Systematic Review

The Accuracy of Third-Trimester Ultrasound in Predicting Large for Gestational Age or Macrosomic Fetuses in Diabetic and Non-Diabetic Pregnant Women: A Systematic Review and Meta-Analysis

Sofia Bussolaro[1], Vincenza Cofini[2], Stefano Necozione[2], Maurizio Guido[2,3], Roberto Rulli[1], Ilaria Fantasia[3],*

[1]Obstetrics & Gynaecology Unit, San Bassiano Hospital, 36061 Bassano Del Grappa, Italy
[2]Department of Life, Health and Environmental Sciences, University of L'Aquila, 67100 L'Aquila, Italy
[3]Obstetrics & Gynaecology Unit, San Salvatore Hospital, 67100 L'Aquila, Italy
*Correspondence: ilariafantasia@gmail.com (Ilaria Fantasia)
Academic Editor: Felix Wong
Submitted: 8 March 2023   Revised: 20 March 2023   Accepted: 20 March 2023   Published: 13 July 2023

Abstract

Background: The accuracy of third-trimester ultrasound in detecting large for gestational age and macrosomic fetuses in diabetic and non-diabetic pregnant women is unclear in the literature. The aim of the study is to examine the precision of the 4-parameter Hadlock formula for the prediction of large fetuses in these two populations. **Methods**: A systematic review and meta-analysis were performed, and only studies evaluating the accuracy of third-trimester ultrasound using the 4-parameter Hadlock formula were included. Data were extracted, and the meta-analysis was performed using STATA software and Meta-disk 2.0 aiming to obtain the pooled sensitivity and specificity. Quality assessment of the risk of bias was performed using the QUADS-2 tool. **Results**: Nine articles were included in the final analysis together with 24,693,702 pregnancies screened and 2336 real large fetuses. The included articles were judged to be at high risk of bias in more than half of the cases and at doubtful risk in the remaining cases. Comparison between diabetic and non-diabetic populations was impossible because the studies considered mixed pregnancies (diabetic and non-diabetic) or only healthy, so the comparison was made between the latter two groups. The pooled sensitivity was 0.54 (95% confidence interval (CI): 0.40–0.68), and the pooled specificity was 0.94 (95% CI: 0.90–0.97). The heterogeneity estimated by the Bivariate I^2 was 0.92, and the area under the summary Receiver Operating Characteristics curve was 0.19. The subgroup analysis revealed a higher level of heterogeneity for the mixed group ($I^2 = 0.92$) and a lower one for the healthy group ($I^2 = 0.67$). The relative sensitivity between the mixed population and the healthy one was 0.85 (95% CI: 0.49–1.45; $p = 0.57$), and the relative specificity between the mixed population and the healthy one was 0.98 (95% CI: 0.91–1.04; $p = 0.54$), the difference between healthy and mixed groups was not significant ($p = 0.11$). **Conclusions**: Despite the high heterogeneity of the data, the overall accuracy of ultrasound is similar in mixed and healthy populations and is overall moderate in predicting large fetuses.

Keywords: large for gestational age; LGA; macrosomia; gestational diabetes; estimated fetal weight

1. Introduction

A large for gestational age (LGA) fetus is defined by the presence of a prenatal abdominal circumference (AC) and/or estimated fetal weight (EFW) $\geq 90°$ percentile [1]. As a result of the increased incidence of obesity in mothers and thus also diabetes [2], the risk for the fetus to be LGA, or being born macrosomic, that is defined by a neonatal weight ≥ 4000 grams, is considerable. Because of the possible perinatal and maternal complications associated with the presence of a large fetus, such as shoulder dystocia and third- and fourth-degree perineal lacerations, the prenatal identification of an LGA fetus may reduce these risks. However, the assessment of the estimated fetal weight by ultrasound has shown a poor prediction rate for LGA and macrosomia, and the likelihood of error is greater, the greater the estimated fetal weight and gestational age [3,4]. Formulas used for calculating the EFW tend to underestimate or overestimate fetal size by a range of 10–15%, making the prenatal estimation of birth-related risks ineffective or inappropriate [5–7]. This effect is secondary to different variable such as the error related to every measured parameter and the large intra- and interobserver variability. Furthermore, it appears that most formulas are mostly accurate for weights up to 3500 grams, albeit tending, in the opinion of some authors [6,8,9], to underestimate large fetuses. For other authors, on the other hand, the overestimation of weight would seem to be all the greater the higher the EFW [10]. Among all available formulas to calculate fetal weight, Hadlock's 4-parameter formula (including biparietal diameter (BPD), head circumference (HC), AC, and femur length (FL)), is the most widely used, and it seems to provide the best predictions of birth weights over 3500 grams [8].

Pregnancies complicated by gestational diabetes mellitus (GDM) have a 2 to 4 times higher risk of having LGA fetuses than non-diabetic women [11], along with a higher

risk of perinatal morbidities related to neonatal macrosomia [12]. It is debated whether the presence of maternal diabetes reduces further the accuracy of ultrasound in estimating fetal weight [13]. In fact, it has been shown that the percentage difference in EFW may be as low as 0.2% in non-diabetic women when fetal biometry is performed within one week before delivery, while it rises to 7.9% in diabetic women [13]. However, the accuracy of ultrasonography in these two groups is not well described in the literature. Therefore, the aim of this review is to define the accuracy of the 4-parameter Hadlock formula in predicting the EFW of LGA fetuses in diabetic and non-diabetic pregnant women.

2. Materials and Methods

A meta-analysis on the accuracy of third-trimester ultrasound in estimating the actual birth weight of suspected LGA and macrosomic fetuses was conducted. The study was registered with the International Prospective Register of Systematic Reviews (PROSPERO) database (CRD42023407146) [14]. The Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines were followed in reporting the results [15].

2.1 Search Strategy

An English literature search was performed from inception until July 2022 in PubMed (Medline). For the purpose of the search, a combination of key terms was used, which, together with the search strategy, is given in the Appendix. Original articles and studies reporting the accuracy of the EFW Hadlock 4 formula [16] in detecting LGA and macrosomic fetuses were considered for inclusion, while literature reviews, meta-analyses, and case reports were not considered eligible. The diabetic patients could have had pregestational type I or II diabetes or GDM. According to the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) and the American College of Obstetricians and Gynecologists (ACOG) [1,5], an LGA fetus was defined by an EFW or AC above the 90th percentile according to gestational age, while a macrosomic fetus was considered a fetus with an EFW above 4000 grams.

Data extracted or derived from the available data of each study included the type of population undergoing ultrasound and the total number of patients scanned, the sensitivity, the specificity, and the total number of true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN) results. A meta-analysis was conducted to present sensitivity and specificity estimates along with 95% confidence intervals (CIs).

2.2 Quality Assessment of Included Studies

The quality assessment of each included study was performed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) criteria [17]. After applying the two separate quality criteria, the Robvis tool web app (Version of 2022, University of Bristol, Bristol, United Kingdom) [18] was then used to visualize the risk-of-bias assessments creating traffic-light plots and weighted bar plots.

2.3 Statistical Analysis

Derived data on TP, FP, TN, and FN were obtained by knowing the number of patients studied and the sensitivity and specificity values through a 2×2 table. The meta-analysis (hierarchical and bivariate models) was performed using the Metandi and Metadata commands on STATA software (Stata 17, StataCorp LLC., College Station, TX, USA) [19–21] and Meta-disk 2.0 (Ramón y Cajal Research Institute, Madrid, Spain) [22], calculating: the pooled accuracy estimation (sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (LR+), negative likelihood ratio (LR–)), and false positive rate (FPR) with their corresponding 95% CIs, the model parameter estimates (logit Sensitivity, logit Specificity, logits variances, and correlation), and the heterogeneity statistics including bivariate I-squared, the median odds ratio (MOR), and the area of 95% prediction ellipse.

For the aim of the study, a subgroup analysis between healthy and mixed populations was performed. A comparative analysis was run using a random effects model with one categorical covariate from Meta-disk 2.0. Summary receiver operating characteristic (ROC) curve and forest plots were also reported.
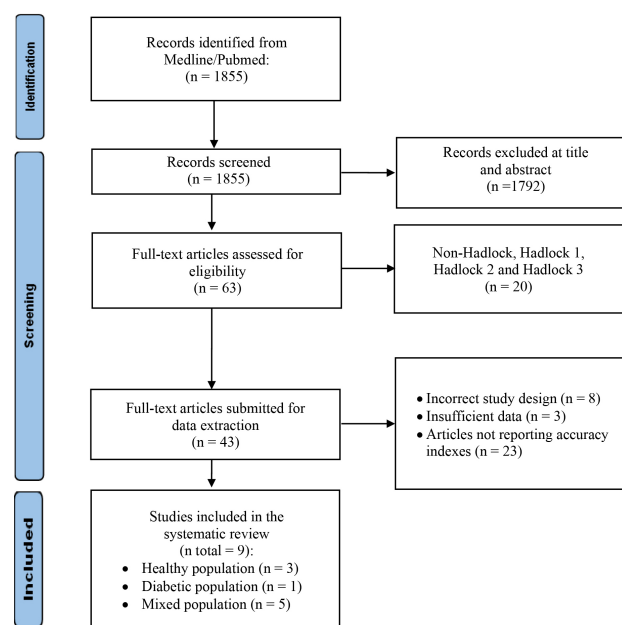

Fig. 1. The selection process of included articles.

Table 1. Main characteristics of the included studies.

Author	Year	Fetal type	N of pregnancy screened	Diabetic women	Non-diabetic women	Timing of US	Sensitivity	Specificity	PPV	NPV	Overall accuracy	LR+	AUC
Melamed [7]	2011	M	4765	NR	NR	within 3 days of delivery	64.6%	94%	53.6%	96.1%	91.1%	11.12	0.92
Scifres [26]	2015	LGA	1374	NR	NR	within 31 days to delivery	75.7%	76.8%	22.6%	97.3%	NR	NR	NR
Aviram [3]	2017	LGA	7996	339	1279	within 7 days to delivery	77.1%	89.5%	67.2%	93.3%	86.9%	6.34	0.95
Shen [27]	2017	LGA	374	NR	NR	within 14 days to delivery	48.1%	97.3%	76.5%	91.1%	NR	NR	NR
Verger [24]	2020	LGA	253	39	214	within 27 days to delivery	66%	82.5%	50%	90%	79%	3.77	NR
Weiss [28]	2018	M	3304	515	2789	within 10 days to delivery	23.4%	96%	64.1%	80.2%	78.8%	5.78	NR
Duncan [29]	2021	LGA	1054	47	76	30–34 weeks	30.1%	97.5%	63.8%	91.4%	NR	12.0	0.64
Bardin [25]	2022	M	5424	NR	NR	within 3 days to delivery	68.1%	93.5%	58.1%	95.7%	90.5%	NR	NR
Roeckner [23]	2022	LGA	630	58	572	26–36 weeks	31.8%	98%	71.1%	90.1%	NR	22.73	0.68

Abbreviations: N, number; US, ultrasound; M, macrosomia; LGA, large for gestational age; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; AUC, area under the Receiver Operating Characteristic (ROC) curve; NR, non-reported.

Risk of bias domains

Study	D1	D2	D3	D4	Overall
Melamed, 2011	–	+	+	+	–
Scifres, 2015	X	X	+	+	X
Aviram, 2017	–	+	+	+	–
Shen, 2017	+	–	+	+	–
Verger, 2020	–	X	+	+	X
Weiss, 2018	–	–	+	+	–
Duncan, 2021	X	X	+	+	X
Bardin, 2022	X	+	+	+	X
Roeckner, 2022	–	X	+	+	X

Domains:
D1: Patient selection.
D2: Index test.
D3: Reference standard.
D4: Flow & timing.

Judgement
X High
– Some concerns
+ Low

Fig. 2. The traffic light plot QUADAS-2 quality evaluation of all included articles.

3. Results

3.1 Search Results

A total of 1855 studies were identified through the search of the literature. 1792 titles and abstracts were screened, resulting in 63 proceeding to the full-text screen. Of these, 20 articles were excluded because of the formula used (non-Hadlock, Hadlock 1, Hadlock 2, or Hadlock 3 formula). Further, eight publications were excluded because of the incorrect study design or outcome, and the other three manuscripts because of insufficient data reported. Finally, 23 articles were excluded because they did not directly report TP, FP, TN, and FN data or could not be derived due to missing data. Thus, nine articles were included in the present meta-analysis [3,7,23–29], of which 3 represented a healthy population [7,25,27], one a diabetic population [26], and 5 a mixed population (healthy and diabetic) [3,23,24,28,29].

The selection process of included articles is presented in Fig. 1, while PRISMA checklist is given in the **Supplementary Materials**.

3.2 Risk of Bias of Included Studies

The risk of bias of included studies was represented with the traffic light plots and weighted bar plots according to the QUADAS-2 criteria (Fig. 2, Ref. [3,7,23–29]; Fig. 3). The rating obtained on "overall" risk of bias was high risk in more than half of the publications included in the review.

In the remaining cases, the overall risk of bias was rated as doubtful. The fields with the highest risk of bias were "index test" because of the interval between ultrasound scans and delivery. Conversely, reference standard bias and flow and timing bias were found to be low risk for all publications.

3.3 Description of Included Studies

For the purpose of the present study, the studies were divided into two major groups: group 1, defined as the "healthy" non-diabetic population; and group 2, defined as the "mixed" population including a population of healthy and diabetic patients [3,23,24,26,28,29].

The total number of patients screened was 24,693,702, comprising 20,770,702 women in group 1 and 3923 women in group 2. The total number of real large fetuses was 2336. Table 1 (Ref. [3,7,23–29]) shows the main characteristics of the included studies.

3.4 Meta-Analysis

Overall, the pooled sensitivity was 0.54 (95% CI: 0.40–0.68), while specificity was 0.94 (95% CI: 0.90–0.96), as reported in Fig. 4, Ref. [3,7,23–29] and Fig. 5. The LR+ was 8.9 (95% CI: 6.2–12.9), and the LR– was 0.49 (95% CI: 0.36–0.65). The between-study heterogeneity statistics estimated by the Bivariate I^2 was 0.92, and the correlation between sensitivities and specificities was negative with a Spearman's rank correlation coefficient, rho = –0.83, evi-
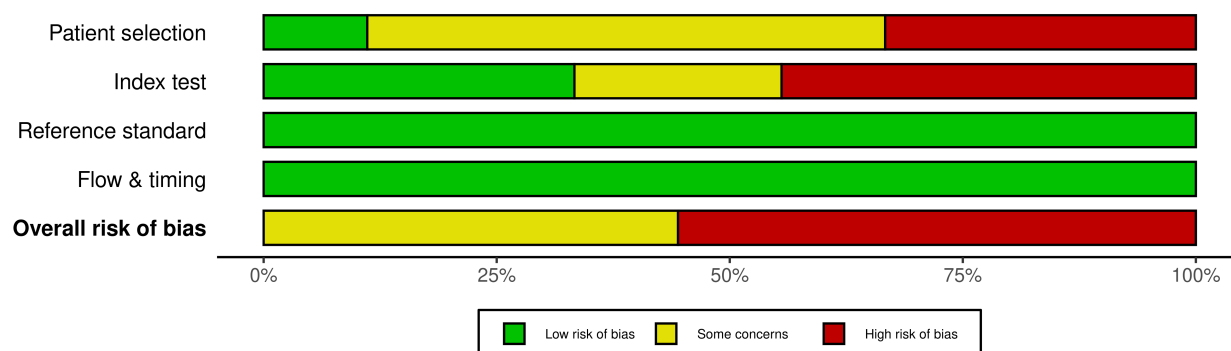
Fig. 3. The weighted bar plot QUADAS-2 quality evaluation of all included articles.
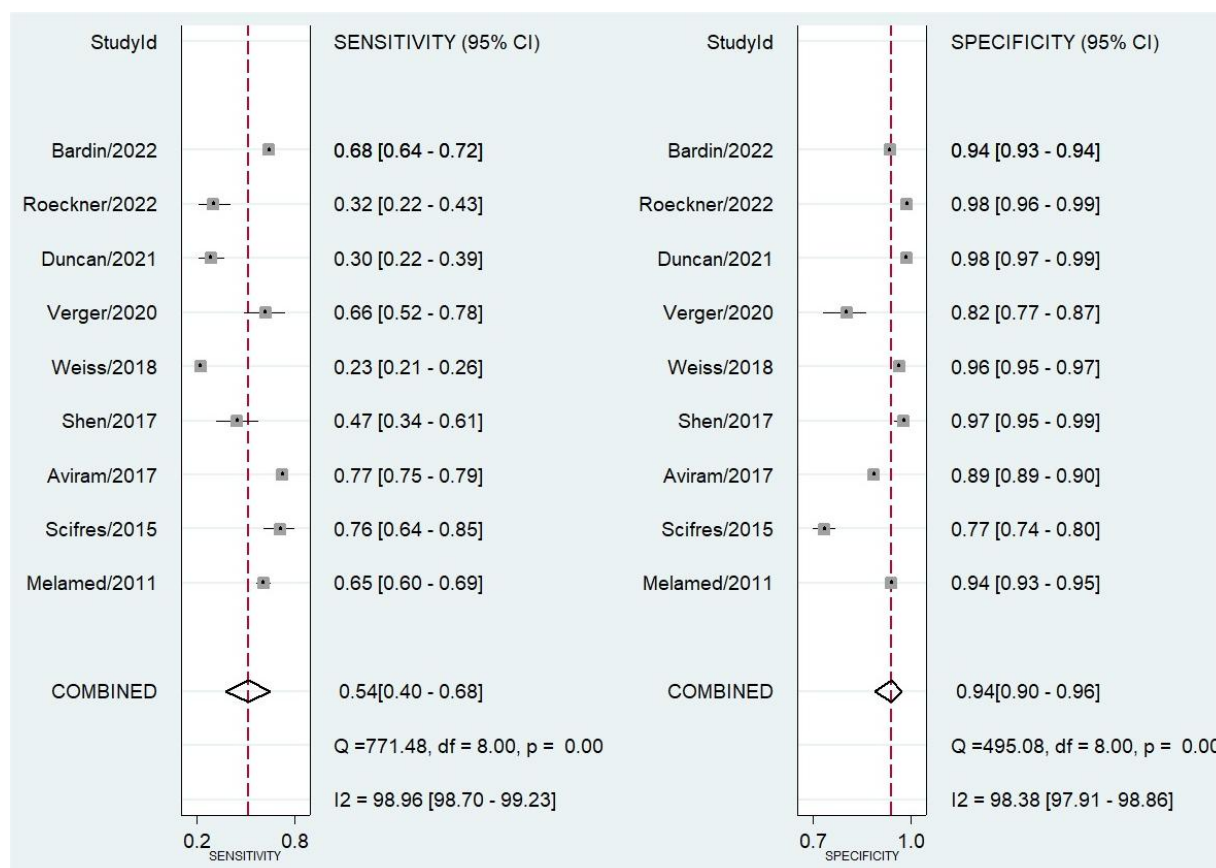

Fig. 4. Forest plot analysis of the overall sensitivity and specificity of ultrasound in predicting large for gestational age fetuses.

dencing high heterogeneity and the threshold effect respectively, as reported in the literature [19,30]. The area of the 95% prediction ellipse in the ROC plane was 0.19, and MOR was about 2, both for sensitivity and specificity [22].

3.5 Subgroup Analysis and Meta-Regression Analysis

The subgroup analysis shows a higher level of heterogeneity for the mixed group ($I^2 = 0.92$) and a lower one for the healthy group ($I^2 = 0.67$). The correlation between the two groups compared was negative (–1.00 and –0.89, respectively). The sensitivity for the healthy population was 0.60 (95% CI: 0.35–0.80), while specificity was assessed

at 0.95 (95% CI: 0.88–0.98). As regards the mixed population, sensitivity was 0.51 (95% CI: 0.33–0.68), while the specificity was 0.93 (95% CI: 0.87–0.96). The ROC curve in the two populations is shown in Fig. 6.

The meta-regression analysis for the subgroup (healthy and mixed populations) assessed that the sensitivity and specificity parameters did not differ between the mixed population and the healthy one ($p = 0.11$). In fact, the relative sensitivity between the mixed population and the healthy one was 0.85 (95% CI: 0.49–1.45; $p = 0.57$), and the relative specificity between the mixed population and the healthy one was 0.98 (95% CI: 0.91–1.04; $p = 0.54$).
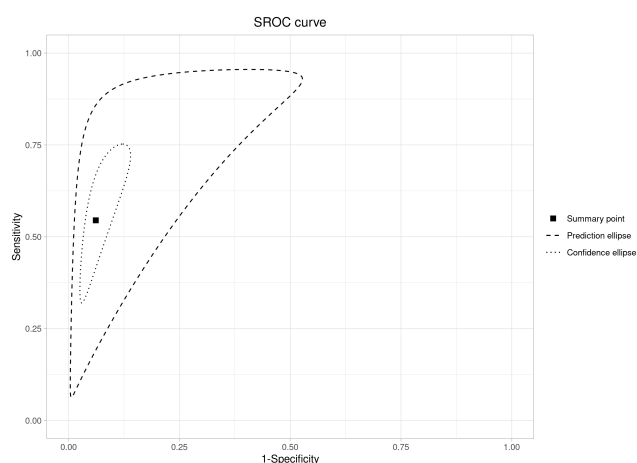
Fig. 5. Summary ROC curve analysis for sensitivity and specificity of the Hadlock-4 in predicting large for gestational age fetuses.
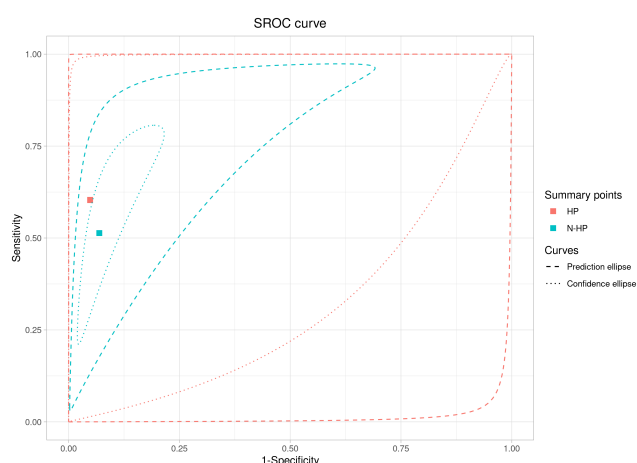

Fig. 6. ROC curve analysis for the ultrasound prediction of large for gestational age fetuses in the healthy and mixed populations.

4. Discussion

The results of this study indicate that the overall accuracy of ultrasound in estimating fetal birth weight is in the range of 54% with a specificity of 94%. The data have not changed much if we consider the sub-analysis of the two population groups. In fact, the sensitivity for the non-diabetic population is 60%, while for the mixed population of diabetic and healthy is 51%. The main problem, however, is the high statistical heterogeneity found among the studies, together with the statistical methodology used, which does not allow to reach conclusive studies on the subject.

Several aspects make approaching this topic complex: first, the choice of growth curves. In planning the study, we had to deal with the question of which growth curve to consider in the systematic review, and we chose to narrow our assessment to that of Hadlock-4 only. In fact, in a

study that evaluated the performance of 36 growth curves on a total population of 350 newborns weighing more than 4000 grams, the Hadlock-4 formula identified 74% of fetuses weighing $\geq$4000 grams with a systematic error not significantly different from zero [31]. However, a false positive rate of 31% was reported. This leads to the second aspect to consider when addressing the issue of weight estimation in large fetuses, which is the decrease in ultrasound accuracy observed for an examination performed in the last weeks of pregnancy and for an EFW $\geq$4000 g. In 2017, the World Health Organization (WHO) published reference ranges for fetal growth charts based on the prospective assessment of 1387 women from 10 different countries [32]. They observed that the growth curve tends to widen towards the end of pregnancy, indicating greater variability in the estimation of fetal weight. Not only but also, this variability seemed to be greater for higher percentiles. In other words, while a small fetus tends to be "more equally small", in the large fetus there is greater variability that makes it difficult to use a standardized cut-off and to make recommendations on it. The reasons why this variability is greater for larger fetuses, especially near the term of pregnancy, have not been elucidated. Several factors have been implicated, such as the technical difficulties in measuring a large fetus at term gestation, with the consequent difficulty of being able to obtain proper imaging plane for the measurement, or the presence of maternal obesity and diabetes, which could reduce the quality of the images in the former and determine different fetal body composition in the latter [33,34]. These aspects lead to the third aspect to consider when estimating the weight of large fetuses: is ultrasound accuracy different in large fetuses of non-diabetic mothers compared to diabetic and/or obese mothers?

The results of our review show that the accuracy of ultrasound in predicting the birth weight of a large fetus in a population of healthy mothers is nearly superimposable to that of diabetic mothers. This result is in line with the previous study that did not find an association between maternal diabetes and poorer accuracy of ultrasound while showing a negative correlation between obesity and performance of the test, although this association is not strong [33,34]. Studies that aimed to improve the accuracy of ultrasound for the diagnosis of LGA or macrosomia by introducing maternal features have failed. Body mass index, fetal sex, and multiparity have no significant influence on measurement error [6] or accuracy. Adding clinical and demographic variables to the ultrasound assessment, including maternal weight and body mass index, does not improve the prediction of macrosomia [35]. In fact, if the ultrasound is performed by an experienced sonographer the impact of maternal body mass index is scarce [36]. This concept is reasonable if we consider that ultrasound is an operator-dependent examination and, therefore, it could indicate that the estimation of fetal weight should be performed by an experienced operator, if LGA is suspected. Although this

is, to our knowledge, the first study that attempts to evaluate the difference in the detection rate of the ultrasound estimation of fetal weight in the diabetic and non-diabetic populations, there are some limitations.

The first is the small number of included studies characterized by high statistical variability. In many studies, the population included a mixed population of diabetic and non-diabetic women, for which it was necessary to merge the data, and this could have had an impact on the results. Moreover, the majority of studies reported on both LGA and macrosomic fetuses, which frequently overlap but maybe two different entities. In addition, the timing of ultrasound performance varies from 7 to more than 30 days contributing to the variability in terms of detection rate.

5. Conclusions

The estimation of the fetal weight in diabetic women is of paramount importance as it may help in identifying the optimal time of delivery of LGA fetuses in an attempt to prevent possible complications related to the birth of a macrosomic fetus. This review confirms that the accuracy of ultrasound in predicting large fetuses at birth is only moderate but that its performance is similar in mixed and non-diabetic populations. However, there is a high heterogeneity between studies that impede the drawing of definitive conclusions. Further studies are needed to establish the exact accuracy of ultrasound estimation of fetal weight.

Abbreviations

LGA, large for gestational age; AC, abdominal circumference; EFW, estimated fetal weight; BPD, biparietal diameter; HC, head circumference; FL, femur length; GDM, gestational diabetes mellitus; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analysis; TP, true-positive; FP, false-positive; TN, true-negative; FN, false-negative; CI, confidence interval; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies; AUC, area under the Receiver Operating Characteristic (ROC) curve; DOR, diagnostic odds ratio; LR+, positive likelihood ratio; LR–, negative likelihood ratio; FPR, false positive rate; MOR, median odds ratio.

Availability of Data and Materials

All data generated or analyzed during this study are included in this published article.

Author Contributions

IF designed the research study. SB performed the research. VC and SN analyzed the data. SB and IF wrote the manuscript. MG and RR provided help and advice on the study design and final draft. All authors contributed to editorial changes in the manuscript. All authors have red and approved the final manuscript.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/j.ceog5007144.

Appendix

Search strategy

(((((((((((obstetric ultrasound) OR (prenatal ultrasound[Title/Abstract])) OR (OB sonography[Title/Abstract])) OR (pregnancy ultrasound[Title/Abstract])) OR (pregnancy echo[Title/Abstract])) OR (pregnant uterus ultrasonography[Title/Abstract])) OR (sonographic estimation[Title/Abstract])) OR (ultrasound estimation[Title/Abstract])) OR (echographic estimation[Title/Abstract])) OR (ultrasonographic estimation[Title/Abstract])) AND ((((((large for gestational age) OR (LGA)) OR (large for date)) OR (large for age)) OR (fetal macrosomia)) OR (macrosomic fetus)).

References

[1] Macrosomia: ACOG Practice Bulletin, Number 216. Obstetrics and Gynecology. 2020; 135: e18–e35.

[2] Lawrence JM, Contreras R, Chen W, Sacks DA. Trends in the prevalence of preexisting diabetes and gestational diabetes mellitus among a racially/ethnically diverse population of pregnant women, 1999-2005. Diabetes Care. 2008; 31: 899–904.

[3] Aviram A, Yogev Y, Ashwal E, Hiersch L, Hadar E, Gabbay-Benziv R. Prediction of large for gestational age by various sonographic fetal weight estimation formulas-which should we use? Journal of Perinatology. 2017; 37: 513–517.

[4] Pressman EK, Bienstock JL, Blakemore KJ, Martin SA, Callan NA. Prediction of birth weight by ultrasound in the third trimester. Obstetrics and Gynecology. 2000; 95: 502–506.

[5] Salomon LJ, Alfirevic Z, Da Silva Costa F, Deter RL, Figueras F, Ghi T, et al. ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound in Obstetrics & Gynecology. 2019; 53: 715–723.

[6] Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. Ultrasound in Obstetrics & Gynecology. 2005; 25: 80–89.

[7] Melamed N, Yogev Y, Meizner I, Mashiach R, Pardo J, Ben-Haroush A. Prediction of fetal macrosomia: effect of sonographic fetal weight-estimation model and threshold used. Ultrasound in Obstetrics & Gynecology. 2011; 38: 74–81.

[8] Scioscia M, Vimercati A, Ceci O, Vicino M, Selvaggi LE. Estimation of birth weight by two-dimensional ultrasonography: a critical appraisal of its accuracy. Obstetrics and Gynecology. 2008; 111: 57–65.

[9] Cesnaite G, Domza G, Ramasauskaite D, Volochovic J. The Accuracy of 22 Fetal Weight Estimation Formulas in Diabetic Pregnancies. Fetal Diagnosis and Therapy. 2020; 47: 54–59.

[10] Zafman KB, Bergh E, Fox NS. Accuracy of sonographic estimated fetal weight in suspected macrosomia: the likelihood of overestimating and underestimating the true birthweight. The Journal of Maternal-fetal & Neonatal Medicine. 2020; 33: 967–972.

[11] Cedergren MI. Maternal morbid obesity and the risk of adverse pregnancy outcome. Obstetrics and Gynecology. 2004; 103: 219–224.

[12] Lorusso L, Kato DMP, Dalla Costa NRA, Araujo Júnior E, Bruns RF. Performance of local reference curve on the diagnosis of large for gestational age fetuses in diabetic pregnant women. The Journal of Maternal-fetal & Neonatal Medicine. 2022; 35: 1899–1906.

[13] Wong SF, Chan FY, Cincotta RB, Oats JJ, McIntyre HD. Sonographic estimation of fetal weight in macrosomic fetuses: diabetic versus non-diabetic pregnancies. The Australian & New Zealand Journal of Obstetrics & Gynaecology. 2001; 41: 429–432.

[14] YORK CRD, AC.uk: National Institute for Health Research (NIHR). PROSPERO, International prospective register of systematic reviews. Available at: https://www.crd.york.ac.uk/prospero (Accessed: 1 March 2023).

[15] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. British Medical Journal. 2021; 372: n160.

[16] Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements–a prospective study. American Journal of Obstetrics and Gynecology. 1985; 151: 333–337.

[17] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine. 2011; 155: 529–536.

[18] McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Research Synthesis Methods. 2021; 12: 55–61.

[19] Kim KW, Lee J, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part I. General Guidance and Tips. Korean Journal of Radiology. 2015; 16: 1175–1187.

[20] Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. Korean Journal of Radiology. 2015; 16: 1188–1196.

[21] Nyaga VN, Arbyn M. Metadta: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data - a tutorial. Archives of Public Health. 2022; 80: 95.

[22] Plana MN, Pérez T, Zamora J. New measures improved the reporting of heterogeneity in diagnostic test accuracy reviews: a metaepidemiological study. Journal of Clinical Epidemiology. 2021; 131: 101–112.

[23] Roeckner JT, Odibo L, Odibo AO. The value of fetal growth biometry velocities to predict large for gestational age (LGA) infants. The Journal of Maternal-fetal & Neonatal Medicine. 2022; 35: 2099–2104.

[24] Verger C, Moraitis AA, Barnfield L, Sovio U, Bamfo JEAK. Performance of different fetal growth charts in prediction of large-for-gestational age and associated neonatal morbidity in multi-ethnic obese population. Ultrasound in Obstetrics & Gynecology. 2020; 56: 73–77.

[25] Bardin R, Aviram A, Hiersch L, Hadar E, Gabbay-Benziv R. False diagnosis of small for gestational age and macrosomia - clinical and sonographic predictors. The Journal of Maternal-Fetal & Neonatal Medicine. 2022; 35: 1539–1545.

[26] Scifres CM, Feghali M, Dumont T, Althouse AD, Speer P, Caritis SN, et al. Large-for-Gestational-Age Ultrasound Diagnosis and Risk for Cesarean Delivery in Women With Gestational Diabetes Mellitus. Obstetrics and Gynecology. 2015; 126: 978–986.

[27] Shen Y, Zhao W, Lin J, Liu F. Accuracy of sonographic fetal weight estimation prior to delivery in a Chinese han population. Journal of Clinical Ultrasound. 2017; 45: 465–471.

[28] Weiss C, Oppelt P, Mayer RB. Disadvantages of a weight estimation formula for macrosomic fetuses: the Hart formula from a clinical perspective. Archives of Gynecology and Obstetrics. 2018; 298: 1101–1106.

[29] Duncan JR, Odibo L, Hoover EA, Odibo AO. Prediction of Large-for-Gestational-Age Neonates by Different Growth Standards. Journal of Ultrasound in Medicine. 2021; 40: 963–970.

[30] Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Medical Research Methodology. 2006; 6: 31.

[31] Hoopmann M, Abele H, Wagner N, Wallwiener D, Kagan KO. Performance of 36 different weight estimation formulae in fetuses with macrosomia. Fetal Diagnosis and Therapy. 2010; 27: 204–213.

[32] Kiserud T, Benachi A, Hecher K, Perez RG, Carvalho J, Piaggio G, et al. The World Health Organization fetal growth charts: concept, findings, interpretation, and application. American Journal of Obstetrics and Gynecology. 2018; 218: S619–S629.

[33] Dude AM, Yee LM. Identifying Fetal Growth Disorders Using Ultrasonography in Women with Diabetes. Journal of Ultrasound in Medicine. 2018; 37: 1103–1108.

[34] Dittkrist L, Vetterlein J, Henrich W, Ramsauer B, Schlembach D, Abou-Dakn M, et al. Percent error of ultrasound examination to estimate fetal weight at term in different categories of birth weight with focus on maternal diabetes and obesity. BMC Pregnancy and Childbirth. 2022; 22: 241.

[35] Balsyte D, Schäffer L, Burkhardt T, Wisser J, Kurmanavicius J. Sonographic prediction of macrosomia cannot be improved by combination with pregnancy-specific characteristics. Ultrasound in Obstetrics & Gynecology. 2009; 33: 453–458.

[36] Gevaerd Martins J, Kawakita T, Jain P, Gurganus M, Baraki D, Barake C, et al. Impact of maternal body mass index on the accuracy of third trimester sonographic estimation of fetal weight. Archives of Gynecology and Obstetrics. 2023; 307: 395–400.