Fetal Diagnosis and Therapy

Prenatal Diagnosis

Fetal Diagn Ther
DOI: 10.1159/000540554

Received: February 11, 2024
Accepted: July 22, 2024
Published online: July 26, 2024

Unisex and Sex-Specific Prescriptive Fetal Growth Charts for Improved Detection of Small-for-Gestational-Age Babies in a Low-Risk Population: A post hoc Analysis of a Cluster-Randomized Study

Mariëlle van Roekel[a, b] Corine J. Verhoeven[a, c] Hester D. Kamphof[d] Sanne J. Gordijn[d] Wessel Ganzevoort[e, f] Arie Franx[g] Wessel van Wieringen[h, i] Ank de Jonge[a, j] Jens Henrichs[a, k]

[a]Amsterdam UMC location Vrije Universiteit Amsterdam, Midwifery Science, Amsterdam, The Netherlands; [b]Midwifery Academy Amsterdam Groningen, Inholland, Amsterdam, The Netherlands; [c]Division of Midwifery, School of Health Sciences, University of Nottingham, Nottingham, UK; [d]Department of Obstetrics and Gynaecology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [e]Department of Obstetrics and Gynaecology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands; [f]Amsterdam Reproduction and Development Research Institute, Amsterdam, The Netherlands; [g]Department of Obstetrics and Gynaecology, Erasmus University Medical Center, Rotterdam, The Netherlands; [h]Department of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC, VU Medical Center, Amsterdam, The Netherlands; [i]Department of Mathematics, VU University Amsterdam, Amsterdam, The Netherlands; [j]Department of Primary and Long-Term Care, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [k]Amsterdam Public Health Research Institute, Quality of Care, Amsterdam, The Netherlands

Keywords
Small-for-gestational age · Detection · Fetal growth · Fetal biometry · Fetal growth chart · Abdominal circumference · Unisex charts · Sex-specific charts

Abstract
Introduction: Our aim was to develop and evaluate the performance of population-based sex-specific and unisex prescriptive fetal abdominal circumference growth charts in predicting small-for-gestational-age (SGA) birthweight, severe SGA (sSGA) birthweight, and severe adverse perinatal outcomes (SAPO) in a low-risk population. ***Methods:*** This is a post hoc analysis of the Dutch nationwide cluster-randomized IRIS study, encompassing ultrasound data of 7,704 low-risk women. IRIS prescriptive unisex and IRIS sex-specific abdominal circumference (AC) fetal growth charts were derived using quantile regression. As a comparison, we used the descriptive unisex Verburg chart, which is commonly applied in the Netherlands. Diagnostic parameters were calculated based on the 34–36 weeks' ultrasound. ***Results:*** Sensitivity rates for predicting SGA and sSGA birthweights were more than twofold higher based on the IRIS prescriptive sex-specific (respectively SGA 43%; sSGA 59%) and unisex (SGA 39%; sSGA 55%) charts, compared to the Verburg chart (SGA 16%; sSGA 23% both $p < 0.01$).

karger@karger.com
www.karger.com/fdt

Correspondence to:
Mariëlle van Roekel, m.vanroekel@amsterdamumc.nl

Specificity rates were highest for Verburg (SGA 99%; sSGA 98%) and lowest for IRIS sex-specific (SGA 94%; sSGA 92%). Results for predicting SGA with SAPO were similar for the prescriptive charts (44%), and again higher than the Verburg chart (20%). The IRIS sex-specific chart identified significantly more males as SGA and sSGA (respectively, 42%; 60%, $p < 0.001$) than the IRIS unisex chart (respectively, 35%; 53% $p < 0.01$). ***Conclusion:*** Our study demonstrates improved performance of both the IRIS sex-specific and unisex prescriptive fetal growth compared to the Verburg descriptive chart, doubling detection rates of SGA, sSGA, and SGA with SAPO. Additionally, the sex-specific chart outperformed the unisex chart in detecting SGA and sSGA. Our findings suggest the potential benefits of using prescriptive AC fetal growth charts in low-risk populations and emphasize the importance of considering customizing fetal growth charts for sex. Nevertheless, the increased sensitivity of these charts should be weighed against the decrease in specificity.

Introduction

Ultrasound assessment of fetal size is a key element of the screening for fetal growth restriction (FGR). FGR is a condition in which the fetus does not reach its intrinsic growth potential, mainly due to placental insufficiency, and is associated with severe short- and long-term health problems [1, 2]. To date, the only effective treatment for FGR is timely delivery [3, 4].

Recognizing that the smallest fetuses are associated with increased morbidity rates has led to the widespread adoption of fetal small-for-gestational-age (SGA) as a proxy for FGR identification. Fetal SGA is commonly defined as the abdominal circumference (AC) and/or estimated fetal weight (EFW) below the 10th percentile (p10) on a population-based growth standard [5]. In the Netherlands, the AC is used as a key indicator, considering it is a crucial component of EFW. Two earlier meta-analyses have shown that third-trimester AC and EFW comparably predict SGA at birth, while for a fixed 10% false-positive rate the extrapolated sensitivity has been reported to be higher for AC [6, 7].

In the IRIS study, a Dutch nationwide clinical trial evaluating the effectiveness of routine third-trimester ultrasound screening among low-risk women, we however found that AC<p10 used as a proxy of SGA, performed poorly in predicting SGA neonates [8]. The suboptimal screening results may arise not only from the low predictive accuracy in fundal height measurement and uncertainty in ultrasound fetal biometry but also from a misalignment between neonatal and antepartum growth charts [9, 10]. In the Netherlands, the commonly used AC-based fetal growth chart by Verburg et al. (2008) is unisex and descriptive, reflecting all risk levels of the population, which might be less appropriate for identifying abnormal growth in the low-risk population [11, 12]. In contrast, the Dutch Hoftiezer birthweight chart is prescriptive, based on a healthy population sample and sex-specific [13]. Earlier research suggests that customizing fetal growth charts for sex is promising [14, 15]. The goal of sex-specific customization is to enhance detection by controlling for sex-specific biologically based factors which are related to the fetal growth potential. Moreover, *not* customizing for fetal sex has been related to underdiagnoses of male SGA fetuses and overtreatment of appropriate for gestational-age females suggesting its relationship with genetic growth potential [16, 17]. The World Health Organization, for example, showed that male EFW is 3.5–4.5% higher than female EFW [18]. Nevertheless, in the Netherlands, fetal growth charts have not adopted customization for fetal sex.

For fetal growth charts to be useful for clinical decision making, charts are needed that adequately identify fetuses that are at risk. We hypothesize that in low-risk populations screening results will improve by optimizing alignment of fetal growth charts with the prescriptive and sex-specific neonatal growth chart [13]. By using data from a large-scale low-risk prospective sample of the IRIS study, we aim to determine whether developing a prescriptive IRIS unisex and IRIS sex-specific AC fetal growth chart improves the prediction of SGA birthweight, severe SGA birthweight, and SGA birthweight with severe adverse perinatal outcomes (SAPO) as compared to the descriptive unisex Verburg AC growth chart [8, 11, 19].

Material and Methods

Study Design

This study is a secondary analysis using data derived from the IUGR Risk Selection (IRIS) study, a nationwide stepped wedge cluster-randomized trial in a low-risk population, evaluating the effectiveness of routine third-trimester ultrasound screening around 28–30 and 34–36 weeks' gestation (intervention strategy) in reducing SAPO, compared to usual care, i.e., symphysis fundal height measurements combined with clinically indicated ultrasound scans (reference strategy) [8, 19]. Both strategies used the same protocol for FGR detection and subsequent obstetric management [19]. A total of 13,046 healthy women with a singleton pregnancy, a reliable gestational age, and a non-anomalous 20 weeks' scan were included. The children were born between

March 2015 and August 2016 [19]. In short, the assessment of pregnant women and data collection consisted of fetal ultrasounds, questionnaires, retrieval of information from medical records, and linkage of perinatal registry data. Written consent was obtained from all participants. The IRIS study was approved by the Institutional Review Board of VU Medical Center (reference number 2013.409). The design of the IRIS study has been described elsewhere [8, 19]

Inclusion and Exclusion Criteria

For the current analyses, we included women with ultrasound data available between 18^{+0}–36^{+6} weeks' gestation, with at least one biometry ultrasound scan in mid-wife-led care after 26^{+0} weeks' gestation. Reasons for lacking ultrasound data comprised early referrals to secondary or tertiary antenatal care due to complications, e.g., pregnancy-related hypertension, early severe SGA or FGR, and preterm birth. Ultrasound data were also absent in women participating in the reference strategy of the IRIS study with reassuring fundal height measurements and thus not having an indication for biometry ultrasounds. Women who were referred for secondary or tertiary care after the second study scan 34–36 weeks' ultrasound scan, were included in this dataset. Data of women with missing values on infant sex, birthweight, and/or gestational age at birth were also excluded.

Data Collection

The data collection process of the demographics, baseline characteristics, maternal and perinatal outcomes, and sonographic quality assessments and guidelines are extensively described elsewhere [8, 19, 20]. Notably, sonographers conducted biometry according to the guidelines of the Dutch Society of Obstetrics and Gynaecology (NVOG) [20]. All sonographers successfully completed the national e-learning training for fetal biometry and either possessed a certificate for structural anomaly screening or were approved as adequate performers of ultrasonography based on 4 cases assessed by two well-experienced sonographers [8]. They also used ultrasound equipment in accordance with the quality standards of the NVOG [20]. Furthermore, quality assessments during the trial were performed by two independent board members of the Dutch Professional Organisation of Sonographers [8, 19]

Measures

Descriptions

Maternal baseline characteristics comprised parity, age, height, pre-pregnancy body mass index (BMI) (in kg/m^2), and ethnicity (Dutch, other western, non-western). Pregnancy and birth characteristics were start of labor in mid-wifery-led care (MLC), mode of birth (i.e., spontaneous birth without interventions, vaginal-assisted birth, pre-labor cesarean, and emergency cesarean), and hypertensive disorders. Neonatal characteristics were gestational age at birth, birthweight (g), birthweight percentile based on the Dutch Hoftiezer birthweight chart [13], Apgar score at 5 min <4, asphyxia (an arterial base excess of cord blood less than −12 mmol/L), >24 h ventilation, transfer to the pediatric unit, preterm birth <37 weeks, NICU admission at term, perinatal death, and SAPO. In the initial IRIS study, SAPO was constructed by perinatologists and neonatologists to optimally capture the morbidity and multifactorial co-morbidity associated with FGR and was defined as a composite of one or more of the following SAPOs: perinatal death between 34 weeks' gestation and 7 days after birth; 5' Apgar score <4; asphyxia, stupor, or decreased response to pain; coma; neonatal seizures; assisted ventilation >24 h; septicemia; meningitis; bronchopulmonary dysplasia; intraventricular hemorrhage; cystic periventricular leukomalacia; and/or necrotizing enterocolitis [8, 19]

Ultrasounds

The biometry ultrasounds used for fetal growth chart modeling comprised the biometric measurements carried out at the fetal anomaly scan between 18^{+0} to 23^{+6} weeks' gestation and ultrasounds that were performed in the IRIS study. Ultrasound data from both the reference group of the IRIS study (clinically indicated biometric measurements) and the intervention group (routine biometric measurements first scan between 28 and 30 and second scan between 34 and 36 weeks' gestation in addition to clinically indicated scans) were used. To develop growth charts all ultrasounds were used from 18 weeks' gestation onward. For our analyses assessing predictive accuracy regarding SGA birthweight <p10, SGA birthweight with SAPO, and severe SGA birthweight <p3, we included data from all women with a biometry ultrasound between 32^{+1} and 36^{+6} weeks' gestation. In case there was more than one ultrasound in that period of time, we chose the last ultrasound. We specifically used this time frame because unlike the 28–30 weeks' gestation time frame, late-onset FGR is more likely to remain undetected, particularly among low-risk women, as it is less frequently associated with hypertensive disorders in pregnancy [21]. AC was measured in mm [19].

Verburg Fetal Growth Chart

The Verburg fetal reference chart is a descriptive unisex chart based on a multi-ethnic urban mixed-risk population of 8,313 pregnant women participating in the Generation R study between 2002 and 2006 in Rotterdam, the Netherlands [11]. Exclusion criteria were enrollment >24^{+0} weeks' gestation, multiple pregnancy, major fetal anomaly, termination of pregnancy for medical reasons, miscarriage, or perinatal death. Women were examined in a research setting, on three occasions, i.e., early pregnancy (≤17^{+6} weeks), mid-pregnancy (≥18^{+0} – <25 weeks), and late pregnancy (≥25 weeks) [11].

Hoftiezer Birthweight Chart

The Hoftiezer birthweight chart is a sex-specific prescriptive chart based on routinely collected data from the Dutch perinatal registry of low-risk mothers between 2000 and 2014 [13]. Exclusion criteria were multiple gestation, pre-existing maternal medical conditions, maternal substance (ab)use, medical conditions related to pregnancy, and congenital malformations. Sex- and gestational-age-specific prescriptive birthweight charts were derived from a total of 16,29,776 life-born singleton infants, born between 23^{+0} and 41^{+6} weeks' gestation to allegedly low-risk healthy mothers [13].

Outcomes

SGA at birth was defined as a birthweight <p10 based on the Hoftiezer chart [13]. SGA birthweight with SAPO and severe SGA birthweight (<p3) were also used as primary outcomes.

Statistical Analyses

Descriptive statistics (means, standard deviations (SD), frequencies, percentages) were calculated for maternal and neonatal characteristics by fetal sex. χ^2 tests (for categorical variables) or

independent *t*-tests (for continuous variables) were applied to test fetal sex differences in maternal and neonatal characteristics.

Growth reference curves for fetal biometry of AC for each fetal sex were fitted to data of fetuses with at least two ultrasound measurements in our study sample. The growth charts are parameterized by a cubic spline of gestational age. Growth was fitted by means of quantile regression, with quantiles 0.10, 0.50, 0.90. In the model of the sex-specific growth chart, we included fetal sex, gestational age (as a cubic spline), and their interaction as explanatory variables. We obtained fetal sex-specific growth curves from the fitted model. Analysis of variance was used to test the contribution of each model term, random and fixed. Finally, from the fitted model the pointwise (at each observed gestational age) quantiles were derived and plotted as curves on the data. Growth curves were also fitted using a fifth-order spline (but visual inspection revealed that it did not result in a substantially better fit). AC <p10 was defined as a gestational-age-specific AC measurement below the p10 using the Dutch unisex fetal growth curve by Verburg et al. [11] (2008) or the newly established IRIS sex-specific and unisex fetal growth reference curves depicted above.

To compare the diagnostic accuracy of the Verburg unisex curve and the newly established IRIS unisex and sex-specific fetal growth charts based on third-trimester AC <p10 to detect SGA birthweight and SGA birthweight with SAPO, we calculated sensitivity, specificity, and positive and negative likelihood ratios with their 95% confidence intervals (CI). The performance of the three fetal growth charts as diagnostic tests was also evaluated by examining the area under the receiver operating characteristic (ROC) curve using the respective continuous percentiles of the charts as predictor variables. McNemar binomial test was used to assess significant differences in sensitivity and specificity between Verburg, and the IRIS sex-specific and IRIS unisex reference curves [22]. All analyses were performed using complete case analyses. The two-sided significance level was set at $p < 0.05$. Analyses were conducted using the Statistical Package for Social Sciences (SPSS) version 26.0 for Windows (SPSS Inc., Chicago, Il, USA) and MedCalc Statistical Software version 19.2.6.

Results

For 13,046 women, data on at least one of the perinatal outcomes were available. For 12,293 of these women, at least two fetal biometry ultrasound measurements >18 weeks' gestation (comprising a total of 28,128 ultrasounds, 12,226 of male and 13,902 of female fetuses) were available to develop the IRIS unisex and IRIS sex-specific fetal growth curves. For measuring third-trimester diagnostic parameters, we excluded subjects without ultrasound measurements 34–36 weeks' gestation (n = 4,508). Another 28 women were excluded due to missing values on birthweight centiles and 53 subjects with missing values on SAPO (n = 53). Thus, 7,704 (3,853 male and 3,851 female) subjects were included in our main analyses (see online suppl. Fig. S1; for all online suppl. material, see https://doi.org/10.1159/000540554; Flowchart).

We observed no fetal sex differences in demographics, maternal anthropometrics, most maternal morbidity- and lifestyle-related factors, and perinatal characteristics (Table 1). Also, we observed no sex differences in SGA prediction using the IRIS sex-specific growth chart; both male and female fetuses were equally often diagnosed with SGA. We did find, however, that boys experienced worse outcomes in various aspects compared to girls. For instance, male fetuses were less likely to be diagnosed with fetal SGA on the unisex fetal growth charts, while male neonates had higher rates of SAPO, assisted ventilation >24 h and were more frequently admitted to the neonatal intensive care unit (NICU) at term, than female counterparts (Table 1). Also, mothers of male fetuses were less likely to experience spontaneous birth without interventions and more likely to give birth through emergency cesarean.

Table 2 shows the diagnostic parameters of antenatal prediction of neonatal SGA based on the derived IRIS sex-specific and unisex growth charts and the Verburg unisex fetal growth chart. SGA detection was highest with the IRIS sex-specific growth chart, predicting a little over four out of ten SGA fetuses during pregnancy (sensitivity 43%) and was followed by the IRIS unisex fetal growth chart with almost four out of ten fetuses (sensitivity 39%). The use of the Verburg unisex fetal growth chart resulted in a significant decrease in prediction rates, effectively more than halving the numbers identified (sensitivity 16%). For male and female fetuses, this pattern was comparable, although sensitivity for females did not differ between the IRIS sex-specific and the IRIS unisex growth chart. For all fetal growth charts, the sensitivity was lower for male fetuses compared to female fetuses.

In contrast, the specificity of the IRIS sex-specific chart was lowest (94%) followed by the IRIS unisex chart (95%). The Verburg chart had a specificity of 99%. Differences between the fetal growth charts were again significant (McNemar p <0.001), except for females between the two IRIS growth charts. For both the IRIS unisex chart and the Verburg unisex chart, the specificity rate was slightly higher for male than for female fetuses. Positive likelihood ratios (LR+) were high for the Verburg chart, increasing the post-test odds of SGA after a positive test (LR+ 19.0), with even higher LR+ for male (LR+ 29.6) than female (LR+ 14.6) fetuses. Conversely, negative likelihood ratios (LR–) were for Verburg almost indiscriminative (LR– overall 0.9). This pattern was similar for the IRIS unisex chart, with much lower LR+ (overall 7.4) and lower LR– (overall 0.6). The IRIS sex-specific growth chart had marginally lower LR+ and similar LR– as compared to the IRIS unisex chart but revealed no

Table 1. Maternal, fetal, and neonatal descriptions by infant sex

	Males, *n* = 3,853	Females, *n* = 3,851	χ^2/ANOVA
	mean (±SD) or *n* (%)	mean (±SD) or *n* (%)	*p* value
Nulliparous[a]	1,859 (48.7)	1,839 (48.2)	0.631
Age[b]	30.7 (±4.5)	30.8 (±4.4)	0.534
Pre-pregnancy BMI, kg/m^{2ab}	24.1 (±4.4)	24.0 (±4.4)	0.338
Height	169.6 (±6.8)	169.7 (±7.0)	0.478
Ethnicity[a]			
Dutch	2,900 (76.1)	2,886 (75.6)	0.373
Other western	405 (10.6)	420 (11.0)	
Non-western	504 (13.2)	513 (13.4)	
Committed relationship[a]	3,715 (97.6)	3,724 (97.8)	0.642
Paid job[a]	3,223 (84.9)	3,173 (83.5)	0.098
Low education[a]	374 (9.9)	386 (10.2)	0.671
Smoking >5 cigarettes after 1st trimester[a]	116 (3.0)	135 (3.5)	0.223
Hypertensive disorder after ultrasound 34–36 weeks	290 (8.5)	270 (7.9)	0.338
AC <p10 IRIS sex-specific chart	381 (9.9)	385 (10)	0.873
AC <p10 IRIS unisex chart	281 (7.3)	372 (9.7)	<0.001
AC <p10 Verburg unisex chart	72 (1.9)	100 (2.6)	0.031
AC percentile IRIS unisex chart[b]	56.8 (±28.4)	51.8 (±28.4)	<0.001
AC percentile IRIS sex-specific chart[b]	50.8 (±28.7)	51.8 (±28.6)	0.136
AC percentile Verburg unisex chart[b]	66.2 (±24.6)	61.9 (±25.2)	<0.001
Premature birth <37 weeks	98 (2.5)	86 (2.2)	0.372
Spontaneous birth without intervention[a]	1,645 (44.6)	1,735 (47.5)	0.015
Vaginal-assisted birth	326 (8.5)	282 (7.3)	0.064
Any cesarean section	525 (13.6)	446 (11.6)	0.007
Pre-labor cesarean	201 (5.2)	192 (5.0)	0.508
Emergency cesarean	324 (8.4)	254 (6.6)	0.008
Gestational age at birth, weeks[b]	39.8 (±1.3)	39.9 (±1.3)	0.02
Birthweight, grams[b]	3,568 (±448)	3,445 (±459)	<0.001
Birthweight percentile[b]	50 (±29)	51 (±29)	0.852
SGA birthweight	380 (9.9)	359 (9.3)	0.421
SAPO[c]	61 (1.6)	41 (1.2)	0.046
SGA birthweight with SAPO	15 (0.4)	10 (0.3)	0.317
5′Apgar <4	11 (0.3)	8 (0.2)	0.491
Perinatal death	4 (0.1)	5 (0.1)	0.738
Asphyxia	47 (1.2)	30 (0.8)	0.052
Assisted ventilation >24 h	13 (0.3)	2 (0.1)	0.004
NICU >37 gestation	28 (0.8)	10 (0.2)	0.003

[a]*n* differs from total *n* due to missing values; nulliparity, *n* = 7,636; BMI and committed relationship, 7,613; height, *n* = 7,664, ethnicity, *n* = 7,628; paid job, *n* = 7,595; low education, *n* = 7,561; smoking >5 cigarettes after 1st trimester, *n* = 7,690; spontaneous birth without intervention, *n* = 7,339. [b]Continuous variable, *p* value refers to ANOVA (for categorial variables *p* values were based on χ^2 tests). [c]SAPO is a composite outcome, i.e., one or more of the following outcomes specified: stillbirth (not due to congenital anomaly), neonatal death at term (not due to congenital anomaly), hypoxic-ischemic encephalopathy at term, use of inotropes at term, mechanical ventilation at term, severe metabolic acidosis at term (defined as pH 12 mmol/L).

Table 2. Test characteristics of small-for-gestational-age (SGA) birthweight detection with AC below the 10th percentile (AC <p10) on the IRIS sex-specific and the IRIS unisex fetal growth charts and the Dutch Verburg unisex fetal growth chart late third trimester between 34 and 36 weeks' gestation

	TP	FN	FP	TN	Sensitivity % (95% CI)	Specificity % (95% CI)	LR+ (95% CI)	LR– (95% CI)	AUC (95% CI)
AC <p10 IRIS sex-specific									
All	319	420	436	6,529	43 (40–47)[a]	94 (93–94)[a]	6.9 (6.1–7.8)	0.6 (0.6–0.7)	0.84 (0.83–0.85)
Males	161	219	209	3,264	42 (37–48)[a]	94 (93–95)[a]	6.7 (5.9–8.4)	0.6 (0.6–0.7)	0.85 (0.83–0.87)
Females	158	201	227	3,265	44 (39–49)[b]	94 (93–94)[b]	6.8 (5.7–8.0)	0.6 (0.6–0.7)	0.83 (0.81–0.85)
AC <p10 IRIS unisex									
All	288	451	365	6,600	39 (35–43)[a]	95 (94–95)[a]	7.4 (6.5–8.5)	0.6 (0.6–0.7)	0.84 (0.82–0.85)
Males	134	246	147	3,326	35 (30–40)[a]	96 (95–96)[a]	8.3 (6.8–10.2)	0.7 (0.6–0.7)	0.85 (0.83–0.87)
Females	154	205	218	3,274	43 (38–48)[b]	94 (93–95)[b]	6.9 (5.8–8.2)	0.6 (0.6–0.7)	0.83 (0.81–0.85)
AC <p10 Verburg unisex									
All	115	624	57	6,908	16 (13–18)[a]	99 (99–99)[a]	19.0 (14.0–25.9)	0.9 (0.8–0.9)	0.84 (0.82–0.85)
Males	55	325	17	3,456	14 (11–18)[a]	100 (99–100)[a]	29.6 (17.4–50.5)	0.9 (0.8–0.9)	0.85 (0.83–0.87)
Females	60	299	40	3,452	17 (13–21)[b]	99 (98–99)[b]	14.6 (9.9–21.5)	0.8 (0.8–0.9)	0.83 (0.81–0.85)

SGA birthweight was defined as birthweight <p10 on the Dutch Hoftiezer birthweight chart. [a]McNemar showed significant differences between IRIS sex-specific, IRIS unisex, and Verburg unisex mutually ($p < 0.001$). [b]McNemar showed significant differences between IRIS sex-specific and Verburg and between IRIS unisex and Verburg unisex ($p < 0.001$).

difference between male and female fetuses. All charts performed equally well in predicting SGA birthweight, with areas under the ROC curves around 0.84.

Repeating these analyses for predicting severe SGA birthweight with AC <p10 on the three fetal growth charts, we found similar patterns, but sensitivity increased and specificity slightly decreased (Table 3). LR– for the IRIS sex-specific chart slightly improved to 0.4, lowering the post-test odds on SGA <p3 more than 50%.

Table 4 presents prediction rates for SGA birthweight with SAPO. Sensitivity doubled with both the prescriptive IRIS sex-specific and IRIS unisex chart (sensitivity overall both 44%) in comparison to the descriptive Verburg chart (sensitivity overall 20%). Detection rates for males were 40%, and for females 50%, but this difference was not statistically significant. Specificities for Verburg were again high, around 98%, resulting in very few false positives.

Figure 1 presents the ROC curves for detecting (a) SGA birthweight <p10; (b) severe SGA birthweight; and (c) SGA birthweight with SAPO using AC percentiles on the IRIS sex-specific, IRIS unisex, and Verburg fetal growth charts, respectively. Online supplementary Figure S2 in the supplementary material shows scatterplots of the individual ultrasound measurements and fitted 10th, 50th, and 90th percentile lines by fetal sex. Fetal sex significantly explained the difference in the regression curves and moderated the effect of gestational age ($p < 0.001$). Nevertheless, the variation of the AC measurements within and between days of gestation exceeds the differences that were observed between the sexes. online supplementary Table S1 outlines the formula of the quantile curves.

Discussion

The current study revealed that in a Dutch low-risk population detection of SGA birthweight, severe SGA birthweight, and SGA birthweight with SAPO more than doubled with both the derived prescriptive unisex and sex-specific IRIS fetal growth charts in comparison to the commonly used descriptive unisex Verburg fetal growth chart. This came at the cost of a decreased specificity. Moreover, utilizing the prescriptive sex-specific IRIS fetal growth chart resulted in slightly higher sensitivity and marginally lower specificity in predicting SGA and severe SGA, compared to the unisex IRIS fetal growth chart. The AUCs of all fetal growth charts were comparable.

Both the unisex and sex-specific IRIS growth charts exhibited significantly improved sensitivity rates compared to the presently utilized Verburg growth chart, resulting in more than doubled predictions for all outcomes. In the current sample, the IRIS fetal growth chart, derived from a healthy low-risk population, aligns more appropriately with the existing prescriptive Dutch

Table 3. Test characteristics of severe SGA birthweight <p3 detection with AC below the 10th percentile (AC <p10) on the IRIS sex-specific and the IRIS unisex fetal growth charts and the Dutch Verburg unisex fetal growth chart late third trimester between 34 and 36 weeks' gestation

	TP	FN	FP	TN	Sensitivity % (95% CI)	Specificity % (95% CI)	LR+ (95% CI)	LR– (95% CI)	AUC (95% CI)
AC<p10 IRIS sex-specific									
All	118	81	637	6,868	59 (52–66)[a]	92 (91–92)[a]	7.0 (6.1–8.0)	0.4 (0.4–0.5)	0.86 (0.84–0.88)
Males	68	46	302	3,437	60 (50–69)[a]	92 (91–93)[a]	7.4 (6.1–8.9)	0.4 (0.4–0.6)	0.88 (0.85–0.91)
Females	50	35	335	3,431	59 (47–69)[b]	91 (90–92)[b]	6.6 (5.4–8.1)	0.5 (0.4–0.6)	0.84 (0.80–0.88)
AC<p10 IRIS unisex									
All	109	90	544	6,961	55 (48–62)[a]	93 (92–93)[a]	7.6 (6.5–8.8)	0.5 (0.4–0.6)	0.86 (0.84–0.88)
Males	60	54	221	3,518	53 (43–62)[a]	94 (93–95)[a]	8.9 (7.2–11.1)	0.5 (0.4–0.6)	0.88 (0.85–0.91)
Females	49	36	323	3,443	58 (46–68)[b]	91 (90–92)[b]	6.7 (5.5–8.3)	0.5 (0.4–0.6)	0.84 (0.80–0.88)
AC<p10 Verburg unisex									
All	46	153	126	7,379	23 (17–30)[a]	98 (98–99)[a]	13.8 (10.1–18.7)	0.8 (0.7–0.8)	0.86 (0.84–0.88)
Males	19	66	81	3,685	22 (14–33)[a]	98 (97–98)[a]	10.4 (6.6–16.3)	0.8 (0.7–0.9)	0.88 (0.85–0.91)
Females	27	87	45	3,694	24 (16–33)[b]	99 (98–99)[b]	19.7 (12.7–31)	0.8 (0.7–0.9)	0.84 (0.80–0.88)

Severe SGA birthweight was defined as birthweight <p3 on the Dutch Hoftiezer birthweight chart. [a]McNemar showed significant differences between IRIS sex-specific, IRIS unisex, and Verburg unisex mutually ($p < 0.01$). [b]McNemar showed significant differences between IRIS sex-specific and Verburg and between IRIS unisex and Verburg ($p < 0.001$) but not between IRIS unisex and IRIS sex-specific.

birthweight chart. This suggests that implementing a prescriptive fetal growth chart is likely to enhance the detection of SGA fetuses in low-risk women, enabling subsequent in-depth FGR diagnostic assessments. The improvement in sensitivity came at the cost of a decreased specificity, which is a crucial consideration in low-risk populations. Lower specificity may contribute to unnecessary healthcare utilization such as medically initiated (preterm) birth resulting in iatrogenic effects and increased healthcare costs associated with the management of false-positive cases [23]. In our sample, the unisex descriptive Verburg growth chart with its high specificity and LR+ provided strong evidence of true SGA status for screen positives, with minimal risk of overutilization. Yet, our findings also suggest that a screen-negative result on the Verburg chart does not rule out SGA in low-risk populations, as illustrated by low sensitivity rates of around 16% and non-discriminative LR– around 0.9. In daily practice, in low-risk populations, where ultrasound biometry screening is commonly based on non-reassuring fundal height measurements, this underdetection may lead to undertreatment of potentially at-risk fetuses requiring additional surveillance and possibly interventions. The LR– of IRIS fetal growth charts for ruling out SGA was better than that of the Verburg chart. Nevertheless, the LR– of the IRIS charts fell short of the desirable <0.1 cut-off, that is considered useful in ruling out a disease [24]. Despite this, we generally argue that the doubled sensitivity and satisfactory specificity of the IRIS charts have the potential to more effectively identify those fetuses that may benefit from additional FGR diagnostics and monitoring [16].

When comparing the IRIS prescriptive charts with each other, the sex-specific IRIS chart demonstrated improved detection rates for SGA and severe SGA in males without a decrease in SGA detection in females. Additionally, overall detection rates of SGA with SAPO were not improved using the IRIS sex-specific chart as compared to the IRIS unisex chart. These findings partially align with recent French population-based research by Monier et al. [16] In contrast to our findings, Monier et al. showed that customizing reference charts for fetal sex increased SGA detection in males but led to a decrease in SGA detection in females. However, similarly to our findings, they did not find improvements in overall sensitivity in detecting SGA fetuses with morbidity. In contrast to this earlier study we used AC instead of EFW, which might account for some variations. Additionally, we speculate that these different findings might be due to differences in sample characteristics as our low-risk sample may be more homogeneous than the population-based French sample that also applied other growth curves, e.g., the WHO fetal growth curve [16, 18]. These differences and the inconsistencies of findings between the study of Monier et al. and ours stress that more research into the performance of sex-specific charts compared to unisex charts is needed. Nevertheless, we emphasize the limitations

Table 4. Test characteristics of small-for-gestational-age (SGA) birthweight with SAPO detection with AC below the 10th percentile (AC <p10) on the IRIS sex-specific and the IRIS unisex fetal growth charts and the Dutch Verburg unisex fetal growth chart late third trimester between 34 and 36 weeks' gestation

	TP	FN	FP	TN	Sensitivity % (95% CI)	Specificity % (95% CI)	LR+ (95% CI)	LR– (95% CI)	AUC (95% CI)
AC<p10 sex-specific									
All	11	14	744	6,935	44 (24–65)[a]	90 (90–91)[b]	4.5 (2.9–7.1)	0.6 (0.4–0.9)	0.79 (0.69–0.88)
Males	6	9	364	3,474	40 (16–68)	91 (90–91)[b]	4.2 (2.3–7.9)	0.7 (0.4–1.0)	0.84 (0.75–0.93)
Females	5	5	380	3,461	50 (19–81)	90 (89–91)[b]	5.1 (2.7–9.5)	0.6 (0.3–1.0)	0.71 (0.52–0.89)
AC<p10 IRIS unisex									
All	11	14	642	7,037	44 (24–65)[a]	92 (91–92)[b]	5.3 (3.4–8.2)	0.6 (0.4–0.9)	0.78 (0.69–0.87)
Males	6	9	275	3,563	40 (16–68)	93 (92–94)[b]	5.6 (3.0–10.4)	0.7 (0.4–1.0)	0.83 (0.75–0.93)
Females	5	5	367	3,474	50 (19–81)	90 (89–91)[b]	5.2 (2.8–9.8)	0.6 (0.3–1.0)	0.71 (0.53–0.89)
AC<p10 Verburg unisex									
All	5	20	167	7,512	20 (7–41)[a]	98 (97–98)[b]	9.2 (4.2–20.5)	0.8 (0.7–1.0)	0.78 (0.69–0.88)
Males	3	12	69	3,769	20 (4–48)	98 (98–99)[b]	11.1 (3.9–31.4)	0.8 (0.6–1.1)	0.84 (0.76–0.93)
Females	2	8	98	3,743	20 (3–56)	97 (97–98)[b]	7.9 (2.2–27.5)	0.8 (0.6–1.1)	0.71 (0.52–0.89)

SGA birthweight was defined as birthweight <p10 on the Dutch Hoftiezer birthweight chart. SAPO is a composite outcome, i.e., one or more of the following outcomes specified: stillbirth (not due to congenital anomaly), neonatal death at term (not due to congenital anomaly), hypoxic-ischemic encephalopathy at term, use of inotropes at term, mechanical ventilation at term, severe metabolic acidosis at term (defined as pH 12 mmol/L). [a]McNemar showed significant difference between IRIS sex-specific and Verburg and between IRIS unisex and Verburg ($p = 0.031$) but not between IRIS sex-specific and IRIS unisex. [b]McNemar showed significant difference between IRIS sex-specific, IRIS unisex, and Verburg mutually ($p < 0.001$).

of relying on a unisex fetal growth chart, which results in cut-off points that particularly lower the sensitivity of detecting SGA males. This latter aspect is important because previous work showed that male neonates were overrepresented in SGA with SAPO compared to female neonates [25].

Importantly, SGA in itself is a poor predictor of SAPO [4]. Also, AC or EFW <p10 are not optimal cut-offs to distinguish between physiologic and pathologic circumstances in utero and even severe SGA is a poor standalone parameter to predict SAPO [4, 7, 26, 27]. Therefore, utilizing the FIGO recommended criteria for the diagnosis of FGR, including standalone AC/EFW <p3, or AC/EFW <p10 combined with maternal hypertensive disease in pregnancy, abnormal maternal uterine arteries and fetal umbilical and middle cerebral artery Dopplers and/or fetal growth declining >2 quartiles, may prevent diagnosing FGR based solely on a single measurement of SGA [4, 26]. If the fetal AC/EFW is between the p3 and the p10 and the growth pattern and other parameters remain normal, the fetus is likely to be constitutionally small [4, 26]. This approach should prevent increased rates of medically indicated (preterm) delivery based on ultrasound-derived FGR <p10 indications only [17]. Recent studies comparing various definitions of FGR and various fetal growth charts have shown that, irrespective of the used definition and growth chart, prenatal prediction of SAPO was low overall [28, 29]. The nature of the disease and the intervention that applies hamper conclusions of the actual relation to SAPO as both the disease (hypoxia and starvation due to placenta insufficiency) and the intervention (relative preterm delivery) result in similar neonatal features [23]. These findings highlight the complexity of daily practice and the ongoing challenge of distinguishing normal from pathological fetal growth. To move forward, future research should focus on developing and evaluating stepwise screening algorithms. These algorithms should integrate ultrasonography, utilizing the best-performing ultrasound growth chart tailored to the specific population, baseline maternal (and paternal) characteristics, sex-specific placental markers [30, 31], Doppler measurements [32, 33], and maternal awareness of fetal well-being [34].

Strengths and Limitations

A strength of this study is that we used a large amount of ultrasound data of low-risk women from a large-scale nationwide trial with reliable prospective data collection, making our results relevant for low-risk populations in other high-income countries. In line with the WHO, we used quantile regression to explain AC by sex and days of gestation. This enables us to particularly estimate the tails of the

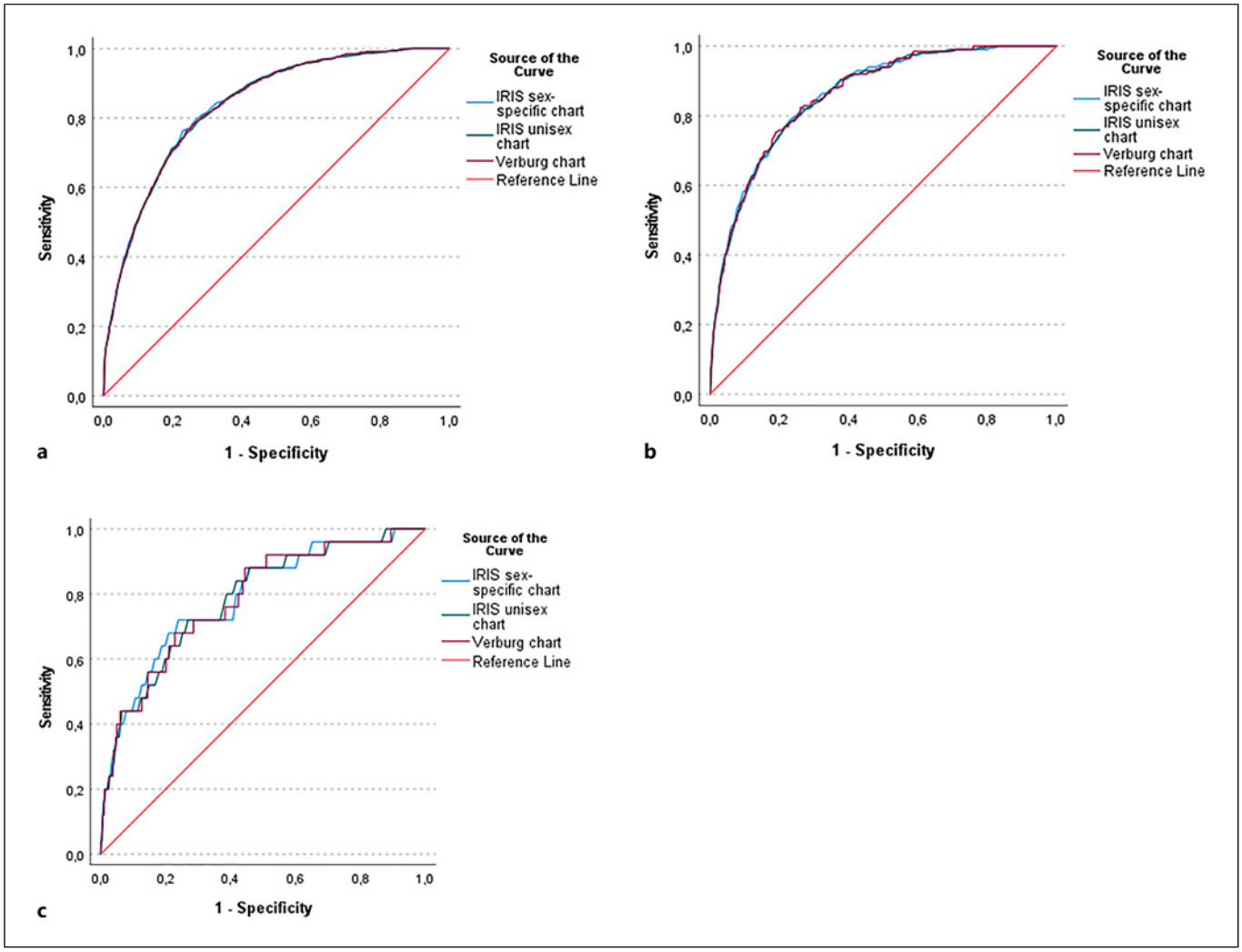

Fig. 1. a Receiver operating characteristic (ROC) curve for detecting SGA birthweight <p10 detection using AC on IRIS sex-specific, IRIS unisex, and Verburg fetal growth charts. **b** ROC curve for detecting severe SGA birthweight <p3 detection using AC on IRIS sex-specific, IRIS unisex, and Verburg fetal growth charts. **c** ROC curve for detecting SGA birthweight <p10 with severe adverse perinatal outcomes (SAPO) detection using AC on IRIS sex-specific, IRIS unisex, and Verburg fetal growth charts.

distribution, e.g., the 10% quantile. Moreover, quantile regression does not rely on the normality assumption of the error. Furthermore, our study was conducted among women who were still at low risk during the third-trimester ultrasound. As diagnostic parameters vary with the prevalence of diseases, a lower sensitivity and higher specificity might be expected than in the general population [35]. Therefore, it would be interesting to evaluate the diagnostic accuracy of the IRIS charts in a clinical population, especially because recent studies in a nationwide cohort [28] and tertiary center [29] did not find the improved performance of prescriptive international standards that were constructed for universal use.

The study was a proof of principle and not purposefully designed with the objective of creating new charts of fetal size and we did not adhere to consensus guidelines regarding data collection and participant recruitment for developing a fetal growth chart [36]. For some gestational-age windows, the number of ultrasound measurements was low. However, as fetal growth follows a continuous trajectory, interpolation using cubic splines between these timeframes is justified. In addition, having data on several timeframes based on many observations contributed to the reliable estimation of the curve's slope using cubic splines. Moreover, clinically indicated ultrasounds, when available, were also used, so

indication bias cannot be ruled out. Nevertheless, women who received clinically indicated ultrasounds were still in mid-wife-led care and ultrasound data were mainly derived from women receiving routine ultrasounds. Therefore, we believe that our sample was adequate to investigate whether developing a prescriptive unisex and/or sex-specific fetal AC growth chart could improve SGA detection in low-risk populations. However, more scientific rigor will have to be applied to develop a more future-proof fetal growth chart. Furthermore, the differences in sensitivity rates between the IRIS and Verburg charts could potentially be overestimated by the differences in methodological approaches, rather than solely reflecting the performance of prescriptive and descriptive growth charts. Although based on expert consensus, the used definition of SGA and SAPO is imperfect and does not uniquely and accurately capture the phenomenon of FGR. To improve the accuracy assessment of FGR, future research should incorporate the markers that may improve phenotyping. These include the earlier mentioned ultrasound and Doppler-based markers and may be expanded with serum biomarkers and placental histological confirmation. Moreover, it is imperative that care is taken to focus on outcomes that represent FGR truly and uniquely, and distinguish them from outcomes that are related to the intervention for suspected FGR [23]. Finally, the current study had the primary aim to explore changes in diagnostic accuracy when customizing for fetal sex and thus did not consider other factors potentially reflecting the actual fetal growth potential, such as maternal height. Future studies should address whether (or not) additionally customizing for maternal height next to customizing for fetal sex might further improve the diagnostic accuracy of fetal growth charts.

Conclusion

Our results suggest that a prescriptive fetal growth chart as opposed to a descriptive fetal growth chart has the potential to improve SGA prediction in low-risk populations. Furthermore, developing a prescriptive sex-specific abdominal circumference growth chart may additionally improve the antenatal diagnoses of SGA and severe SGA fetuses but not that of SGA with SAPO. These findings shed light on the potential benefits of utilizing prescriptive fetal growth charts in low-risk populations for the detection of fetuses in need of subsequent in-depth diagnostic assessments and emphasize the significance of considering customizing fetal growth charts for sex. Nevertheless to prevent overutilization, such a strategy would require a balanced approach to weigh increased sensitivity against decreased specificity, considering the implications of the FIGO diagnostic criteria for FGR.

Acknowledgments

We thank the women and care professionals who participated in this study. We also thank the Perinatal Registry of the Netherlands (Perined) for the use of their database. The IRIS study was approved by the Dutch Institutional Review Board of the Amsterdam UMC, VU Medical University Center Amsterdam (Reference No. 2013.409) [19]. Written consent was obtained from all participants. As this secondary analysis is encompassed by the original study's ethical approval, no additional consent was required.

Conflict of Interest Statement

The authors have declared no competing interests.

Funding Sources

The IRIS study was supported by a grant from the Netherlands Organisation for Health Research and Development (ZonMw; Grant No. 209030001). The work for the current study was funded within the research program Doctoral Grant for Teachers with project no. 023.018.034, financed by the Netherlands Organization for Scientific Research (NWO). The funding organizations had no role in study design, data collection, data analysis, data interpretation, writing of the scientific article, or the decision to submit the paper for publication.

Author Contributions

All authors contributed to the conceptualization of the current study and the revision of the manuscript and approved the final manuscript as submitted. C.V. and M.R. developed the study idea. M.R. and J.H. drafted and revised the manuscript. H.K. was involved in drafting the manuscript and in the interpretation of the results. W.W., M.R., and J.H. analyzed the data. A.-D.J. conceptualized, designed, and supervised the IRIS study and acts as guarantor. J.H. and A.F. were also involved in the conceptualization of the IRIS study. C.V., A.F., S.J., and W.G. reviewed the manuscript for its intellectual content, aided in the interpretation of results, and revised the manuscript.

Data Availability Statement

The ethical approval for the IRIS study does not permit the publication of individual participant level data. Reasonable requests for a de-identified dataset need a Data Transfer Agreement that is in line with the European Union´s General Data Protection Regulation (GDPR). Further inquiries can be directed to the corresponding author.

References

1 Barker DJ. Adult consequences of fetal growth restriction. Clin Obstet Gynecol. 2006;49(2):270–83. https://doi.org/10.1097/00003081-200606000-00009
2 Malhotra A, Allison BJ, Castillo-Melendez M, Jenkin G, Polglase GR, Miller SL. Neonatal morbidities of fetal growth restriction: pathophysiology and impact. Front Endocrinol. 2019;10:55. https://doi.org/10.3389/fendo.2019.00055
3 Bricker L, Medley N, Pratt JJ. Routine ultrasound in late pregnancy (after 24 weeks' gestation). Cochrane Database Syst Rev. 2015;2015(6):CD001451. https://doi.org/10.1002/14651858.CD001451.pub4
4 Melamed N, Baschat A, Yinon Y, Athanasiadis A, Mecacci F, Figueras F, et al. FIGO (international Federation of Gynecology and obstetrics) initiative on fetal growth: best practice advice for screening, diagnosis, and management of fetal growth restriction. Int J Gynaecol Obstet. 2021;152(Suppl 1):3–57. https://doi.org/10.1002/ijgo.13522
5 ACOG Practice Bulletin No. 204 summary: fetal growth restriction. Obstet Gynecol. 2019;133(2):390–2.
6 Blue NR, Yordan JMP, Holbrook BD, Nirgudkar PA, Mozurkewich EL. Abdominal circumference alone versus estimated fetal weight after 24 Weeks to predict small or large for gestational age at birth: a meta-analysis. Am J Perinatol. 2017;34(11):1115–24. https://doi.org/10.1055/s-0037-1604059
7 Caradeux J, Martinez-Portilla RJ, Peguero A, Sotiriadis A, Figueras F. Diagnostic performance of third-trimester ultrasound for the prediction of late-onset fetal growth restriction: a systematic review and meta-analysis. Am J Obstet Gynecol. 2019;220(5):449–59.e19. https://doi.org/10.1016/j.ajog.2018.09.043
8 Henrichs J, Verfaille V, Jellema P, Viester L, Pajkrt E, Wilschut J, et al. Effectiveness of routine third trimester ultrasonography to reduce adverse perinatal outcomes in low risk pregnancy (the IRIS study): nationwide, pragmatic, multicentre, stepped wedge cluster randomised trial. BMJ. 2019;367:l5517. https://doi.org/10.1136/bmj.l5517
9 Pay ASD, Wiik J, Backe B, Jacobsson B, Strandell A, Klovning A. Symphysis-fundus height measurement to predict small-for-gestational-age status at birth: a systematic review. BMC Pregnancy Childbirth. 2015;15(1):22. https://doi.org/10.1186/s12884-015-0461-z
10 Dudley NJ. The management of error in ultrasound fetal growth monitoring. Ultrasound. 2021;29(1):4–9. https://doi.org/10.1177/1742271X20945749
11 Verburg BO, Steegers EA, De Ridder M, Snijders RJ, Smith E, Hofman A, et al. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. Ultrasound Obstet Gynecol. 2008;31(4):388–96. https://doi.org/10.1002/uog.5225
12 Hoftiezer L, Hukkelhoven CW, Hogeveen M, Straatman HM, van Lingen RA. Defining small-for-gestational-age: prescriptive versus descriptive birthweight standards. Eur J Pediatr. 2016;175(8):1047–57. https://doi.org/10.1007/s00431-016-2740-8
13 Hoftiezer L, Hof MHP, Dijs-Elsinga J, Hogeveen M, Hukkelhoven C, van Lingen RA. From population reference to national standard: new and improved birthweight charts. Am J Obstet Gynecol. 2019;220(4):383 e1–7. https://doi.org/10.1016/j.ajog.2018.12.023
14 Pang MW, Leung TN, Sahota DS, Lau TK, Chang AM. Customizing fetal biometric charts. Ultrasound Obstet Gynecol. 2003;22(3):271–6. https://doi.org/10.1002/uog.196
15 Melamed N, Meizner I, Mashiach R, Wiznitzer A, Glezerman M, Yogev Y. Fetal sex and intrauterine growth patterns. J Ultrasound Med. 2013;32(1):35–43. https://doi.org/10.7863/jum.2013.32.1.35
16 Monier I, Ego A, Benachi A, Hocquette A, Blondel B, Goffinet F, et al. Unisex vs sex-specific estimated fetal weight charts for fetal growth monitoring: a population-based study. Am J Obstet Gynecol MFM. 2022;4(1):100527. https://doi.org/10.1016/j.ajogmf.2021.100527
17 van Roekel M, Henrichs J, Franx A, Verhoeven CJ, de Jonge A. Implication of third-trimester screening accuracy for small-for-gestational age and additive value of third-trimester growth-trajectory indicators in predicting severe adverse perinatal outcome in low-risk population: pragmatic secondary analysis of IRIS study. Ultrasound Obstet Gynecol. 2023;62(2):209–18. https://doi.org/10.1002/uog.26167
18 Kiserud T, Piaggio G, Carroli G, Widmer M, Carvalho J, Neerup Jensen L, et al. The World health organization fetal growth charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. PLoS Med. 2017;14(1):e1002220. https://doi.org/10.1371/journal.pmed.1002220
19 Henrichs J, Verfaille V, Viester L, Westerneng M, Molewijk B, Franx A, et al. Effectiveness and cost-effectiveness of routine third trimester ultrasound screening for intrauterine growth restriction: study protocol of a nationwide stepped wedge cluster-randomized trial in The Netherlands (The IRIS Study). BMC Pregnancy Childbirth. 2016;16(1):310. https://doi.org/10.1186/s12884-016-1104-8
20 Nederlandse Vereniging voor Obstetrie en Gynacologie (NVOG). NVOG-richtlijn foetale groeirestrictie (FGR) [Dutch Society of Obstetrics and Gynaecology (NVOG)-guideline fetal growth restriction (FGR)]. Utrecht: NVOG; 2017.
21 Figueras F, Gratacos E. An integrated approach to fetal growth restriction. Best Pract Res Clin Obstet Gynaecol. 2017;38:48–58. https://doi.org/10.1016/j.bpobgyn.2016.10.006
22 Kim S, Lee W. Does McNemar's test compare the sensitivities and specificities of two diagnostic tests? Stat Methods Med Res. 2017;26(1):142–54. https://doi.org/10.1177/0962280214541852
23 Gordijn SJ, Ganzevoort W. Search for the best prediction model, definition and growth charts for fetal growth restriction using a composite of adverse perinatal outcomes: a catch-22? Ultrasound Obstet Gynecol. 2022;60(3):305–6. https://doi.org/10.1002/uog.26037
24 Power M, Fell G, Wright M. Principles for high-quality, high-value testing. Evid Based Med. 2013;18(1):5–10. https://doi.org/10.1136/eb-2012-100645
25 Di Renzo GC, Rosati A, Sarti RD, Cruciani L, Cutuli AM. Does fetal sex affect pregnancy outcome? Gend Med. 2007;4(1):19–30. https://doi.org/10.1016/s1550-8579(07)80004-0
26 Gordijn SJ, Beune IM, Thilaganathan B, Papageorghiou A, Baschat AA, Baker PN, et al. Consensus definition of fetal growth restriction: a Delphi procedure. Ultrasound Obstet Gynecol. 2016;48(3):333–9. https://doi.org/10.1002/uog.15884
27 Meler E, Martinez-Portilla RJ, Caradeux J, Mazarico E, Gil-Armas C, Boada D, et al. Severe smallness as predictor of adverse perinatal outcome in suspected late small-for-gestational-age fetuses: systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2022;60(3):328–37. https://doi.org/10.1002/uog.24977
28 Choi SKY, Gordon A, Hilder L, Henry A, Hyett JA, Brew BK, et al. Performance of six birth-weight and estimated-fetal-weight standards for predicting adverse perinatal outcome: a 10-year nationwide population-based study. Ultrasound Obstet Gynecol. 2021;58(2):264–77. https://doi.org/10.1002/uog.22151
29 Mascherpa M, Pegoire C, Meroni A, Minopoli M, Thilaganathan B, Frick A, et al. Prenatal prediction of adverse outcome using different charts and definitions of fetal growth restriction. Ultrasound Obstet Gynecol. 2024;63(5):605–12. https://doi.org/10.1002/uog.27568
30 Gong S, Sovio U, Aye IL, Gaccioli F, Dopierala J, Johnson MD, et al. Placental polyamine metabolism differs by fetal sex, fetal growth restriction, and preeclampsia. JCI Insight. 2018;3(13):e120723. https://doi.org/10.1172/jci.insight.120723
31 Inkster AM, Fernandez-Boyano I, Robinson WP. Sex differences are here to stay: relevance to prenatal care. J Clin Med. 2021;10(13):3000. https://doi.org/10.3390/jcm10133000

32 Ciobanu A, Wright A, Syngelaki A, Wright D, Akolekar R, Nicolaides KH. Fetal Medicine Foundation reference ranges for umbilical artery and middle cerebral artery pulsatility index and cerebroplacental ratio. Ultrasound Obstet Gynecol. 2019;53(4):465–72. https://doi.org/10.1002/uog.20157
33 Marijnen MC, Kamphof HD, Damhuis SE, Smies M, Leemhuis AG, Wolf H, et al. Doppler ultrasound of umbilical and middle cerebral artery in third trimester small-for-gestational age fetuses to decide on timing of delivery for suspected fetal growth restriction: a cohort with nested RCT (DRIGITAT). BJOG An Int J Obstet Gynaecol. 2024;131(8):1042–53. https://doi.org/10.1111/1471-0528.17770
34 Thompson JMD, Wilson J, Bradford BF, Li M, Cronin RS, Gordon A, et al. A better understanding of the association between maternal perception of foetal movements and late stillbirth-findings from an individual participant data meta-analysis. BMC Med. 2021;19(1):267. https://doi.org/10.1186/s12916-021-02140-z
35 Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med. 1997;16(9):981–91. https://doi.org/10.1002/(sici)1097-0258(19970515)16:9<981::aid-sim510>3.0.co;2-n
36 Ioannou C, Talbot K, Ohuma E, Sarris I, Villar J, Conde-Agudelo A, et al. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. BJOG. 2012;119(12):1425–39. https://doi.org/10.1111/j.1471-0528.2012.03451.x