Small-for-gestational age and large-for-gestational age thresholds to predict infants at risk of adverse delivery and neonatal outcomes: are current charts adequate? An observational study from the Born in Bradford cohort

T Norris, W Johnson, D Farrar, D Tuffnell, J Wright, N Cameron

ABSTRACT

Objectives: Construct an ethnic-specific chart and compare the prediction of adverse outcomes using this chart with the clinically recommended UK-WHO and customised birth weight charts using cut-offs for small-for-gestational age (SGA: birth weight <10th centile) and large-for-gestational age (LGA: birth weight >90th centile).

Design: Prospective cohort study.

Setting: Born in Bradford (BiB) study, UK.

Participants: 3980 White British and 4448 Pakistani infants with complete data for gestational age, birth weight, ethnicity, maternal height, weight and parity.

Main outcome measures: Prevalence of SGA and LGA, using the three charts and indicators of diagnostic utility (sensitivity, specificity and area under the receiver operating characteristic (AUROC)) of these chart-specific cut-offs to predict delivery and neonatal outcomes and a composite outcome.

Results: In White British and Pakistani infants, the prevalence of SGA and LGA differed depending on the chart used. Increased risk of SGA was observed when using the UK-WHO and customised charts as opposed to the ethnic-specific chart, while the opposite was apparent when classifying LGA infants. However, the predictive utility of all three charts to identify adverse clinical outcomes was poor, with only the prediction of shoulder dystocia achieving an AUROC>0.62 on all three charts.

Conclusions: Despite being recommended in national clinical guidelines, the UK-WHO and customised birth weight charts perform poorly at identifying infants at risk of adverse neonatal outcomes. Being small or large may increase the risk of an adverse outcome; however, size alone is not sensitive or specific enough with current detection to be useful. However, a significant amount of missing data for some of the outcomes may have limited the power needed to determine true associations.
subgroups when using a population-derived birth weight chart, such as the UK-WHO.

Since the development of the UK90 references and in response to this variation, birth weight charts have been developed that are tailored for some ethnic minority groups.\textsuperscript{8, 7} These charts allow a more personalised assessment of size at birth, helping to determine whether an infant is small or large as a result of a pathological growth perturbation and therefore at risk of neonatal morbidity/mortality, or whether the infant is constitutionally small or large and therefore healthy. Using the conventional cut-offs to identify small-for-gestational age (SGA; <10th centile) and large-for-gestational age (LGA; >90th centile), ethnic-specific charts have been shown to perform significantly better than population references at identifying infants at risk of neonatal morbidity and mortality.\textsuperscript{8, 10} This finding may be unsurprising as the ethnic-specific cut-off typically classifies the more extreme infants as SGA (lower birth weight).

Importantly (and what the above studies did not report), for ethnic-specific cut-offs to be adopted into clinical practice, they need to demonstrate a clinical benefit, that is, the ability to accurately predict adverse neonatal outcomes by differentiating between the small and unhealthy infant and the small and healthy one.

A step further than adjusting for ethnicity are the ‘customised’ (gestation-related optimal weight: GROW) charts developed by Gardosi et al.\textsuperscript{11} which additionally adjust for maternal characteristics (height, body mass index, age and parity) known to have physiological effects on fetal growth.\textsuperscript{12} These charts have been recommended for clinical practice by the Royal College of Obstetricians and Gynaecologists.\textsuperscript{13} However, the evidence is inconsistent with regard to their clinical utility.\textsuperscript{14–17}

Whether the use of an ethnic-specific or customised birth weight chart improves the detection of infants at risk of adverse delivery, neonatal and infant outcomes is a discussion relevant to the issue of overmedicalisation in healthcare. If the production of ethnicity-specific or customised birth weight distributions provide no greater predictive benefit than a population-based distribution, then the use of a single tool for everyone is sufficient.

No published studies have assessed whether the UK-WHO birth weight chart predicts neonatal outcomes better than an ethnic-specific one. Therefore, the objectives of the study were to produce a birth weight chart adjusting for ethnicity and compare this with the UK-WHO and GROW birth weight charts to determine which chart better identifies neonates at risk of the adverse delivery and neonatal outcomes associated with SGA and LGA.

**METHODS**

The starting sample comprised 9102 (51.41% male, 53.34% Pakistani) singleton live births enrolled in the Born in Bradford (BiB) study. BiB is a longitudinal multi-ethnic birth cohort study aiming to examine the impact of environmental, psychological and genetic factors on maternal and child health and well-being.\textsuperscript{18} Bradford is a city in the North of England with high levels of socioeconomic deprivation and ethnic diversity. Approximately half of the births in the city are to mothers of South Asian origin. Women were recruited while waiting for their glucose tolerance test, a routine procedure offered to all pregnant women registered at the Bradford Royal Infirmary, at 26–28 weeks gestation. For those consenting, a baseline questionnaire was completed via an interview with a study administrator. All babies born to women who agreed to participate in the cohort study were eligible for recruitment. The full BiB cohort recruited 12 453 women during 13 776 pregnancies between 2007 and 2010 and the cohort is broadly characteristic of the city’s maternal population. All participants provided written informed consent before inclusion in the research. Birth data were extracted from the maternity information system. Infants were weighed naked within 24 h of birth to the last completed 10 g using Seca baby scales. Gestational age was determined in accordance with guidelines issued by the National Institute for Health and Care Excellence (NICE); crown-rump length up to 13 weeks 6 days and head circumference thereafter.\textsuperscript{19} Categorisation of infant ethnicity (White British and Pakistani) was based on maternal self-reported ethnicity at interview, with response options selected based on guidance from the Office of National Statistics (ONS).\textsuperscript{20}

**Centile and z score production**

Centiles were produced for live-born singleton infants born between 32 and 42 weeks gestation without congenital anomalies, of either White British (n=4247) or Pakistani origin (n=4855), who had complete data for weight and exact decimal age at birth. Sex-specific and ethnic-specific scatterplots were produced to visually identify any outlying cases that may also have substantial influence on the centiles (checked by two reviewers to reduce bias). Sex-specific and ethnic-specific centiles were produced using the LMS method. Briefly, this technique summarises the distribution of birth weights at each gestational age by its median (M), variability (S) and measure of skewness (L) required to transform the distribution to normality.\textsuperscript{21} With these parameters, any birth weight centile can be generated. These centiles are not exact, but are rounded from z scores −2, −1.33, −0.67, 0, 0.67, 1.33 and 2.\textsuperscript{22}

As well as producing centiles, the LMS method can be used to convert the centiles into z scores, using the following formula:

\[ Z = \frac{\text{Measurement} - M(t)}{S(t)L(t)} \]

where measurement is the infant’s birth weight and L(t), M(t) and S(t) are values read from the smooth curves for the infant’s ethnicity, age and sex.\textsuperscript{21}
Once these ethnic-specific and sex-specific z scores were produced, z scores were produced based on the LMS values used to construct the revised UK-WHO charts. Therefore, each infant had two z scores, one ethnic-specific, sex-specific and age-specific z score (BiB z score) and another based on the UK-WHO chart (UK-WHO z score).

GROW centiles, additionally adjusting for maternal weight, height and parity, were also produced for each infant. These charts use coefficients obtained from a customisation model (regression of birth weight on fetal sex, gestational age and the maternal variables aforementioned) to obtain an ‘optimal’ birth weight, given the maternal characteristics and based on a gestational length of 280 days. Using a proportionality formula derived from an intrauterine fetal weight standard, optimal weights for births occurring prior to 280 days are calculated. The actual birth weight of the infant is compared with its optimal birth weight, and infants whose actual birth weight falls below the 10th or above the 90th centile of the assumed distribution around its optimal weight are classified as SGA and LGA, respectively. Only those pregnancies with complete data necessary for customisation were included in the final analysis (n=8428). Analysis of those with and without the necessary maternal variables revealed that those with the necessary variables had babies which were 11 g lighter (p=0.59) and born 1 day later (p=0.02) than those without and with no significant differences in parity, pregnancy and existing hypertension and gestational diabetes. Those women who did not have the necessary variables did have significantly less pre-eclampsia 2.49% vs 3.76%, p=0.04) and diabetes (0.23% vs 1.15%, p<0.01).

Relative risk (RR) of being classified as SGA or LGA
Using the conventional cut-off for SGA as a z score less than −1.88 (10th centile) or, in the case of the GROW charts, a centile <10, the RR of being classified as SGA based on the BiB, UK-WHO and GROW charts were calculated. Furthermore, classifications of SGA and LGA were calculated using other commonly used thresholds, 5th/95th (z<−1.645 or >1.645) and 3rd/97th (z<−1.88 or >1.88) centiles, to see if these increased prediction of adverse delivery characteristics and neonatal morbidity. In White British infants across the three charts, using the 5th/95th centiles only provided significantly improved prediction (over the 10th/90th) for two outcomes, with the 10th/90th providing significantly better prediction for three outcomes (the 3rd/97th centiles provided improved prediction for none). In Pakistani infants, however, across the three charts, the 10th/90th centiles out predicted the 5th/95th by 8:3 and 3rd/97th by 11:1. Therefore, the discussion of which of the three charts performs better at predicting outcomes will be restricted to the 10th and 90th centiles only.

Outcome data
Outcomes were chosen which had previously been shown to be associated with SGA or LGA, either in the literature or from consultations with clinical experts. Delivery outcomes were: the need for induction of labour; assistance during labour (ventouse, forceps or both); and caesarean section, while the neonatal outcomes included: shoulder dystocia; hypothermia (axillary temperature ≤36.5°C); low Apgar score at 1 min (Apgar score ≤5); low Apgar score at 5 min (Apgar score ≤5); and admission to the neonatal unit. A composite outcome was also generated to capture all adverse outcome events and improve power. Furthermore, as many of the outcomes could be a result of the same underlying cause which manifests as either SGA or LGA, for example, an SGA infant could have both a low Apgar score and hypothermia, this composite would also serve to reduce risks associated with multiple testing. Composites were specific to either the SGA or LGA cut-off; however, some outcomes are associated with both SGA and LGA deliveries, including: admission to neonatal unit (NNU); induction of labour; assistance during labour and caesarean section. For the SGA-specific composite outcome: low Apgar score at 1 min; low Apgar score at 5 min and hypothermia were also added. A pregnancy with any of these outcomes present was scored with a value of 1 (0 if none were present). The LGA-specific composite outcome variable included shoulder dystocia in addition to the above four variables. A pregnancy with any of these outcomes present was scored with a value of 1 (0 if none was present). The completeness of data for each outcome was variable and as it is not recommended to impute outcomes, a complete case approach was adopted (n with complete data for all outcomes was 5684).

Diagnostic accuracy of SGA and LGA
Within each ethnic group, the ability of SGA and LGA (exposure) to predict the outcomes was assessed. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated for each of the three charts’ SGA and LGA cut-offs. Receiver operating characteristics (ROC) and area under the ROC (AUROC) were produced to summarise the diagnostic performance of the respective cut-offs. As more than one diagnostic test (BiB vs UK-WHO cut-off; BiB vs GROW, etc) was tested on the same individual, a comparison of AUROCs between charts was made using Stata’s correlated ROC method, “roccomp”.

The production of the ethnic-specific centiles/z scores was done using ‘LMSchartmaker Light’ V2.5.1. GROW centiles were produced using the ‘Customised Centile Calculator’ V6.7 software from the Perinatal Institute. All other analyses were conducted in Stata/IC V.12.1.

RESULTS
Table 1 provides descriptive statistics for the 8428 infants and mothers in the sample who had cut-offs from all three charts. There was a similar distribution of males that were included in the analysis.
and females within each of the ethnic groups. Pakistani infants were born, on average, approximately 1.25 days earlier than White British infants and were around 220 g lighter. Pakistani mothers were approximately 4.4 cm smaller, 7 kg lighter and 1 year older than White British mothers.

Online supplementary table S1 lists the LMS values by gestation, sex and ethnic group, with median birth weights for Pakistani males being 4.8–11.6% lighter than those for White British males. For females, the differences between ethnicities ranged from 5.1% to 15%. The numbers of births delivered at each gestational age (to the nearest week) are shown in online supplementary table S2 (based on n=9102). To see how well the BiB centiles summarised the distribution, the expected and observed percentages of infants lying above each centile can be seen in online supplementary table S3. The maximum deviation from the expected percentages was 0.78% (Pakistani infants above the 75th centile) and so the centiles summarised the distribution well.

### Table 1 Maternal and infant characteristics (percentages rounded to nearest whole number (1 dp. when close to zero))

<table>
<thead>
<tr>
<th></th>
<th>White British (n=3980)</th>
<th>Pakistani (n=4448)</th>
<th>p Value (for difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2054 (52%)</td>
<td>2279 (51%)</td>
<td>NS</td>
</tr>
<tr>
<td>Female</td>
<td>1926 (48%)</td>
<td>2169 (49%)</td>
<td></td>
</tr>
<tr>
<td><strong>Gestational age (weeks) (mean (SD))</strong></td>
<td>39.7 (1.6)</td>
<td>39.6 (1.5)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Preterm (&lt;37 weeks)</td>
<td>199 (5%)</td>
<td>181 (4%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Term (37–41 weeks)</td>
<td>3711 (93%)</td>
<td>4221 (95%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Post-term (≥42 weeks)</td>
<td>70 (2%)</td>
<td>46 (1%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Birth weight (g) (mean (SD))</td>
<td>3373.3 (522.3)</td>
<td>3157.7 (494.8)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>AGA (all three cut-offs by BiB reference)</td>
<td>3223 (81%)</td>
<td>3562 (80%)</td>
<td>NA</td>
</tr>
<tr>
<td>LGA</td>
<td>376 (9%)</td>
<td>457 (10%)</td>
<td></td>
</tr>
<tr>
<td>Low birth weight (&lt;2500 g)</td>
<td>178 (4%)</td>
<td>365 (8%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Maternal height (cm) (mean (SD))</td>
<td>164.1 (6.2)</td>
<td>159.7 (5.7)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Maternal weight (kg) (mean (SD))</td>
<td>78.1 (16.6)</td>
<td>71.1 (14.0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Maternal age (mean (SD))</td>
<td>26.9 (6.1)</td>
<td>27.9 (5.2)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Parity (median range)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Para 0</td>
<td>1933 (49%)</td>
<td>1425 (32%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Para 1</td>
<td>1247 (31%)</td>
<td>1188 (27%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Para 2</td>
<td>514 (13%)</td>
<td>914 (21%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Para &gt;2</td>
<td>286 (7%)</td>
<td>921 (21%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Pre-existing hypertension</td>
<td>46 (1%)</td>
<td>30 (0.7%)</td>
<td>0.02</td>
</tr>
<tr>
<td>Pregnancy-induced hypertension</td>
<td>256 (7%)</td>
<td>205 (5%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Pre-eclampsia</td>
<td>99 (3%)</td>
<td>103 (2%)</td>
<td>NS</td>
</tr>
<tr>
<td>Pre-existing diabetes</td>
<td>14 (0.4%)</td>
<td>7 (0.2%)</td>
<td>NS</td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td>191 (5%)</td>
<td>479 (11%)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

AGA, appropriate-for-gestational age; BiB, Born in Bradford; LGA, large-for-gestational age; NA, not applicable; NS, not significant; SGA, small-for-gestational age.

**Prevalence of SGA/LGA and RR**

Figures 1–4 display Venn diagrams showing the frequencies of SGA and LGA classifications for both Pakistani and White British infants, with each circle representing a different chart.

**BiB versus UK-WHO**

Using the UK-WHO centile charts compared with the BiB, a further 133% (n=593) of Pakistani and 2.0% (n=78) of White British infants born between 32 and
42 weeks would have been classified as SGA. At the other end of the distribution, 6.8% (n=304) and 1.6% (n=64) of LGA Pakistani and White British infants, respectively, would not have been identified using the UK-WHO thresholds. Online supplementary tables S4 and S5 list the prevalence of SGA and LGA by week for each ethnic group (sexes pooled).

### BiB versus GROW

Using the GROW centile charts compared with the BiB, 407 (9.2%) and 355 (8.9%) more Pakistani and White British infants, respectively, would have been classified as SGA. In terms of LGA, 245 (5.5%) and 203 (5.2%) fewer Pakistani and White British infants, respectively, would have been identified using the GROW charts. Online supplementary tables S6 and S7 list the prevalence of SGA and LGA by week for each ethnic group.

### Diagnostic accuracy of the BiB, UK-WHO and GROW SGA and LGA cut-offs to predict adverse delivery and neonatal outcomes

Table 2 shows the prevalence of the neonatal and delivery outcomes, along with numbers with missing data. The prevalence of shoulder dystocia, low Apgar scores (1 and 5 min) and admission to the NNU did not differ significantly between the two ethnic groups. Hypothermia was significantly more common in Pakistani infants, however. The three delivery outcomes (induction of labour, caesarian section and assistance during labour) were either borderline or significantly different between the two groups, with Pakistani gravidas displaying lower prevalence of all three. White British infants also had a significantly higher prevalence of both the SGA-specific and LGA-specific composites.

### Pakistani infants

Of the 14 outcomes, only the prediction of shoulder dystocia was predicted with an AUROC greater than 0.6 on each of the three charts (BiB 0.70, UK-WHO 0.62, GROW 0.67). For the 13 other outcomes, AUROCs ranged from 0.48 to 0.61. Table 3 lists sensitivity, specificity, PPV, NPV and AUROCs for those outcomes which were predicted significantly or borderline significantly (p<0.1) differently between the charts, at the 10th and 90th centile.

### BiB versus UK-WHO

The BiB chart achieved AUROC values which were statistically significantly (or borderline significantly) better than the UK-WHO chart for six of the outcomes: shoulder dystocia; low Apgar score at 1 min; induction of
labor for LGA infants; caesarean section for LGA infants; admission to the NNU for SGA infants and the LGA-specific composite outcome. For the prediction of hypothermia, the UK-WHO 10th centile obtained a significantly better AUROC than that of the BiB chart.

### BiB versus GROW

The BiB chart significantly improved prediction of: low Apgar score at 1 min; induction of labour for LGA infants; and assistance during labour for SGA deliveries, compared with the GROW chart. The GROW chart provided significantly (or borderline significantly) better prediction for hypothermia and assistance during labour for LGA deliveries.

In summary, for Pakistani infants the BiB chart had a greater predictive ability than both the GROW charts and the UK-WHO. However, the actual predictive ability of the chart was weak for the majority of the outcomes, as evidenced by the negligible increases in AUROC above 0.5. Limiting the sample to only term infants marginally improved the predictive ability of the BiB chart, increasing the average AUROC across the 14 outcomes by 1.71% (SD 2.17%).

### White British Infants

As was the case for Pakistani infants, shoulder dystocia was the only outcome which was predicted with an AUROC greater than 0.6 on each of the three charts (BiB 0.66, UK-WHO 0.62, GROW 0.61). For the 13 other outcomes, AUROCs ranged from 0.48 to 0.60. Table 4 provides details of the diagnostic accuracy for those outcomes which were predicted significantly or borderline significantly (p<0.1) differently between the charts.

### BiB versus UK-WHO

The BiB chart achieved AUROC values which were statistically significantly (or borderline significantly) better than the UK-WHO chart for four of the outcomes: shoulder dystocia; low Apgar score at 1 min; caesarean section for SGA infants; and admission to the NNU for SGA infants. The UK-WHO cut-offs did not predict any of the outcomes significantly better than the ethnic-specific BiB chart.

### DISCUSSION

Classifying infants as SGA or LGA by the BiB, UK-WHO or GROW charts had low predictive utility for the outcomes under investigation. Despite the fact that the BiB ethnic-specific birth weight reference provided significantly better prediction for more outcomes than the UK-WHO and GROW charts in both White British and Pakistani infants, the lack of predictive ability of the charts for the majority of the outcomes is of concern.
Pakistani infants, with the exception of shoulder dystocia, AUROC values for all three charts were all below 0.61. This represents a diagnostic tool with little or no discriminatory power, substantially below the value of 0.75 which is deemed clinically useful.34 As expected, there was an increased risk of being classified as SGA when using the clinically recommended UK-WHO chart as opposed to an ethnic-specific chart. This was predominantly observed in Pakistani infants, with an RR of 2.42 after 37 weeks gestation, though a significantly increased risk was also observed in White British infants at 40 and 41 weeks. The risk of classifying Pakistani infants as LGA was around one-third if using the UK-WHO chart as opposed to the ethnic-specific chart from 36 weeks, with a reduced risk also observed in White British infants, though this reduced risk was smaller and mostly non-significant. In the South Asian context, an ethnic-specific chart serves to decrease the risk of classifying Pakistani infants as SGA and increase the risk for LGA as lower weight cut-offs are used. It was assumed that this would lead to increased predictive ability as it would greatly reduce the amount of ‘small but healthy’ infants classified as SGA and increase the amount of ‘less big but unhealthy’ classified as LGA, thereby improving the sensitivity and specificity of the tool. Although the ethnic-specific chart performed

### Table 3  Pakistani origin infants

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Chart</th>
<th>Prevalence (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>AUROC</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>50.00</td>
<td>90.40</td>
<td>5.30</td>
<td>99.40</td>
<td>0.70</td>
<td>p=0.01</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>50.00</td>
<td>90.40</td>
<td>5.30</td>
<td>99.40</td>
<td>0.70</td>
<td>p=0.01</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>UK-WHO</td>
<td>27.30</td>
<td>96.90</td>
<td>8.60</td>
<td>99.20</td>
<td>0.62</td>
<td>p=0.01</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>GROW</td>
<td>32.50</td>
<td>82.40</td>
<td>14.00</td>
<td>93.30</td>
<td>0.57</td>
<td>p=0.07</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>GROW</td>
<td>32.50</td>
<td>82.40</td>
<td>14.00</td>
<td>93.30</td>
<td>0.57</td>
<td>p=0.07</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>GROW</td>
<td>32.50</td>
<td>82.40</td>
<td>14.00</td>
<td>93.30</td>
<td>0.57</td>
<td>p=0.07</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>GROW</td>
<td>32.50</td>
<td>82.40</td>
<td>14.00</td>
<td>93.30</td>
<td>0.57</td>
<td>p=0.07</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>BiB</td>
<td>19.70</td>
<td>91.20</td>
<td>16.50</td>
<td>92.80</td>
<td>0.55</td>
<td>p=0.03</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia BiB 1.10</td>
<td>GROW</td>
<td>32.50</td>
<td>82.40</td>
<td>14.00</td>
<td>93.30</td>
<td>0.57</td>
<td>p=0.07</td>
<td></td>
</tr>
</tbody>
</table>

AUROC, area under the receiver operating characteristic; BiB, Born in Bradford; GROW, gestation-related optimal weight; LGA, large-for-gestational age; NNU, neonatal unit; NPV, negative predictive value; PPV, positive predictive value; SGA, small-for-gestational age.
significantly better than the UK-WHO chart for a variety of the outcomes, we were unable to demonstrate that these lower cut-offs would provide any clinically important increases in the identification of those at risk of adverse neonatal outcomes. Adjustment for maternal characteristics in the GROW charts also increased the risk of classifying infants as SGA. This was especially apparent in White British infants and at earlier gestational ages. A number of studies that found that adjusting for maternal characteristics did not improve detection of neonatal morbidity,\(^17\)\(^35\)\(^36\) has led to an argument that the perceived benefits in the customised model are not actually a result of customising for maternal characteristics. Rather, it is the use of an intrauterine standard for the reference values at gestational ages younger than 280 days that provides the added benefit.\(^{37}\) The use of an intrauterine standard for preterm deliveries serves to address the bias apparent at these ages due to the association between fetal growth restriction and preterm birth. However, our results suggest that the correction of this bias does not improve the prediction of infants at risk of adverse outcomes.

There is now a large body of evidence suggesting that ethnic inequalities in health are in some part determined by differences in socioeconomic status (SES).\(^{38\text{-}41}\) Indeed, Pakistani-born infants in Bradford are much more likely to have parents living in the most deprived areas of the city, and it has been observed that this increased deprivation mediates up to 13% of the ethnic difference observed in fetal weight.\(^{42}\) Caution is

<table>
<thead>
<tr>
<th>Table 4</th>
<th>White British origin infants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>Chart</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>BiB versus UK-WHO</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>UK-WHO</td>
</tr>
<tr>
<td>Low Apgar 1 min</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>UK-WHO</td>
</tr>
<tr>
<td>Caesarean section (SGA)</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>UK-WHO</td>
</tr>
<tr>
<td>NNU (SGA)</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>UK-WHO</td>
</tr>
<tr>
<td>BiB versus customised</td>
<td></td>
</tr>
<tr>
<td>Shoulder dystocia</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>Low Apgar 5 min</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>Caesarean section (LGA)</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>Assistance (SGA)</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>NNU (SGA)</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>NNU (LGA)</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>Composite (LGA)</td>
<td>BiB</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>UK-WHO versus customised</td>
<td></td>
</tr>
<tr>
<td>Low Apgar 5 min</td>
<td>UK-WHO</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>Caesarean section (LGA)</td>
<td>UK-WHO</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>Assistance (SGA)</td>
<td>UK-WHO</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
<tr>
<td>NNU (SGA)</td>
<td>UK-WHO</td>
</tr>
<tr>
<td></td>
<td>GROW</td>
</tr>
</tbody>
</table>

AUROC, area under the receiver operating characteristic; BiB, Born in Bradford; GROW, gestation-related optimal weight; LGA, large-for-gestational age; NNU, neonatal unit; NPV, negative predictive value; PPV, positive predictive value; SGA, small-for-gestational age.
outcomes were prevalent at a rate of less than 5% could average sample size was around 4200 for each ethnic lysis was strati fi longer term outcomes were not available. outcomes used here. However, as is discussed below, to management, which may occur later in time than the viewpoint, one would ideally like to use outcomes where comes were included as it was felt that it was important into the cohort, robust analyses could not be completed on these infants and they were therefore removed. Generalisability to other South Asian populations is therefore not possible. By assessing the ability of the GROW centiles, we had to limit the sample to those with the necessary maternal variables. In comparison to the larger sample who had BiB centiles produced, the smaller sample had a higher prevalence of pre-eclampsia and pre-existing diabetes, suggesting that we may have analysed a less healthy sample of women. The reasoning behind including a composite outcome has been addressed, but the clinical relevance of the composites created is uncertain as there is no general consensus in the literature as to what should be included. This is apparent from the large number of trials including neo-natal morbidity composite outcomes, each of which varies in composition.9 50 46 47 Including preterm births in the analysis may have introduced a bias, as some outcomes may be more strongly associated with this delivery at these ages. However, when limited to term births, the change in predictive ability of the charts was negligible. The authors are aware of the UK-WHO Neonatal and Infant close monitoring charts for very preterm infants and those with early health problems. However, a comparison against this chart was not conducted as the amount of data at these very early ages was limited and would most likely have resulted in unreliable estimates at these ages.

Despite the ethnic-specific chart demonstrating a marginally better predictive ability than the UK-WHO chart and customised chart, all three charts performed poorly at predicting neonatal and delivery outcomes. A recommendation for which chart should be used is therefore unclear. For example, NICE suggests that recommendations should be based on the “estimated costs of the service in relation to their expected health benefits”.48 Despite incurring minimal expenditure in the production and implementation of the charts (particularly for an ethnic-specific and the UK-WHO), the poor clinical utility of the charts suggests that their cost-effectiveness would be low. A further cost-effectiveness analysis could provide a more definitive conclusion.

However, we have shown that although being small or large may increase the risk of an adverse outcome, size alone is not sensitive or specific enough with current detection to be a useful clinical tool and poses the question of whether classification of size helps with the

Strengths and weaknesses
This study is the first to provide evidence relating the utility of two nationally recommended birth weight charts for the prediction of clinical outcomes. This evidence is based on a large sample of approximately 9000 mother-infant dyads and is a further strength of the study. Additionally, the diagnostic test analysis employed provides more information about these charts’ predictive ability than previous studies conducted.9 10 14 15 The BiB cohort itself has a number of strengths. The large bi-ethnic sample of White British and Pakistani-origin families allows a detailed analysis of associations and causal pathways for differences between these two ethnic groups. The composition within each ethnic group is relatively homogeneous and the cohort overall is characteristic of the population of Bradford, suggesting minimal selection bias. However, the study does have some limitations. The classification of SGA/LGA occurs after delivery and thus any diagnosis occurs too late to influence delivery management. However, these outcomes were included as it was felt that it was important to highlight their association with size. From a clinical viewpoint, one would ideally like to use outcomes where the diagnosis of SGA/LGA might make a real difference to management, which may occur later in time than the outcomes used here. However, as is discussed below, longer term outcomes were not available.

Missing data in the cohort and the fact that the analysis was stratified by ethnic group meant that the average sample size was around 4200 for each ethnic group. Paired with the fact that most of the neonatal outcomes were prevalent at a rate of less than 5% could have resulted in the study being underpowered to detect significant differences in the predictive ability of the two charts. Furthermore, missing data occurred at variable rates depending on the outcome, which may result in differences in power for each outcome; for example, in the calculation of SGA and LGA composite outcomes, missing data meant that more than 20% of the sample were without this composite measure. Neonatal outcomes which are commonly observed in SGA and LGA infants and which may have been predicted well by the respective cut-offs, for example, hypoglycaemia, necrotising enterocolitis, assisted ventilation, seizures and hypercoagulability, were not available. However, these outcomes are likely to result in admission to the neonatal unit and are thus captured in this single outcome measure. An absence of outcomes occurring in infancy and childhood which have been reported to be associated with size at birth, for example, motor and cognitive development,43–45 means that we are unable to speculate as to the longer term utility of the ethnic-specific cut-offs. Owing to a much smaller number of families from other countries within South Asia (India and Bangladesh) resident in Bradford and thus enrolled into the cohort, robust analyses could not be completed on these infants and they were therefore removed. Generalisability to other South Asian populations is therefore not possible. By assessing the ability of the GROW centiles, we had to limit the sample to those with the necessary maternal variables. In comparison to the larger sample who had BiB centiles produced, the smaller sample had a higher prevalence of pre-eclampsia and pre-existing diabetes, suggesting that we may have analysed a less healthy sample of women. The reasoning behind including a composite outcome has been addressed, but the clinical relevance of the composites created is uncertain as there is no general consensus in the literature as to what should be included. This is apparent from the large number of trials including neonatal morbidity composite outcomes, each of which varies in composition.9 50 46 47 Including preterm births in the analysis may have introduced a bias, as some outcomes may be more strongly associated with this delivery at these ages. However, when limited to term births, the change in predictive ability of the charts was negligible. The authors are aware of the UK-WHO Neonatal and Infant close monitoring charts for very preterm infants and those with early health problems. However, a comparison against this chart was not conducted as the amount of data at these very early ages was limited and would most likely have resulted in unreliable estimates at these ages.

Despite the ethnic-specific chart demonstrating a marginally better predictive ability than the UK-WHO chart and customised chart, all three charts performed poorly at predicting neonatal and delivery outcomes. A recommendation for which chart should be used is therefore unclear. For example, NICE suggests that recommendations should be based on the “estimated costs of the service in relation to their expected health benefits”.48 Despite incurring minimal expenditure in the production and implementation of the charts (particularly for an ethnic-specific and the UK-WHO), the poor clinical utility of the charts suggests that their cost-effectiveness would be low. A further cost-effectiveness analysis could provide a more definitive conclusion.

However, we have shown that although being small or large may increase the risk of an adverse outcome, size alone is not sensitive or specific enough with current detection to be a useful clinical tool and poses the question of whether classification of size helps with the
management at all. Perhaps, as suggested by Hutcheon, the focus should be on developing more accurate methods such as the combination of weight alongside assessments of placental health (placental weight; birth weight to placental weight ratio) or pregnancy biomarkers (placental growth factor; pregnancy-associated plasma protein A).

BiB is only possible because of the enthusiasm and commitment of the Children and Parents in BiB. We are grateful to all the participants, health professionals and researchers who have made BiB happen.

Author affiliations
1Centre for Global Health and Human Development, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK
2MRC Unit for Lifelong Health & Ageing, University College London, London, UK
3Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK
4Bradford Teaching Hospitals NHS Foundation Trust, Bradford Royal Infirmary, Bradford, UK
5Department of Health and Social Care, Public Health England, London, UK
6British Heart Foundation, London, UK
7Charity Commission for England and Wales, London, UK
8Playgroups and Pre-schools Network, London, UK
9University of Oxford, Oxford, UK
10University of Cambridge, Cambridge, UK
11Guy’s and St Thomas’ NHS Foundation Trust, London, UK
12Department of Obstetrics and Gynaecology, King’s College Hospital, London, UK
13University Hospitals of Leicester NHS Trust, Leicester, UK
14Leicester Diabetes Centre, University of Leicester, Leicester, UK
15University of Warwick, Coventry, UK
16University of East Anglia, Norwich, UK
17Babraham Institute, Cambridgeshire, UK
18Ministry of Defence, Aldershot, UK
19Rothamsted Research, Harpenden, Hertfordshire, UK
20University of Northampton, Northampton, UK
21University of Loughborough, Loughborough, UK

Contributors
TN was involved in the conception of the study, was the lead author of the manuscript and conducted the data cleaning and analysis. WJ provided guidance with data analysis, was involved in interpreting the data and contributed to editing of the manuscript. DF was involved in data collection, study development and editing of the manuscript. JL is the lead investigator of the Born in Bradford cohort and was involved in parts of the data preparation, interpretation of data and editing of the manuscript. NC was involved in the conception of the study, the analysis and interpretation of data and made extensive contributions to the drafting and revision of the manuscript.

Funding
This research was funded by an NIHR CLAHRC implementation grant and an NIHR applied programme grant (RP-PG-0407–10044). This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under the CLAHRC programme.

Competing interests
None.

Ethics approval
Bradford Research Ethics Committee (Ref 07/H1302/112).

Provenance and peer review
Not commissioned; externally peer reviewed.

Data sharing statement
Patient level data and full data set or technical appendix and statistical code are available with open access from TN.

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES


Small-for-gestational age and large-for-gestational age thresholds to predict infants at risk of adverse delivery and neonatal outcomes: are current charts adequate? An observational study from the Born in Bradford cohort

T Norris, W Johnson, D Farrar, D Tuffnell, J Wright and N Cameron

*BMJ Open* 2015 5:
doi: 10.1136/bmjopen-2014-006743

Updated information and services can be found at:
http://bmjopen.bmj.com/content/5/3/e006743

These include:

**Supplementary Material**
Supplementary material can be found at:
http://bmjopen.bmj.com/content/suppl/2015/03/16/bmjopen-2014-006743.DC1

**References**
This article cites 41 articles, 7 of which you can access for free at:
http://bmjopen.bmj.com/content/5/3/e006743#BIBL

**Open Access**
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections
- Epidemiology (2235)
- Obgyn (352)
- Research methods (631)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/