ground for research, with implementation of Doppler technology a top priority to lessen the burden of stillbirth and perinatal complications. Here we report on the quality of fetal biometry and pulsed-wave Doppler ultrasound measurements and the quality control strategies employed in a large prospective cohort study conducted in a rural community in Uganda.

METHODS

Design and participants

This quality assessment study was designed and reported in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS)[11]. We report on the ultrasound measurement quality control protocols followed in the Ending Preventable Stillbirths by Improving Diagnosis of Babies at Risk (EPID) project, a prospective cohort study implemented between 2018 and 2020 in a rural obstetric care facility in western Uganda[12–14]. The primary aim of the EPID study was to determine the predictive performance of Doppler ultrasound for adverse perinatal outcome in a low-resource setting. Pregnant women were enroled at < 23 weeks' gestation and were offered fetal growth and Doppler scans in the third trimester (between 32 and 40 weeks). Perinatal and maternal outcomes were assessed at the time of birth and up to 28 days postnatally. Additional information on the EPID study is published elsewhere[12–14]. All women recruited to the main prospective cohort provided informed and written consent, with illiterate participants signing using a thumbprint. This study was approved by the Makerere University School of Medicine Research and Ethics Committee (SOMREC) (reference 2018-090) and the Uganda National Council for Science and Technology (UNCST) (reference HS 2459). We obtained permission to work within the Kagadi region from the Kagadi District Health Team and local authorities.

Setting

The EPID study was implemented in Kagadi Hospital, a secondary healthcare facility handling around 4000 births annually and serving women in the rural communities of the greater Kibaale region in midwestern Uganda, approximately 215 km from the national capital, Kampala. In 2016, the Uganda Demographic and Health Survey estimated the total population of Kibaale at 788 714, with 389 278 (49.4%) males and 399 436 (50.6%) females[15].

Data collection

The EPID project adapted the strategies for standardization and quality control of ultrasound measurements from the methodological recommendations of the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st) project[16]. The training and quality control program was led and co-ordinated by a maternal–fetal ultrasound specialist and research scientist (S.A.). We created a training and quality control team (S.A., E.A.B., A.N.K., A.M., M.M., S.K., C.S. and M.J.R.) that supported the training of sonographers and the continuous assessment of the quality and consistency of ultrasound measurements obtained during the study. The EPID study quality control team had 8–15 years' experience in obstetric ultrasound as of 2020.

The training components included, but were not limited to: pregnancy dating using ultrasound; recognition of common fetal anomalies; principles and safety concerns associated with Doppler ultrasound and interpretation of results; scanning techniques; and general data collection procedures. To ensure continuity of the service beyond the termination of the study, two of the trainees were onsite residential sonographers at the time of study implementation. We evaluated the trainees' knowledge and skills using pre- and post-training theoretical and practical assessments. Even though the trainee sonographers had been practising basic obstetric ultrasound for 11–15 years as of 2018, they had little-to-no experience in performing Doppler scans. Their preassessment results also indicated the need for thorough training. Thus, the prestudy training exercise lasted for 6 weeks: 1 week was dedicated to the review of theoretical materials, 3 weeks to hands-on practice at training centers and 2 weeks were spent at the study site.

Preimplementation training took place at Ernest Cook Ultrasound Research and Education Institute (ECUREI), Mengo Hospital, Kampala, Uganda and The Woman's Place, Kampala, Uganda. ECUREI is a local tertiary teaching institution that was nominated as a center of excellence by the World Federation for Ultrasound in Medicine and Biology in 2007 and has participated previously in ultrasound capacity-building projects. The Woman's Place is a local maternal–fetal ultrasound clinic equipped with state-of-the-art ultrasound systems for specialized obstetric and gynecological imaging.

Before commencing participant enrolment, we conducted additional onsite training to familiarize the sonographers with the study equipment and data collection procedures. We undertook spot reviews of a few randomly selected images on three occasions during implementation (once during the initial phase of participant enrolment and twice at the start of follow-up examinations in the third trimester), of which the results highlighted areas for onsite retraining. Monthly site visits were organized to address implementation issues, such as equipment servicing, adherence to study protocol and refresher training for the entire study team.

All fetal biometry and Doppler scans in the EPID study were performed by two Kagadi Hospital resident ultrasonographers and the first author (S.A.), following prescribed standards[17,18], using a Voluson™ e (GE Healthcare, Zipf, Austria) or Philips HD9 (Philips Ultrasound, Amsterdam, The Netherlands) machine. Head circumference (HC) was acquired in the transthalamic plane with calipers placed on the outer border of the skull (Figure S1). Abdominal circumference (AC) was measured in an axial plane, with the umbilical vein in the anterior third of the fetal abdomen (at the level of the portal sinus) and the stomach bubble visible. The femur length (FL) of the limb closest to the transducer was obtained with the calipers placed on the outer borders of the diaphysis of the femoral bone. The UA was examined in a free loop of the umbilical cord, with measurements taken in the absence of fetal movement, while keeping the insonation angle at $< 30^\circ$ (Figure S2). The fetal middle cerebral artery (MCA) was examined at its proximal third, close to its origin in the internal carotid artery, with the angle of insonation kept as close as possible to 0°. The UtAs were recorded transabdominally, with the angle of insonation maintained at $< 30^\circ$.

Ultrasound images were scored according to a set of published criteria (Appendices S1 and S2)[7–9] by two independent raters (Raters A and B), who were blinded to each other's results. Each criterion was assigned 1 point when satisfied and 0 when not satisfied, and was equally weighted towards the total score, which was calculated as the sum of the points. A maximum of 4 points were available for FL and 6 points for HC, AC, UA, MCA and UtA. Women for whom all fetal biometry and Doppler measurements had been obtained were selected randomly for inclusion in this study. Based on previous studies[7,8], a minimum of 125 images would be sufficient to detect a 10% difference in agreement between two raters with 90% power, assuming an inter-rater agreement of 80%. Therefore, we randomly selected 125 images per ultrasound examination for this analysis, using the set.seed function in R (R Foundation for Statistical Computing Platform, Vienna, Austria) to ensure that our random sample was reproducible.

Statistical analysis

The total image scores of each rater were dichotomized into acceptable and unacceptable. For FL, images with a score of ≥ 3 points were classified as acceptable, while for HC, AC, UA, MCA and UtA images, a threshold of ≥ 4 points was considered acceptable. These classification schemes were based on published studies and the recommendations of the developers of the scoring tools[7,8,19]. Descriptive statistics for image quality scores were reported as *n* (%) and were presented graphically using bar graphs.

To determine inter-rater agreement, we used the s* statistic, a modified Fleiss' kappa for nominal variables that is not affected by the paradoxes of Cohen's and Fleiss' kappa statistics[20]. The 95% bootstrap CI of the s* statistic was calculated using Monte Carlo simulations with 1000 iterations. *P*-values were also approximated using the Monte Carlo procedure at a 5% level of significance. For interpretation of kappa, we used the following cut-off values: 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, good agreement; and > 0.80, very good agreement[11].

Analysis of the *Z*-score distribution is also a recommended approach for quality assessment[21]. The *Z*-score distribution is expected to follow the properties of a standard normal distribution; the mean and SD should be approximately 0 and 1, respectively. We first transformed the HC, AC, FL and UA pulsatility index (PI) measurements into gestational-age-specific *Z*-scores using the INTERGROWTH-21st fetal-growth standards[22] embedded in the R package healthy birth, growth & development (hbgd)[23] and INTERGROWTH-21st Doppler charts[24]. We then constructed quantile–quantile (Q–Q) plots for the HC, AC, FL and UA-PI *Z*-scores to allow for visual assessment of their distributions. The mean and SD of the *Z*-score distributions were compared with the standard normal distribution. We also used the Shapiro–Wilk test to assess normality at a 5% level of significance. Data were managed in STATA version 14.0 (StataCorp., College Station, TX, USA) and analyzed using the package raters in R version 4.0.4 (R Foundation for Statistical Computing Platform; https://www.r-project.org/).

RESULTS

Image quality

A total of 875 ultrasound images (125 each for the HC, AC, FL, UA, MCA and each UtA) scored by two raters were considered for this analysis. The image-quality scores for pulsed-wave Doppler and fetal biometry awarded by Raters A and B are reported in Table 1 and Figures S3 and S4. For the MCA, 119/125 (95.2%) images were classified as acceptable by Rater A, 111/125 (88.8%) by Rater B and 106/125 (84.8%) by both raters. For the UA, 125/125 (100%) images were classified as acceptable by Rater A, 121/125 (96.8%) by Rater B and 121/125 (96.8%) by both raters. For biometry measurements, 124/125 (99.2%), 120/125 (96.0%) and 120/125 (96.0%) AC images were classified as acceptable by Rater A, Rater B and both raters, respectively. For the FL, Rater A scored 114/125 (91.2%) images, Rater B scored 120/125 (96.0%) images and both raters scored 110/125 (88.0%) images as acceptable. The proportion of images meeting each constituent criterion of the overall quality score is reported for fetal biometry in Table S1 and Doppler ultrasound in Table S2.

Inter-rater agreement

The overall inter-rater percentage agreement for objective evaluation of UA Doppler images was 96.8% (modified kappa, 0.94 (95% CI, 0.87–0.99); $P < 0.001$) (Table 2). For the MCA, inter-rater agreement was 85.6% (modified kappa, 0.71 (95% CI, 0.58–0.82); $P < 0.001$), while for the UtAs, inter-rater agreement was 93.6% (modified kappa, 0.87 (95% CI, 0.78–0.95); $P < 0.001$) and 91.2% (modified kappa, 0.82 (95% CI, 0.71–0.92); $P < 0.001$), for the right and left sides, respectively. The inter-rater percentage agreement for objective evaluation of AC measurement was 96.8% (modified kappa, 0.93 (95% CI, 0.87–0.98), $P < 0.001$), and for HC measurement was 96.8% (modified kappa, 0.94 (95% CI, 0.87–0.98), $P < 0.001$).

Table 1 Proportion of fetal biometry and pulsed-wave Doppler ultrasound images in pregnant women at 32–40 weeks' gestation that were classified as acceptable, according to Raters A and B

Parameter	*Rater A*	*Rater B*	*Both raters*
Biometry			
HC*	120 (96.0)	124 (99.2)	120 (96.0)
AC*	124 (99.2)	120 (96.0)	120 (96.0)
FL†	114 (91.2)	120 (96.0)	110 (88.0)
Doppler			
UA*	125 (100)	121 (96.8)	121 (96.8)
MCA*	119 (95.2)	111 (88.8)	106 (84.8)
RUtA*	124 (99.2)	118 (94.4)	117 (93.6)
LUtA*	123 (98.4)	116 (92.8)	114 (91.2)

Data are given as *n* (%). For each parameter, 125 images were assessed. *Acceptable score cut-off is ≥ 4. †Acceptable score cut-off is ≥ 3. AC, abdominal circumference; FL, femur length; HC, head circumference; LUtA, left uterine artery; MCA, middle cerebral artery; RUtA, right uterine artery; UA, umbilical artery.

Table 2 Percentage agreement and modified Fleiss' kappa between Raters A and B for quality of fetal biometry and pulsed-wave Doppler measurements in pregnant women at 32–40 weeks' gestation

Parameter	*Agreement (%)*	*Modified Fleiss' kappa (95% CI)*	P
Biometry			
HC	96.8	0.94 (0.87–0.98)	< 0.001
AC	96.8	0.93 (0.87–0.98)	< 0.001
FL	88.8	0.78 (0.66–0.88)	< 0.001
Doppler			
UA	96.8	0.94 (0.87–0.99)	< 0.001
MCA	85.6	0.71 (0.58–0.82)	< 0.001
RUtA	93.6	0.87 (0.78–0.95)	< 0.001
LUtA	91.2	0.82 (0.71–0.92)	< 0.001

For each parameter, 125 images were assessed. AC, abdominal circumference; FL, femur length; HC, head circumference; LUtA, left uterine artery; MCA, middle cerebral artery; RUtA, right uterine artery; UA, umbilical artery.

Z-score distributions

The Q–Q plots in Figure 1 demonstrate normal distributions of the HC, AC, FL and UA-PI gestational-age-specific *Z*-scores. The mean ± SD was −0.02 ± 1.00 for HC *Z*-scores, 0.03 ± 0.92 for AC *Z*-scores, 0.53 ± 0.82 for FL *Z*-scores and −0.27 ± 1.05 for UA-PI *Z*-scores in the sample. The corresponding Shapiro–Wilk test *P*-values for the HC *Z*-scores, AC *Z*-scores, FL *Z*-scores and UA-PI *Z*-scores were 0.955, 0.409, 0.416 and 0.286, respectively.

DISCUSSION

Summary of key findings

The quality of ultrasound measurements obtained in the EPID study was high, with ≥ 84.8% of pulsed-wave Doppler images and ≥ 88.0% of fetal biometry images scored as acceptable by both raters. Inter-rater agreement was good or very good for Doppler and biometry images, with a modified kappa of 0.94 (95% CI, 0.87–0.99) for the UA and 0.94 (95% CI, 0.87–0.98) for the HC. All fetal biometry and UA-PI *Z*-scores had normal distributions, implying negligible influence of systematic error in our measurements.

Strengths and limitations

Strengths of this study include the fact that ultrasound measurements in the EPID study were acquired using standardized techniques and that the sonographers underwent a longer course of preimplementation training (up to 6 weeks) compared with that in most published studies, supplemented by audits and refresher training. Although we did not measure the impact of the retraining exercise, it could have enhanced the quality of measurements. Quality assessment was performed by experienced reviewers who were blinded to each other's ratings and to the findings of the study sonographer. Further, they were not involved in data collection or retraining, allowing for independence in their reviews.

Although we used arbitrary thresholds to classify images as acceptable or unacceptable, this approach is recommended in the literature to simplify the interpretation of findings[7,8,19]. According to the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) clinical standards committee, a comprehensive quality control strategy should also involve assessment of caliper placement bias and the limits of agreement of the actual measurements. These are not reported as we did not obtain multiple measurements in a repeated fashion from each woman, for reasons of feasibility. However, we used the *Z*-score distribution method to assess for systematic bias in the measurements. Another limitation was the resource-intensive nature of the training exercise. However, based on our experience conducting ultrasound studies in low-resource contexts, similar results are achievable within a shorter time period. Future studies should also include ultrasound-naïve practitioners.

Interpretation

This study shows that it is possible to train ultrasonographers in underprivileged regions to perform fetal biometry and Doppler scans with consistency. The quality of ultrasound measurements can be assessed reliably using freely available objective evaluation tools[7,8,19]. In this study, ≥ 85% of acquired Doppler images were judged to be of an acceptable standard by two independent raters. Similar findings were reported in a multicenter Doppler study conducted in a HIC, in which 89.2% of MCA images and 85.0% of UA images were of acceptable quality[25]. Moreover, nearly 90% of images of each fetal biometric parameter in the present study were deemed acceptable. In comparison, over 98% of biometry images were of very high quality in a large multicenter international project[19].

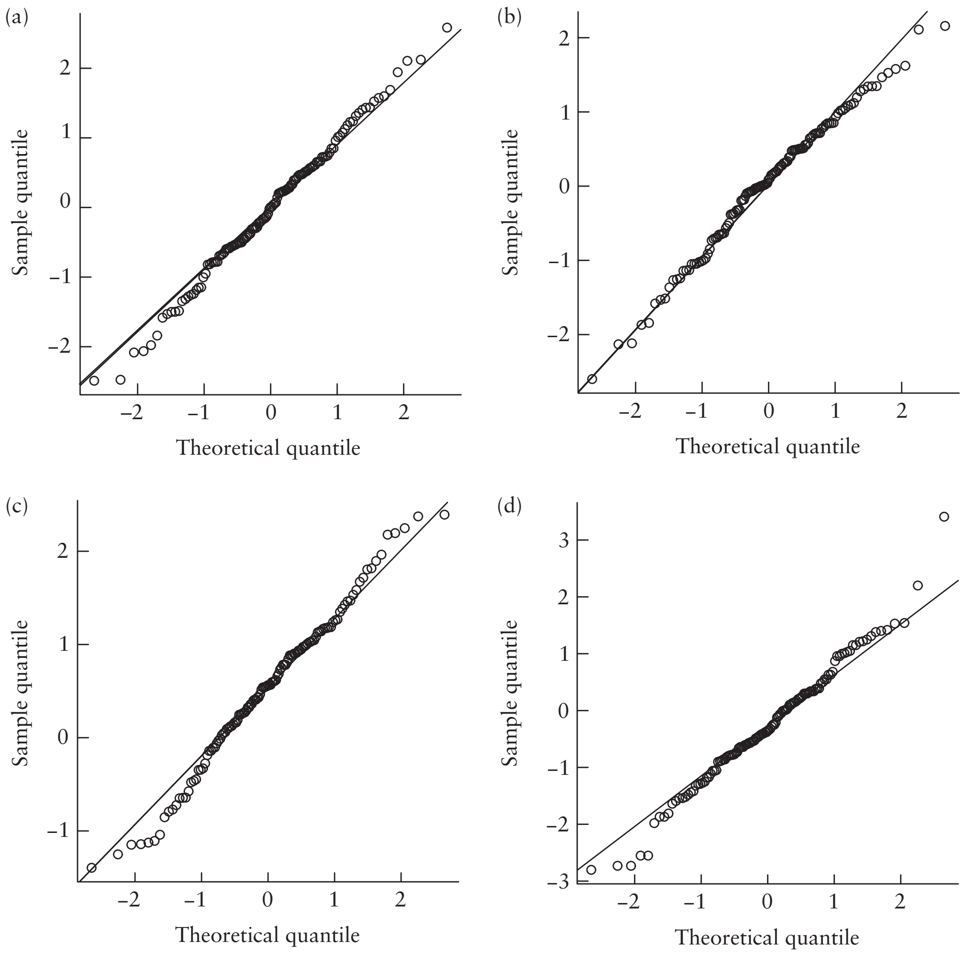

Figure 1 Quantile–quantile plots for fetal head circumference (a), abdominal circumference (b), femur length (c) and umbilical artery pulsatility index (d) *Z*-scores in pregnant women at 32–40 weeks' gestation.

The inter-rater agreement in this study was good or very good for all Doppler images, with an agreement rate of 96.8% (modified kappa, 0.94) for the UA and 85.6% (modified kappa, 0.71) for the MCA. Previous studies from HICs have demonstrated that interobserver agreement for Doppler image assessment is very good when using an objective scoring system, and higher than that achieved by subjective assessment, in both clinical and research settings[7,8,25]. The INTERGROWTH-21st group reported overall agreement of 85% (adjusted kappa, 0.70) when using an objective scale compared with 70% (adjusted kappa, 0.47) for subjective assessment of UA and UtA Doppler images[8]. Likewise, objective assessment with a 6-point scoring system had higher interobserver agreement (91.9%; kappa, 0.839) compared with subjective agreement (75.8%; kappa, 0.516) for MCA images[7]. In a multicenter randomized controlled trial, kappa values for inter-rater reliability of quality assessment were 0.85 (95% CI, 0.81–0.89) and 0.84 (95% CI, 0.80–0.89) for the MCA and UA, respectively[25]. Further, a high level of inter-rater agreement was reported for HC (kappa, 0.99 (95% CI, 0.98–0.99)), AC (kappa, 0.98 (95% CI, 0.97–0.99)) and FL (kappa, 0.96 (95% CI, 0.95–0.98)) measurements across multiple INTERGROWTH-21st study sites adhering to strict quality control measures[19]. Similarly, the inter-rater agreement for all fetal biometry measurements obtained in this study was very good.

It is not surprising that most of our findings were comparable with those from HICs. Acceptable and accurate Doppler scans are achievable when performed by adequately trained ultrasonographers observing strict examination protocols[6]. We demonstrated previously similar results for CRL measurements in a Ugandan clinical setting[26] and for fetal biometry performed by local healthcare workers in a refugee camp on the Thai–Burmese border[27]. Standardization of obstetric ultrasound practice in LMICs is important for clinical care and research, particularly in multicenter studies in which a broad range of settings, women and practitioners are involved. Although there was no evidence of systematic error in our measurements, differences in reviewer scores for individual elements of the scoring system, such as UA image clarity (Table S2), emphasize the need to adequately orient practitioners using clinical tools to ensure uniform interpretation.

As the World Health Organization now recommends the use of UA Doppler to manage high-risk pregnancies[28], it is imperative that quality control systems are established and adhered to in obstetric ultrasound units in LMICs. Local guidelines for the management of suspected growth-restricted fetuses should be developed using a bottom-up collaborative approach, and should emphasize the use of similar and context-appropriate reference standards constructed using robust methodologies, such as the INTERGROWTH-21st charts, to deliver appropriate care to women[29,30].

Increasing access and promoting the efficient use of ultrasound technology in LMICs will require commitments from governments, funding agencies and international communities to improve the quality of antenatal care. A structured training program in obstetric ultrasound, taking into account the local context and available cadres at the frontline, may be a constructive strategy[27,31]. With the advent of artificial intelligence, there is hope for future commercial products with the ability to support practitioners to undertake and interpret complex ultrasound procedures with high precision[32,33]. Such modern clinical decision support tools would be of great benefit in high-burden settings in which the number of highly skilled fetal medicine specialists falls far short of demand.

Conclusions

Training healthcare providers in underserved regions to undertake Doppler ultrasound examinations with consistency is feasible. Implementation of quality control systems, using freely available objective ultrasound image scoring tools, is recommended in clinical and research settings in low-resource communities. Future studies should seek to evaluate the impact of regular in-service audits and retraining.

ACKNOWLEDGMENTS

We thank the study participants, Kagadi District Health Team, The Woman's Place, ECUREI and Makerere University College of Health Sciences for supporting the implementation of the EPID project. We are indebted to the University of Oxford, Oxford, UK, for contributing the ultrasound equipment for data collection. Further, our gratitude goes to Fosca Tumushabe who assisted with the English language review. We thank the INTERGROWTH-21st consortium for assistance and for providing freely available training materials and protocols (https://intergrowth21.tghn.org/) and Lior Drukker for his help in calculating umbilical artery Doppler centiles. This project was funded by Grand Challenges Canada through a generous grant (number: R-ST-POC-1808-17038) to Makerere University College of Health Sciences. S.A. is supported by the UMC Utrecht Global Health Support Program (reference: FM/ADB/D-18-015006), University Medical Center Utrecht, Utrecht, The Netherlands. A.T.P. is funded by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the NIHR Biomedical Research Centre funding scheme.

REFERENCES

1. Mills TA, Ayebare E, Mweteise J, Nabisere A, Mukhwana R, Nendela A, Omoni G, Wakasiaka S, Lavender T. 'There is trauma all round': A qualitative study of health workers' experiences of caring for parents after stillbirth in Kenya and Uganda. *Women Birth* 2023; **36**:56–62.
2. Heazell AEP, Siassakos D, Blencowe H, Burden C, Bhutta ZA, Cacciatore J, Dang N, Das J, Flenady V, Gold KJ, Mensah OK, Millum J, Nuzum D, O'Donoghue K, Redshaw M, Rizvi A, Roberts T, Toyin Saraki HE, Storey C, Wojcieszek AM, Downe S, Flenady V, Frøen JF, Kinney M V, de Bernis L, Lawn JE, Blencowe H, Heazell AEP, Leisher SH, Radestad I, Jackson L, Ogwulu C, Hills A, Bradley S, Taylor W, Budd J. Stillbirths: economic and psychosocial consequences. *Lancet* 2016; **387**: 604–616.

3. Hug L, You D, Blencowe H, Mishra A, Wang Z, Fix MJ, Wakefield J, Moran AC, Gaigbe-Togbe V, Suzuki E, Blau DM, Cousens S, Creanga A, Croft T, Hill K, Joseph KS, Maswime S, McClure EM, Pattinson R, Pedersen J, Smith LK, Zeitlin J, Alkema L. Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: a systematic assessment. *Lancet* 2021; **398**: 772–785.
4. Lawn JE, Blencowe H, Waiswa P, Amouzou A, Mathers C, Hogan D, Flenady V, Frøen JF, Qureshi ZU, Calderwood C, Shiekh S, Jassir FB *et al.* Stillbirths: Rates, risk factors, and acceleration towards 2030. *Lancet* 2016; **387**: 587–603.
5. Burton GJ, Jauniaux E. Pathophysiology of placental-derived fetal growth restriction. *Am J Obstet Gynecol* 2018; **218**: S745–761.
6. Rolnik DL, da Silva Costa F, Sahota D, Hyett J, McLennan A. Quality assessment of uterine artery Doppler measurement in first-trimester combined screening for pre-eclampsia. *Ultrasound Obstet Gynecol* 2019; **53**: 245–250.
7. Ruiz-Martinez S, Volpe G, Vannuccini S, Cavallaro A, Impey L, Ioannou C. An objective scoring method to evaluate image quality of middle cerebral artery Doppler. *J Matern Neonatal Med* 2020; **33**: 421–426.
8. Molloholli M, Napolitano R, Ohuma EO, Ash S, Wanyonyi SZ, Cavallaro A, Giudicepietro A, Barros F, Carvalho M, Norris S, Min AM, Zainab G, Papageorghiou AT. Image-scoring system for umbilical and uterine artery pulsed-wave Doppler ultrasound measurement. *Ultrasound Obstet Gynecol* 2019; **53**: 251–255.
9. 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**: 34–40.
10. Wanyonyi SZ, Napolitano R, Ohuma EO, Salomon LJ, Papageorghiou AT. Image-scoring system for crown–rump length measurement. *Ultrasound Obstet Gynecol* 2014; **44**: 649–654.
11. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; **64**: 96–106.
12. Ali S, Kawooya MG, Byamugisha J, Kakibogo IM, Biira EA, Kagimu AN, Grobbee DE, Zakus D, Papageorghiou AT, Klipstein-Grobusch K, Rijken MJ. Middle cerebral arterial flow redistribution is an indicator for intrauterine fetal compromise in late pregnancy in low-resource settings: a prospective cohort study. *BJOG* 2022; **129**: 1712–1720.
13. Ali S, Kabajaasi O, Kawooya MG, Byamugisha J, Zakus D, Papageorghiou AT, Klipstein-Grobusch K, Rijken MJ. Antenatal Doppler ultrasound implementation in a rural sub-Saharan African setting: exploring the perspectives of women and healthcare providers. *Reprod Health* 2021; **18**: 199.
14. Ali S, Heuving S, Kawooya MG, Byamugisha J, Grobbee DE, Papageorghiou AT, Klipstein-Grobusch K, Rijken MJ. Prognostic accuracy of antenatal Doppler ultrasound for adverse perinatal outcomes in low-income and middle-income countries: a systematic review. *BMJ Open*. 2021; **11**: e049799.
15. Uganda Bureau of Statistcs (UBOS) and ICF. Uganda Demographic and Health Survey 2016: Key Indicators Report. https://www.ubos.org/wp-content/uploads/publications/03_2018Uganda_DHS_2016_KIR.pdf
16. Sarris I, Ioannou C, Ohuma EO, Altman DG, Hoch L, Cosgrove C, Fathima S, Salomon LJ, Papageorghiou AT; International Fetal and Newborn Growth Consortium for the 21st Century. Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project. *BJOG* 2013; **120**: 33–37.
17. Salomon LJ, Alfirevic Z, Da Silva Costa F, Deter RL, Figueras F, Ghi T, Glanc P, Khalil A, Lee W, Napolitano R, Papageorghiou A, Sotiriadis A, Stirnemann J, Toi A, Yeo G. ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. *Ultrasound Obstet Gynecol* 2019; **53**: 715–723.
18. Bhide A, Acharya G, Baschat A, Bilardo CM, Brezinka C, Cafici D, Ebbing C, Hernandez-Andrade E, Kalache K, Kingdom J, Kiserud T, Kumar S, Lee W, Lees C, Leung KY, Malinger G, Mari G, Prefumo F, Sepulveda W, Trudinger B. ISUOG Practice Guidelines (updated): use of Doppler velocimetry in obstetrics. *Ultrasound Obstet Gynecol* 2021; **58**: 331–339.
19. Cavallaro A, Ash ST, Napolitano R, Wanyonyi S, Ohuma EO, Molloholli M, Sande J, Sarris I, Ioannou C, Norris T, Donadono V, Carvalho M, Purwar M, Barros FC, Jaffer YA, Bertino E, Pang R, Gravett MG, Salomon LJ, Noble JA, Altman DG, Papageorghiou AT. Quality control of ultrasound for fetal biometry: results from the INTERGROWTH-21st Project. *Ultrasound Obstet Gynecol* 2018; **52**: 332–339.
20. Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. *Stat Methods Med Res* 2014; **25**: 2611–2633.
21. Salomon LJ, Bernard JP, Ville Y. Analysis of Z-score distribution for the quality control of fetal ultrasound measurements at 20–24 weeks. *Ultrasound Obstet Gynecol*. 2005; **26**: 750–754.
22. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Ismail LC, Lambert A, Jaffer YA, Bertino E, Gravett MG, Purwar M, Noble JA, Pang R, Victora CG, Barros FC, Carvalho M, Salomon LJ, Bhutta ZA, Kennedy SH, Villar J. International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet*. 2014; **384**: 869–879.
23. Hafen R, Anderson C, Schloerke B. hbgd: healthy birth, growth & development. 2016. https://github.com/HBGDki/hbgd.
24. Drukker L, Staines-Urias E, Villar J, Barros FC, Carvalho M, Munim S, McGready R, Nosten F, Berkley JA, Norris SA, Uauy R, Kennedy SH, Papageorghiou AT. International gestational age-specific centiles for umbilical artery Doppler indices: a longitudinal prospective cohort study of the INTERGROWTH-21st Project. *Am J Obstet Gynecol* 2020; **222**: 602.e1–15.
25. Rial-Crestelo M, Morales-Roselló J, Hernández-Andrade E, Prefumo F, Oros D, Caffici D, Sotiriadis A, Zohav E, Cruz-Martinez R, Parra-Cordero M, Lubusky M, Kacerovsky M, Figueras F. Quality assessment of fetal middle cerebral and umbilical artery Doppler images using an objective scale within an international randomized controlled trial. *Ultrasound Obstet Gynecol* 2020; **56**: 182–186.
26. Ali S, Byanyima RK, Ononge S, Ictho J, Nyamwiza J, Loro ELE, Mukisa J, Musewa A, Nalutaaya A, Ssenyonga R, Kawooya I, Temper B, Katamba A, Kalyango J, Karamagi C. Measurement error of mean sac diameter and crown–rump length among pregnant women at Mulago hospital, Uganda. *BMC Pregnancy Childbirth* 2018; **18**: 129.
27. Rijken MJ, Lee SJ, Boel ME, Papageorghiou AT, Visser GHA, Dwell SLM, Kennedy SH, Singhasivanon P, White NJ, Nosten F, McGready R. Obstetric ultrasound scanning by local health workers in a refugee camp on the Thai–Burmese border. *Ultrasound Obstet Gynecol* 2009; **34**: 395–403.
28. World Health Organization. *WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience*. 2016. https://www.who.int/publications/i/item/9789241549912.
29. Papageorghiou AT, Kennedy SH, Salomon LJ, Altman DG, Ohuma EO, Stones W, Gravett MG, Barros FC, Victora C, Purwar M, Jaffer Y, Noble JA, Bertino E, Pang R, Cheikh Ismail L, Lambert A, Bhutta ZA, Villar J. The INTERGROWTH-21st fetal growth standards: toward the global integration of pregnancy and pediatric care. *Am J Obstet Gynecol* 2018; **218**: S630–640.
30. Drukker L, Staines-Urias E, Papageorghiou AT. The INTERGROWTH-21st Doppler centile charts: complementing tools for monitoring of growth and development from pregnancy to childhood. *Am J Obstet Gynecol* 2021; **224**: 249–250.
31. Kawooya MG. Training for Rural Radiology and Imaging in Sub-Saharan Africa: Addressing the Mismatch Between Services and Population. *J Clin Imaging Sci* 2012; **2**: 37.
32. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol* 2020; **56**: 498–505.
33. Drukker L. Real-time identification of fetal anomalies on ultrasound using artificial intelligence: what's next? *Ultrasound Obstet Gynecol* 2022; **59**: 285–287.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

Figure S1 Fetal biometry images showing correct measurement of head circumference (a), abdominal circumference (b) and femur length (c).

Figure S2 Color and pulsed-wave Doppler images showing correct measurement of umbilical artery (a), middle cerebral artery (b) and uterine artery (c).

Figure S3 Distribution of the total quality scores for pulsed-wave Doppler images of the umbilical artery (a), middle cerebral artery (b), right uterine artery (c) and left uterine artery (d), according to Raters A and B.

Figure S4 Distribution of the total quality scores for ultrasound images of head circumference (a), abdominal circumference (b) and femur length (c), according to Raters A and B.

Appendix S1 Image-scoring criteria for umbilical artery, middle cerebral artery and uterine artery pulsed-wave Doppler ultrasound measurements

Appendix S2 Image-scoring criteria for head circumference, abdominal circumference and femur length ultrasound measurements

Table S1 Proportion of fetal biometric ultrasound images meeting each constituent criterion of the quality score, overall and according to rater

Table S2 Proportion of pulsed-wave Doppler ultrasound images meeting each constituent criterion of the quality score, overall and according to rater