BMJ Open

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,[1] W Johnson,[2] D Farrar,[3] D Tuffnell,[4] J Wright,[3] N Cameron[1]

For numbered affiliations see end of article.

Correspondence to
T Norris;
T.Norris2@lboro.ac.uk

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.

Strengths and limitations of this study

- This study is the first to provide evidence relating the utility of two nationally recommended birth weight charts for predicting clinical outcomes.
- A further strength is that the diagnostic test analysis employed provides more information about these charts' predictive ability than previous studies.
- However, large amounts of missing data for some of the outcomes may result in the analysis being underpowered to detect true associations.
- Longer term outcomes were not available.

INTRODUCTION

Since 2009, the UK-WHO growth chart for children aged 0–4 years has been used in the UK. The WHO chart was based on children born at term, in six different countries, to non-smoking mothers whose socioeconomic environment would not constrain their growth.[1] In the UK, however, it was necessary to retain the former charts, the UK90 references, for assessment at birth, as not only did the UK-WHO charts have no preterm section (by design) but also the WHO mean birth weight for term births was significantly lower than in the UK.[2] The UK90 references, however, were constructed using a sample of White British infants only, as it was thought that 'ethnic non-white children' may grow differently.[3] Indeed, there are substantial ethnic variations in the distribution of birth weights. Studies have shown that UK-born South Asians, for example, are 200–300 g lighter at birth compared with White British infants,[4 5] and this fact may have implications for the assessment of health in these

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.[6] [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.[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,[11] which additionally adjust for maternal characteristics (height, body mass index, age and parity) known to have physiological effects on fetal growth.[12] These charts have been recommended for clinical practice by the Royal College of Obstetricians and Gynaecologists.[13] However, the evidence is inconsistent with regard to their clinical utility.[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 ethnic-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.[18] Bradford is a city in the North of England with high levels of socio-economic 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.[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).[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.[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.[22]

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

$$\text{Z score} \left(\left[\frac{\text{Measurement}}{\text{M}(t)} \right]^{\text{L}(t)} - 1 \right) / \text{S}(t)\text{L}(t)$$

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.[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.[23] 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 afore-mentioned) to obtain an 'optimal' birth weight, given the maternal characteristics and based on a gestational length of 280 days.[24] Using a proportionality formula derived from an intrauterine fetal weight standard,[25] 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 target 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 −1.28 (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,[13 19 26–28] 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[29–31] 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' V.2.54.[32] GROW centiles were produced using the 'Customised Centile Calculator' V.6.7 software from the Perinatal Institute.[33] 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

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

	White British (n=3980) n (%)	Pakistani (n=4448) n (%)	p Value (for difference)
Sex			
Male	2054 (52%)	2279 (51%)	NS
Female	1926 (48%)	2169 (49%)	
Gestational age (weeks) (mean (SD))	39.7 (1.6)	39.6 (1.5)	<0.01
Preterm (<37 weeks)	199 (5%)	181 (4%)	0.04
Term (37–41 weeks)	3711 (93%)	4221 (95%)	<0.01
Post-term (≥42 weeks)	70 (2%)	46 (1%)	<0.01
Birth weight (g) (mean (SD))	3373.3 (522.3)	3157.7 (494.8)	<0.01
SGA	381 (10%)	429 (10%)	
AGA (all three cut-offs by BiB reference)	3223 (81%)	3562 (80%)	NA
LGA	376 (9%)	457 (10%)	
Low birth weight (<2500 g)	178 (4%)	365 (8%)	<0.01
Maternal height (cm) (mean (SD))	164.1 (6.2)	159.7 (5.7)	<0.01
Maternal weight (kg) (mean (SD))	78.1 (16.6)	71.1 (14.0)	<0.01
Maternal age (mean (SD))	26.9 (6.1)	27.9 (5.2)	<0.01
Parity (median range)	1 (10)	1 (10)	NS
Para 0	1933 (49%)	1425 (32%)	<0.01
Para 1	1247 (31%)	1188 (27%)	<0.01
Para 2	514 (13%)	914 (21%)	<0.01
Para >2	286 (7%)	921 (21%)	<0.01
Pre-existing hypertension	46 (1%)	30 (0.7)%	0.02
Pregnancy-induced hypertension	256 (7%)	205 (5%)	<0.01
Pre-eclampsia	99 (3%)	103 (2%)	NS
Pre-existing diabetes	14 (0.4%)	7 (0.2%)	NS
Gestational diabetes	191 (5%)	479 (11%)	<0.01

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.

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.

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
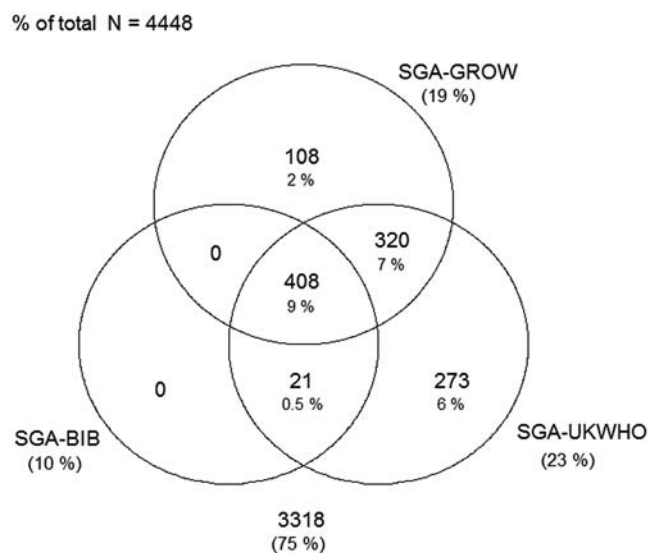

Figure 1 Venn diagram of SGA classification in Pakistani infants (SGA, small-for-gestational age; BiB, Born in Bradford).
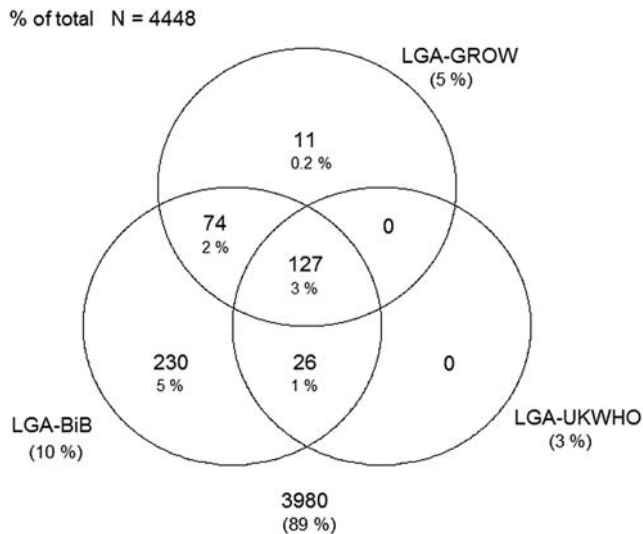
% of total   N = 4448

Figure 2 Venn diagram of LGA classification in Pakistani infants (LGA, large-for-gestational age; GROW, gestation-related optimal weight; BiB, Born in Bradford).

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,
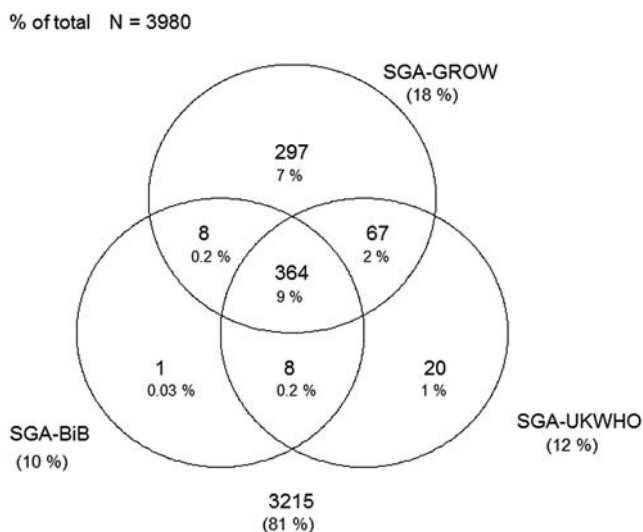
% of total   N = 3980

Figure 3 Venn diagram of SGA classification in White British infants (SGA, small-for-gestational age; GROW, gestation-related optimal weight; BiB, Born in Bradford).
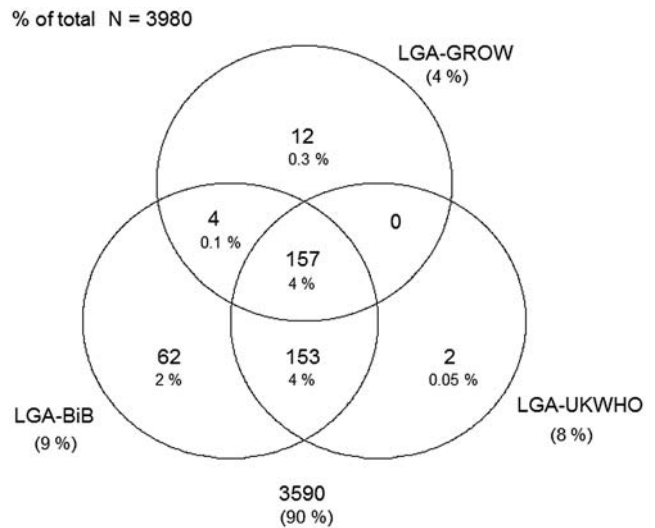
% of total   N = 3980

Figure 4 Venn diagram of LGA classification in White British infants (LGA, large-for-gestational age; GROW, gestation-related optimal weight; BiB, Born in Bradford).

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, caesarean 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

Open Access

Table 2 Prevalence of outcomes and numbers of overall sample with missing data (n, with % to 1 dp.)

Outcome	Overall prevalence n (%*)	White British prevalence (n=3980) n (%*)	Pakistani prevalence (n=4448) n (%*)	p Value (for difference)	Number with missing data (%, of 8428)
Neonatal outcome					
Shoulder dystocia	85 (1.1)	41 (1.1)	44 (1.1)	0.8	592 (7.0)
Hypothermia	504 (6.9)	190 (5.5)	314 (8.1)	<0.01	1080 (12.8)
Low Apgar score at 1 min	414 (5.0)	201 (5.1)	213 (4.8)	0.6	93 (1.1)
Low Apgar score at 5 min	44 (0.5)	23 (0.6)	21 (0.5)	0.5	102 (1.2)
NNU admission	348 (4.9)	160 (4.7)	188 (5.1)	0.4	1313 (15.6)
Delivery characteristic					
Induction of labour	1734 (20.6)	853 (21.5)	881 (19.8)	0.07	8 (0.1)
Caesarean section	1788 (21.2)	906 (22.8)	882 (19.8)	<0.01	0 (0)
Assistance during labour	864 (10.3)	471 (11.9)	393 (8.8)	<0.01	4 (0.1)
Composites					
Composite-SGA	3100 (50.5)	1567 (53.3)	1533 (47.8)	<0.01	2283 (27.1)
Composite-LGA	2981 (45.2)	1532 (48.6)	1449 (42.1)	<0.01	1830 (21.7)

*Prevalence calculated based on sample size of those with available data for outcome and not all those who had cut-offs produced for.
LGA, large-for-gestational age; NNU, neonatal unit; SGA, small-for-gestational age.

labour 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.

BiB versus GROW
The prediction of shoulder dystocia was also (borderline) significantly better when using the BiB chart compared with the GROW chart. Caesarean section for LGA infants, assistance during labour for SGA deliveries and the LGA-specific composite outcome were also significantly (or borderline significantly) better predicted when using the BiB chart compared with the GROW chart. The GROW chart provided a significantly (or borderline significantly) better prediction for low Apgar score at 5 min admission to NNU for both SGA and LGA infants.

As was the case in Pakistani infants, in White British infants, the BiB chart had a greater predictive ability than both the GROW chart and the UK-WHO chart. Limiting the sample to only term infants had a minimal effect on the predictive ability of the BiB chart, increasing the average AUROC across the 14 outcomes by 0.14% (SD 2.49%) However, as in the Pakistani infants, the lack of predictive ability of all three charts is the principal finding.

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

Table 3 Pakistani origin infants

Outcome	Chart	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUROC	Significance
BiB versus UK-WHO								
Shoulder dystocia	BiB	1.10	50.00	90.40	5.30	99.40	0.70	p=0.01
	UK-WHO		27.30	96.90	8.60	99.20	0.62	
Hypothermia	BiB	8.10	19.70	91.20	16.50	92.80	0.55	p=0.03
	UK-WHO		37.60	78.30	13.20	93.50	0.58	
Low Apgar 1 min	BiB	4.80	13.60	90.60	6.90	95.40	0.52	p=0.02
	UK-WHO		22.50	77.10	4.80	95.10	0.50	
Induction LGA	BiB	19.80	12.70	90.30	24.50	80.70	0.52	p=0.01
	UK-WHO		3.70	96.60	21.60	80.20	0.50	
Caesarean section (LGA)	BiB	19.80	13.70	90.60	26.50	80.90	0.52	p=0.06
	UK-WHO		5.30	97.00	30.70	80.60	0.51	
NNU (SGA)	BiB	5.10	27.10	91.20	14.20	95.90	0.59	p=0.08
	UK-WHO		36.70	77.70	8.10	95.80	0.57	
Composite (LGA)	BiB	42.10	12.80	91.70	52.80	59.10	0.52	p=0.02
	UK-WHO		4.80	97.50	58.80	58.50	0.51	
BiB versus customised								
Hypothermia	BiB	8.10	19.70	91.20	16.50	92.80	0.55	p=0.07
	GROW		32.50	82.40	14.00	93.30	0.57	
Low Apgar 1 min	BiB	4.80	13.60	90.60	6.90	95.40	0.52	p<0.01
	GROW		17.40	81.30	4.50	95.10	0.49	
Induction LGA	BiB	19.80	12.70	90.30	24.50	80.70	0.52	p=0.05
	GROW		5.70	95.50	23.60	80.40	0.51	
Assistance (SGA)	BiB	8.80	10.20	90.40	9.30	91.20	0.50	p<0.01
	GROW		15.00	80.80	7.10	90.80	0.48	
Assistance (LGA)	BiB	8.80	10.40	89.70	9.00	91.20	0.50	p=0.02
	GROW		7.10	95.50	13.20	91.40	0.51	
UK-WHO versus customised								
Shoulder dystocia	UK-WHO	1.10	27.30	96.90	8.60	99.20	0.62	p=0.04
	GROW		38.60	95.70	8.70	99.30	0.67	
Low Apgar 5 min	UK-WHO	0.50	42.90	77.20	0.90	99.60	0.60	p=0.09
	GROW		23.80	81.40	0.60	99.60	0.53	
Assistance (SGA)	UK-WHO	8.80	22.40	77.00	8.60	91.10	0.50	p=0.03
	GROW		15.00	80.80	7.10	90.80	0.48	
Assistance (LGA)	UK-WHO	8.80	3.80	96.60	9.80	91.20	0.50	p=0.02
	GROW		7.10	95.50	13.20	91.40	0.51	
NNU (SGA)	UK-WHO	5.10	36.70	77.70	8.10	95.80	0.57	p<0.01
	GROW		39.40	81.80	10.40	96.20	0.61	
Composite (LGA)	UK-WHO	42.10	4.80	97.50	58.80	58.50	0.51	p=0.03
	GROW		7.00	96.60	60.10	58.80	0.52	

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.

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 4 White British origin infants

Outcome	Chart	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUROC	Significance
BiB versus UK-WHO								
Shoulder dystocia	BiB	1.10	41.50	90.80	4.80	99.30	0.66	p=0.08
	UK-WHO		31.70	92.40	4.50	99.20	0.62	
Low Apgar 1 min	BiB	5.10	10.00	90.40	5.30	94.90	0.50	p<0.01
	UK-WHO		10.40	88.40	4.60	94.80	0.49	
Caesarean section (SGA)	BiB	22.80	9.30	90.30	22.00	77.20	0.50	p=0.06
	UK-WHO		10.50	88.20	20.70	77.00	0.49	
NNU (SGA)	BiB	4.70	19.40	91.00	9.50	95.80	0.55	p=0.02
	UK-WHO		18.80	89.00	7.70	95.70	0.54	
BiB versus customised								
Shoulder dystocia	BiB	1.10	41.50	90.80	4.80	99.30	0.66	p=0.08
	GROW		26.80	95.70	6.50	99.10	0.61	
Low Apgar 5 min	BiB	0.60	13.00	90.40	0.80	99.40	0.51	p=0.07
	GROW		39.10	81.60	1.20	99.60	0.60	
Caesarean section (LGA)	BiB	22.80	14.80	92.10	35.60	78.60	0.53	p=0.08
	GROW		8.40	96.80	43.90	78.20	0.53	
Assistance (SGA)	BiB	11.80	10.00	90.50	12.40	88.20	0.50	p<0.01
	GROW		15.30	81.10	9.80	87.70	0.48	
NNU (SGA)	BiB	4.70	19.40	91.00	9.50	95.80	0.55	p=0.03
	GROW		34.40	82.30	8.70	96.20	0.58	
NNU (LGA)	BiB	4.70	9.40	90.60	4.70	95.30	0.50	p=0.08
	GROW		7.50	95.90	8.20	95.50	0.52	
Composite (LGA)	BiB	48.60	11.70	18.40	59.50	52.60	0.52	p=0.07
	GROW		5.90	96.80	63.40	52.10	0.51	
UK-WHO versus customised								
Low Apgar 5 min	UK-WHO	0.60	17.40	88.40	0.90	99.50	0.53	p=0.09
	GROW		39.10	81.60	1.20	99.60	0.60	
Caesarean section (LGA)	UK-WHO	22.80	13.10	93.70	38.10	78.50	0.53	p=0.06
	GROW		8.40	96.80	43.90	78.20	0.53	
Assistance (SGA)	UK-WHO	11.80	11.90	88.50	12.20	88.20	0.50	p<0.01
	GROW		15.30	81.10	9.80	87.70	0.48	
Assistance (LGA)	UK-WHO	11.80	5.90	91.90	9.00	87.90	0.49	p=0.09
	GROW		3.80	95.60	10.40	88.10	0.50	
NNU (SGA)	UK-WHO	4.70	18.80	89.00	7.70	95.70	0.54	p<0.01
	GROW		34.40	82.30	8.70	96.20	0.58	
NNU (LGA)	UK-WHO	4.70	8.10	92.40	5.00	95.30	0.50	p=0.09
	GROW		7.50	95.90	8.20	95.50	0.52	

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).[38–41] 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

therefore warranted when constructing ethnicity-specific charts because if Pakistani infants are smaller as a result of a deficient in utero environment, resulting from factors associated with a lower SES, producing an ethnic-specific chart may serve to normalise the weight of this group which has experienced a potentially SES-induced suboptimal fetal milieu and is therefore at risk of adverse neonatal outcomes. However, as our analysis revealed, the UK-WHO chart (which would classify more Pakistani infants as at 'risk') did not increase the prediction of adverse outcomes in this group, suggesting that the increased risk of a suboptimal fetal environment did not translate into an increased risk of adverse outcomes.

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 30 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,[49] 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
[1]Centre for Global Health and Human Development, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK
[2]MRC Unit for Lifelong Health & Ageing, University College London, London, UK
[3]Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK
[4]Bradford Teaching Hospitals NHS Foundation Trust, Bradford Royal Infirmary, Bradford, 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 contributed towards data preparation and manuscript editing. DT was involved in data collection, study development and editing of the manuscript. JW is the chief 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.

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. Participants gave informed consent for data sharing and the presented data are anonymised and risk of identification is low.

REFERENCES

1. de Onis M, Garza C, Victora CG, et al. The WHO Multicentre Growth Reference Study: planning, study design, and methodology. *Food Nutr Bull* 2004;25(Supple 1):15S–26S.
2. SACN/RCPCH Expert Group. Application of WHO Growth Standards in UK. 2007.
3. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* 1998;17:407–29.
4. Kelly Y, Panico L, Bartley M, et al. Why does birthweight vary among ethnic groups in the UK? Findings from the Millennium Cohort Study. *J Public Health (Oxf)* 2009;31:131–7.
5. West J, Lawlor DA, Fairley L, et al. UK-born Pakistani-origin infants are relatively more adipose than white British infants: findings from
6. 8704 mother-offspring pairs in the Born-in-Bradford prospective birth cohort. *J Epidemiol Community Health* 2013;67:544–51.
6. Seaton SE, Yadav KD, Field DJ, et al. Birthweight centile charts for South Asian infants born in the UK. *Neonatology* 2011;100:398–403.
7. Ray JG, Sgro M, Mamdani M, et al. Birth weight curves tailored to maternal world region. *J Obstet Gynaecol Can* 2012;34:159–71.
8. Kierans WJ, Joseph K, Luo Z, et al. Does one size fit all? The case for ethnic-specific standards of fetal growth. *BMC Pregnancy Childbirth* 2008;8:1.
9. Hanley GE, Janssen PA. Ethnicity-specific growth distributions for prediction of newborn morbidity. *J Obstet Gynaecol Can* 2012;34:826–9.
10. Hanley GE, Janssen PA. Ethnicity-specific birthweight distributions improve identification of term newborns at risk for short-term morbidity. *Obstet Gynecol* 2013;209:428.e1–e6.
11. Gardosi J, Chang A, Kalyan B, et al. Customised antenatal growth charts. *Lancet* 1992;339:283–7.
12. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull World Health Organ* 1987;65:663–737.
13. Royal College of Obstetricians and Gynaecologists. The investigation and management of the small-for-gestational-age fetus. 2014. http://www.rcog.org.uk/files/rcog-corp/GTG31SGA23012013.pdf (accessed 24 Feb 2014).
14. Clausson B, Gardosi J, Francis A, et al. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. *BJOG* 2001;108:830–4.
15. McCowan LM, Harding JE, Stewart AW. Customised birthweight centiles predict SGA pregnancies with perinatal morbidity. *BJOG* 2005;112:1026–33.
16. Ego A, Subtil D, Grange G, et al. Customized versus population-based birth weight standards for identifying growth restricted infants: a French multicenter study. *Obstet Gynecol* 2006;194:1042–9.
17. Zhang X, Platt R, Cnattingius S, et al. The use of customised versus population-based birthweight standards in predicting perinatal mortality. *BJOG* 2007;114:474–7.
18. Wright J, Small N, Raynor P, et al. Cohort profile: the Born in Bradford multi-ethnic family cohort study. *Int J Epidemiol* 2013;42:978–91.
19. National Institute for Health and Clinical Excellence. Antenatal Care: NICE clinical guideline 62. 2010. http://www.nice.org.uk/nicemedia/live/11947/40115/40115.pdf (accessed 24 Feb).
20. Office for National Statistics. Ethnic group statistics: a guide for the collection and classification of ethnicity data. 2003.
21. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.
22. Cole TJ. Do growth chart centiles need a face lift? *BMJ* 1994;308:641–2.
23. Cole TJ, Williams AF, Wright CM. Revised birth centiles for weight, length and head circumference in the UK-WHO growth charts. *Ann Hum Biol* 2011;38:7–11.
24. Gardosi J, Mongelli M, Wilcox M, et al. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 1995;6:168–74.
25. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991;181:129–33.
26. Cnossen JS, Morris RK, ter Riet G, et al. Use of uterine artery Doppler ultrasonography to predict pre-eclampsia and intrauterine growth restriction: a systematic review and bivariable meta-analysis. *CMAJ* 2008;178:701–11.
27. Challa AS, Evageliadou EN, Cholevas VI, et al. Growth factors and adipocytokines in prepubertal children born small for gestational age: relation to insulin resistance. *Diabetes Care* 2009;32:714–19.
28. Moore GS, Kneitel AW, Walker CK, et al. Autism risk in small- and large-for-gestational-age infants. *Obstet Gynecol* 2012;206:314.e1–e9.
29. Doctor BA, O'Riordan MA, Kirchner HL, et al. Perinatal correlates and neonatal outcomes of small for gestational age infants born at term gestation. *Obstet Gynecol* 2001;185:652–9.
30. Crowther CA, Hiller JE, Moss JR, et al. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med* 2005;352:2477–86.
31. Hung T, Lo L, Chiu T, et al. Risk factors and perinatal outcomes associated with idiopathic small for gestational age Taiwanese newborns. *Int J Gynecol Obstet* 2013;122:212–15.
32. Pan H, Cole TJ. LMSchartmaker Light. 2011.
33. Gardosi J, Francis A. Customised weight centile calculator. 2013.
34. Worster A, Fan J, Upadhye S. Understanding receiver operating characteristic (ROC) curves. *CJEM* 2006;8:19–20.
35. Cha H, Lee S, Park J, et al. Comparison of perinatal outcomes in small-for-gestational-age infants classified by population-based versus customised birth weight standards. *Aust N Z J Obstet Gynaecol* 2012;52:348–55.

36. Carberry AE, Raynes-Greenow CH, Turner RM, *et al*. 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.

37. Hutcheon J, Zhang X, Cnattingius S, *et al*. Customised birthweight percentiles: does adjusting for maternal characteristics matter? *BJOG* 2008;115:1397–404.

38. Davey Smith G, Neaton JD, Wentworth D, *et al*. Mortality differences between black and white men in the USA: contribution of income and other risk factors among men screened for the MRFIT. MRFIT Research Group. Multiple Risk Factor Intervention Trial. *Lancet* 1998;351:934–9.

39. Nazroo JY. South Asian people and heart disease: an assessment of the importance of socioeconomic position. *Ethn Dis* 2001;11:401–11.

40. Karlsen S, Nazroo JY. Relation between racial discrimination, social class, and health among ethnic minority groups. *Am J Public Health* 2002;92:624–31.

41. Nazroo JY. The structuring of ethnic inequalities in health: economic position, racial discrimination, and racism. *Am J Public Health* 2003;93:277–84.

42. Norris T, Tuffnell D, Wright J, *et al*. Modelling foetal growth in a bi-ethnic sample: results from the Born in Bradford (BiB) birth cohort. *Ann Hum Biol* 2014:41:481–7.

43. Pena IC, Teberg AJ, Finello KM. The premature small-for-gestational-age infant during the first year of life: comparison by birth weight and gestational age. *J Pediatr* 1988;113:1066–73.

44. McCarton CM, Wallace IF, Divon M, *et al*. Cognitive and neurologic development of the premature, small for gestational age infant through age 6: comparison by birth weight and gestational age. *Pediatrics* 1996;98(6 Pt 1):1167–78.

45. Gutbrod T, Wolke D, Soehne B, *et al*. Effects of gestation and birth weight on the growth and development of very low birthweight small for gestational age infants: a matched group comparison. *Arch Dis Child Fetal Neonatal Ed* 2000;82:F208–14.

46. Twin Birth Study Collaborative group. The Twin Birth Study: a multicentre randomised controlled trial comparing planned caesarean section with planned vaginal birth for twins at 32–38 weeks gestation. 2005.

47. Catalano PM, McIntyre HD, Cruickshank JK, *et al*. The hyperglycemia and adverse pregnancy outcome study: associations of GDM and obesity with pregnancy outcomes. *Diabetes Care* 2012;35:780–6.

48. National Institute for Health and Clinical Excellence. The guidelines manual. 2012;2014(06/02).

49. Hutcheon J. Do customized birth weight charts add anything but complexity to the assessment of fetal growth? *J Obstet Gynaecol Can* 2014;36:107–13.

BMJ Open

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

---

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-2014-006 743.DC1.html**
References	This article cites 41 articles, 10 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 (980) Obgyn (151) Research methods (240)

Notes