



# HHS Public Access

Author manuscript

*Am J Perinatol.* Author manuscript; available in PMC 2022 August 01.

Published in final edited form as:

*Am J Perinatol.* 2021 August ; 38(Suppl 1): e46–e56. doi:10.1055/s-0040-1705114.

## Customized versus population growth standards for morbidity and mortality risk stratification using ultrasonographic fetal growth assessment at 22-29 weeks' gestation

**Nathan R. Blue, MD,**

University of Utah Health, Dept. of Obstetrics and Gynecology. Salt Lake City, UT.

**William A. Grobman, MD, MBA,**

Northwestern University, Dept. of Obstetrics and Gynecology. Chicago, IL.

**Jacob C. Larkin, MD,**

Magee-Womens Hospital of University of Pittsburgh School of Medicine, Department of Obstetrics, Gynecology, and Reproductive Sciences, Pittsburgh, PA

**Christina M. Scifres, MD,**

Indiana University School of Medicine, Department of Obstetrics and Gynecology, Indianapolis, IN

**Hyagriv N. Simhan, MD, MS,**

Magee-Womens Hospital of University of Pittsburgh School of Medicine, Dept. of Obstetrics, Gynecology, and Reproductive Sciences. Pittsburgh, PA

**Judith H. Chung, MD, PhD,**

University of California, Irvine, Dept. of Obstetrics and Gynecology. Orange, CA.

**George R. Saade, MD,**

University of Texas Medical Branch, Dept. of Obstetrics and Gynecology. Galveston, TX.

**David M. Haas, MD, MS,**

Indiana University School of Medicine, Dept. of Obstetrics and Gynecology. Indianapolis, IN.

**Ronald Wapner, MD,**

Columbia University Irving Medical Center, Department of Obstetrics and Gynecology, New York, NY

**Uma M. Reddy, MD,**

Yale School of Medicine, Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, New Haven, CT

**Brian Mercer, MD,**

MetroHealth Medical Center, Department of Obstetrics and Gynecology. Cleveland, OH.

**Samuel I. Parry, MD,**

---

Corresponding author: Nathan R. Blue. 30 North 1900 East, Rm 2B200. Salt Lake City, UT 84132. Nblue1297@gmail.com. Phone: 801-581-8425.

Disclosures: the authors report no disclosures.

University of Pennsylvania, Department of Obstetrics and Gynecology, Philadelphia, Pennsylvania.

**Robert M. Silver, MD**

University of Utah Health, Dept. of Obstetrics and Gynecology. Salt Lake City, UT.

## Abstract

**Objective:** To compare the performance of ultrasonographic customized and population fetal growth standards for prediction adverse perinatal outcomes.

**Study design:** This was a secondary analysis of the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be, in which 1 data were collected at visits throughout pregnancy and after delivery. Percentiles were assigned to estimated fetal weights (EFWs) measured at 22-29 weeks using the Hadlock population standard and a customized standard ([www.gestation.net](http://www.gestation.net)). Areas under the curve (AUCs) were compared for prediction of composite and severe composite perinatal morbidity using EFW percentile.

**Results:** Among 8,701 eligible study participants, the population standard diagnosed more fetuses with FGR than the customized standard (5.5% vs 3.5%,  $p<0.001$ ). Neither standard performed better than chance to predict composite perinatal morbidity. Although the customized performed better than the population standard to predict severe perinatal morbidity (AUC 0.56 vs 0.54,  $p=0.003$ ), both were poor. Fetuses considered FGR by the population standard but normal by the customized standard had morbidity rates similar to fetuses considered normally grown by both standards.

The population standard diagnosed FGR among black women and Hispanic women at nearly double the rate it did among white women ( $p<0.001$  for both comparisons), even though morbidity was not different across racial/ethnic groups. The customized standard diagnosed FGR at similar rates across groups. Using the population standard, 77% of FGR cases were diagnosed among female fetuses even though morbidity among females was lower ( $p<0.001$ ). The customized model diagnosed FGR at similar rates in male and female fetuses.

**Conclusion:** At 22-29 weeks gestation, EFW percentile alone poorly predicts perinatal morbidity, whether using customized or population fetal growth standards. The population standard diagnoses FGR at increased rates in sub-groups not at increased risk of morbidity and at lower rates in sub-groups at increased risk of morbidity, whereas the customized standard does not.

## Keywords

customized fetal growth standard; fetal growth restriction; intrauterine growth curve; perinatal morbidity

## Introduction:

Fetal growth restriction (FGR) occurs when intrinsic or extrinsic pathology prevents a fetus from meeting its inherent growth potential, with a resultant increased risk of adverse perinatal outcomes.<sup>1,2</sup> The observation that fetuses at the lowest end of the weight-for-

gestational age spectrum experience the highest rates of morbidity has led FGR to be widely defined as estimated fetal weight (EFW) <10<sup>th</sup> percentile based on the mean weights of fetuses from the population at a given gestational age (“population standard”).<sup>3</sup> This is problematic, however, as comparing fetal size to the population mean does not take into account that all fetuses do not have the same growth potential. One fetus of a given small size may be normally grown with a low risk of morbidity, whereas another fetus of the same size and gestational age may be suffering from FGR and a higher risk of morbidity. Because escalated prenatal surveillance of growth-restricted fetuses has been shown to reduce perinatal mortality, the inability to accurately identify FGR prenatally represents a critical obstacle to reducing morbidity and mortality from FGR.<sup>4</sup>

Efforts have been made to identify factors which contribute to non-pathologic variation in fetal growth and integrate them into a model to project individualized growth potential.<sup>5</sup> Analyses to date have focused primarily on using customized standards to assess birth weights rather than EFWs during pregnancy, when there is still time to adjust prenatal care and delivery planning to reduce perinatal mortality. Even so, results from studies of customized birth weight assessments are mixed with regard to whether their use improves stratification of morbidity risk over population based standards.<sup>6-14</sup> Accordingly, U.S. professional societies and clinicians have not adopted routine use of customization, which is more complex than population-based approaches. Therefore, our objective was to compare a leading population-based, ultrasound-derived fetal growth standard (Hadlock, 1991) with a well-developed, customized standard (gestation-related optimal weight, GROW) for the performance in identifying fetuses perinatal morbidity.<sup>15,16</sup>

### Study design:

This was a secondary analysis of the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b). In this study, 10,038 nulliparous women with a singleton gestation had comprehensive demographic, clinical, and biomarker data prospectively collected across pregnancy, with a goal of developing an observational data set that could be used to predict adverse pregnancy outcomes. Women were recruited from hospitals affiliated with any of the following eight academic centers: Case Western University, Columbia University, Indiana University, University of Pittsburgh, Northwestern University, University of California at Irvine, University of Pennsylvania, and University of Utah. Participation included study visits during each of 3 separate epochs during pregnancy (6 weeks 0 days – 13 weeks 6 days, 16 weeks 0 days – 21 weeks 6 days, and 22 weeks 0 days – 29 weeks 6 days). Outcomes were abstracted from the medical records after delivery. Prior to study initiation, institutional review board approval was obtained at each site and the data coordinating center and all participants gave written informed consent.<sup>17</sup>

In this analysis, we included women who gave birth to non-anomalous singletons, and who had well-dated pregnancies during which ultrasonographic fetal measurements were taken at visit 3. In some cases, a visit 3 occurred after the target window. A crown-rump length (CRL) measurement by a certified study sonographer was required for participation, and gestational age was determined using a standardized protocol (previously published) based on the degree of concordance between the LMP and the CRL (previously

published).<sup>17</sup> Nevertheless, women whose visit 3 ultrasound took place after 29 weeks 6 days were not excluded from our analysis on this basis alone since our hypothesis was not related to a specific gestational period. We also excluded neonates with major anomalies, delivery prior to 24 weeks' gestation, stillbirths occurring prior to EFW measurement, or when information required to calculate EFWs, EFW percentile, or to assess the perinatal composite outcome was missing. Major anomalies were defined as any malformation predisposing to adverse perinatal outcomes or neonatal surgical intervention (e.g., cardiac anomaly, abdominal wall defect, suspected skeletal dysplasia, neural tube defect, brain anomaly). Clinicians were not informed of ultrasound findings unless pre-specified disclosure criteria were met, which included major fetal structural malformation, hydrops, death, EFW <5<sup>th</sup> percentile (according to the growth standard in use at each local site), oligohydramnios, fetal tachycardia or bradycardia, placenta previa, vasa previa, or cervical length <15mm.<sup>17</sup> Measurements from indication-driven ultrasounds performed as part of clinical care were not collected for this study. Umbilical artery Doppler velocimetry was not assessed during the study visit. Sonographers were certified to perform ultrasound for the nuMoM2b study by a standardized process that included review of a webinar and submission of images for review and approval, which has been previously described.<sup>17</sup> EFWs were calculated using the Hadlock formula that incorporates all four biometric parameters (biparietal diameter, head circumference, abdominal circumference, and femur length).<sup>18</sup>

## Analysis

To account for potential differences in EFW formula use by site, we used measured fetal biometric parameters from visit 3 to re-calculate EFWs before assigning percentiles using both the Hadlock population standard<sup>16</sup> and the GROW customized standard, which integrates parity, maternal race/ethnicity, maternal height and weight, and fetal sex.<sup>15</sup> We chose the Hadlock standard for comparison against the customized model as it has the best performance of ultrasound-derived fetal growth standards for the U.S. population.<sup>14,19</sup> In order to calculate a customized EFW percentile, self-identified primary racial/ethnic categories from the nuMoM2b data set were designated as one of 4 categories for which coefficients have been developed in the United States population: European, Hispanic, black, and Native American. For women with missing racial/ethnic information or who self-identified as another group, the global average was substituted. Because more specific racial/ethnic information was collected and coefficients exist for several other racial and ethnic groups, we also performed the analysis using the most specific designations possible. For this analysis, we used more general designations (such as those published specifically for the U.S. population)<sup>5</sup> only when a more specific designation could not be made or when women self-identified using multiple categories. FGR was defined as EFW <10<sup>th</sup> percentile.<sup>3</sup>

Our primary outcome was composite perinatal morbidity. Composite perinatal morbidity was defined as delivery room resuscitation with bag mask ventilation, intubation, chest compressions, or medications, NICU or intermediate care nursery admission, meconium aspiration syndrome, stillbirth, or neonatal death prior to discharge. Secondary outcomes included severe composite morbidity, which was defined as delivery room chest

compression or medication use, sepsis, high-frequency ventilation, persistent pulmonary hypertension of the newborn (PPHN), NEC, seizures, grade III/IV intraventricular hemorrhage (IVH), periventricular leukomalacia (PVL), length of NICU stay > 30 days, stillbirth, or death prior to discharge.<sup>14</sup> The areas under the receiver-operator curve (AUCs) were used to assess the performance of EFW percentiles according to each growth standard to identify fetuses that would later experience the primary and secondary outcomes.

### Comparison groups and subanalyses

We compared the proportion of neonates who experienced the primary and secondary outcomes among three groups: fetuses with normal growth according to both standards, fetuses with FGR according to the customized standard, and fetuses with FGR according to the Hadlock standard but considered normal by the customized standard. We defined the groups this way because preliminary data suggested that the customized standard diagnosed fewer cases of FGR than the population standard. In this case, the relevant clinical question is whether pregnancies with FGR by the more “inclusive” standard but not by the more restrictive standard experience morbidity at rates similar to or higher than the baseline.

The primary analysis was stratified by gestational age at the time of ultrasound (<26 weeks or 26 weeks’ gestation), and sub-analyses of women at elevated risk for growth restriction (current tobacco use, age  $\geq$  35, BMI  $\geq$  30, chronic hypertension, pre-gestational diabetes) and among those with an EFW-delivery interval <60 days were performed. We chose the cutoff of 26 weeks because this represents a reasonable time after which serial fetal growth surveillance might begin in routine clinical practice and because it represented the mid-point of the study visit 3 epoch. Because research umbilical artery Doppler velocimetry values were not performed in the study cohort, we performed a sub-analysis among fetuses with EFW <5<sup>th</sup> and 5<sup>th</sup>-10<sup>th</sup> percentiles to better stratify morbidity risk among FGR cases. We also described the proportions of composite and severe composite morbidity and FGR by each standard according to fetal sex, maternal race/ethnicity and BMI. Rates of composite and severe composite morbidity were compared between the three FGR groups according to each standard, as described above.

AUCs were compared using the Delong method, chi-square and McNemar tests were used to compare independent and correlated proportions, respectively, and ANOVA was used for continuous variables. The p values for pairwise comparisons were adjusted for multiplicity using the Hommel multiple comparison procedure.<sup>20,21</sup> For AUC analyses, a cutoff of 0.8 was considered to represent good discriminatory value, and significance was set at a two-sided p value of 0.05. Statistical analyses were generated using NCSS 12 Statistical Software.<sup>22</sup> IRB approval was obtained for each participating institution.

### Results:

Of the 10,038 women in the nuMoM2b study, 8,701 were eligible for inclusion in this analysis (Fig. 1). The majority of study ultrasounds (90%) took place at or beyond 26 weeks’ gestation, with 2.0% occurring between 30-34 weeks. The mean interval between EFW and delivery was  $89 \pm 42$  days. Only 37 met fetal size criteria for disclosure of study ultrasound findings to managing clinicians. According to the visit 3 ultrasound results, 6.0%

of fetuses (n=519) had an EFW consistent with FGR by either the Hadlock or customized standard. Nineteen percent of neonates (n=1,654) experienced composite morbidity, and 2.2% (n=191) had severe composite morbidity.

### Comparison of ultrasound fetal growth standards

Analysis of baseline characteristics demonstrated that non-white racial/ethnic groups, women with low or normal BMI, and female fetal sex were overrepresented among those diagnosed with FGR using the Hadlock population standard only. In contrast, overweight and obese women were overrepresented in those with FGR using the customized standard (Table 1). More fetuses were diagnosed with FGR by the Hadlock standard compared with the customized standard (Table 2). Fetuses with FGR by the customized standard had a higher rate of composite and severe composite morbidity than fetuses designated as normally grown by both standards ( $p<0.001$ ), whereas fetuses diagnosed with FGR by Hadlock only did not experience morbidity at a higher rate than fetuses considered normally grown by both methods (Table 3). The degree of overlap and discordance between the two standards is illustrated in Figure 2.

For the prediction of severe composite morbidity, the customized standard performed marginally better than the Hadlock standard and was the only standard to perform better than chance. Nevertheless, it remained a very poor classifier of fetuses destined to have severe composite morbidity. For overall performance of morbidity prediction, neither the Hadlock nor the customized standard performed better than chance to predict composite perinatal morbidity. These results are summarized in Table 4. These findings were unchanged when the analysis was limited to women with risk factors for FGR (tobacco use, age  $\geq 35$ , BMI  $\geq 30$ , chronic hypertension, pre-gestational diabetes) stratified according to gestational age at EFW, or when the most detailed racial/ethnic information was used in the customization model (data not shown). However, prediction did improve among deliveries occurring  $<60$  days after EFW (Table 4).

### Outcomes of fetuses with EFW $<5^{\text{th}}$ or $5^{\text{th}}-10^{\text{th}}$ percentile

Fetuses with EFW  $<5^{\text{th}}$  percentile by either the customized or Hadlock standards experienced higher rates of composite and severe composite morbidity than those with EFW  $>5^{\text{th}}$  percentile according to both standards (Table 5). Fetuses with EFW at the  $5^{\text{th}}-10^{\text{th}}$  percentile by either standard experienced similar rates of composite morbidity compared to fetuses with EFW  $>10^{\text{th}}$  percentile by both methods. Fetuses with EFW  $5^{\text{th}}-10^{\text{th}}$  percentile by the customized standard were more likely to experience severe composite morbidity than fetuses with EFW  $>10^{\text{th}}$  percentile by both methods. In contrast, fetuses with EFW  $5^{\text{th}}-10^{\text{th}}$  percentile by Hadlock that were considered normal by the GROW standard did not have more severe composite morbidity than fetuses at baseline risk. These results are summarized in Table 6.

### Patterns of FGR diagnosis and perinatal morbidity in population sub-groups

In our sub-analysis of FGR and morbidity rates among sub-groups defined by BMI, race/ethnicity, and fetal sex, both composite and severe composite morbidity increased with increasing BMI class ( $p<0.001$  for both comparisons). The rate of FGR using the customized



standard followed a similar pattern, with the rates of FGR with increasing BMI class ( $p=0.003$ ). The Hadlock population standard did not follow this pattern and instead had the highest FGR rate among low BMI women ( $p=0.08$ , Figure 3).

Among racial/ethnic groups, there was no difference in either composite or severe composite morbidity across groups ( $p=0.4$  for both comparisons). Using the customized standard, rates of FGR were also similar across groups ( $p=0.4$ ). Using the Hadlock population standard, however, black, Hispanic, and “other” groups had significantly higher rates of FGR compared to white women ( $p<0.001$ , Figure 4). The Hadlock standard diagnosed FGR at more than twice the rate of the customized standard among non-white groups (7.0 vs 3.5%,  $p<0.0001$ ), though neither composite nor severe composite morbidity among non-white groups was higher than among white women ( $p=0.2$  for both comparisons).

Both composite and severe composite morbidity were significantly higher in pregnancies with male compared with female fetal sex ( $p=0.03$  for both comparisons). The rate of FGR by the customized standard was similar for both fetal sexes ( $p=0.9$ ), whereas the rate of FGR by the Hadlock population standard among female fetuses was almost twice that among male fetuses ( $p<0.001$ , Figure 5). Put another way, female fetuses accounted for 77% of FGR diagnoses by the population standard while comprising only 45% of composite morbidity and 41% of severe composite morbidity cases. Using the customized standard, they accounted for 49% of FGR diagnoses.

## Comment:

### Summary of findings

We evaluated and compared the performance of the leading ultrasound-derived population fetal growth standard in the United States with a well-developed customized fetal growth standard for the prediction of perinatal morbidity.<sup>15,16</sup> When applied to fetal ultrasonographic measurements taken at 22-29 weeks, neither method performed well to predict composite morbidity in our cohort, though the customized method performed slightly better than the Hadlock population-based method to predict severe morbidity. We also found that the population standard diagnosed FGR at a rate nearly 60% higher than the customized standard, but that the rate of morbidity among fetuses considered FGR by the population standard but normal by the customized standard was similar to fetuses with normal growth according to both standards.

In the sub-analysis of fetuses with EFW  $<5^{\text{th}}$  or  $5^{\text{th}}-10^{\text{th}}$  percentile, EFW  $<5^{\text{th}}$  percentile by either standard had a strong association with both composite and severe composite morbidity. In contrast, while an EFW at the  $5^{\text{th}}-10^{\text{th}}$  percentile using Hadlock was not associated with morbidity, severe morbidity occurred more often among those with EFW at  $5^{\text{th}}-10^{\text{th}}$  percentile using the customized standard than in those with normal fetal growth. Finally, rates of FGR by the customized standard followed a similar pattern as the rates of morbidity among sub-groups of race/ethnicity, BMI class, and fetal sex, whereas the population standard tended to diagnose FGR at the highest rates in the groups with lowest morbidity and at the lowest rates in groups with higher morbidity.

## Comparison with existing data

While customized fetal growth standards have been tested in a variety of populations, the majority of studies focused on customization of birth weight assessment only and did not address the more relevant question of how these models correlate with risk stratification in ongoing pregnancies.<sup>7-10,12,13,23-29</sup> Three studies have addressed this in a more limited fashion.<sup>30-32</sup> The first tested the GROW customized standard for prediction of neonatal anthropometric measurements consistent with malnutrition in 258 fetuses.<sup>32</sup> They found the customized model to be moderately useful, but did not compare the model with a population standard and did not assess neonatal morbidity or mortality. The second applied the GROW customized model to 109 pregnancies diagnosed with FGR by a population standard.<sup>30</sup> They found that if the customized model had been used on prenatal ultrasounds, the rate of FGR would have been reduced by more than half without adversely affecting the prediction of NICU admission, especially among non-white women. Our finding that the population model diagnosed FGR among non-white groups at more than twice the rate of FGR among white women is consistent with these. The third study, which assessed a variety of population and customized fetal growth standards to predict adverse outcomes in over 3,000 African-American women, found that all methods had similar performance in prediction of adverse outcomes.<sup>31</sup> Though the study found that the rate of FGR was different using each standard, the authors did not assess the frequency of morbidity in fetuses diagnosed with FGR by more inclusive standards but considered normal by more stringent standards. Our analysis extends these findings with a large cohort, detailed analysis of neonatal outcomes, and application to ultrasounds performed at a standardized time earlier in pregnancy.

Previous comparisons of customized and population standards, applied to birth weights, have compared outcomes among newborns designated as small by one standard and normal by the other (“SGA by population standard only” or “SGA by customized standard only”). In contrast, we compared outcomes of fetuses with FGR by whichever standard was more restrictive (identified as the customized standard by our analysis) with those of fetuses the more restrictive standard labeled as normal but the more generous (population) standard diagnosed with FGR. We arranged the comparisons this way to address the practical question facing clinicians of whether additional FGR cases diagnosed by a more inclusive standard over the more restrictive standard are at elevated or baseline risk of morbidity. In other words, can the standard with a lower FGR rate and therefore a lower surveillance and follow-up burden be safely adopted without adversely affecting identification of fetuses at risk of morbidity? In our analysis, cases of FGR diagnosed by the population standard that would not have been diagnosed by the customized standard did not experience higher rates of morbidity.

## Strengths and Weaknesses

Our study had multiple strengths. It made use of a large, prospectively-collected data set. The blinding of clinicians to data from study ultrasounds and standardized data collection methods all served to minimize bias. Detailed collection of self-described maternal racial and ethnic data enhanced our ability to identify the optimal customization approach.



Also, recalculation of EFWs and EFW population percentiles from individual biometric parameters standardized EFW percentile assignment from data across study sites.

Our study also had limitations. Because only one EFW was measured for each study participant between 22-29 weeks with a long mean interval until delivery, our findings can only be applied to ultrasound measurements from this time period and do not allow for assessment of performance of serial EFWs, EFWs in the late third trimester, or EFWs performed close to delivery. Additionally, the inability to stratify using fetal Doppler examinations means that our stratification of morbidity risk was based only on fetal size. This may underestimate the performance of ultrasound to identify at-risk fetuses and does not reflect contemporary practice in the U.S.. The disclosure of fetuses with EFW <5<sup>th</sup> percentile (by the standards in local use at that time) mean we cannot rule out bias from altered clinical management. The four racial/ethnic designations used by the GROW customized model to generate coefficients for the U.S. population are inherently limited as they do not allow for overlap between race and ethnicity and lack data for large portions of the U.S. population, such as Asians. Despite this, using the most specific country-of-origin designations did not improve performance of the customized model. One of the parameters accounted for in the GROW model is parity, and because our cohort was exclusively nulliparous, the full effect of customization may not have been realized. Finally, our analysis of a U.S. study population and a U.S. population-derived fetal growth standard, limits generalizability to non-United States populations.

### Clinical implications and future directions

Though neither standard had good performance to predict morbidity, it is telling that the population standard diagnosed FGR at a significantly higher rate among sub-groups that did not experience higher rates of morbidity. In fact, in groupings by BMI and fetal sex, the Hadlock standard made the highest rate of FGR diagnoses among the groups at lowest risk of morbidity. In contrast, the rates of FGR using the customized standard were similar across fetal sex and racial/ethnic groups and followed the pattern of morbidity in BMI groups, which increased with increasing BMI class. In an era when health disparities are increasingly a focus of research and quality improvement efforts, our study results show that the population-based fetal growth standard in current use diagnoses FGR at a disproportionately high rate among female fetuses and racial/ethnic minorities.

There has been recent controversy about whether additional study into fetal growth standards is worthwhile.<sup>33,34</sup> Inherent limitations of ultrasound and the many causes for morbidity limit the utility of isolated fetal size assessment for accurate risk stratification. Indeed, when fetal size is used in isolation, as in our analysis, the prediction of morbidity and mortality is poor. These findings must be taken in context, however, since several characteristics of the nuMoM2b data set may underestimate the performance of ultrasound fetal growth assessment. The lack of umbilical artery Doppler assessments and the early gestational period when EFWs were measured likely precluded identification of most late-onset FGR. Nonetheless, our data do not support the routine use of ultrasound for fetal growth surveillance at 22-29 weeks.

A variety of published norms and conceptual frameworks are available for clinicians to choose from, and they do not perform equally well.<sup>14,19</sup> If fetal size is to be one component of a multi-modal risk prediction model, identifying the best available ultrasound standard for use in such a model is worthwhile. Finally, because adopting an alternative fetal growth standard is usually an approachable adjustment to office workflow that can have an immediate impact on clinical care, identifying the optimal standard for a given population represents an attractive target for improving care. While our data suggest that customized standards may be superior to the population standard, the improvement in morbidity prediction is marginal. Therefore, customized standards need to be assessed in the context of sonograms performed later in pregnancy and in concert with Doppler studies. Only if they demonstrate superior performance in these additional contexts would their use justify the additional effort required for their implementation.

In summary, we found that when used at 22-29 weeks, both the customized and population fetal growth standards performed poorly to identify fetuses at risk of perinatal morbidity. When comparing the two methods, the customized standard had marginally better performance than the population standard to stratify severe perinatal morbidity risk, while diagnosing fewer fetuses with FGR and better reflecting morbidity patterns in population sub-groups. While these findings need to be tested prospectively and replicated using ultrasounds performed later in the 3<sup>rd</sup> trimester, they suggest that application of customized models have potential to improve the value of care by diminishing false positive results and redirecting FGR diagnoses to population groups at higher risk of morbidity.

## Acknowledgments:

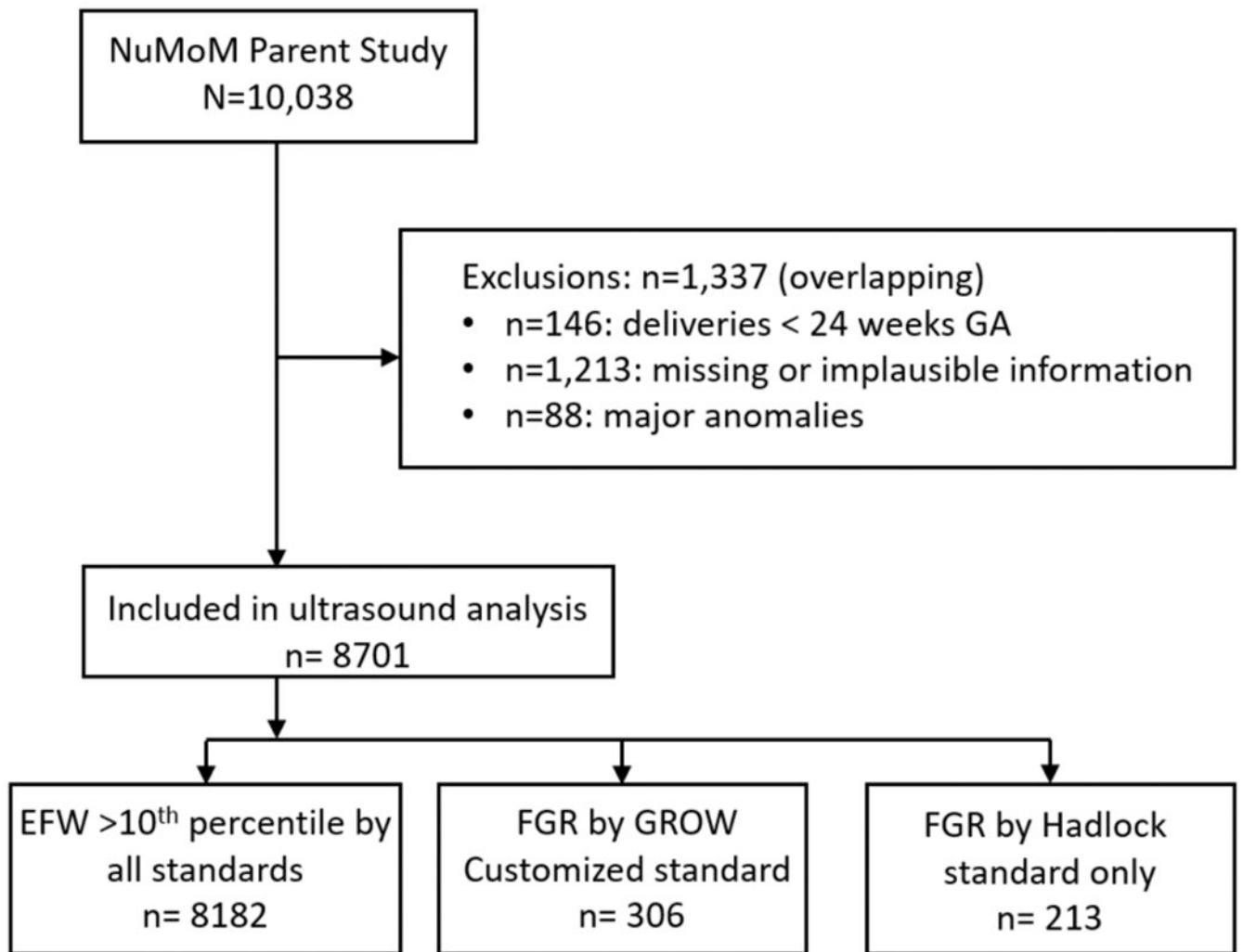
This work was funded by the following NICHD grant awards: U10 HD063020, U10 HD063037, U10 HD063041, U10 HD063046, U10 HD063047, U10 HD063048, U10 HD063053, U10 HD063072, 1K12 HD085816.

## References

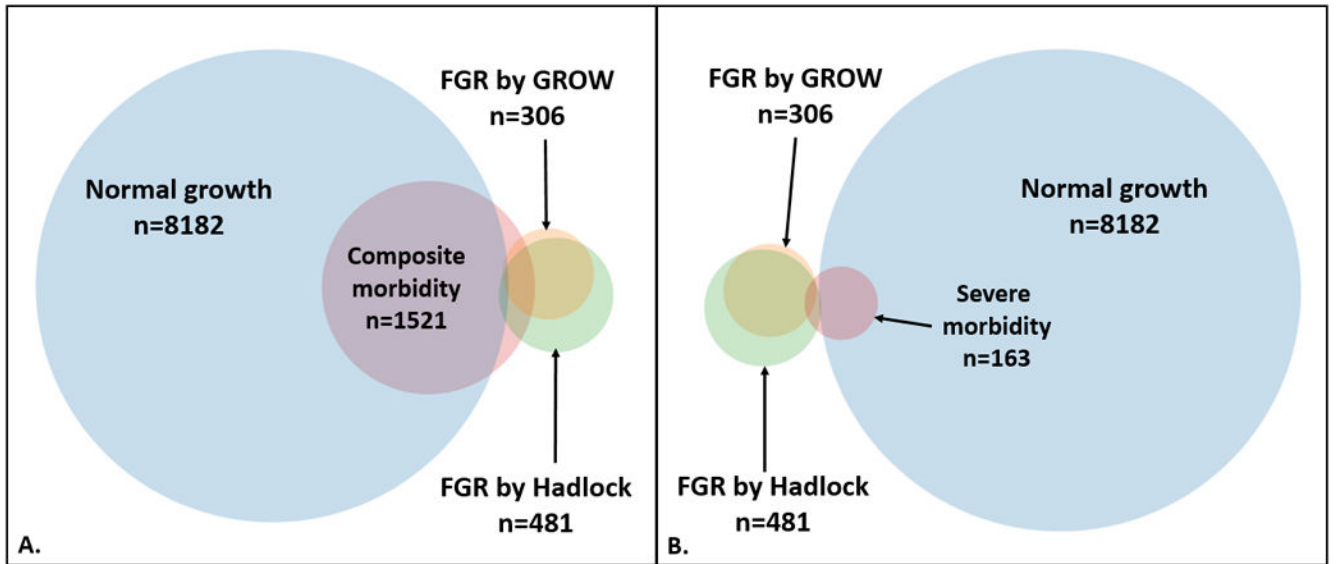
1. Crispi F, Miranda J, Gratacos E. Long-term cardiovascular consequences of fetal growth restriction: biology, clinical implications, and opportunities for prevention of adult disease. *Am J Obstet Gynecol*2018;218:S869–S79. [PubMed: 29422215]
2. McIntire DD, Bloom SL, Casey BM, Leveno KJ. Birth weight in relation to morbidity and mortality among newborn infants. *N Engl J Med*1999;340:1234–8. [PubMed: 10210706]
3. Fetal Growth Restriction. Practice Bulletin No. 204. American College of Obstetricians and Gynecologists. *Obstet Gynecol*2019;133:e97–e109. [PubMed: 30681542]
4. Alfirevic Z, Stampalija T, Dowswell T. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. *Cochrane Database Syst Rev*2017;6:CD007529. [PubMed: 28613398]
5. Gardosi J, Francis A. A customized standard to assess fetal growth in a US population. *2009;201:25.e1–e7.*
6. Carberry AE, Raynes-Greenow CH, Turner RM, Jeffery HE. Customized versus population-based birth weight charts for the detection of neonatal growth and perinatal morbidity in a cross-sectional study of term neonates. *Am J Epidemiol*2013;178:1301–8. [PubMed: 23966560]
7. Costantine MM, Lai Y, Bloom SL, et al. Population versus customized fetal growth norms and adverse outcomes in an intrapartum cohort. *Am J Perinatol*2013;30:335–41. [PubMed: 22893556]
8. Francis A, Hugh O, Gardosi J. Customized vs INTERGROWTH-21(st) standards for the assessment of birthweight and stillbirth risk at term. *Am J Obstet Gynecol*2018;218:S692–S9. [PubMed: 29422208]

9. Gardosi J, Clausson B, Francis A. The value of customised centiles in assessing perinatal mortality risk associated with parity and maternal size. *BJOG*2009;116:1356–63. [PubMed: 19538413]
10. Gardosi J, Francis A. Adverse pregnancy outcome and association with small for gestational age birthweight by customized and population-based percentiles. *Am J Obstet Gynecol*2009;201:28 e1–8. [PubMed: 19576372]
11. Hutcheon JA, Zhang X, Cnattingius S, Kramer MS, Platt RW. Customised birthweight percentiles: does adjusting for maternal characteristics matter? *BJOG*2008;115:1397–404. [PubMed: 18823489]
12. Iliodromiti S, Mackay DF, Smith GC, et al. Customised and Noncustomised Birth Weight Centiles and Prediction of Stillbirth and Infant Mortality and Morbidity: A Cohort Study of 979,912 Term Singleton Pregnancies in Scotland. *PLoS Med*2017;14:e1002228. [PubMed: 28141865]
13. Sovio U, Smith GCS. The effect of customization and use of a fetal growth standard on the association between birthweight percentile and adverse perinatal outcome. *Am J Obstet Gynecol*2018;218:S738–S44. [PubMed: 29199029]
14. Blue NR, Beddow ME, Savabi M, Katukuri VR, Chao CR. Comparing the Hadlock fetal growth standard to the Eunice Kennedy Shriver National Institute of Child Health and Human Development racial/ethnic standard for the prediction of neonatal morbidity and small for gestational age. *American Journal of Obstetrics and Gynecology*2018;219:474.e1–e12. [PubMed: 30118689]
15. Gardosi J, Francis A, Williams M, Hugh O, Loi S. Customized Centile Calculator GROW. 8.0.2 ed: Gestation network; 2018.
16. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology*1991;181:129–33. [PubMed: 1887021]
17. Haas DM, Parker CB, Wing DA, et al. A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). *Am J Obstet Gynecol*2015;212:539 e1–e24. [PubMed: 25648779]
18. 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. *Am J Obstet Gynecol*1985;151:333–7. [PubMed: 3881966]
19. Blue NR, Savabi M, Beddow ME, et al. The Hadlock Method Is Superior to Newer Methods for the Prediction of the Birth Weight Percentile. *Journal of Ultrasound in Medicine*2019;38:587–96. [PubMed: 30244476]
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*1988;44:837–45. [PubMed: 3203132]
21. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med*1997;16:2529–42. [PubMed: 9403954]
22. NCSS 12 Statistical Software. 2018.
23. Anderson NH, Sadler LC, McKinlay CJD, McCowan LME. INTERGROWTH-21st vs customized birthweight standards for identification of perinatal mortality and morbidity. *Am J Obstet Gynecol*2016;214:509 e1–e7. [PubMed: 26546850]
24. Chioffi G, Pedroza C, Costantine MM, Truong VTT, Gargano G, Saade GR. Customized vs population-based growth charts to identify neonates at risk of adverse outcome: systematic review and Bayesian meta-analysis of observational studies. *Ultrasound in Obstetrics & Gynecology*2017;50:156–66. [PubMed: 27935148]
25. Groom KM, Poppe KK, North RA, McCowan LM. Small-for-gestational-age infants classified by customized or population birthweight centiles: impact of gestational age at delivery. *Am J Obstet Gynecol*2007;197:239 e1–5. [PubMed: 17826403]
26. Larkin JC, Hill LM, Speer PD, Simhan HN. Risk of morbid perinatal outcomes in small-for-gestational-age pregnancies: customized compared with conventional standards of fetal growth. *Obstet Gynecol*2012;119:21–7. [PubMed: 22183207]
27. McCowan LM, Harding JE, Stewart AW. Customized birthweight centiles predict SGA pregnancies with perinatal morbidity. *BJOG*2005;112:1026–33. [PubMed: 16045513]

28. Moussa HN, Wu ZH, Han Y, et al. Customized versus Population Fetal Growth Norms and Adverse Outcomes Associated with Small for Gestational Age Infants in a High-Risk Cohort. *Am J Perinatol* 2015;32:621–6. [PubMed: 25262454]
29. Zhang X, Platt RW, Cnattingius S, Joseph KS, Kramer MS. The use of customised versus population-based birthweight standards in predicting perinatal mortality. *BJOG* 2007;114:474–7. [PubMed: 17378820]
30. Dua A, Schram C. An investigation into the applicability of customised charts for the assessment of fetal growth in antenatal population at Blackburn, Lancashire, UK. *J Obstet Gynaecol* 2006;26:411–3. [PubMed: 16846865]
31. Kabiri D, Romero R, Gudicha DW, et al. Prediction of adverse perinatal outcomes by fetal biometry: a comparison of customized and population-based standards. *Ultrasound in Obstetrics & Gynecology* 2019.
32. Owen P, Ogah J, Bachmann LM, Khan KS. Prediction of intrauterine growth restriction with customised estimated fetal weight centiles. *BJOG: An International Journal of Obstetrics and Gynaecology* 2003;110:411–5. [PubMed: 12699804]
33. Ganzevoort W, Thilaganathan B, Baschat A, Gordijn SJ. Point. *Am J Obstet Gynecol* 2019;220:74–82.
34. Gardosi J. Counterpoint. *Am J Obstet Gynecol* 2019;220:74–82.



**Figure 1: Study inclusion flow diagram**  
GA, gestational age.



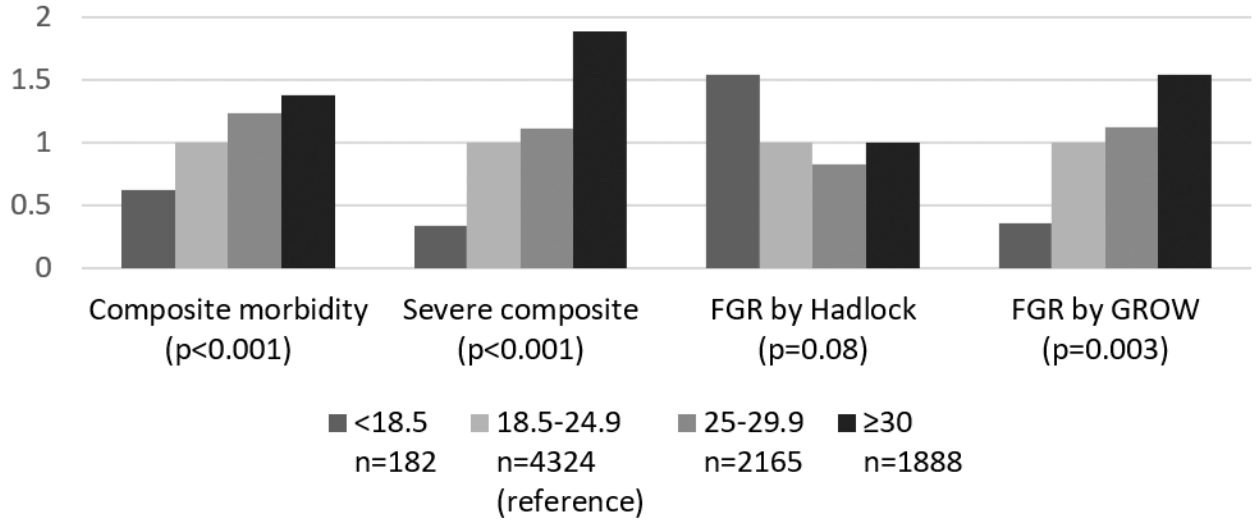
**Figure 2: Overlap and discordance between population and customized ultrasound growth standards and perinatal morbidity.**

Panel A: Growth status and overlap with composite morbidity; Panel B: growth status and overlap with severe morbidity. The circles and areas of overlap are proportional to their respective groups' sizes and degrees to which they overlap, except where FGR and normal growth groups overlap, which is an artifact required to maintain each groups' circular shape. Fetuses in the "normal growth" category are those with estimated fetal weights > 10<sup>th</sup> percentile by both population and customized standards.

FGR, fetal growth restriction; GROW, gestation-related optimal weight.



# Relative risk of morbidity and FGR diagnosis by BMI category

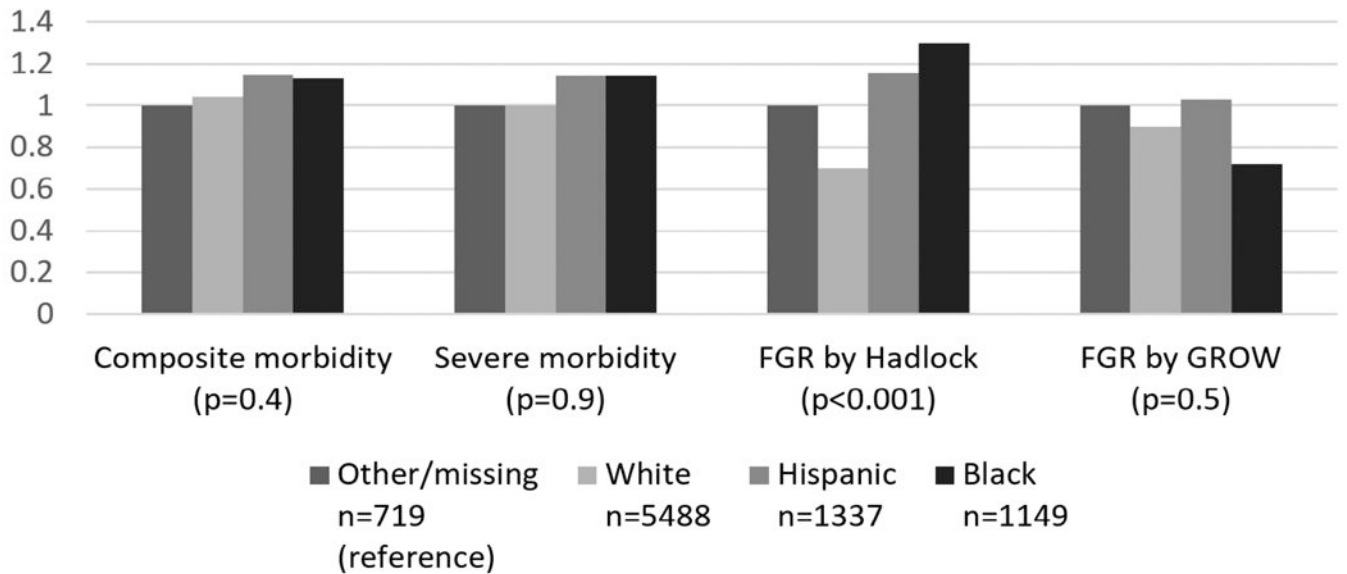


**Figure 3. Relative risk of composite morbidity and FGR by BMI category.**

This figure demonstrates how the relative risk increases with increasing BMI category for all variables except for FGR by Hadlock. P values are for the distribution of proportions across groups (chi-square). Relative risks by ascending BMI category are as follows: for composite morbidity: 0.63, 1 (reference), 1.2, 1.4; for severe composite morbidity: 0.33, 1 (reference), 1.1, 1.9; for FGR by Hadlock: 1.5, 1 (reference), 0.8, 1.0; FGR by GROW: 0.35, 1 (reference); 1.1, 1.5.

BMI, body mass index; FGR, fetal growth restriction; GROW, gestation-related optimal weight.

## Relative risk of composite morbidity and FGR by race/ethnicity



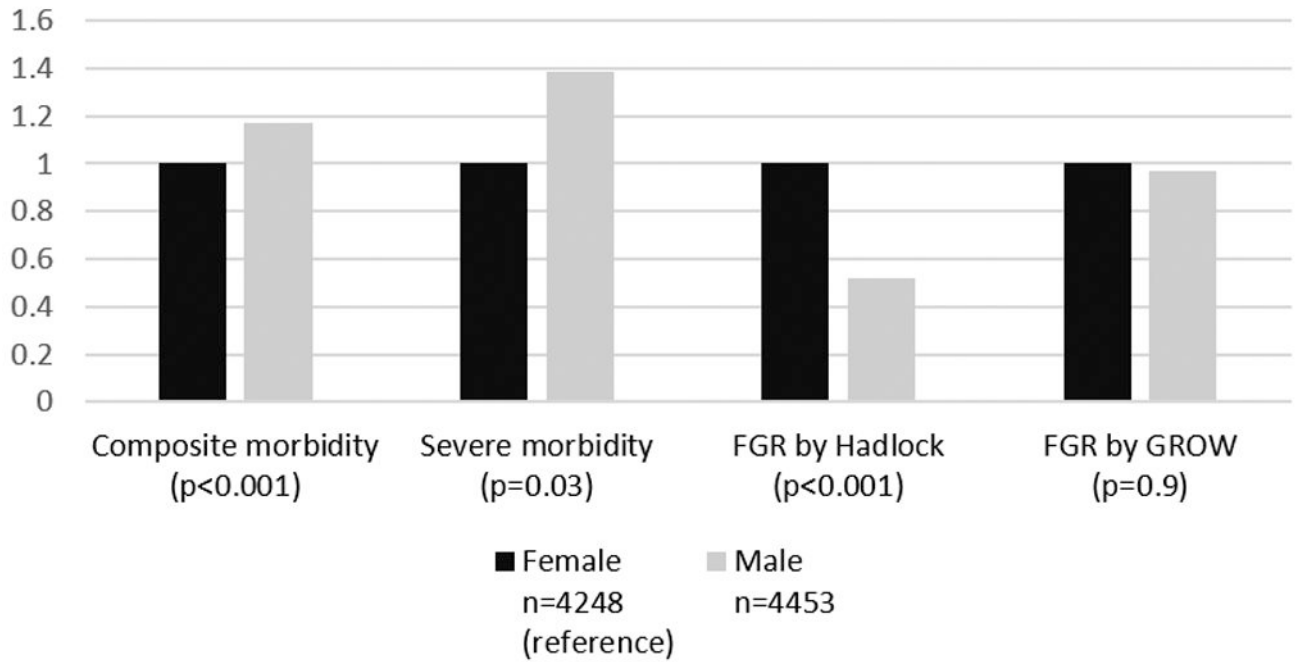
**Figure 4. Relative risk of composite morbidity and FGR diagnosis by race/ethnicity**

This figure demonstrates how the relative risks are similar among race/ethnicity categories for all variables except for FGR by Hadlock. P values are for the distribution of proportions across groups (chi-square). The “other/missing” group is chosen as the reference group because it has the lowest rate of composite morbidity. Relative risks for “other/missin”, White, Hispanic, and Black groups, respectively, are as follows: for composite morbidity: 1, 1.0, 1.1, 1.1; for severe composite morbidity: 1, 1.0, 1.1, 1.1; for FGR by Hadlock: 1, 0.7, 1.2, 1.3; FGR by GROW: 1, 0.9, 1.0, 0.7. For FGR by Hadlock,  $p < 0.03$  for all pairwise comparisons of Hispanic, Black, and other/missing vs White.

FGR, fetal growth restriction; GROW, gestation-related optimal weight.

The number of Native Americans was too small ( $n=8$ ) for sub-analysis and so are not included in this table.

## Relative risk of morbidity and FGR diagnosis by neonatal sex



**Figure 5. Relative risk of composite morbidity and FGR diagnosis by neonatal sex**

This figure demonstrates how the relative risks of morbidity are higher for male than female fetuses, lower for FGR by Hadlock, and similar for FGR by GROW. P values are for the distribution of proportions between groups (chi-square). The female group is chosen as the reference group because it has the lowest rate of composite morbidity. Relative risks for female and male groups, respectively, are as follows: 1, 1.2; for severe composite morbidity: 1, 1.4; for FGR by Hadlock: 1, 0.5; FGR by GROW: 1, 1.0.

FGR, fetal growth restriction; GROW, gestation-related optimal weight.

**Table 1:**

Study population characteristics by growth status

	Overall cohort N=8701	EFW >10% by both standards n=8182	FGR by GROW n=306	FGR by Hadlock only n=213 <sup>a</sup>	p <sup>b</sup>
Maternal age (years)	27.05 ±5.6	27.1 ±5.6	26.2 ±5.2	24.7 ±5.8	<0.001
BMI					
< 18.5	182 (2.1)	166 (2.1)	2 (0.8)	14 (6.7)	
18.5-24.9	4324 (50.0)	4067 (50.5)	133 (43.5)	124 (59.1)	<0.001
25-29.9	2165 (24.9)	2050 (25.5)	76 (24.8)	39 (18.6)	
≥30.0	1888 (21.7)	1765 (21.9)	90 (29.4)	33 (15.7)	
Race/Ethnicity					
Non-Hispanic white	5488 (63.1)	5211 (63.7)	193 (63.1)	84 (39.4)	
Hispanic	1337 (15.4)	1235 (15.1)	53 (17.3)	49 (23.0)	<0.001
Non-Hispanic Black	1149 (13.2)	1055 (12.9)	32 (10.5)	62 (29.1)	
Native American	8 (0.1)	8 (0.1)	0 (0)	0 (0)	
Other/missing	719 (8.3)	673 (8.2)	28 (9.2)	18 (8.5)	
Insurance provider <sup>c</sup>					
Public	2323 (26.9)	2135 (26.3)	97 (32.1)	91 (43.1)	
Military	59 (0.7)	56 (0.7)	2 (0.7)	1 (0.5)	<0.001
Commercial	6041 (70.0)	5736 (70.6)	197 (65.2)	108 (51.2)	
Uninsured	1547 (17.9)	1484 (18.3)	39 (12.9)	24 (11.4)	
Other	117 (1.4)	104 (1.3)	6 (2.0)	7 (3.3)	
Poverty <sup>d</sup>					
<100% FPL	1072 (15.0)	977 (14.5)	56 (23.0)	39 (25.8)	<0.001
100-200% FPL	1005 (14.1)	942 (13.9)	35 (14.4)	28 (18.5)	
>200% FPL	5071 (70.9)	4835 (71.6)	152 (62.6)	84 (55.6)	
Diabetes					
Pre-gestational DM	125 (1.4)	114 (1.4)	6 (1.9)	5 (2.4)	0.5
Gestational DM	191 (2.2)	180 (2.2)	5 (1.9)	6 (2.8)	
Chronic hypertension	211 (2.4)	192 (2.3)	9 (2.9)	10 (4.7)	0.04
Tobacco use	361 (4.1)	357 (4.4)	25 (8.2)	16 (7.5)	0.001
Alcohol use <sup>e</sup>	965 (11.1)	463 (5.7)	18 (5.9)	8 (3.8)	0.5
Female fetal sex	4248 (48.8)	3932 (48.1)	151 (49.3)	165 (77.4)	<0.001

EFW, estimated fetal weight; FGR, fetal growth restriction; GROW, gestation-related optimal weight; BMI, body mass index; FPL, federal poverty level; DM, diabetes mellitus. Data are expressed as either n(%) or mean ± standard deviation.

<sup>a</sup>Number of fetuses in FGR by Hadlock only group in this table reflects the number of fetuses diagnosed with FGR by the Hadlock standard but that were considered normal by the customized standard. It differs from the corresponding column in Table 2, which is a total count of fetuses diagnosed with FGR using the Hadlock standard.

<sup>b</sup>P values are for comparisons of values between growth status groups and do not include the “overall cohort” column in the comparison, as indicated by box groupings.

<sup>c</sup>Insurance information available for n=8641.

<sup>d</sup>Household income information available for n=7148.

<sup>e</sup>No women reported using opioid replacement, amphetamine, or cocaine, and only 2 reported heroin use.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Rates of fetal growth restriction according to fetal growth standard

<b>N=8701</b>	<b>GROW customized standard</b>	<b>Hadlock population standard</b>	<b>p</b>
FGR (EFW < 10 <sup>th</sup> percentile), n (%)	306 (3.5)	481 (5.5)	< 0.001
EFW < 5 <sup>th</sup> percentile	121 (1.4)	183 (2.1)	< 0.001
EFW 5-10 <sup>th</sup> percentile	185 (2.1)	298 (3.4)	< 0.001

GROW, gestation-related optimal weight; FGR, fetal growth restriction; EFW, estimated fetal weight.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 3:**

Perinatal morbidity and mortality rates according to fetal growth status by customized or population standards.

	EFW >10 <sup>th</sup> percentile by all standards N=8182 <sup>a</sup>	FGR by GROW customized standard n=306		FGR by Hadlock standard only <sup>b</sup> n=213	
	n(%)	n(%)	RR (95% CI) <sup>c</sup>	n(%)	RR (95% CI) <sup>c</sup>
Composite morbidity	1521 (18.6)	86 (28.1)	1.5 (1.3-1.8)	47 (22.1)	1.2 (0.9-1.5)
Severe morbidity	163 (1.99)	25 (8.2)	4.1 (2.7-6.1)	3 (1.4)	0.7 (0.2-2.1)
GA at delivery	39.3 ±1.7	38.8 ±2.8	p = 0.002	39.2 ±1.8	p = 0.4
Preterm birth (all)	582 (7.1)	52 (17.0)	2.4 (1.8-3.1)	20 (9.4)	1.3 (0.9-2.0)
Preterm birth (indicated)	348 (4.3)	20 (6.5)	1.5 (1.0-2.4)	6 (2.8)	0.7 (0.3-1.4)
BW (g)	3231 ±508	2800 ±701	p < 0.001 <sup>c</sup>	2824 ±483	p < 0.001 <sup>c</sup>
Delivery room resusc.					
Bag mask ventilation	364 (4.5)	22 (7.2)	1.6 (1.1-2.4)	12 (5.6)	1.3 (0.7-2.2)
Intubation	143 (1.7)	7 (2.3)	1.3 (0.6-2.7)	3 (1.4)	0.8 (0.3-2.4)
Chest compressions	19 (0.2)	7 (2.3)	9.9 (4.3-22.6)	0	--
Medications	4 (0.05)	7 (2.3)	46.8 (14.7-148.7)	0	--
NICU admission	1076 (13.2)	74 (24.2)	1.8 (1.5-2.3)	37 (17.4)	1.3 (1.0-1.8)
Level II nursery <sup>d</sup>	228 (2.8)	8 (2.6)	0.9 (0.5-1.8)	6 (2.8)	1.0 (0.5-2.2)
Meconium aspiration syndrome	48 (0.6)	1 (0.3)	0.6 (0.1-3.2)	0	--
Sepsis	22 (0.3)	2 (0.7)	2.4 (0.6-9.2)	0	--
High-frequency ventilation	23 (0.3)	3 (1.0)	3.5 (1.1-10.8)	0	--
PPHN	11 (0.1)	1 (0.3)	2.4 (0.4-14.5)	0	--
NEC	6 (0.07)	1 (0.3)	4.5 (0.7-28.0)	0	--
Seizures	18 (0.2)	1 (0.3)	1.5 (0.3-8.7)	0	--
IVH grade III/IV	3 (0.04)	1 (0.3)	8.9 (1.3-61.9)	0	--
PVL	4 (0.05)	1 (0.3)	6.7 (1.0-44.2)	0	--
NICU stay >30d	61 (0.7)	17 (5.6)	7.5 (4.4-12.5)	3 (1.4)	1.9 (0.6-5.6)
Stillbirth	2 (0.02)	1 (0.3)	13.4 (1.8-101.6)	0	--
Death	4 (0.05)	1 (0.3)	6.7 (1.0-44.2)	0	--

EFW, estimated fetal weight; FGR, fetal growth restriction; GROW, gestation-related optimal weight; RR, relative risk; GA, gestational age; BW, birth weight; NICU, neonatal intensive care unit; PPHN, persistent pulmonary hypertension of the newborn; NEC, necrotizing enterocolitis; IVH, intraventricular hemorrhage; PVL, periventricular leukomalacia.

Measures of statistical significance are reported as relative risks for categorical variables and p values for continuous variables.

<sup>a</sup>Referent group

<sup>b</sup>Number of fetuses in FGR by Hadlock only group differs from that in Table 2 to reflect the number of fetuses diagnosed with FGR by the Hadlock standard but that were considered normal by the customized standard.

<sup>c</sup>Compared to the referent group

<sup>d</sup>highest level of care

**Table 4:**

Overall performance of population vs customized fetal growth standards to predict perinatal morbidity

Entire cohort, N=8701	GROW customized standard AUC	Hadlock population standard AUC	95% CI for the difference	p
Composite morbidity	0.50 <sup>a</sup>	0.50 <sup>a</sup>	-0.005, 0.007	0.6
Severe morbidity	0.56 <sup>b</sup>	0.54 <sup>c</sup>	0.01, 0.03	0.003
Deliveries occurring < 60 days from EFW, n=661				
Composite morbidity	0.59 <sup>b</sup>	0.58 <sup>b</sup>	-0.01, 0.03	0.2
Severe morbidity	0.65 <sup>b</sup>	0.62 <sup>b</sup>	0.01, 0.05	0.002

GROW, gestation-related optimal weight; AUC, area under the curve; CI, confidence interval.

<sup>a</sup> p is non-significant when compared against the null of AUC = 0.5<sup>b</sup> p<0.01 when compared against the null of AUC = 0.5<sup>c</sup> p=0.09 when compared against the null of AUC = 0.5

**Table 5.**Morbidity according to FGR <5<sup>th</sup> percentile by fetal growth standard

	EFW >5 <sup>th</sup> by both n=8500	EFW <5 <sup>th</sup> by GROW n=121	EFW <5 <sup>th</sup> by Hadlock only, n=80 <sup>a</sup>	P
Composite morbidity	1588 (18.7)	40 (33.1) <sup>b</sup>	26 (32.5) <sup>b</sup>	<0.001
Severe morbidity	172 (2.0)	13 (10.7) <sup>b</sup>	6 (7.5) <sup>b</sup>	<0.001 <sup>c</sup>

FGR, fetal growth restriction; EFW, estimated fetal weight; GROW, gestation-related optimal weight.

P &gt;0.05 for pairwise comparisons except where otherwise specified.

<sup>a</sup>This includes 59 fetuses that are otherwise excluded from “EFW >10<sup>th</sup> by Hadlock only” groups because their EFWs are <10<sup>th</sup> by GROW. They are included here, however, because the reference group in this stratus is “EFW >5<sup>th</sup> percentile by both”.<sup>b</sup>p <0.01 when compared to EFW >5<sup>th</sup> percentile by both standards<sup>c</sup>Fisher-Freeman-Halton exact test

**Table 6.**Morbidity according to FGR between the 5-10<sup>th</sup> percentiles by fetal growth standard

	EFW >10 <sup>th</sup> by both, n=8182	EFW 5-10 <sup>th</sup> by GROW n=185	EFW 5-10 <sup>th</sup> by Hadlock only, n=192	P
Composite morbidity	1521 (18.6)	46 (24.9)	39 (20.3)	0.08
Severe morbidity	163 (2.0)	12 (6.4) <sup>a</sup>	2 (1.0)	0.001 <sup>b</sup>

FGR, fetal growth restriction; EFW, estimated fetal weight; GROW, gestation-related optimal weight.

P &gt;0.05 for pairwise comparisons except where otherwise specified.

<sup>a</sup>Fisher-Freeman-Halton exact test<sup>b</sup>p <0.001 when compared to EFW >10<sup>th</sup> percentile by both standards