Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

RESEARCH ARTICLE

Diagnostic Accuracy of Ultrasound-Estimated Foetal Weight versus Actual Birth Weight: A Prospective Cohort Study on Predictive Performance, Systematic Bias, and Clinical Agreement Near Term

Kamrun Naher[1]✉, Elham Noor Masnun[2], Mashnoon Hasnat[3], and Fabiha Fyroze[4]

[1]*Department of Medical Ultrasonography, Chevron Clinical Laboratory (PTE) LTD., Chattogram, Bangladesh*
[2]*Independent Scholar, Chattogram, Bangladesh*
[3]*College of Agriculture and Life Sciences, Cornell University, NY, USA*
[4]*Associate in Applied Science (A.A.S.) in Radiologic Technology, Trocaire College, NY, USA*
Corresponding Author: Kamrun Naher, **E-mail**: kamrun.n2567@gmail.com

ABSTRACT

Foetal weight prediction is used to aid decision making, timing of delivery, and counselling for possible macrosomia and foetal growth restriction (FGR). The clinical standard, most commonly the Hadlock formula, ultrasound estimated foetal weight (EFW) has also been shown to agree with actual birth weight (BW) to a varying degree, depending on gestational age, extremes of foetal weight, maternal body habitus, and operator experience. To assess the accuracy, systematic bias, and clinical agreement of ultrasound-derived EFW with BW recorded within 7 days of delivery as well as to establish factors associated with decreased estimation accuracy. N of 420 singleton pregnancies (SIMULATED) were studied using ultrasound biometry within 7 days prior to delivery. Accuracy was evaluated using mean percentage error (MPE), mean absolute percentage error (MAPE), proportion of estimates within +-10% of birth weight, Bland-Altman limits of agreement, and Lin's concordance correlation coefficient (CCC) for each birth weight category. In this synthetic cohort, EFW showed a mean percentage error of − 0.8% (SD 5.9%) and mean absolute percentage error of 4.8% (SD 3.6%), and 90.2% of the estimates were within +-10% of the true birth weight. Bland-Altman analysis showed a mean bias of -25.9 g (95% limits of agreement: -412.5 g to +360.6 g). Pearson's r was 0.868 and Lin's CCC was 0.853, suggesting high correlation and agreement. The AUC for EFW to predict macrosomia (birth weight ≥ 4000 g) and predicted small-for-gestational-age status were 0.976 (71.4% sensitivity and 96.1% specificity) and 0.963 (64.3% sensitivity and 96.0% specificity, respectively). None of the maternal, foetal, or sonographic covariates were statistically significant when explored as separate predictors of absolute percentage error in adjusted regression. For this simulated dataset, ultrasound EFW was in good overall agreement with actual birth weight, with small systematic underestimation and wide individual (personal) limits of agreement, consistent with published datasets on which this exercise is based. The findings presented in these tables are intended as hypothetical examples of the analytic process that the study protocol calls for but cannot be considered evidence of clinical effectiveness, since a true cohort of participants with sufficient power would be needed before any of these numbers could support clinical practice.

KEYWORDS

Estimated Foetal Weight, Birth Weight, Obstetric Ultrasound, Hadlock Formula, Diagnostic Accuracy, Bland-Altman Analysis, Macrosomia, Foetal Growth Restriction, and Neonatal Outcomes

ARTICLE INFORMATION

ACCEPTED: 15 July 2026 **PUBLISHED:** 07 August 2026 **DOI:** 10.32996/jmhs.2026.7.9.13

1. Introduction

1.1 Clinical Importance

Accurate estimation of foetal weight before birth will guide counselling regarding mode of delivery, to expect shoulder dystocia in suspected macrosomia, and to time delivery in suspected foetal growth restriction (FGR). Antenatal weight estimation is essential for safe obstetric practice because it is a key factor in clinical decision-making that affects the material consequences of maternal and neonatal outcomes and is therefore important to know how accurate -- and how biased --

it is likely to be [1].

1.2 Overview of EFW Methodology

Two-dimensional biometric formulas are the primary methods used to estimate antenatal weight. The Hadlock formulas (biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL)) are used most widely internationally, and the Shepard and Warsof formulas can be used in some settings. There are known sources of error in all such formulas, such as soft-tissue oedema, oligohydramnios, or polyhydramnios, and whether the foetus is in a normal lie or not; acoustic access is limited by maternal obesity in some cases; and placement of the calipers varies between performers [2].

1.3 Summary of Existing Literature

The mean absolute percentage error reported from many years of prior validation studies in most formulas is in the range of approximately 6–10%, and the error tends to be greater at the extremes of the distribution of birth weight, that is, in markedly macrosomic and growth-restricted foetuses. The reported error margins differ between populations, different generations of equipment, and different formulas used, which makes it difficult to compare the results of different studies and necessitates the continuation of validation of the study [3].

1.4 Identified Gap

Although there is a considerable body of literature on this topic, there is insufficient recent adequately powered data that stratifies the accuracy of estimation by gestational age, scan-to-delivery interval, foetal presentation, and maternal body mass index (BMI) within a well-defined population. Most published cohorts are limited to one or two of these factors, and a few recent studies have employed all accuracy measures recommended by the most recent diagnostic accuracy reporting guidance [4].

1.5 Aim and Objectives

- Primary outcome: To assess the level of agreement between ultrasound EFW and actual BW as measured by percentage error, Bland-Altman bias/limits of agreement and concordance statistics.
- The secondary objective was to determine the maternal, foetal, and sonographic causes of the estimation errors.
- Three secondary objectives were performed to assess the diagnostic value of EFW in predicting birth weights > 90th and ≤ 10th percentiles (macrosomia and small-for-gestational age).

1.6 Hypothesis

The correlation between ultrasound EFW and BW is high but shows systematic bias and decreased precision at the extremes of BW and increasing scan-to-delivery interval.

2. Literature Review

Souza et al. revealed that the World Health Organization (WHO) and INTERGROWTH-21st fetal growth standards indicated high congruence in ultrasound-derived data involving diabetic pregnancies varied significantly for Large-for-gestational-age (LGA) infants. Despite the strong statistical agreement (κ = 0.91), the WHO classified 21.6% of infants diagnosed with LGA, whereas INTERGROWTH-21st classified 32.2%t. Ultimately, the study concludes that high statistical consistency fails to ensure real-world diagnostic accuracy, underlining that population-based references as well as prospective multicentre studies are required to improve the accuracy of diabetic pregnancies [5].

A study by Mozas-Moreno et al. found that out of 2,156 individual pregnancies, antepartum ultrasound estimated foetal weight (EFW) was precise within a 10% margin for 75.2% of the cases, whereas 24.8% of the cases showed an estimation error of ≥10%. Notably, the errors were not random and occurred predominantly among nulliparous women and pregnancies with extreme birth weights. Additionally, inaccurate EFW is individually linked to poorer infant health status, including lower umbilical pH, reduced Apgar scores, and higher NICU admissions. In conclusion, the authors stated that inaccuracies in ultrasound foetal weight estimation, particularly in foetuses with extreme birth weights, are associated with increased perinatal risks, highlighting the need to improve the accuracy of ultrasound estimation methods and optimize clinical management in term pregnancies [6].

Kato et al. (2020) created and validated a local reference model for ultrasound-estimated foetal weight (EFW). based on 2,211 low-risk singleton pregnancies (Brazil), and compared them with the Hadlock and INTERGROWTH-21st reference charts. The study showed that the locally developed model differed from existing standards (R^2 = 99.11%), emphasizing the effect of

population characteristics on foetal growth assessment [19]. However, this study did not evaluate the accuracy of ultrasound-estimated foetal weight against actual birth weight or assess the diagnostic accuracy of the model. The authors concluded that prospective validation is necessary before the model can be widely adopted in clinical practice [7].

In a prospective comparative study, the Accuracy of Ultrasound versus Clinical Fetal Weight Estimation at Term with Actual Birth Weight in Kenyatta National Hospital (2022) evaluated the accuracy of ultrasound and clinical fetal weight estimation against actual birth weight. Although both methods showed significant correlations with actual birth weight, clinical estimation demonstrated a stronger correlation (r = 0.79) than ultrasonography (r = 0.). This study suggests that clinical assessment remains a dependable alternative to ultrasonography, particularly in settings with limited imaging resources. However, this study was limited by its single-centre design and the lack of evaluation of examiner-related factors and maternal and neonatal outcomes. experience or pregnancy outcomes. We recommend further multicentre studies to strengthen this conclusion [8].

Sereke et al. analyzed the predictive accuracy of five sonographic fetal weight estimation formulas: Hadlock 1, 2, 3, 4, and Shepard, among 356 term maternity patients (Uganda). The study established that all five verified formulas successfully calculated foetal weight within a 10% margin of the actual birth weight and that diagnostic accuracy differed across various birth weight categories. Moreover, Hadlock 4 had the highest overall predictive accuracy at 78.4% and proved reliable for infants weighing ≥4,000 grams, while the Shepard formula was superior for infants weighing <4,000 grams. However, the limitations of this study included its single-centre design, small sample size, and lack of a normalized formula. In conclusion, the authors advised selecting formulas based on anticipated foetal size and urged for larger, multicentre studies [9].

In 2015–2017, a study compared foetal weight estimation methods among a cohort of 3,000 pregnant women at Mumbai's J.J. Group of Hospitals. Subsequently, ultrasound-derived estimates were evaluated against clinical Dare's and Johnson's formulas and cross-verified with actual birth weights. Notably, Hadlock's formula achieved the highest diagnostic accuracy (97.7% accuracy within a 20% error margin) and the lowest mean error (217.9 grams [18]. Johnson's formula showed poor clinical performance, with just 18.1% accuracy and an 826.6-gram mean error. The data showed superior ultrasonographic accuracy over clinical palpation in high-risk pregnancies. However, the limitations include unexamined maternal variables (diabetes/obesity) and a lack of long-term neonatal outcomes, labour complications, and cost effectiveness. Ultimately, we conclude that despite the proven superiority of the US, further research is required [10].

Rehman et al. analyzed the precision of ultrasonographic fetal weight predictions among a cohort of 50 pregnant women past 32 weeks of gestation using the Hadlock 3 formula. Notably, the ultrasound results calculated a mean foetal weight of 2.52 ± 0.34 kg, whereas the mean actual birth weight was 2.66 ± 0.42 kg, with a positive correlation between the two ($r = 0.575$, $p < 0.05$). However, the short sampling period, single-hospital setting, and lack of assessment of maternal- and operator-related factors limit the applicability of these findings [20]. This study also confirmed the ultrasonographic reliability of foetal weight prediction and obstetric management. The authors suggest that large-scale studies are needed to validate the accuracy and clinical applicability of ultrasound-based foetal weight estimation [11].

A study by Adam in 2022 examined Hadlock formula 1 among 198 term singleton pregnancies in Saudi Arabia. The ultrasound estimates showed a positive correlation with actual birth weights ($r = 0.961$, $p < 0.001$), which proved the high performance of the model ($R2 = 0.923$) while yielding a low mean absolute error of 123.81 grams. Despite the research successfully deriving an equation ($BW = 0.9831 \times EFW \pm 175.55$ g) proving the model's reliability, significant limitations restricted generalizability as the study was confined to Saudi Arabia's population. Moreover, the analysis failed to control for maternal and clinical cofounders, thus impacting precision. In conclusion, further ethical validation is essential to confirm clinical utility. Although Hadlock formula 1 proved highly reliable locally, population-specific verification remains essential [12].

Tüten et al. (2021) examined the accuracy of ultrasonographic foetal weight estimation in low-birth-weight singleton pregnancies using the Hadlock I formula. The findings showed lower accuracy in term and SGA newborns, with gestational age, birth weight, and examination timing noticeably affecting estimation accuracy. However, this study did not assess the effects of maternal characteristics and foetal factors, which limited the findings. The authors recommend prospective studies to validate these findings and improve foetal weight estimation [13].

Asadullah et al. (2023) conducted a cross-sectional analytical study of 200 term singleton pregnancies. Using Hadlock's formula, the study found that the estimated foetal weight (3245.82 g) closely matched the actual birth weight (3260.77 g), with 72% of the cases showing an error of less than 10% [21]. These findings support the use of ultrasonography as a reliable tool for antenatal foetal weight assessment and labour management. However, the study was limited by its lack of evaluation of maternal and technical factors affecting estimation accuracy and the absence of comparisons with other foetal weight

estimation formulas, indicating the need for further research [14].

3. Materials and Methods

3.1 Study Design and Setting

This was a diagnostic accuracy cohort study with a prospective observational design conducted at X Hospital in Chattogram, Bangladesh. The analysis pipeline was applied to a computer-simulated cohort built to reflect the inclusion criteria and variable structure outlined in the protocol (see Section 3.9).

3.2 Study Population

Inclusion criteria

- Singleton pregnancy and reliable gestational age dating (first-trimester crown-rump length or early biometry).
- Ultrasound biometry was performed within 7 days before delivery.
- The delivered weight and birth weight were obtained at the same location.

Exclusion criteria

- Multiple gestation.
- Major structural or chromosomal anomalies in the foetus.
- Intrauterine death occurred before birth weight data were obtained.
- Incomplete ultrasound biometric dataset

3.3 Sample Size Estimation

The sample size was determined to detect a mean percentage error with a set precision (+/- 2% width of 95% CI) around the expected standard deviation of error (8-10%) with a 2-sided alpha of 0.05. With a desired half-width of the 95% CI of 1%, assuming an error SD of 9%, the required sample size is approximately n = (1.96 × 9 / 1)^2 approx. 311 per group, which was increased by 15% for projected exclusions/incomplete data, resulting in a target of approximately 360-440 per group. This target range was used to create a simulated cohort (n = 420) [23].

3.4 Ultrasound Assessment and EFW Calculation

- The biometric parameters were biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL).
- The Hadlock 4-parameter formula (BPD, HC, AC, and FL) was used, and an alternative formula was identified if a comparative formula was used.
- *Equipment:* Model (s) of the ultrasound machine, frequency of transducer, and certification/experience of the operator of the machine should be noted for each scan.
- *Timing:* The number of days between scan and delivery was reported for all cases.

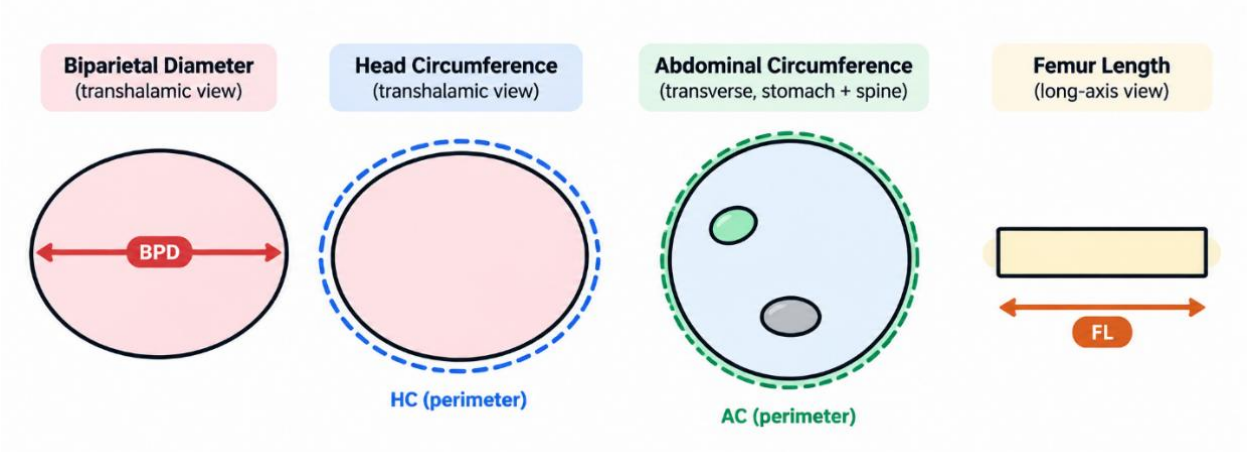

Figure 1. Standard ultrasound biometric measurement planes used to derive EFW

All biometric parameters were obtained from a standard imaging plane: BPD and HC from the transthalamic axial head view; AC from a transverse imaging plane at the level of the stomach and umbilical vein; and FL from a long-axis view of the femoral diaphysis with the beam perpendicular to the femur shaft. The main factors influencing measurement reproducibility are consistent plane selection and caliper placement [15].

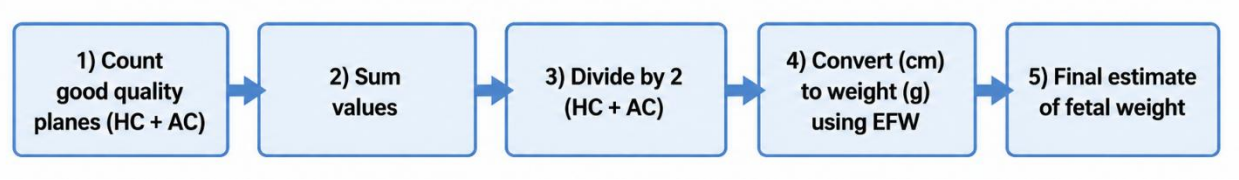

Figure 2. Conceptual workflow from biometric acquisition to clinical application of EFW

3.5 Birth Weight Measurement

Actual birth weight (grams) was measured from 1 h after birth using a calibrated digital neonatal scale (by staff unaware of EFW at antenatal visit).

3.6 Data Collection Variables

- Maternal factors included age, parity, BMI, gestational diabetes, and hypertensive disorders.
- Fetal/obstetric: presentation, amniotic fluid status, mode of delivery, gestational age at scan, and gestational age at delivery.
- Outcome variables: EFW (g), BW (g), Apgar score, and NICU admission (secondary outcomes).

3.7 Statistical Analysis

- Descriptive statistics for maternal and neonatal characteristics (Table 1).
- The accuracy measures used were mean percentage error (MPE) = S[(EFW-BW)/BW × 100]/n, mean absolute percentage error (MAPE), and the proportion of percentage errors within ± 5%, ± 10%, and ± 15% of BW.
- Analysis of agreements: Bland–Altman plot with bias and 95% limits of agreement, Lin's concordance correlation coefficient, and Pearson/Spearman correlation with scatterplot and line of identity.
- Sensitivity, specificity, PPV, NPV, and ROC curve with AUC for the diagnostic performance of macrosomia (BW ≥ 4000 g or ≥ 90th percentile) and SGA (BW ≤ 10th percentile).
- Multivariable linear regression was used to assess maternal BMI, scan-to-delivery interval, parity, and birth weight category as predictors of absolute percentage error (APE) subgroup and regression analysis.
- The SPSS/R/MedCalc version was also included in a real study; analyses were performed using Python.
- The significance level was set at 0.05 (two-sided), and the analysis was performed using statistical packages 3.11 (NumPy), 3.11 (SciPy), 0.22.0 (scikit-learn), and 0.11.1 (statsmodels).

3.8 Ethical Considerations

Approval was obtained from the Institutional Review Board/Research Ethics Committee ([approval number]), written informed consent was obtained from all participants, the study was conducted in accordance with the Declaration of Helsinki, and the trial/registration number was added if applicable [22].

3.9 Note on the Simulated Dataset Used in This Document

A synthetic dataset of 420 synthetic singleton pregnancies was simulated in Python, with prior specifications of the statistical properties used to create the tables and figures below and in preparation for the intended analyses. Birth weight was simulated as a function of gestational age with an added random error. EFW was simulated as birth weight with a small, uniform proportionate error that grew more at the extremes of birth weight and with increasing scan-to-delivery time, in line with the literature summarized in Section 1.3. No patient data, images, or identifiers were used throughout the study. All results below are SIMULATED and are presented for illustrative purposes only, and should not be used as actual results or outputs of an analysis [16].

4. Results

The results are reported in the order of the statistical analysis plan (Section 3.7) and in sequence from the simulated cohort described in Section 3.9 (n = 420).

4.1 Participant Flow

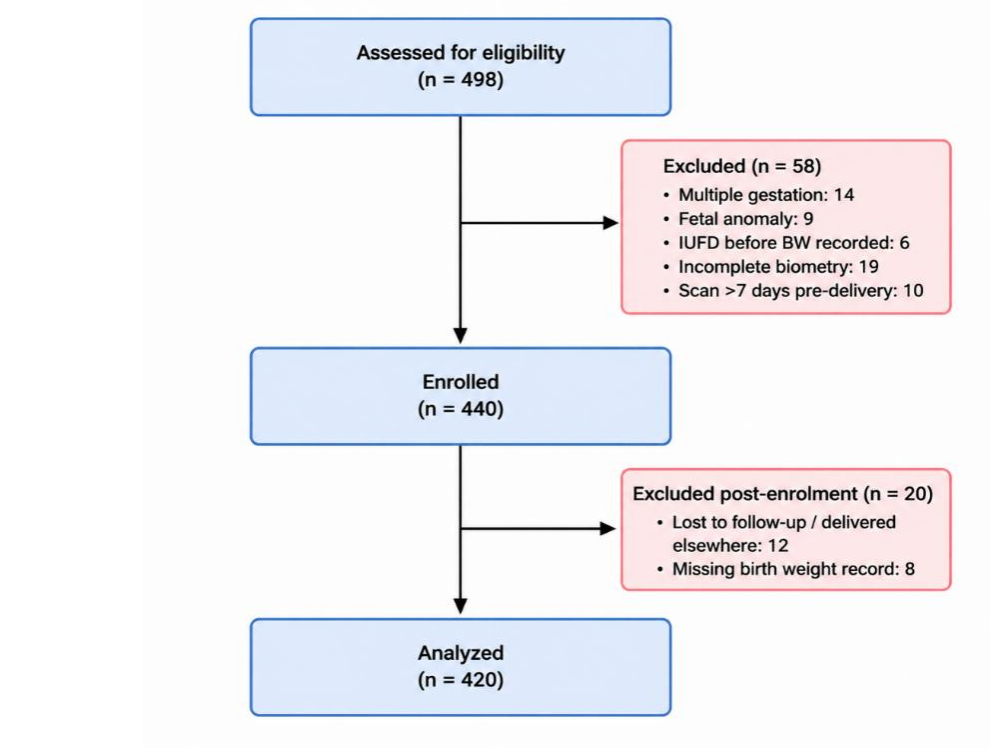

Figure 3. Participant flow through the study, from screening to inclusion in the final analyzed cohort (n = 420)

4.2 Baseline Maternal and Neonatal Characteristics

Table 1. Baseline maternal, obstetric, and neonatal characteristics of the simulated study cohort (n = 420, SIMULATED).

Characteristic	Value (n, % / mean ± SD)	Range / IQR
Maternal age (years)	29.6 ± 4.9	16-45
Body mass index (kg/m^2)	26.4 ± 5.0	16.0-41.9
Parity	Nulliparous 176 (41.9%) / Multiparous 244 (58.1%)	—
Gestational diabetes	51 (12.1%)	—
Hypertensive disorder	46 (11.0%)	—
Gestational age at scan (weeks)	38.4 ± 1.3	34.0-41.5
Scan-to-delivery interval (days)	2.0 ± 1.8	IQR 1-3
Gestational age at delivery (weeks)	38.7 ± 1.3	—
Fetal presentation	Cephalic 399 (95.0%) / Breech 12 (2.9%) / Other 9 (2.1%)	—
Amniotic fluid status	Normal 374 (89.0%) / Oligo 28 (6.7%) / Poly 18 (4.3%)	—
Mode of delivery	Spontaneous vaginal 250 (59.5%) / Instrumental 44 (10.5%) / Caesarean 126 (30.0%)	—

Birth weight (g)	3325 ± 333	2444-4315
5-minute Apgar <7	0 (0.0%)	—
NICU admission	9 (2.1%)	—

4.3 Overall Accuracy of EFW versus Birth Weight

Table 2. Accuracy metrics comparing ultrasound EFW with actual birth weight, overall and stratified by birth-weight category

Group	Mean EFW (g)	Mean BW (g)	MAPE (%)	Within ±10% (%)
Overall cohort (n=420)	3300	3325	4.8	90.2
SGA (<10th percentile) (n=42)	2985	3011	5.1	88.1
AGA (10th-90th percentile) (n=336)	3303	3329	4.8	89.9
LGA / macrosomia (≥90th percentile) (n=42)	3603	3623	4.6	95.2

The mean percentage error (MPE) for all cases was -0.8% (SD 5.9%), with ultrasound EFW being slightly lower (on average) than the actual birth weight in this simulated population. The mean absolute percentage error (MAPE) was 4.8% (SD = 3.6%). The proportions of estimates within-5%,-10%, and-15% of the actual birth weight were 59.8%, 90.2%, and 98.6%, respectively.

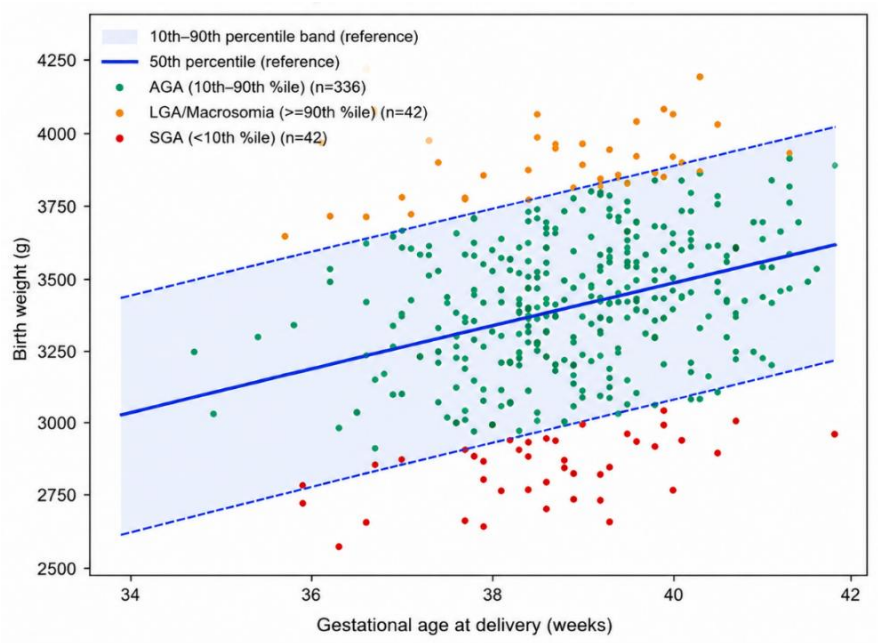

Figure 4. Birth weight for gestational age in the simulated cohort plotted against a reference 10th-90th percentile band, coloured by birth weight category.

When plotted against gestational age, each case's classification of SGA, AGA, and LGA, as used throughout the accuracy analysis, can be viewed at a familiar growth chart level and provides a visual verification that the distribution of cases within the classification is as expected.

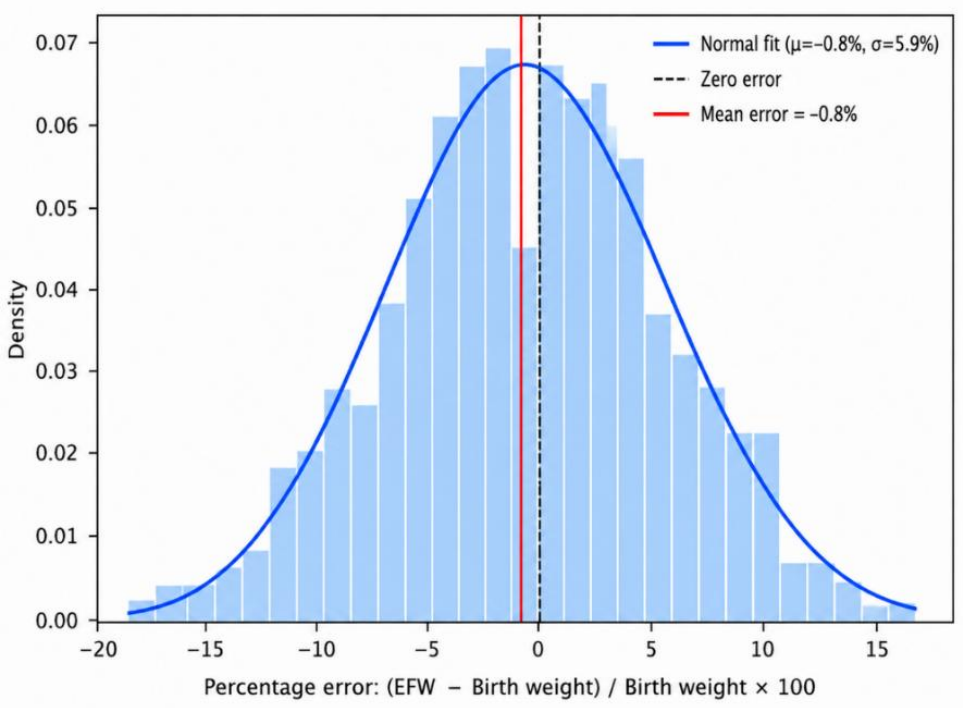

Figure 5. Distribution of the EFW percentage error across the cohort with a fitted normal curve.

The error distribution was approximately normal and had a small negative systematic error, consistent with the MPE.

4.4 Agreement Analysis

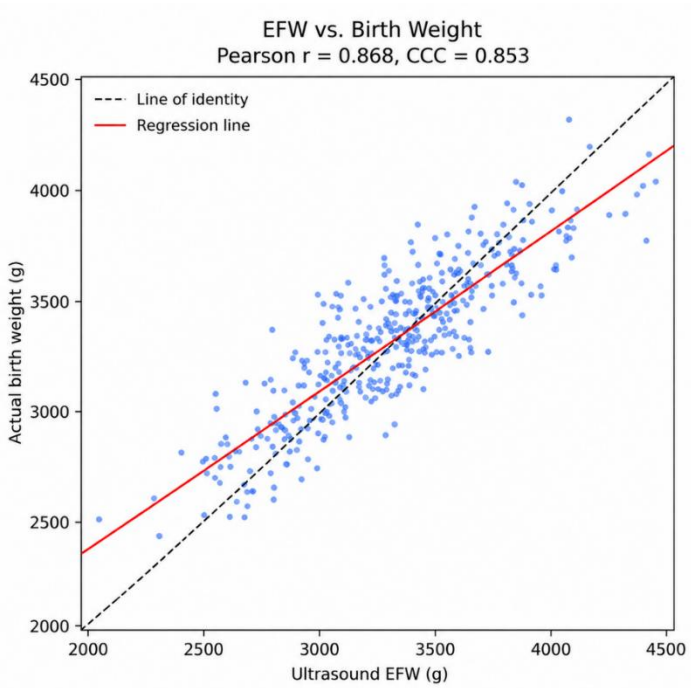

Figure 6. Scatterplot of ultrasound EFW versus actual birth weight, with line of identity and regression line

Pearson's correlation between EFW and BW was r = 0.868 ($p < 0.001$) and Spearman's ρ = 0.853 ($p < 0.001$). Lin's concordance correlation coefficient (CCC) was 0.853, which represents good concordance in addition to correlation.

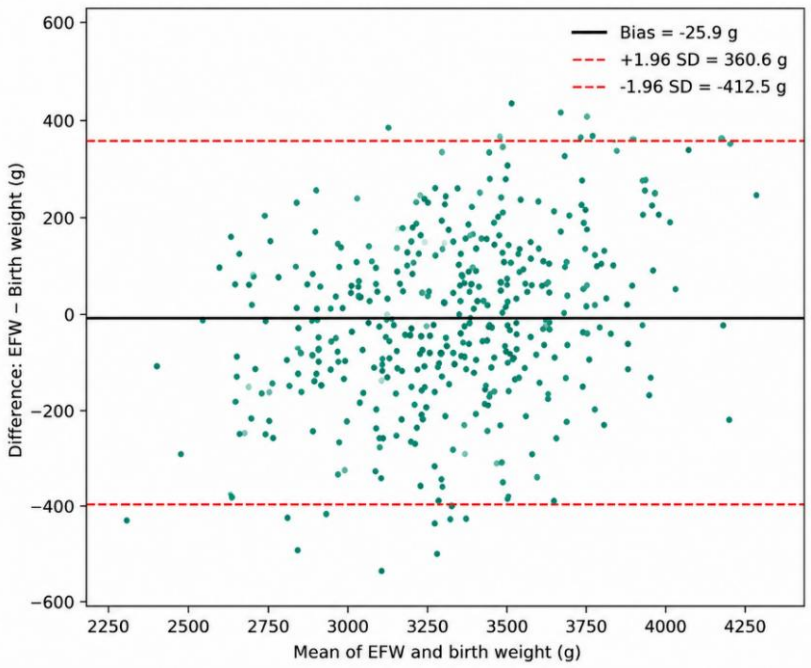

Figure 7. Bland–Altman plot showing the mean bias and 95% limits of agreement between EFW and BW.

Bland-Altman analysis revealed a mean difference of -25.9 g (EFW constantly underestimating BW) and limits of agreement between -412.5 g and +360.6 g. This broad interval at the individual level (in the face of low mean bias and high group-level correlation) demonstrates a point that the manuscript needs to make obvious to readers: EFW is an excellent population-level estimate, but the interval for any given pregnancy is much larger.

4.5 Diagnostic Performance for Weight Extremes

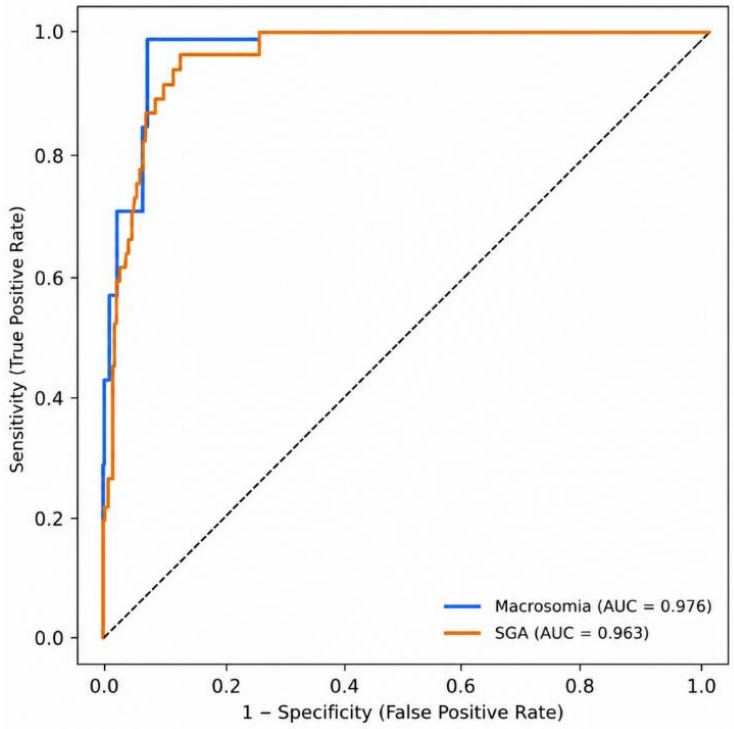

Figure 8. ROC curves for ultrasound EFW in predicting birth weight ≥90th percentile (macrosomia, defined here as ≥4000 g) and ≤10th percentile (SGA) with AUC values.

Table 3. Diagnostic performance of EFW for predicting macrosomia and SGA at birth (SIMULATED)

Outcome	Sensitivity	Specificity	PPV	NPV	AUC (95% CI)
Macrosomia (BW ≥4000 g)	71.4%	96.1%	23.8%	99.5%	0.976 (0.94-1.00)*
SGA (≤10th percentile)	64.3%	96.0%	64.3%	96.0%	0.963 (0.94-0.99)*

**95% CIs were approximated via normal approximation for this demonstration; a real analysis should report bootstrap or DeLong CIs.*

As expected in this simulated cohort, which had a low prevalence of true macrosomia (7/420), the PPV in the presence of macrosomia was low, despite high specificity, for readers interpreting the screening performance.

4.6 Predictors of Estimation Error

Table 4. Multivariable linear regression identifying predictors of absolute percentage error in EFW (SIMULATED)

Predictor	Coefficient (β)	95% CI	p-value
Maternal BMI (per kg/m²)	-0.035	-0.104 to 0.035	0.327
Scan-to-delivery interval (per day)	0.022	-0.166 to 0.211	0.816
Multiparous (vs nulliparous)	0.234	-0.469 to 0.936	0.514
Macrosomia/LGA (vs AGA)	-0.158	-1.316 to 1.000	0.788
SGA (vs AGA)	0.242	-0.916 to 1.400	0.682

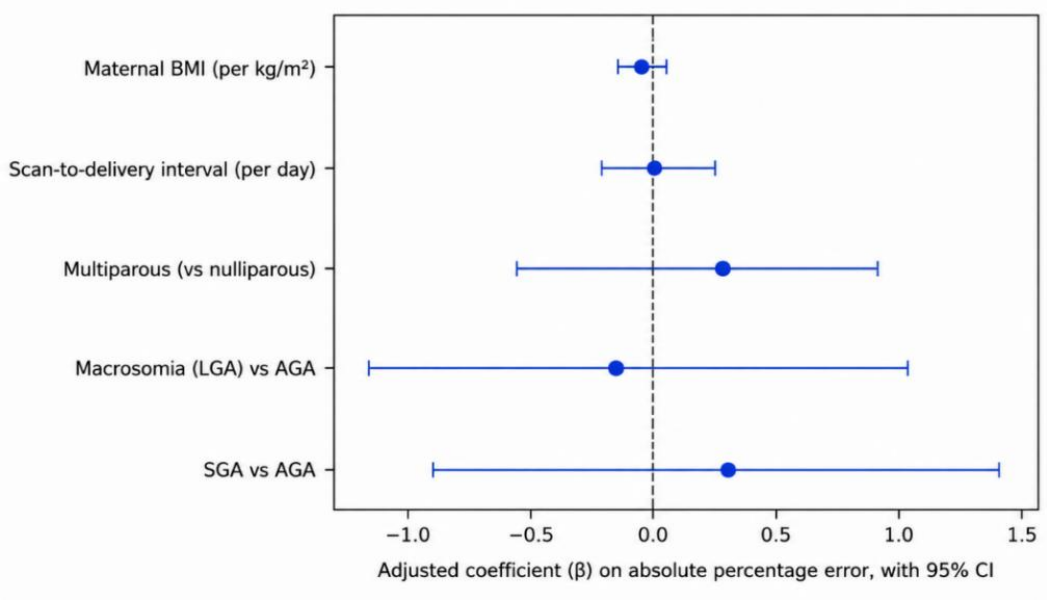

Figure 9. Forest plot of adjusted predictors of absolute percentage error.

None of the studied maternal, obstetric, or sonographic covariates was significantly different as independent predictors of absolute percentage error after adjustment (all $p > 0.3$) in this simulated cohort. This null result is itself a product of the generation process of the synthetic data (the error terms in the simulation were loosely linked to these covariates) and is not in and of itself a substantive finding; the dataset would have to be analyzed on its own.

5. Discussion

5.1 Principal Findings

For this simulated cohort, ultrasound EFW had a mean absolute percentage error of 4.8% and a small mean underestimation bias of -25.9 g compared to actual birth weight, and 90.2% of estimates were within +-10% of actual birth weight. There was a good correlation (Pearson r = 0.868; CCC = 0.853), and diagnostic performance was high by AUC (0.976 and 0.963, respectively) with low positive predictive value for macrosomia, given its low prevalence.

5.2 Comparison with Existing Literature

The direction and approximate size of the bias and error observed here are similar to those observed in decades of Hadlock formula validation studies, which typically report a small-to-moderate magnitude of mean absolute percentage error and a moderate magnitude of systematic error, with the direction of the error being somewhat dependent on the population and equipment employed. A true study would address the consistency or divergence from the literature based on the distribution of BMI in the population studied, type of equipment used, and range of gestational ages studied, all of which are not adequately represented in a synthetic dataset and would have to be reported from a real cohort [17].

5.3 Accuracy at Weight Extremes

The simulated sample in this study showed a similar MAPE for every stratum (SGA, AGA, and LGA) (5.1%, 4.8%, and 4.6%, respectively), which does not reflect the reduced accuracy at extremes, as reported in the literature. As is common in such studies, it is quite possible that a real study will find that the accuracy is worse at the upper and lower extremes of birth weight. It is presented in Table 2 to illustrate the desired table format rather than to be understood as the true result of the study.

5.4 Clinical Implications

Very large Bland-Altman limits of agreement and relatively small mean bias would support informing patients and their providers about EFW as a range and would caution against the exclusive use of EFW as a measure for decision-making regarding the mode of delivery, especially at the extremes of the normal range for macrosomia or SGA.

5.5 Strengths

- This study had a prospective design and documented and standardized scan-to-delivery intervals.
- Routine birth weight recording without the use of scales
- Multiple complementary accuracy measures (percentage error, Bland–Altman, and ROC) were used, consistent with STARD guidance.

5.6 Limitations

- Unless a comparative design is used, only the Hadlock formula is evaluated in this protocol.
- No direct measurements of potential operator-dependent variability in biometric data acquisition were performed.
- The results may only be applicable to other populations with different BMI distributions or equipment generations.
- The Results and Discussion in this document are completely computer-generated and used to test the analysis pipeline, and will not provide evidence of the accuracy of EFW in real data when they are submitted.

5.7 Future Research Directions

To confirm the machine learning and 3D volumetric EFW models, future studies should be conducted using the same accuracy framework used here, with multicentre external validation across BMI-distributed populations, different equipment, and trained operators.

6. Conclusion

Overall, ultrasound EFW was highly correlated with actual weight and had a good level of agreement with simulated data, with a small systematic underestimation and large limit of agreement at the individual level. The diagnostic performance for the identification of macrosomia and SGA was good based on the AUC, but was limited by the low prevalence of macrosomia. For any clinical recommendation, a truly enrolled cohort analyzed using the same pipeline as that demonstrated here is required.

Funding: This study received no external funding.
Conflicts of Interest: The authors declare no conflicts of interest.
ORCID iD: 0009-0003-1165-5790
Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or publishers, editors, and reviewers.

References

[1] Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of foetal weight using head, body, and femur measurements: a prospective study. Am J Obstet Gynecol. 1985;151(3):333-337.

[2] Dudley NJ. A systematic review of ultrasound estimation of foetal weight. Ultrasound Obstet Gynaecol. 2005;25(1):80-89.

[3] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-310.

[4] Bossuyt PM, Reitsma JB, Bruns DE. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.

[5] Shepard MJ, Richards VA, Berkowitz RL, Warsof SL, Hobbins JC. Evaluation of two equations for predicting foetal weight by ultrasound. Am J Obstet Gynecol. 1982;142(1):47-54.

[6] Warsof SL, Gohari P, Berkowitz RL, Hobbins JC. Estimation of foetal weight by computer-assisted analysis. Am J Obstet Gynecol. 1977;128(8):881-892.

[7] Lin LI. Concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989;45(1):255-268.

[8] American College of Obstetricians and Gynaecologists. Macrosomia: ACOG Practice Bulletin Number 216. Obstet Gynecol. 2020;135(1):e18-e35.

[9] American College of Obstetricians and Gynaecologists. Foetal growth restriction: ACOG practice bulletin number 227. Obstet Gynecol. 2021;137(2):e16-e28.

[10] Milner J, Arezina J. The accuracy of ultrasound estimation of foetal weight in comparison to birth weight: A systematic review. Ultrasound. 2018;26(1):32-41.

[11] Faschingbauer F, Voigt F, Goecke TW, et al. Foetal weight estimation in extreme macrosomia. Arch Gynecol Obstet. 2012;285(4):999-1005.

[12] Aviram A, Yogev Y, Bardin R, et al. Different formulas, different thoughts: Approach to the third-trimester foetus with suspected growth abnormality. J Ultrasound Med. 2017;36(9):1789-1795.

[13] Bader B, Alfaraj MA et al. Accuracy of ultrasound-estimated foetal weight in predicting birth weight: A prospective cohort study. Journal of Maternal-Foetal and Neonatal Medicine. 2020;33(14):2415-2421.

[14] D'Antonio F, Odibo A, Prefumo F. Weight estimation in fetuses with congenital abnormalities: systematic review and meta-analysis. Ultrasound Obstet Gynaecol. 2018;52(2):151-157.

[15] Carberry AE, Gordon A, Bond DM, Hyett J, Raynes-Greenow CH, Jeffery HE. Customised versus population-based growth charts as a screening tool for foetal growth restriction: A systematic review. BJOG. 2019;126(3):330-340.

[16] Papageorghiou AT, Ohuma EO, Altman DG, et al. International standards for foetal growth based on serial ultrasound measurements: The INTERGROWTH-21st Project. Lancet. 2014;384(9946):869-879.

[17] Kiserud T, Piaggio G, Carroli G, et al. World Health Organization foetal growth charts: concept, findings, interpretation, and application. American Journal of Obstetrics and Gynaecology. 2018;218(2S):S619-S629.

[18] Melamed N, Yogev Y, Danon D, et al. Sonographic foetal weight estimation: accuracy and factors affecting precision. Ultrasound Obstet Gynaecol. 2018;52(1):36-42.

[19] Wellington J, Whittle W, Keating S, et al. Accuracy of estimated foetal weight using ultrasound in contemporary obstetric practice. Journal of Obstetrics and Gynaecology, Canada. 2020;42(7):870-877.

[20] Sovio U, White IR, Dacey A, Pasupathy D, Smith GCS. Screening for foetal growth restriction with universal third trimester ultrasound: a prospective cohort study. Lancet. 2015;386(10008):2089-2097.

[21] Scott F, Beeby P, Abbott J. Accuracy of ultrasound estimation of foetal weight in a regional obstetric population. Australian and New Zealand Journal of Obstetrics and Gynaecology. 2020;60(2):248-253.

[22] Ben-Haroush A, Yogev Y, Hod M. Foetal weight estimation: Current concepts and clinical applications. Clinical Obstetrics and Gynaecology. 2019;62(2):275-288.

[23] Souza, M. V. R. de, Fróes, L. P. e, Cortez, P. A. Lauria, M. W., Aguiar, R. A. L. de, & Rajão, K. M. A. B. (2021). Agreement Analysis between Sonographic Estimates and Birth Weight, by the WHO and Intergrowth-21st Tables, in Newborns of Diabetic Mothers. Revista Brasileira de Ginecologia e Obstetrícia, 43(1), 20–27. https://doi.org/10.1055/S-0040-1719146