Association between perinatal mortality and morbidity and customised and non-customised birthweight centiles: a comparative record-linkage study in Denmark, Finland, Norway, Wales and England

Fanny Kilpi [1,2], Hayley E Jones [2], Maria C Magnus [1,2,3], Gillian Santorelli [4], Lise Kristine Højsgaard Schmidt [5], Stine Kjaer Urhoj [5], Scott M Nelson [6,7], Derek Tuffnell [4], Robert French [8], Per Magnus [3], Anne-Marie Nybo Andersen [5], Pekka Martikainen [9,10], Kate Tilling [1,2,7], Deborah A Lawlor [1,2,7]*

Author affiliations
1 Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK
2 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
3 Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway
4 Bradford Institute for Health Research, Bradford, UK
5 Department of Public Health, University of Copenhagen, Copenhagen, Denmark
6 School of Medicine, Dentistry and Nursing, University of Glasgow, Glasgow, UK
7 NIHR Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, UK
8 School of Medicine, Cardiff University, Cardiff, UK
9 Population Research Unit, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland
10 The Max Planck Institute for Demographic Research, Rostock, Germany

Fanny Kilpi, research fellow, fanny.kilpi@bristol.ac.uk
Hayley E Jones, associate professor, Hayley.Jones@bristol.ac.uk
Maria C Magnus, researcher, MariaChristine.Magnus@fhi.no
Gillian Santorelli, principal statistician, gillian.santorelli@bthft.nhs.uk
Lise Kristine Højsgaard Schmidt, data manager, liks@sund.ku.dk
Stine Kjaer Urhoj, researcher, stur@sund.ku.dk
Scott M Nelson, professor, scott.nelson@glasgow.ac.uk
Derek Tuffnell, professor, Derek.Tuffnell@bthft.nhs.uk
Robert French, senior research fellow, FrenchR3@cardiff.ac.uk
Per Magnus, professor, PerMinor.Magnus@fhi.no
Anne-Marie Nybo Andersen, professor, amny@sund.ku.dk
Pekka Martikainen, professor, pekka.martikainen@helsinki.fi
Kate Tilling, professor, kate.tilling@bristol.ac.uk
Deborah A Lawlor, professor, d.a.lawlor@bristol.ac.uk

Corresponding author
* Professor Deborah Lawlor, MRC Integrative Epidemiology Unit, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

Abstract

Objectives To compare the risk of adverse perinatal outcomes according to small-for-gestational-age (SGA, <10th centile) and large-for-gestational-age (LGA, >90th centile) defined using non-customised (standardised by sex and gestational age only) and customised (by sex, gestational age, maternal weight, height, parity and ethnicity) birthweight centiles.

Design A comparative population-based record-linkage study with meta-analysis of results from five countries.

Setting Denmark, Finland, Norway, Wales, and England (city of Bradford).

Participants 2,129,782 infants born at term in birth registries.

Main outcome measures Stillbirth, neonatal death, infant death, admission to neonatal intensive care unit (NICU) and low Apgar score (<7) at 5 minutes.

Results Relative to those born average for gestational age (AGA), both SGA and LGA were at increased risk of all five outcomes, but observed relative risks were similar whether non-customised or customised charts were used. For example, for SGA vs AGA, when non-customised and customised charts were used, the relative risks pooled over countries were, (i) for stillbirth 3.60 (95% CI 3.29 to 3.93) vs 3.58 (3.02 to 4.24), (ii) for neonatal death 2.83 (2.18 to 3.67) vs 3.32 (2.05 to 5.56), (iii) for infant death 2.82 (2.07 to 3.83) vs 3.17 (2.20 to 4.56), (iv) for low Apgar score at five minutes 1.66 (1.49 to 1.86) vs 1.54 (1.30 to 1.81) and, based on Bradford data only, (v) for admission to NICU 1.97 (1.74 to 2.22) vs 1.94 (1.70 to 2.21). The estimated sensitivity of combined SGA or LGA to identify the three mortality outcomes ranged from 31 to 34% for non-customised and 34 to 38% for customised charts, with a specificity of 82% and 80% with non-customised and customised charts, respectively.

Conclusions This study found increased risk of adverse perinatal outcomes among SGA or LGA term infants of a similar magnitude when using customised and non-customised centiles. Use of customised SGA/LGA - over and above non-customised SGA/LGA - is unlikely to provide benefits in terms of identifying term-births at risk of these outcomes.

Word count: 324/400

Key words

birthweight; small-for-gestational age, large-for-gestational age; perinatal mortality; stillbirth

What is already known on this topic

- SGA (<10th centile) and LGA (>90th centile) are used to detect fetal growth problems.

- Fetal growth problems may identify fetuses at a higher risk of stillbirth, neonatal mortality and severe morbidity.

What this study adds

- Our study suggests that incorporating maternal weight, height, parity and ethnicity to define SGA or LGA for term-births has no discernible benefit for detecting associations with perinatal mortality or morbidity compared with standardizing for gestational age and infant sex only.

Introduction

Fetal growth problems underlie a high proportion of stillbirths, neonatal mortality and severe morbidity.(1–5) An important aim of antenatal care is the detection and management of such problems to minimise adverse outcomes.(6,7) Birthweight centile charts aim to reflect 'healthy', or 'physiological', growth and these are converted to estimated fetal growth charts used antenatally. Traditionally such charts control for sex and gestational age as these are seen as physiological aspects of fetal growth. However, customised charts adjust for additional characteristics – maternal height, weight, parity and ethnicity – considering these to reflect physiological rather than pathological differences in fetal growth.(8,9) Customised charts aim to indicate the optimal fetal growth potential, against which maternal fundal height measurements and estimated fetal weight (from ultrasound scan) may be plotted throughout pregnancy.

A recent systematic review and meta-analysis that included up to 20 studies, with up to 1,095,589 pregnancies, concluded that both customised and non-customised charts identified increased risk of adverse outcomes, particularly for small-for-gestational-age (SGA) infants, but had insufficient power to robustly determine the difference between the two.(10) Individual studies published since have varied in size from 4,095 to 1.25 million and have come to varying conclusions. Three of these studies, a record-linkage study of ~1 million births from Scotland, a UK pregnancy cohort of ~4000 births to nulliparous women, and a Swedish record-linkage study of over 200,000 births, did not find evidence of differences between non-customised and customised charts in predicting adverse perinatal outcomes in singleton births.(2,11,12) In comparison, a study of 1.25 million singleton term births from 10 countries and a study of ~53,000 singleton pregnancies from New Zealand found that customised centile charts identified more SGA infants and more of those who were at risk of stillbirth than the Intergrowth-21st standards.(13,14) These previous studies have a number of limitations. The large Scottish record linkage study did not have information on maternal ethnicity or weight and was only able to partially customise.(2) This is important as the other large study,(13) and the smaller New Zealand study,(14) suggested that better prediction with customised charts compared to Intergrowth-21st charts may reflect the latter not accounting for the physiological effect of ethnicity on fetal growth. However, the larger of those two studies only explored stillbirth as an outcome,(13) and the smaller had a composite neonatal outcome combining mortality and morbidity.(14)

The aim of our study is to compare how SGA and large-for-gestational-age (LGA) defined according to customised (by fetal sex, gestational age, maternal weight, height, parity and ethnicity) and non-customised birthweight centiles (adjusted for fetal sex and gestational age only) are associated with perinatal mortality and morbidity. We do this using record-linkage data from Denmark, Finland, Norway, Wales, and England, and explore consistency of findings across different populations.

Data and Methods

Data

We used record-linkage data from five countries: national data from Denmark (2004-2010), Finland (2004-2014), Norway (2012-2016) and Wales, UK (1986-2016), and regional data from the city of Bradford in the North of England, UK (2010-2019). Full details of each dataset are available in the Supplementary Material. Information from the separate countries was harmonised, and the analysis sample was restricted to singleton births which occurred between 37 and 43 gestational weeks as in most previous studies to distinguish fetal growth problems from preterm birth effects. We excluded infants with known major congenital anomalies (CA) as causes of death or identified within the first year of life to exclude cases where CA may be the cause of fetal growth restriction, and influenced by some of the customised variables (e.g. BMI). We excluded observations with missing data on sex, birthweight or gestational age at delivery, and with birthweight outside the 0.15-8.00 kg range. We also excluded observations with customisation variables outside the ranges previously suggested for customised charts (maternal height 100-213cm, maternal pre- or early pregnancy weight 30-300kg, parity 0-15).(15) Table S1 (supplementary material) describes the sample selection (number (%) overall excluded: Bradford 3,919 (8), Denmark 49,437 (11), Finland 73,715 (11), Norway 17,628 (6), Wales 94,327 (10)).

Predictors

Non-customised birthweight centiles

As in most previous research, birthweight was used as a proxy to reflect fetal growth across pregnancy. We generated non-customised birthweight centiles that were standardised for infant sex and gestational age using the UK1990 reference population within the 'zanthro' package in Stata to derive z-scores, and then created within-country centiles from the ordering of those scores. For both customised and non-customised centiles, we categorised infants into SGA (<10^{th} centile), average-for-gestational age (AGA; 10^{th}-90^{th} centile) and LGA (> 90^{th} centile).

Customised birthweight centiles

In addition to infant sex and gestational age, the variables used for customisation are maternal ethnicity, height, weight, and parity.(15) Nordic countries in general collect information on country of birth/origin but not self-reported ethnicity. Thus, we based ethnicity on the mother's self-reported ethnicity for Bradford and Wales and on the mother's country of birth/origin for Denmark, Finland, and Norway. Ethnicity categories with small numbers of observations (<100) were collapsed into higher category levels. For example, if there were fewer than 100 pregnancies in women of Bangladeshi origin in a dataset they would be classified as South Asian, or as 'other' if numbers were still <100. Maternal height was reported in centimetres (100-213cm), maternal pre- or early pregnancy weight in kilograms (30-300kg), and parity categorised as (i) 0, (ii) 1, (iii) 2, (iv) 3, and (v) 4 or more.

We used the Perinatal Institute's global centile bulk calculator (version 8.0.4), to calculate customised centiles in Bradford data.(15) Use of the calculator required inputting data into an external server, which was not feasible for other countries due to data governance reasons, so the customised centiles for the other countries were calculated using equations based on the methods published by the Perinatal Institute.(16) Customised birthweight centiles were generated within

each dataset by first regressing birthweight on gestational age at delivery, infant sex, maternal height, weight, parity and ethnicity in complete-case data (i.e. no missing data on variables used in the customisation). Then, we calculated a predicted birthweight for all observations based on gestational age, sex, and available maternal characteristics. If a customisation variable was missing, median weight or height, majority ethnic group or parity = 0 were assumed for the purpose of these predictions. We then calculated whether the observed birthweight fell within the calculated customised SGA and LGA thresholds for each observation's predicted birthweight. Full details of the methods are provided in supplementary methods and the customised coefficients for each dataset in Table S2.

Outcome

We examined the following outcomes in all countries when available: (i) stillbirth (birth of an infant without signs of life), (ii) neonatal death (death within first 28 days of life for liveborn infants), (iii) infant death (death within first year of life for liveborn infants), (iv) low Apgar score at 5 minutes for livebirths (score<7, which includes 4-6= moderately abnormal and 0-3= low, to increase power and precision), and (v) admission to neonatal intensive care unit (NICU) within the first week of life.

Analysis strategy

Main analyses

We used log binomial regression to separately estimate the relative risk (RR) of each outcome comparing SGA and LGA to AGA using the non-customised and customised centiles in each country. We meta-analysed the results across the five countries with a restricted maximum likelihood random-effects model using the 'meta' command in Stata (version 16). As being diagnosed as SGA or LGA alters obstetric care pathways we also calculated the sensitivity and specificity of combined SGA/LGA versus AGA to predict each outcome again based on non-customised and on customised centiles. Sensitivity is calculated by the number of outcome cases identified as SGA/LGA divided by the total number of outcome cases (i.e. true positives/(true positives + false negatives)), and specificity by the number of cases without the outcome identified as AGA divided by the total number of cases without the outcome (i.e. true negatives/(true negatives+false positives)). The command 'metandi' (17) was used to perform bivariate random-effects meta-analysis of sensitivity and specificity for each outcome.

Supplementary analyses

We performed a number of sensitivity analyses: (1) We explored the sensitivity of the results from each dataset to their main sources of missing data either by using more years of data available with only partial customisation (e.g. only customising for ethnicity and parity in data from Denmark and Norway) or checking results when using more restricted datasets with less missingness in customisation variables (Bradford and Wales) (detailed in Supplementary Methods). (2) We meta-analysed data from the three Nordic countries separately from the two UK data sets, on the basis that ethnicity was collected differently between these countries. (3) At the suggestion of a reviewer, we performed a meta-analysis excluding data from Norway as we were not able to exclude those with CA in this data. (4) To explore whether generating our own customised charts within each study might have influenced results, we repeated analyses for Bradford using within-study customisation equations and compared results using those to our main results using the bulk calculator.

Patient and public involvement

The current research uses secondary data and was not informed by patient and public involvement. No patients were involved in setting the research question, developing the study design or analysis plan or setting the outcomes measures. The results will be disseminated to stakeholders and the broader public as relevant. There are no plans to disseminate the results of the research to individual study participants because all participants are deidentified.

Results

A total of 2,129,782 term births were included, with 191,923 identified as SGA and 212,732 as LGA with non-customised centiles and 215,719 as SGA and 217,836 as LGA with customised centiles. Table 1 details the characteristics of the five datasets. Median birthweight was lowest in Bradford (3,300g) and highest in Denmark and Finland (3,570g). The proportions of stillbirth, neonatal and infant death were highest in Bradford (0.24%, 0.10%, 0.20% respectively), and low Apgar score at 5 minutes had highest prevalence in Wales (3.25%). The Welsh data had the highest proportion of nulliparous mothers (45%), and Bradford the lowest (33%).

Figure 1 presents the pooled RR of each adverse perinatal outcome for infants born SGA versus AGA by non-customised and customised definitions of SGA (meta-analysis forest plots are provided in Figure S1 and detailed country-specific estimates in Tables S3A-S3E). For stillbirth (RR 3.60 (95% confidence interval 3.29 to 3.93) vs 3.58 (3.02 to 4.24) for non-customised and customised SGA, respectively), neonatal death (2.83 (2.18 to 3.67) vs 3.32 (2.05 to 5.36)), infant death (2.82 (2.07 to 3.83) vs 3.17 (2.20 to 4.56)) and low Apgar score (1.66 (1.49 to 1.86) vs 1.54 (1.30 to 1.81)) the pooled RR were consistent for non-customised and customised definitions of SGA. In Bradford (the only study with data on NICU admission), the RR of admission to NICU was 2.48 (2.20 to 2.79) with SGA based on non-customised centiles and 2.04 (1.83 to 2.27) based on customised centiles (Table S3A). In Danish analyses, customised SGA had consistently higher relative risks than non-customised SGA for all outcomes, while the reverse was the case for Bradford. In other countries, the results were more mixed. There were varying amounts of heterogeneity in the pooled analyses with heterogeneity statistics up to I^2=94.10% for the analyses of Apgar score by customised SGA.

Figure 2 shows the pooled RRs by non-customised and customised definitions of LGA versus AGA (meta-analysis forest plots are found in Figure S2 and detailed study-specific estimates in Tables S3A-S3E). For stillbirth (1.00 (0.82 to 1.23) vs 1.05 (0.86 to 1.29)) results for both charts were close to the null and consistent with each other. By contrast, for neonatal death (1.54 (0.76 to 3.03) vs 1.11 (0.83 to 1.49)) and infant death (1.27 (0.78 to 2.07) vs 0.90 (0.75 to 1.09)) the RR was greater with the non-customised definition of LGA than for customised LGA, though for both the confidence intervals were wide and included the null value. For Apgar scores (1.29 (1.02 to 1.64) vs 1.33 (1.04 to 1.69)) and NICU admission (in Bradford only), (1.97 (1.74 to 2.22) vs 1.94 (1.70 to 2.21)) LGA was associated with increased risk to a similar magnitude with non-customised and customised centiles. Neither customised nor non-customised LGA was consistently associated with higher risk of the outcomes than the other in any country. Greatest amount of heterogeneity was found for Apgar score by customised LGA (I^2=97.07%).

Figures 3 and 4 (Table S4) show the pooled summary estimates of sensitivity and specificity respectively for combined SGA or LGA versus AGA, based on non-customised and customised centiles and (detailed country-specific estimates are found in Tables S5A – S5E). Results were similar for non-customised and customised charts, with confidence intervals overlapping. For example, for stillbirths, pooled sensitivity with non-customised centiles was 0.34 (0.33 to 0.36) and with customised centiles 0.38 (0.36 to 0.40), and specificity was 0.82 (0.81 to 0.84) and 0.80 (0.78 to 0.82) respectively. Estimated sensitivity to the 5 outcomes ranged from 26-38%. Point estimates of sensitivity tended to be somewhat higher for customised centiles for all mortality outcomes. Slightly higher specificity was observed for non-customised centiles than customised centiles for all outcomes, particularly for NICU (based on Bradford data only).

Supplementary analyses exploring the sensitivity of the results from each dataset to their main sources of missing data (Tables S3A-S3E and S5A-S4E) showed highly similar results to the main analyses. Pooled results comparing the three Nordic datasets to the UK datasets were also similar with the exceptions of associations with neonatal and infant death (Figures S3 and S4). Associations of SGA with these two outcomes were stronger in Nordic countries than in the UK for customised charts, while the associations of LGA with the two outcomes were stronger In Nordic countries than UK for non-customised charts. However, results from these subgroup analyses were imprecise, with wide confidence intervals. Pooled results excluding Norway were similar to the main results (Figures S5 and S6). Comparing two different customisation methods to assign SGA/LGA values in Bradford data, we found that the results from within-study customisation were consistent with those from the bulk calculator, but the wide confidence intervals prevented making strong conclusions (Table S3A).

Discussion

Summary of main findings

In this comparative record-linkage study, we did not find evidence that SGA or LGA identified with customised birthweight centiles differed from non-customised centiles in their association with mortality (stillbirth, neonatal and infant mortality) or morbidity (low Apgar and admission to NICU). However, the low sensitivity for both implies that a high proportion of infants at risk of these outcomes would not be identified by SGA or LGA with either chart.

Study strengths and limitations

We used data from four nationally representative datasets of births (Denmark, Finland, Norway, and Wales) and one regional dataset (Bradford, England), enhancing statistical precision. The record-linkage datasets were large and covered the whole population, so we were able to compare associations between non-customised and customised for rare outcomes such as perinatal mortality with greater precision than previous studies. The nature of our outcomes (mortality and admission to NICU) and the data from high income countries means it is likely that there will be very little, if any, outcome misclassification. To our knowledge this is the largest study to date of these associations, with over two million births it is double the size of a previous systematic review (10) and two previous record linkage studies of ~1million births.(2,13) However, we acknowledge that we still had limited precision for some estimates, with associations of LGA with neonatal and perinatal deaths being particularly imprecise with wide confidence intervals. NICU was also only available for Bradford data and customisation variables were missing for 33% of that data. We restricted our analyses to term births to better distinguish fetal growth problems from preterm birth effects, and to facilitate comparisons with previous studies. However, we acknowledge this could introduce one form of selection bias (healthier sample than those born prematurely) over another (we do not have weight for those continuing in utero and centiles derived from babies born preterm do not reflect healthy fetal growth).(18) There was varying amounts of missing data across different datasets and for some this was considerable (e.g. data from Wales had missing weight for 54%). Nevertheless, results were very similar in sensitivity analyses where we used either the restricted the data or partial customization, and the customisation equation coefficients were similar across different datasets with different amounts of missing data (Table S2).

We focused here on the pooled analyses across all data sources but acknowledge that there were some heterogeneity in estimates across locations. This could be due to differences in ethnicity or other maternal characteristics or may reflect between-country differences in fetal growth monitoring, how adverse perinatal outcomes are screened for, and their risk managed. The data come from different years within each country spanning 1986 to 2019 during which time perinatal mortality has decreased globally.(19) A change over time in the outcome is unlikely to affect associations of exposures with that outcome and neither would it affect the within-country comparisons of non-customised and customised SGA/LGA with the outcomes we have explored here. However, there may be within-country differences in how strongly customised/non-customised centiles associate with the outcomes that are influenced by the different time periods that are covered in each study, for example due to differences and changes in the ethnic composition of the country. This could bias our (pooled) analyses, but seems unlikely as we do not see differences between non-customised and customised SGA/LGA in their associations with outcomes within countries. In the future, if countries become e.g. more ethnically diverse, it is

possible that between-country differences in the strength of the association of customised centile derived SGA/LGA will emerge and that may result in customised SGA/LGA having stronger or weaker associations with adverse perinatal outcomes in different countries.

Comparison with other studies

Consistent with our findings, a systematic review and meta-analysis did not find conclusive evidence of differences between SGA and LGA defined with customised and non-customised charts in predicting several important perinatal outcomes,(10) and three subsequent within-country studies of perinatal morbidity or mortality have also shown no differences.(2,11,12) In the previous meta-analysis,(10) infants defined as SGA according to non-customised and customised charts had increased pooled risks for perinatal death (odds ratio (OR) 4.0 (2.8 to 5.1) vs 5.8 (3.8 to 7.8)), neonatal death (2.9 (1.2 to 4.5) vs 3.5 (1.1 to 8.0)) and NICU admission (2.4 (1.7 to 3.2) vs 3.6 (2.0 to 5.5)) compared with non-SGA infants. The authors concluded that the overlap of confidence intervals suggested little evidence of difference between the two charts but acknowledged that with larger studies differences may become apparent. With double the sample size of that previous study, we had considerably more precision to estimate associations with SGA but found the confidence intervals for non-customised and customised estimates still overlapped. One consistency across all studies to date, including ours, is that associations of SGA, defined by any of the charts, with adverse outcomes are stronger and more consistent than those of LGA.

Two studies examining associations of SGA and/or LGA with perinatal outcomes have compared customised centile charts to the Intergrowth-21st standards, with both concluding that the customised charts identified at risk SGA better than the Intergrowth-21st charts and that the differences between the two were likely related to physiological differences by ethnicity.(13,14) The largest of the two included 1.25 million births from 10 different countries, including Bhutan, China, India, the US and six European countries, with the vast majority (97%) from the US or European countries.(13) The results in relation to SGA were driven by results from India (n=6,436) and Bhutan (n=2,779), which reflect < 1% of the whole cohort. We cannot directly compare our results to those from these studies as we were comparing to general population non-customised reference charts rather than Intergrowth-21st. However, there may be some value in comparing these results to those from our study in the Bradford population (the most ethnically diverse of our data sources), where we found that SGA based on non-customised charts had consistently higher relative risks for all outcomes compared with customised charts, though with overlapping confidence intervals suggesting no statistical difference in associations. Based on these data, birthweight customisation did not appear to better identify a more at-risk group in Bradford, which had the highest proportions of infants with non-White maternal ethnicity and the lowest median birthweight compared to the other countries. The role of ethnicity in customising fetal growth charts remains unclear, and its use in algorithms guiding clinical practice may 'normalise' non-physiological differences that may even result in increasing ethnic inequalities in health.(20) There is a lack of clarity in the customisation literature over how ethnicity should be conceptualized (self-identified ethnicity versus country of origin/birth), which may also be important in relation to the extent to which ethnicity is appropriate to customise on. However, the promotion of customised charts is wide-spread globally and assumed to be measuring the same exposure/predictor. In the Nordic countries, parental country of birth/origin is often used as a measure of ethnicity, which is different to the concept of self-reported ethnicity (or race e.g. in the US) used in other countries. We feel strongly that in any research and

development of global tools, such as customised centile charts, is important to be clear about the meaning and justification of the measure of ethnicity.

Implications of our and other studies to date

The systematic review concluded that, given the limitations of observational studies, randomised controlled trials (RCTs) of the use of customised and non-customised charts in monitoring infant growth and deciding when to intervene were needed.(10) However, given the low incidence of severe adverse outcomes such as stillbirth and neonatal mortality, alternative approaches including natural experiments (21) or smaller adequately powered trials for surrogates (22) are required to determine the efficacy of using customised versus non-customised charts in clinical settings. For example, a natural experiment comparing change in stillbirth rates over time between Scotland, where very few units have adopted 'GAP' (Growth Assessment Protocol), which uses customised charts, to the same in England concluded that there was little evidence to support GAP being more effective at reducing stillbirth than standard care.(21) The first large RCT to compare GAP to standard care, including over 180,000 deliveries from 13 obstetric units in the UK found no difference between GAP and standard care for the primary outcome of antenatal detection of SGA, the proposed route through which GAP is hypothesised to reduce stillbirth.(23) There was evidence of a lower stillbirth rate, which may be a chance finding given no differences in the other 24 secondary outcomes and no effect on the primary outcome that would be the mechanism for reducing stillbirth.(23) It is important to note that the natural experiment and the RCTs do not test whether customised compared to non-customised charts are superior in identifying at risk pregnancies and results for both will be influenced by the effects of 'standard care' in the comparison group. Thus, large observational studies remain the most feasible design for addressing questions about which centile chart best predicts adverse perinatal outcomes.

Our results and those of previous studies suggest that SGA and LGA defined by the bottom and top deciles of birthweight distribution in general have poor general predictive ability for these rare outcomes whether using non-customised or customised charts. We found similar high specificity and low sensitivity for non-customised and customised charts. Taking the pooled point estimate of stillbirths as an example, our findings suggest that non-customised and customised SGA/LGA could identify, on average, 34% and 38% respectively, of those who went on to have a stillbirth, missing over 60% of cases. Recent literature on birthweight centiles shows that no cut-off point for SGA performs well for predicting neonatal morbidity and mortality.(24–27) Our aim was not to develop models to predict outcomes on an individual basis, but to assess whether customised centiles of birthweight are more strongly associated with outcomes than non-customised centiles, at the population level. On an individual level, continuous birthweight combined with multivariable predictive models may have improved predictive accuracy. Our analyses also did not assess wider programs of plotting fetal growth and the benefits of reacting to changes in expected growth rates, such as faltering fundal height measurements, which remains an important area of research. Nevertheless, the use of customised charts in clinical practice requires more resources, so their clinical benefit would need to be demonstrable and balanced with their cost.

In conclusion, we did not find SGA or LGA identified with customised birthweight centiles in term-births to increase the strength of association with adverse perinatal outcomes such as stillbirth and infant mortality over non-customised centiles.

Supporting statements

Ethical approval

Approval to compete this study using the pseudonymized Danish data was approved by the Danish Data Protection Agency through the joint notification of The Faculty of Health and Medical Sciences at The University of Copenhagen (record no 514-0230/18-3000). The data were collected for routine administrative registration purposes and, therefore, informed consent of the participants was not obtained. These register data can be used for scientific purposes under the Danish Data Act. Approval to complete this study in the anonymised Finnish data was provided by the Statistics Finland Board of Ethics (permit TK-53-339- 13). The data were collected for routine administrative registration purposes and, therefore, informed consent of the participants was not obtained. These register data can be used for scientific purposes under the Personal Data Act and the Statistics Act. Statistics Finland anonymised the data prior to providing them to researchers. The use of the Norwegian register data was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway (ref. 2018/1114). The committee provided a waiver of the need for consent for the use of data from registered individuals on the basis of Norwegian legislation. Separate ethical approval was not required for the use of the deidentified data from Bradford by the authorized data analyst. The use of the Welsh data in SAIL was approved by the independent Information Governance Review Panel (IGRP) to ensure the proposal was appropriate and in the public interest. This work uses data provided by patients and collected by the NHS as part of their care and support.

Contributors

DAL conceived the study, and DAL, FK, KT, PM, RF, PM, and AMNY developed the analysis plan. GS conducted the analysis of Bradford data and LKHS and SKU conducted the analysis of Denmark data, and FK conducted analyses in the other datasets and the meta-analyses. HEJ and KT provided advice for the meta-analysis statistical methods. FK and DAL drafted the initial manuscript. All authors contributed to the acquisition, analysis, or interpretation of the data, commented on drafts of the manuscript, and approved the final version. DAL and FK are the guarantors and attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Transparency statement

The lead author (FK) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Funding and role of funding source

This work was supported by the Bristol National Institute of Health Biomedical Research Centre. FK, KT and DAL work in or are affiliated to a Unit that is funded by the UK Medical Research Council (Grant Refs: MC_UU_00011/6 and MC_UU_00011/3) and University of Bristol, and DAL is a National Institute for Health Research Senior Investigator (NF-0616-10102). FK is supported by Economic and Social Research Council grant (ES/T013923/1). PM, PK, AMNA, LKHS and SKU were partly supported by grant 83540 from the NordForsk Research Programme: ”Nordic Programme on Health and Welfare (Wellife)." PM was supported by the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 101019329), the Strategic Research Council (SRC) within the Academy of Finland grants for ACElife (#352543-352572)

and LIFECON (#308247), and grants to the Max Planck – University of Helsinki Center from the Jane and Aatos Erkko Foundation, the Max Planck Society, University of Helsinki, and Cities of Helsinki, Vantaa and Espoo. RF is supported by the UK Administrative Data Research grant ES/V017314/1.

The study was partly funded by the Research Council of Norway through its Centres of Excellence funding scheme (project number 262700). M.C.M. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 947684).

This publication is the work of the authors and DAL and FK will serve as guarantors for the contents of this paper. The funders had no role in designing the study, collecting or analysing data or contributing to writing the paper. The views expressed in this paper are those of the authors and not necessarily any funding body.

Data sharing

The register data can be used for scientific purposes by approved researchers by application to the relevant data-holding authorities of each country. For statistical code, please contact fanny.kilpi@bristol.ac.uk

Competing interests

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/disclosure-of-interest/ and declare: no support from any other organisation for the submitted work aside from the listed funding sources; DJT reports medicolegal work not related to this specific study; KT reports acting as Expert Witness to the High Court in England, called by the UK MHRA, defendants in a case on hormonal pregnancy tests and congenital anomalies 2021/22; SMN reports consulting and presenting fees from Merck, Modern Fertility, Ferring Pharmaceuticals, TFP, Roche Diagnostics, Access Fertility and payment for expert testimony from various legal practices, and participation on board for HELLP Trial and stocks or stock options for TFP; no other relationships or activities that could appear to have influenced the submitted work.

Copyright statement

Acknowledgments

This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who make anonymised data available for research.

TABLE 1. Descriptive characteristics of the study populations

	BRADFORD, ENGLAND	DENMARK	FINLAND	NORWAY	WALES
N	47,583	384,885	576,758	276,078	844,478
Years	2010-2019	2004-2010	2004-2014	2012-2016	1986-2016
Median birthweight in g (IQR)	3,300 (2,990-3,630)	3,570(3,250-3,900)	3,570 (3,270-3,890)	3,570 (3,260-3,890)	3,430 (3,120-3,760)
Median gestational age in days (IQR)	279 (273-285)	281 (274-287)	281 (275-287)	281 (275-287)	280 (273-287)
OUTCOMES					
Stillbirth, N (%)	112 (0.24)	516 (0.13)	538 (0.09)	370 (0.13)	1,102 (0.13)
Neonatal death, N (%)	47 (0.10)	236 (0.06)	139 (0.02)	129 (0.05)	244 (0.03)
Infant death, N (%)	93 (0.20)	365 (0.09)	351 (0.06)	242 (0.09)	625 (0.07)
Apgar score <7 at 5 min, N (%)	427 (0.90)	1,874 (0.49)	7,498 (1.53)	2,699 (0.98)	19,030 (3.25)
NICU admission, N (%)	1,849 (3.89)	NA	NA	NA	NA
MATERNAL CHARACTERISTICS					
Minority ethnicity, N (%)	21,126 (44)	53,609 (14)	52,532 (9)	75,421 (28)	29,875 (8)
Missing ethnicity (%)	26*	<1	<1	1	53
Median height in cm (IQR)	162 (157-166)	168 (164-172)	165 (162-170)	167 (163-171)	163 (159-168)
Missing height (%)	8	6	2	26	53
Median weight in kg (IQR)	66 (58-76)	65 (59-75)	64 (57-73)	65 (58-74)	64 (57-74)
Missing weight (%)	6	6	5	29	54
No previous births, N (%)	15,748 (33)	164,484 (43)	238,462 (41)	115,763 (42)	320,532 (45)
Missing parity (%)	1	0	<1	0	15

IQR – Interquartile range, NA – Not available * Assigned “global average” in customisation main analyses.

FIGURE 1. Pooled risk ratio estimates for perinatal adverse outcomes by SGA (<10th vs 10-90th) with non-customised and customised birthweight centiles (N=2,129,782 births, 1986-2019)

Note: Pooled results from Bradford, England (N=47,583, 2010-2019), Denmark (N=384,885, 2004-2010), Finland (N=576,758, 2004-2014), Norway (N=276,078, 2012-2016) and Wales (N=844,478, 1986-2016).Number of cases/birth included: stillbirth 2,683/2,129,782, neonatal death 795/2,127,697, infant death 1,676/2,127,697, Apgar score 31,633/1,780,158, NICU 1,849/47,471.

FIGURE 2. Pooled risk ratio estimates for perinatal adverse outcomes by LGA (>90th vs 10-90th) with non-customised and customised birthweight centiles (N=2,129,782 births, 1986-2019)

Note: Pooled results from Bradford, England (N=47,583, 2010-2019), Denmark (N=384,885, 2004-2010), Finland (N=576,758, 2004-2014), Norway (N=276,078, 2012-2016) and Wales (N=844,478, 1986-2016). Number of cases/birth included: stillbirth 2,683/2,129,782, neonatal death 795/2,127,099, infant death 1,676/2,127,099, Apgar score 31,633/1,780,158, NICU 1,849/47,471.

FIGURE 3. Pooled sensitivity of SGA and LGA (<10th and >90th vs 10-90th) for adverse perinatal outcomes with non-customised and customised birthweight centiles (N=2,129,782 births, 1986-2019)

Note: Pooled results from Bradford, England (N=47,583, 2010-2019), Denmark (N=384,885, 2004-2010), Finland (N=576,758, 2004-2014), Norway (N=276,078, 2012-2016) and Wales (N=844,478, 1986-2016). Number of cases/birth included: stillbirth 2,683/2,129,782, neonatal death 795/2,127,099, infant death 1,676/2,127,099, Apgar score 31,633/1,780,158, NICU 1,849/47,471.

FIGURE 4. Pooled specificity of SGA and LGA (<10th and >90th vs 10-90th) for adverse perinatal outcomes with non-customised and customised birthweight centiles (N=2,129,782 births, 1986-2019)

Note: Pooled results from Bradford, England (N=47,583, 2010-2019), Denmark (N=384,885, 2004-2010), Finland (N=576,758, 2004-2014), Norway (N=276,078, 2012-2016) and Wales (N=844,478, 1986-2016). Number of cases/birth included: stillbirth 2,683/2,129,782, neonatal death 795/2,127,099, infant death 1,676/2,127,099, Apgar score 31,633/1,780,158, NICU 1,849/47,471.

References

1. Flenady V, Koopmans L, Middleton P, Frøen JF, Smith GC, Gibbons K, et al. Major risk factors for stillbirth in high-income countries: A systematic review and meta-analysis. Lancet. 2011 Apr 16;377(9774):1331–40.

2. Iliodromiti S, Mackay DF, Smith GCS, Pell JP, Sattar N, Lawlor DA, et al. Customised and Noncustomised Birth Weight Centiles and Prediction of Stillbirth and Infant Mortality and Morbidity: A Cohort Study of 979,912 Term Singleton Pregnancies in Scotland. Myers JE, editor. PLOS Med. 2017 Jan 31;14(1):e1002228.

3. Wilcox AJ. On the importance—and the unimportance— of birthweight. Int J Epidemiol. 2001 Dec;30(6):1233–41.

4. Katz J, Lee ACC, Kozuki N, Lawn JE, Cousens S, Blencowe H, et al. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: A pooled country analysis. Lancet. 2013;382(9890):417–25.

5. Battaglia FC, Lubchenco LO. A practical classification of newborn infants by weight and gestational age. J Pediatr. 1967;71(2):159–63.

6. Intrapartum care for women with existing medical conditions or obstetric complications and their babies [P] Evidence review for small-for-gestational age baby NICE guideline NG121 [Internet]. 2019 [cited 2021 Apr 27]. Available from: https://www.nice.org.uk/guidance/ng121/evidence/evidence-review-p-smallforgestationalage-baby-pdf-241806242779

7. Intrapartum care for women with existing medical conditions or obstetric complications and their babies [Q] Evidence review for large-for-gestational age baby NICE guideline [Internet]. 2018 [cited 2021 Apr 27]. Available from: https://www.nice.org.uk/guidance/ng121/documents/evidence-review-17

8. Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. Lancet. 1992 Feb 1;339(8788):283–7.

9. Gardosi J, Francis A, Turner S, Williams M. Customized growth charts: rationale, validation and clinical benefits. Am J Obstet Gynecol. 2018 Feb 1;218(2):S609–18.

10. Chiossi G, Pedroza C, Costantine MM, Truong VTT, Gargano G, Saade GR. Customized vs population-based growth charts to identify neonates at risk of adverse outcome: systematic review and Bayesian meta-analysis of observational studies. Vol. 50, Ultrasound in Obstetrics and Gynecology. 2017. p. 156–66.

11. Sovio U, Smith GCS. The effect of customization and use of a fetal growth standard on the association between birthweight percentile and adverse perinatal outcome. Am J Obstet Gynecol. 2018 Feb 1;218(2):S738–44.

12. Vieira MC, Relph S, Persson M, Seed PT, Pasupathy D. Determination of birth-weight centile thresholds associated with adverse perinatal outcomes using population, customised, and Intergrowth charts: A Swedish population-based cohort study. Smith GC, editor. PLOS Med. 2019 Sep 20;16(9):e1002902.

13. Francis A, Hugh O, Gardosi J. Customized vs INTERGROWTH-21 st standards for the assessment of birthweight and stillbirth risk at term. Am J Obstet Gynecol. 2018 Feb 1;218(2):S692–9.

14. Anderson NH, Sadler LC, McKinlay CJD, McCowan LME. INTERGROWTH-21st vs customized

birthweight standards for identification of perinatal mortality and morbidity. Am J Obstet Gynecol. 2016 Apr 1;214(4):509.e1-509.e7.

15. Gardosi J, Williams A, Hugh O, Francis A. Customised Centile Calculator. GROW version 8.0.4. Gestation Network;

16. GROW documentation [Internet]. [cited 2021 Apr 29]. Available from: https://www.gestation.net/GROW_documentation.pdf

17. Harbord RM, Whiting P. Metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. Stata J. 2009;9(2):211–29.

18. Pearce N, Lawlor DA. Causal inference-so much more than statistics. Vol. 45, International Journal of Epidemiology. Oxford University Press; 2016. p. 1895–903.

19. Hug L, You D, Blencowe H, Mishra A, Wang Z, Fix MJ, et al. Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: a systematic assessment. Lancet. 2021 Aug 28;398(10302):772–85.

20. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. N Engl J Med. 2020 Aug 27;383(9):874–82.

21. Iliodromiti S, Smith GCS, Lawlor DA, Pell JP, Nelson SM. UK stillbirth trends in over 11 million births provide no evidence to support effectiveness of Growth Assessment Protocol program. Ultrasound Obstet Gynecol. 2020;55(5).

22. Relph S, Elstad M, Coker B, Vieira MC, Moitt N, Gutierrez WM, et al. Using electronic patient records to assess the effect of a complex antenatal intervention in a cluster randomised controlled trial—data management experience from the DESiGN Trial team. Trials. 2021 Dec 1;22(1):12.

23. Vieira MC, Relph S, Muruet-Gutierrez W, Elstad M, Coker B, Moitt N, et al. Evaluation of the Growth Assessment Protocol (GAP) for antenatal detection of small for gestational age: The DESiGN cluster randomised trial. PLOS Med. 2022 Jun 1;19(6):e1004004.

24. Hua X, Shen M, Reddy UM, Buck Louis G, Souza JP, Gülmezoglu AM, et al. Comparison of the INTERGROWTH-21st, National Institute of Child Health and Human Development, and WHO fetal growth standards. Int J Gynecol Obstet. 2018 Nov 1;143(2):156–63.

25. Wilcox AJ, Cortese M, McConnaughey DR, Moster D, Basso O. The limits of small-for-gestational-age as a high-risk category. Eur J Epidemiol. 2021 Oct 1;36(10):985–91.

26. Yang L, Waldhoer T. Re: The limits of small-for-gestational-age as a high-risk category. Eur J Epidemiol. 2022 Mar 1;37(3):305–6.

27. Kamphof HD, Gordijn SJ, Ganzevoort W, Verfaille V, Offerhaus PM, Franx A, et al. Associations of severe adverse perinatal outcomes among continuous birth weight percentiles on different birth weight charts: a secondary analysis of a cluster randomized trial. BMC Pregnancy Childbirth. 2022 Dec 1;22(1):1–11.