Practice of Epidemiology

Assessing the Value of Customized Birth Weight Percentiles

Jennifer A. Hutcheon*, Mark Walker, and Robert W. Platt

* Correspondence to Dr. Jennifer Hutcheon, Department of Obstetrics and Gynaecology, BC Women’s Hospital and Health Centre, Room 2H30, 4500 Oak Street, Vancouver, British Columbia, Canada V6H 3N1 (e-mail: jhutcheon@cfri.ca).

Initially submitted June 10, 2010; accepted for publication October 21, 2010.

Customized birth weight percentiles are weight-for-gestational-age percentiles that account for the influence of maternal characteristics on fetal growth. Although intuitively appealing, the incremental value they provide in the identification of intrauterine growth restriction (IUGR) over conventional birth weight percentiles is controversial. The objective of this study was to assess the value of customized birth weight percentiles in a simulated cohort of 100,000 infants aged 37 weeks whose IUGR status was known. A cohort of infants with a range of healthy birth weights was first simulated on the basis of the distributions of maternal/fetal characteristics observed in births at the Royal Victoria Hospital in Montreal, Canada, between 2000 and 2006. The occurrence of IUGR was re-created by reducing the observed birth weights of a small percentage of these infants. The value of customized percentiles was assessed by calculating true and false positive rates. Customizing birth weight percentiles for maternal characteristics added very little information to the identification of IUGR beyond that obtained from conventional weight-for-gestational-age percentiles (true positive rates of 61.8% and 61.1%, respectively, and false positive rates of 7.9% and 8.5%, respectively). For the process of customization to be worthwhile, maternal characteristics in the customization model were shown through simulation to require an unrealistically strong association with birth weight.

Abbreviations: IUGR, intrauterine growth restriction; SGA, small for gestational age; reference standards

The need to differentiate between infants whose growth has been pathologically restricted in utero and infants who are physiologically small is a key concern in the classification of fetal growth and the clinical management of these pregnancies. Customized birth weight percentiles were designed to improve the assessment of fetal growth by incorporating information on maternal influences on birth weight (1). By use of maternal height, ethnicity, parity, prepregnancy weight, and fetal sex, a physiologically optimal or “target” birth weight is predicted for each infant through a multivariable linear regression “customization” model. The infant’s individually predicted target birth weight is then compared with its actual birth weight, and any infant whose observed weight is below an 80% normal range around its target weight (i.e., below the 10th percentile) is classified as small for gestational age (SGA) by the customized standard (2). Customized birth weight percentiles are recommended for clinical use by practice guidelines of the British Royal College of Obstetricians and Gynaecologists (3), and an editorial in a leading American obstetric journal concluded, “… it would seem to be an appropriate time for American obstetricians to adopt the use of customized fetal growth standards” (4, p. 221).

Although intuitively appealing, evidence on the extent to which customized birth weight percentiles actually improve the identification of infants with intrauterine growth restriction (IUGR) compared with conventional population weight-for-age percentiles is inconsistent (5–10). A major limitation in evaluating customized percentiles is that population-level data on true IUGR, defined as the failure of an infant to meet its own growth potential because of placental dysfunction as manifested by clinical findings including abnormal umbilical artery Doppler studies or amniotic fluid volume reduction (11, 12), are unavailable (13). Without accurate information on whether a given infant is small as a result of IUGR or “small but healthy,” it is difficult to properly evaluate if
MATERIALS AND METHODS

Simulation of study population

Our simulated cohort was created in 2 steps. First, we simulated a cohort of “healthy” pregnancies based on maternal and fetal characteristics observed in a cohort of 9,697 healthy singleton births at the Royal Victoria Hospital, a McGill University teaching hospital in Montreal, Canada, between 2002 and 2006 contained in the McGill Obstetric and Neonatal Database. We defined “healthy” births as term live-births (37–42 weeks’ gestation) with a 5-minute Apgar score greater than 7 and no admission to the neonatal intensive care unit. We further excluded pregnancies with any noted maternal comorbidities of diabetes in pregnancy (gestational, type 1, or type 2); cardiac, pulmonary, or renal disease; or hypertension in pregnancy (chronic or pregnancy induced).

Finally, we restricted the cohort to 100,000 births born at 37 weeks to ensure the clinical relevance of our study. Prior to 37 weeks, the neonatal risks associated with preterm birth mean that iatrogenic delivery of a fetus with suspected IUGR (i.e., <10th percentile) would not be routine but would, instead, be based on additional indicators of fetal compromise, such as abnormal Doppler ultrasound findings or biophysical profile assessments. Once term gestational ages are reached at 37 weeks, however, neonatal risks are considerably reduced, and risk of stillbirth begins to rise sharply. Iatrogenic delivery of infants with suspected IUGR is therefore routinely performed once 37 weeks are reached in order to prevent stillbirth. An assessment of the ability of a tool such as customized percentiles to distinguish small-but-healthy infants from infants with true IUGR at this age is therefore extremely pertinent for clinical care. Further details on the simulation of the cohort are provided in Appendix 1.

Having simulated a population of healthy births, we then simulated the occurrence of IUGR in selected pregnancies. Our simulation of IUGR was guided by the literature and expert clinical opinion. We assumed that 1) the prevalence of IUGR is 5% (13); 2) one third of small-for-gestational-age infants (<10th population percentile) are growth restricted, while the remaining two thirds are small-but-healthy or small as the result of congenital anomalies (14, 15); 3) IUGR decreases birth weight by 10%–20% from the optimal weight, on the basis of birth weight reductions observed in pregnancies complicated by preeclampsia compared with controls (16), and the percent difference in birth weight between infants with and without an elevated head circumference/abdominal circumference ratio on ultrasound or low ponderal index in singleton births in the Royal Victoria Hospital between 2002 and 2006 (data available upon request). The lowest plausible weight of an IUGR infant (at term) was considered to be 1,500 g, because infants weighing less than this would likely either be delivered prior to 37 weeks or would not survive until term.

We therefore selected at random 5% of infants from the healthy cohort to have IUGR, and a new, “growth-restricted” birth weight was created by reducing their “healthy” birth weight by an amount drawn at random from a normal distribution with a mean reduction of 20% and a standard deviation of 10%. The distribution was truncated so that the reduction in birth weight due to IUGR was a minimum of 5%. These values were chosen after sensitivity analyses comparing different possible mean reductions in birth weight and variances demonstrated that the values produced a cohort that best met our prespecified criteria on IUGR and maximized face validity based on clinical opinion. Our assumptions on the reduction in birth weight due to IUGR were explored through sensitivity analyses in which the mean percent birth weight was decreased by up to one half (i.e., mean reduction in birth weight of 10%) and doubled (mean reduction of 40%).

Calculation of customized and population-based birth weight percentiles

Customized birth weight percentiles were calculated by using a customized standard developed for the US population (17). Briefly, the previously published coefficients for maternal height, weight, ethnicity, parity, prepregnancy weight, and fetal sex were used to predict a target birth weight at 280 days (40 weeks) for each infant. The 280-day target weight was then extrapolated backward to obtain a target weight for the infant’s age at birth by use of an ultrasound standard (18). Infants below the 10th percentile according to the customized standard were classified as small for gestational age, denoted “SGAcustomized.” A detailed description of the methodology of customized percentiles (1, 2) is provided in Appendix 2.

Conventional population birth-weight-for-gestational-age percentiles were calculated in a similar manner as the customized percentiles, but with the use of only the coefficient
for sex in the prediction of target birth weight at 280 days (i.e., all coefficients for maternal characteristics were excluded in the prediction). The proportionality curves were then used to adjust the target birth weight for the gestational age at birth of the infant. Infants below the 10th percentile for sex and gestational age were classified as “SGApopulation.”

Assessment of the value of customized percentiles

We examined the classification accuracy of the customized percentiles in distinguishing infants with IUGR from non-IUGR infants. Using the 10th percentile as the threshold for “increased risk,” we calculated the true positive rate to determine how many infants with IUGR were correctly identified as being in the high risk group (SGA) and the false positive rate to determine how many small-birth infants (infants with a birth weight in the smallest 10% of the population but with no IUGR) were incorrectly identified as being in the high risk group. The true positive rate and the false positive rate were calculated as the number of infants with IUGR classified as SGA divided by the total number of infants with IUGR (i.e., sensitivity) and the number of infants without IUGR classified as SGA divided by the number of infants without IUGR, respectively (19).

Because the inclusion of information on maternal characteristics is believed to better identify infants who are small but have reached their own growth potential, we calculated the percentage of SGA infants who were small but healthy (i.e., <10th percentile according to each of the customized or population standards but did not have IUGR). Finally, we calculated the difference between the percentile assigned by the customized standard and that of the population standard in these infants (difference = customized percentile – population percentile).

Table 1. Descriptive Characteristics for the Simulated Study Cohort of 100,000 Births

<table>
<thead>
<tr>
<th>Maternal characteristics</th>
<th>No.</th>
<th>Mean (SD)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height, cm</td>
<td>163</td>
<td>(7)</td>
<td>1.0</td>
</tr>
<tr>
<td>Prepregnancy weight, kg</td>
<td>63</td>
<td>(12)</td>
<td>7.6</td>
</tr>
<tr>
<td>Parity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>40,236</td>
<td></td>
<td>40.2</td>
</tr>
<tr>
<td>1</td>
<td>37,416</td>
<td></td>
<td>37.4</td>
</tr>
<tr>
<td>2</td>
<td>14,770</td>
<td></td>
<td>14.8</td>
</tr>
<tr>
<td>≥3</td>
<td>7,578</td>
<td></td>
<td>7.6</td>
</tr>
<tr>
<td>Ethnicity*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>63,581</td>
<td></td>
<td>63.6</td>
</tr>
<tr>
<td>Asian</td>
<td>7,199</td>
<td></td>
<td>7.2</td>
</tr>
<tr>
<td>Black</td>
<td>13,545</td>
<td></td>
<td>13.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2,484</td>
<td></td>
<td>2.5</td>
</tr>
<tr>
<td>Other</td>
<td>13,191</td>
<td></td>
<td>13.2</td>
</tr>
<tr>
<td>Fetal characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth weight, g</td>
<td>3,137</td>
<td>(398)</td>
<td>3.1</td>
</tr>
<tr>
<td>Infants with IUGR, g</td>
<td>2,496</td>
<td>(435)</td>
<td>2.5</td>
</tr>
<tr>
<td>Male sex</td>
<td>50,101</td>
<td></td>
<td>50.0</td>
</tr>
</tbody>
</table>

Abbreviations: IUGR, Intrauterine growth restriction; SD, standard deviation.

* Ethnicity groupings based on those reported in the customization model for a US population by Gardosi and Francis (17).

RESULTS

The maternal and fetal characteristics of our simulated cohort are shown in Table 1. The mean birth weight of infants simulated to have IUGR corresponded to the fifth percentile of the birth weight distribution of healthy infants aged 37 weeks in this population. The maternal characteristics used in the customization model (i.e., excluding fetal sex and gestational age) explained 10% of the variability in birth weight.

The accuracy of customized and population percentiles in identifying infants with IUGR is shown in Table 2. The true positive rate and the false positive rate for the customized and population percentiles were highly similar, with 61.1% of infants with IUGR identified by the SGApopulation classification and 61.8% of infants with IUGR identified by the SGAcustomized classification. Sensitivity analyses in which the mean percent reduction associated with birth weight was varied also showed minimal differences between the population and customized percentiles. For example, with a mean reduction in birth weight of 40%, the true positive rate of the SGApopulation classification was 95.5%, and that of the SGAcustomized classification was 95.9%. Halving the mean reduction in birth weight to 10% produced true positive rates of 35.3% for both classifications.

We then focused on infants classified as SGA by the conventional standard, but who had reached their individual growth potential (small-but-healthy infants). If customization for maternal characteristics is effective, these infants should be assigned a percentile close to the 50th customized
Table 2. Classification Accuracy in Identifying Intrauterine Growth Restriction of a Conventional Population Birth Weight Standard, an Existing Customized Birth Weight Standard, and Customized Standards With Hypothetical Maternal Characteristics Predictive of Birth Weight Added Among a Simulated Study Cohort

<table>
<thead>
<tr>
<th>Birth Weight Standard</th>
<th>True Positive Rate, %</th>
<th>False Positive Rate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGA_{population}</td>
<td>61.1</td>
<td>8.5</td>
</tr>
<tr>
<td>SGA_{customized}</td>
<td>61.8</td>
<td>7.9</td>
</tr>
</tbody>
</table>

SGA_{customized} was based on customization models with the addition of hypothetical predictors of birth weight:

- $r = 0.5$: 61.6%, 5.0%
- $r = 0.75$: 62.4%, 1.4%
- $r = 0.95$: 61.7%, 0.1%

Abbreviation: SGA, small for gestational age.

In Figure 1, we explored the target weights predicted by the customization model for small-but-healthy infants compared with infants with an optimal weight above the 10th population percentile. With a perfect customization model, 2 distinct distributions would be observed, with the target weights of infants with an optimal weight above the 10th percentile centered around the population 50th percentile. The target weights predicted for the small-but-healthy infants would be centered at a distinctly lower value (below the population 10th percentile), recognizing that their small size was expected given their mother’s characteristics. However, a target weight close to the population average was expected for most infants, regardless of their true growth potential.

Very few (<0.1%) of the small-but-healthy infants were actually predicted to have an optimal weight below the 10th percentile of healthy weights in the population. The weights of the small-but-healthy infants were systematically smaller (82 g) than those of the infants with an optimal weight above the population 10th percentile, but this difference was small in relation to the total variability in birth weight among healthy births at 40 weeks (e.g., standard deviation of 495 g between the 50th and 10th population percentiles).

Our simulation of hypothetical predictors of birth weight showed that as the ability to predict birth weight in the customization model improved, so did the extent to which customized percentiles were able to improve identification of IUGR (Table 2). In the case of a very strong predictor of birth weight (i.e., correlation with birth weight = 0.95), the false positive rate was close to zero. However, regardless of how strong a predictor was included in the customization model, there remained a proportion of infants with IUGR classified as non-SGA. As a result, the true positive rate did not increase to a meaningful degree. These infants were more likely to be female (56% vs. 46%) and with mothers who were smaller (60 kg vs. 64.6 kg), shorter (162 cm vs. 164 cm), nulliparous (47% vs. 38%), and less likely to be white (57% vs. 67%) than women with IUGR infants who were identified as SGA by the customized percentiles, suggesting that characteristics associated with a smaller target birth weight make it more difficult to distinguish between small-but-healthy infants and infants who are small because of IUGR.

As the predictive ability of the customization model increased, the customized percentiles assigned to small-but-healthy infants became meaningfully different from the percentile assigned by the conventional approach (Figure 1). With the inclusion of a hypothetical predictor correlated with birth weight at a value of $r = 0.95$, the customized percentile was over 15 percentage points larger than the percentile assigned by conventional methods for 95% of infants. The reason for this can be seen in Figure 2, B–D, where as the predictive ability of the customization model increased, the target weights predicted for small-but-healthy infants became much more distinct from the target weights predicted for non-SGA infants. In other words, the customization model became better able to recognize that the target weight

![Figure 1](https://example.com/image.png)

**Figure 1.** Difference between the percentile assigned by the customized and the population standards to “small-but-healthy” infants (birth weight below the population 10th percentile but no intrauterine growth restriction), for a simulated study cohort, where black = within 5 percentiles, hatched lines = 5.1–10 percentiles, dots = 10.1–15 percentiles, and white = more than 15 percentiles. The customized percentiles were produced by an existing customization model and customization models with the addition of a hypothetical maternal characteristic correlated with birth weight at $r = 0.5$, $r = 0.75$, and $r = 0.95$. 

**Table 2.** Classification Accuracy in Identifying Intrauterine Growth Restriction of a Conventional Population Birth Weight Standard, an Existing Customized Birth Weight Standard, and Customized Standards With Hypothetical Maternal Characteristics Predictive of Birth Weight Added Among a Simulated Study Cohort.

<table>
<thead>
<tr>
<th>Birth Weight Standard</th>
<th>True Positive Rate, %</th>
<th>False Positive Rate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGA_{population}</td>
<td>61.1</td>
<td>8.5</td>
</tr>
<tr>
<td>SGA_{customized}</td>
<td>61.8</td>
<td>7.9</td>
</tr>
</tbody>
</table>

SGA_{customized} was based on customization models with the addition of hypothetical predictors of birth weight:

- $r = 0.5$: 61.6%, 5.0%
- $r = 0.75$: 62.4%, 1.4%
- $r = 0.95$: 61.7%, 0.1%

Abbreviation: SGA, small for gestational age.

In Figure 1, we explored the target weights predicted by the customization model for small-but-healthy infants compared with infants with an optimal weight above the 10th population percentile. With a perfect customization model, 2 distinct distributions would be observed, with the target weights of infants with an optimal weight above the 10th percentile centered around the population 50th percentile. The target weights predicted for the small-but-healthy infants would be centered at a distinctly lower value (below the population 10th percentile), recognizing that their small size was expected given their mother’s characteristics. However, a target weight close to the population average was expected for most infants, regardless of their true growth potential.

Very few (<0.1%) of the small-but-healthy infants were actually predicted to have an optimal weight below the 10th percentile of healthy weights in the population. The weights of the small-but-healthy infants were systematically smaller (82 g) than those of the infants with an optimal weight above the population 10th percentile, but this difference was small in relation to the total variability in birth weight among healthy births at 40 weeks (e.g., standard deviation of 495 g between the 50th and 10th population percentiles).

Our simulation of hypothetical predictors of birth weight showed that as the ability to predict birth weight in the customization model improved, so did the extent to which customized percentiles were able to improve identification of IUGR (Table 2). In the case of a very strong predictor of birth weight (i.e., correlation with birth weight = 0.95), the false positive rate was close to zero. However, regardless of how strong a predictor was included in the customization model, there remained a proportion of infants with IUGR classified as non-SGA. As a result, the true positive rate did not increase to a meaningful degree. These infants were more likely to be female (56% vs. 46%) and with mothers who were smaller (60 kg vs. 64.6 kg), shorter (162 cm vs. 164 cm), nulliparous (47% vs. 38%), and less likely to be white (57% vs. 67%) than women with IUGR infants who were identified as SGA by the customized percentiles, suggesting that characteristics associated with a smaller target birth weight make it more difficult to distinguish between small-but-healthy infants and infants who are small because of IUGR.

As the predictive ability of the customization model increased, the customized percentiles assigned to small-but-healthy infants became meaningfully different from the percentile assigned by the conventional approach (Figure 1). With the inclusion of a hypothetical predictor correlated with birth weight at a value of $r = 0.95$, the customized percentile was over 15 percentage points larger than the percentile assigned by conventional methods for 95% of infants. The reason for this can be seen in Figure 2, B–D, where as the predictive ability of the customization model increased, the target weights predicted for small-but-healthy infants became much more distinct from the target weights predicted for non-SGA infants. In other words, the customization model became better able to recognize that the target weight
DISCUSSION

In this study, we used simulation to first create a cohort of newborn infants with healthy birth weights, given their mothers' characteristics, and then manipulated the weights of a small percentage of infants to re-create the effects of intrauterine growth restriction. With knowledge of the infants' true growth restriction status, we were able to then evaluate the extent to which customization of birth weight percentiles for maternal characteristics was able to improve identification of IUGR compared with conventional population percentiles. We found that, contrary to widely held perceptions (3, 4, 17), customized percentiles produced by existing customization models did little to improve the accuracy of identification of IUGR or to reduce the percentage of small-but-healthy infants misclassified as high risk. Only with a customization model simulated to include predictors with an unrealistically high correlation with birth weight (e.g., \( r \geq 0.75 \)) were the widely believed benefits of incorporating information on maternal characteristics in distinguishing small-but-healthy infants from infants with IUGR observed.

Because all maternal characteristics combined explain only a small percentage of the variability in birth weight (roughly 9%–10%) (10, 17), these results should not be surprising. In the prediction of target birth weights in the absence of any other information (e.g., predicting birth weight by using a null regression model), the birth weight predicted for every infant would be the population average weight. This, in essence, is the approach taken by population weight-for-gestational-age percentiles, where the target birth weight for all infants is simply assumed to be that of the sex- and gestational age-specific population 50th percentile weight (and the threshold for the 10th percentile calculated on the basis of the entire distribution). With a regression model that is only modestly better than a null regression model in explaining variability in birth weight, the customized target weights predicted for each infant (and the corresponding 10th percentile threshold) will also be only modestly different from the population sex- and age-specific 50th percentile, as observed in our work. These results also help to explain the findings of past work, which suggests that the apparent benefits of customization may be primarily attributable to their use of an intrauterine standard at preterm gestational ages rather than to any adjustment for maternal characteristics (8, 10).

Our simulation of hypothetical predictors of birth weight demonstrated that, in order for customized birth weight percentiles to be useful, the maternal characteristics must have an extremely high correlation with birth weight. It was only when characteristics had a correlation with birth weight of \( r = 0.75 \) or higher that the process of customization resulted in percentiles that were meaningfully better than the conventional percentiles in distinguishing small-but-healthy infants from infants who were small because of IUGR. This observation is compatible with work from other areas of epidemiology, which has shown that, in order for individualized risk prediction to be successful, a given marker or predictor must have an association with the outcome of interest at a degree rarely encountered in epidemiologic research (e.g., an odds ratio \( > 16.0 \)) (21). An ability to explain population-level differences in an outcome (i.e., a statistically significant predictor in a regression model) does not necessarily translate into successful individual-level prediction, for which much higher associations with the outcome are needed. Although maternal characteristics in the customization model are all significant predictors of population-level differences in birth weight, our simulations show that there are simply too many other unexplained influences on birth weight for meaningfully accurate prediction of birth weight at the individual level.

A major difficulty in the study of intrauterine growth restriction is that available population data do not allow us to distinguish between infants who have failed to reach their own growth potential and infants who are small-but-healthy. The conventional classification of SGA is inadequate for this purpose, because it inappropriately assumes that all infants in the smallest 10% of the population are growth restricted, and that no infants above the 10th percentile had their growth restricted in utero. In the absence of a diagnostic

---

**Table 3. Percentage of Small-for-Gestational-Age Infants Who Are “Small but Healthy” Versus Small From Intrauterine Growth Restriction as Classified by a Conventional Population Birth Weight Standard, an Existing Customized Birth Weight Standard, and Customized Standards With Hypothetical Maternal Characteristics Predictive of Birth Weight Added for a Simulated Study Cohort**

<table>
<thead>
<tr>
<th>Birth Weight Standard</th>
<th>SGA Infants Who Are Small but Healthy*</th>
<th>SGA Infants Who Are Small From IUGR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>SGApopulation (( n = 11,080 ))</td>
<td>8,053</td>
<td>72.7</td>
</tr>
<tr>
<td>SGAcustomized (( n = 10,604 ))</td>
<td>7,544</td>
<td>71.1</td>
</tr>
<tr>
<td>SGAcustomized based on customization models with addition of hypothetical predictors of birth weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r = 0.50 (n = 7,818) )</td>
<td>4,766</td>
<td>61</td>
</tr>
<tr>
<td>( r = 0.75 (n = 4,396) )</td>
<td>1,306</td>
<td>29.7</td>
</tr>
<tr>
<td>( r = 0.95 (n = 3,138) )</td>
<td>81</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Abbreviations: IUGR, intrauterine growth restriction; SGA, small for gestational age.

*“Small-but-healthy” infants (i.e., infants in the <10th percentile but no IUGR).
test at birth for IUGR, outcomes such as stillbirth, neonatal mortality, and serious neonatal morbidity have been used as markers of intrauterine growth restriction (6–8, 10). However, although these outcomes occur more commonly in growth-restricted infants, they also occur for a variety of reasons unrelated to IUGR. For example, although intrauterine growth restriction is believed to be one of the most common causes of stillbirth, placental abruption and infection are also important causes, and the largest cause of death grouping for stillbirths is “unexplained” (22). A novel aspect of our study was our use of simulation to establish IUGR status in our cohort, which allowed us to overcome this limitation. Although “failure of the fetus to reach its growth potential” (23) is widely used to define the condition of intrauterine growth restriction, in the absence of a biomarker or other measure of individual growth potential, an artificial manipulation of the birth weights of infants with healthy growth may be the only viable approach to quantify the concept.

Simulation has previously been used in the study of fetal growth restriction, including studies by Basso et al. (24) and Basso and Wilcox (25) on fetal growth and mortality. By simulating a population of fetuses who had reached their target birth weight and then introducing a hypothetical third factor (X1) that decreased birth weight and was causally linked to mortality, they were able to explore the extent to which the birth weight–neonatal mortality association might be explained by confounding. Whitcomb et al. (26) used simulation to illustrate that adjustment for birth weight in the presence of unmeasured factors that affect both birth weight and neonatal mortality will create bias, and to explore the potential magnitude of this bias. However, this is the first study of which we are aware to use simulation to evaluate the effectiveness of a screening tool such as customized percentiles in identifying intrauterine growth restriction.

Our study is not without limitations. The validity of our evaluation of customized percentiles is dependent on the assumptions made when simulating IUGR in this cohort, such as the prevalence and percent reduction in birth weight associated with IUGR. An underestimation of the percent reduction in birth weight, for example, could have lead to an

Figure 2. Distributions of 280-day target birth weights for “small-but-healthy” infants (birth weight below the population 10th percentile but no intrauterine growth restriction; dashed line) compared with target birth weights predicted for infants whose optimal weight was above the population 10th percentile (solid line) for a simulated study cohort. The target birth weights were produced by an existing customization model (A) and customization models that additionally included a hypothetical maternal characteristic correlated with birth weight at \( r = 0.50 \) (B), \( r = 0.75 \) (C), and \( r = 0.95 \) (D). The vertical line indicates the population 10th percentile weight for 280 days.
underestimation of the potential benefits of customization, because smaller reductions in birth weight would be more difficult to detect. For this reason, we chose to simulate a percent reduction in birth weight from the larger end of the range of estimates generated from the literature and clinical data (20%, from a range of 10%–20%). We considered this to be a conservative choice, because the larger the effect size, the easier it should be to detect by a process such as customization. As a result, it is unlikely that the lack of benefits from customization demonstrated in this study was attributable to incorrect assumptions regarding the simulation of IUGR. Further, sensitivity analyses varying the percent reduction in birth weight had no impact on our conclusion that customized percentiles provided only minimal improvements over conventional population-based percentiles.

Although intuitively appealing, the process of customizing birth weight percentiles to account for maternal characteristics does not meaningfully improve identification of infants with IUGR or distinguish SGA infants who are small-but-healthy from SGA infants who are small because of IUGR. For the process of customization to be worthwhile, predictors in the customization model would need to be able to explain an amount of the variability in birth weight that is unrealistically high on the basis of current knowledge. In their current form, customized percentiles appear to be more of an unnecessary complication than a meaningful improvement in the identification of intrauterine growth restriction.

ACKNOWLEDGMENTS

Author affiliations: Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, British Columbia, Canada (Jennifer A. Hutchon); Department of Obstetrics and Gynaecology, University of Ottawa, Ottawa, Ontario, Canada (Mark Walker); Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada (Robert W. Platt); and Department of Pediatrics, McGill University, Montreal, Quebec, Canada (Robert W. Platt).

J. A. H. is supported by a postdoctoral fellowship award from the Canadian Institutes of Health Research (CIHR); M. W. is a CIHR New Investigator; R. W. P. holds a Chercheur-from the Canadian Institutes of Health Research (CIHR); the FRSQ. This work was supported by grant MOP-84379 from the FRSQ. In their current form, customized percentiles appear to be more of an unnecessary complication than a meaningful improvement in the identification of intrauterine growth restriction.

REFERENCES

APPENDIX 1

Simulation of the Healthy Cohort

A cohort of infants with healthy birth weights was simulated on the basis of the characteristics of 9,697 “healthy” births at the Royal Victoria Hospital, a McGill University teaching hospital in Montreal, Canada, between 2002 and 2006. We defined “healthy” births as term livebirths (37–42 weeks’ gestation), a 5-minute Apgar score greater than 7, and no admission to the neonatal intensive care unit. We further excluded pregnancies with any noted maternal comorbidities of diabetes in pregnancy (gestational, type 1, or type 2); cardiac, pulmonary, or renal disease; or hypertension in pregnancy (chronic or pregnancy induced). Information on ethnicity in the Royal Victoria Hospital population was obtained from data collected during a gestational diabetes screening study with 5,487 women from 2001 to 2004 (27). This sample is well representative of the institution’s general obstetric population, because universal screening for gestational diabetes at 24–28 weeks’ gestation is conducted at our institution, and the study had a high (84%) participation rate.

First, a correlation matrix was created of the intercorrelations between maternal/fetal characteristics of births observed in the Royal Victoria Hospital cohort (birth weight, gestational age, sex, maternal height, prepregnancy weight). Prepregnancy weight was log transformed because of a non-Gaussian distribution. Standard normal random variables were simulated by drawing a sample of 1,661,000 observations from a multivariate normal distribution established from the correlation matrix. The variables were then converted to follow the means and standard deviations (for birth weight, gestational age, maternal height, prepregnancy weight) or proportions (fetal sex) of the characteristics in the Royal Victoria cohort. Specifically, let \( M \) be the desired matrix of means, and let \( V \) be the desired covariance matrix (consisting of the product of \( A \) and \( A’ \)). We first generated \( X \), such that \( X \sim N(0,I) \). Generating \( Y \) such that \( Y = A’X + M \) produced our final multivariate normal sample, where \( Y \sim N(M,V) \). The sample was then restricted to 100,000 births at 37 weeks (from 37\(^0\) to 37\(^6\) weeks). Simulations were performed in STATA SE, version 10 (Statacorp LP, College Station, Texas), by using the .matrix and .drawnorm commands.

For each of the 2 categorical variables (parity, ethnicity), a multinomial regression model predicting the probability of each category level based on maternal and fetal characteristics was built by using either all the Royal Victoria Hospital births (for parity) or the subsample of women in the gestational diabetes screening study (for ethnicity). The regression equation was then applied to calculate the predicted probabilities of each level of the characteristic for each pregnancy in our simulated data set (which sum to 1 for each woman). For example, given her characteristics, a woman might be predicted to have a 0.45 probability of being nulliparous, a 0.30 probability of having a parity of 1, a 0.2 probability of having a parity of 2, and a 0.05 probability of having a parity of 3 or greater. A uniform random number was generated for each woman, and the value of the uniform random number was used to assign a specific parity or ethnic group to each woman in the simulated cohort. For example, if the uniform random number had a value between 0 and 0.45, the aforementioned woman would be assigned a parity of 0; if it had a value between 0.45 and 0.75, she would be assigned a parity of 1, and so forth. Parity was simulated first, and information on each woman’s parity was subsequently included in the prediction of ethnicity. This produced a final population that had the appropriate proportions of each ethnic group and parity status, as well as correlations with other maternal and fetal characteristics.

APPENDIX 2

Calculation of Customized Percentiles

Customized birth weight percentiles were calculated by standard methods (2). First, coefficients from a customization model predicting birth weight in a US population (17) are used to predict a target weight at 280 days (40 weeks) for each infant in the simulated cohort. The target weights are predicted by using only those variables considered to have a physiologic influence on birth weight (fetal sex, maternal height, prepregnancy weight, ethnicity, and parity), because the inclusion of coefficients for characteristics known to have a pathologic association with birth weight (such as preeclampsia) would normalize the low birth weight of infants with intrauterine growth restriction. Although the validity of the assumption that the variables have a purely physiologic influence on fetal growth is debatable (28), they are nevertheless currently used in the calculation of customized percentiles for clinical practice and research use (3). After having predicted a 280-day target weight for each infant, we then adjusted the target weight to correspond to the actual gestational age at birth of the infants. This is achieved by using Gardosi’s proportionality formula, which expresses fetal weight at younger gestational ages as a proportion of its target weight at 280 days. The proportionality formula, which is derived from the intrauterine (ultrasound) standard of Hadlock et al. (18), is specified as follows: % weight = 299.1 – 31.85 × gestational age in weeks (GA) + 1.094 × GA\(^2\) – 0.01055 × GA\(^3\). For example, the target weight of an infant born at 259 days (37 weeks) would be...
84% of its 280-day target weight $(299.1 - 31.85 \times 37 + 1.094 \times 37^2 - 0.01055 \times 37^3 = 84)$ (2).

In the population used to create the customized standard for the US population, the coefficient of variation for birth weight at 280 days was 11% $(382.6 \text{ g} / 3453.4 \text{ g})$ (17). The 10th and 90th customized percentiles for 280 days are therefore calculated by multiplying the 280-day 50th percentile weight by 1.28 standard deviations $\times$ the coefficient of variation for birth weight of 11% (i.e., the 50th percentile weight $\pm 14\%$) (2). Once the upper and lower limits of the normal range at 280 days have been established, they are combined with the proportionality formula to produce 2 new equations and curves for the 10th and 90th customized percentiles (i.e., 114% and 86% of target weights at each age) (2). Infants with a birth weight below this 10th percentile curve are classified as small for gestational age by the customized standard, denoted as “SGA$_{customized}$.”

In our study, we used the same methods previously outlined to additionally calculate an exact customized percentile by replacing the $z$ score of 1.28 in the calculation of the proportionality curve for the 10th percentile with the $z$ value of the Gaussian distribution corresponding to each percentile; for example, the threshold for the fifth percentile was calculated as $-1.64 \times$ the coefficient of variation, the threshold for the third percentile was $-1.88 \times$ the coefficient of variation, and so on).