Fitness of INTERGROWTH-21st birthweight standard for Chinese-ethnicity babies

Authors: Xue-Lian Wang[1,2], PhD, Lai Ling Hui[3], PhD, Tim J. Cole[4], PhD, Edmund Anthony Severn Nelson[5,2], MD, Hugh Simon Lam[2,6], MD

Affiliations: [1]Department of Neonatology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangdong Provincial Clinical Research Center for Child Health, Guangzhou, China; [2]Department of Paediatrics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China; [3]Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hong Kong SAR, China; [4]University College London Great Ormond Street Institute of Child Health, London, United Kingdom; [5]School of Medicine, The Chinese University of Hong Kong Shenzhen, Guangdong, China; and [6]Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong, Hong Kong SAR, China.

Address correspondence to: Prof. Lam Hugh Simon, Department of Paediatrics, 6/F Lui Che Woo Clinical Sciences Building, Prince of Wales Hospital, Sha Tin, New Territories, Hong Kong SAR, China, [hshslam@cuhk.edu.hk], +852 35052851

Conflict of interest:

No financial or non-financial benefits have been received or will be received from any party related directly or indirectly to the subject of this article. None of the authors was involved in the journal's review of or decisions related to this manuscript.

Data availability statement:

The data that support the findings of this study are available from the Hong Kong Hospital Authority Data Collaboration Laboratory, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of the Hong Kong Hospital Authority Data Collaboration Laboratory.

CRediT authorship contribution statement:

LHS: conceptualization, methodology, project administration, resources, writing - review &

Abstract

Objective. To determine the fitness of the INTERGROWTH-21st birthweight standard (INTERGROWTH21) for ethnic Chinese babies compared to a local reference (FOK2003).

Design. Population-based analysis of territory-wide birth data.

Setting. All public hospitals in Hong Kong.

Participants. Live births between 24 and 42 complete weeks' gestation during 2006–2017.

Main outcome measures. Babies' birthweight Z-scores were calculated using published methods. The two references were compared in three aspects: 1) the proportions of large- or small-for-gestational-age (SGA) infants, 2) the gestation- and sex-specific mean birthweight Z-scores, and 3) the predictive power for SGA-related complications.

Results. 488,896 infants were included. Using INTERGROWTH21, among neonates born <33 weeks' gestation, the mean birthweight Z-scores per week were closer to zero (-0.2 to 0.05), while most of them were further from zero (0.06 to 0.34) after excluding infants with a high risk of abnormal intrauterine growth. Compared to FOK2003, INTERGROWTH21 classified smaller proportions of infants as SGA (8.3% vs. 9.6%) and LGA (6.6% vs. 7.9%), especially SGA among preterm infants (13.1% vs. 17.0%). The area under the receiver operating characteristic curve for predicting SGA-related complications was larger with FOK2003 (0.674, 95% CI: 0.670-0.677) than INTERGROWTH21 (0.658, 95% CI: 0.655-0.661) (P<0.001).

Conclusions. The INTERGROWTH21 performed less well than FOK2003, a local reference for ethnic Chinese babies, especially in infants born <33 weeks' gestation. Although the differences are clinically small, both these references performed poorly for extremely preterm infants, and thus a more robust chart based on a larger sample of appropriately selected infants

is needed.

What is already known on this topic?

The INTERGROWTH-21st Project has published newborn size standards and encouraged diverse localities to adopt universal growth standards.

There may be factors beyond optimal nutrition and environmental constraints that influence birth weight in different populations.

The performance of the INTERGROWTH-21st birthweight standard should be evaluated in a contemporary local population before implementation is considered.

What this study adds?

Our local reference based on data from 20 years ago statistically outperforms the INTERGROWTH-21st birthweight standard in a contemporary cohort of babies born in Hong Kong, although the differences are clinically small.

Neither the local reference nor the INTERGROWTH-21st standard performed well for extremely preterm infants.

How this study might affect research, practice or policy?

Apart from the impact of local factors, the small sample sizes involved for reference construction for extremely preterm infants can lead to large standard errors at the extreme centiles and affect performance.

It is unclear whether the local factors affecting birth weight, such as ethnicity, optimal nutrition and environmental constraints, are modifiable, and thus a local reference based on a large contemporary cohort should be explored.

Introduction

Birth weight is the most important anthropometric measure assessed in newborns. Abnormal birth weight is not only associated with neonatal mortality and morbidity but is also linked to long-term health[1-9]. Birth weight charts are used for the classification of neonatal size and identification of high-risk infants, i.e., those born small for gestational age (SGA, <10th percentile) or large for gestational age (LGA, >90th percentile). SGA babies are at a higher risk of neonatal death, low Apgar scores, bronchopulmonary dysplasia, short-term complications (hypoglycemia, hypothermia, polycythemia, thrombocytopenia, and neutropenia), and long-term complications (suboptimal growth, neurodevelopmental impairment, cardiovascular diseases, and metabolic syndrome)[5-18]. LGA babies have higher risks of suffering from birth injuries, birth asphyxia, persistent pulmonary hypertension of the newborn, and childhood and adulthood overweight or obesity[1-4,19-21]. Appropriately classifying these high-risk infants helps ensure they receive appropriate clinical care.

In Hong Kong, clinicians have been using a local birth weight reference developed by Fok *et al.* (FOK2003) for the past 20 years[22]. Recently, the INTERGROWTH-21st Project has published newborn size standards (INTERGROWTH21) through prospective studies among a multiethnic population and encouraged diverse localities to adopt the universal growth standards[23,24]. Conceptually, with externally validated growth standards, suboptimal birth weight and subsequent growth of infants born into impoverished populations could be more easily identified[23-26]. Further, poor/excessive growth could be compared between populations, guiding research into associated factors and possible interventions that could be applied on a public health scale. However, many investigators have demonstrated ethnic and geographic

differences in birth weight, which were also seen in INTERGROWTH-21st project participants[24]. In a study in New Zealand, INTERGROWTH21 classified significantly fewer infants as SGA (4.5% vs. 11.6%) than a customized birth weight standard and performed worse in predicting neonatal outcomes of death, neonatal intensive care unit (NICU) admission, positive pressure respiratory support, and 5-minute Apgar score<7[26]. Similar studies have not been performed studying the fitness of INTERGROWTH21 for ethnic Chinese newborns. In the present study, we aim to use population-based big data to determine how well INTERGROWTH21 statistically "fits" a contemporary cohort of Hong Kong-born neonates compared with a local reference.

Methods

Ethics approval was obtained from the Joint The Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee. Neonates born between 24 and 42 complete weeks' gestation during 01/01/2006-31/12/2017 were included if they were 1) born in the public hospitals in Hong Kong or 2) born outside hospital (e.g., at home, in transit) but admitted to one of the public hospitals in Hong Kong shortly after birth, with birth weight recorded in the obstetrics dataset. Babies were excluded if 1) the gestational age was uncertain, or 2) the gestational age estimated by maternal self-report last menstrual period differed from the estimation by antenatal ultrasound or postnatal new Ballard score by more than one week. The following data were extracted from the electronic health records: sex, year of birth, maturity, mode of delivery, birth weight, Apgar scores at 1st and 5th minutes, diagnoses, laboratory results of complete blood count (CBC) and plasma glucose, and clinical notes. Maternal data were extracted and linked to the infants' data, including age, diagnoses, and CBC results.

FOK2003 and INTERGROWTH21 were developed based on the LMS (Lambda-Mu-Sigma) method or its variations. To calculate the birth weight Z-scores, the sex- and gestational week-specific L (power in Box-Cox transformation), M (median), and S (coefficient of variation) values of the two references were calculated using the methods provided by Cole[27]. The L, M, and S values within each gestational week (gestational day-specific) were calculated using 'aspline' applying the 'akima' package in R. Individual birth weight Z-scores were calculated using the following equation:

$$Z\text{-score} = ((\text{individual birth weight value}/M)^L - 1)/(L*S)$$

The birth weight Z-score was calculated for each baby using FOK2003 and INTERGROWTH21, respectively. Infants with Z-scores less than -1.282 (10th percentile) were classified as SGA, and those with Z-scores over 1.282 (90th percentile) were classified as LGA.

Infants were considered to have complications associated with SGA if one or more of the following conditions were identified: 1) at least one episode of hypoglycemia before the first discharge home; 2) neonatal polycythemia detected within 48 hours of birth; 3) history of IUGR; 4) hypothermia; 5) neutropenia in the first week; 6) thrombocytopenia in the first week. Neonatal hypoglycemia was defined as a plasma glucose <1.65mmol/L in the first 24 hours of life and <2.5mmol/L thereafter[28]. Neonatal polycythaemia was defined as a venous haematocrit ≥65%. Neutropenia was defined as an absolute neutrophil count $<1\times10^9$/L. Thrombocytopenia was defined as a platelet count $<100\times10^9$/L. IUGR and hypothermia were identified based on maternal and infant diagnostic information.

R version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) was used for

statistical analysis. Data were summarized by mean (standard deviation (SD)), median (interquartile range (IQR)), or number (percentage), as appropriate. The Chi-square test was used to compare proportions, and the t-test was used to compare means. The sensitivity, specificity, diagnostic accuracy, and area under the receiver operating characteristics curves (AUROC) were compared between the two references in predicting SGA-related complications. Net reclassification improvement (NRI) is a statistical measure that represents the proportion of items that have been correctly reclassified. Integrated discrimination improvement (IDI) measures the impact of new classification methods on a prediction model. In our study, NRI and IDI were calculated to quantitatively assess the improvement in risk prediction of SGA-related complications by one reference compared to another[29]. P<0.05 was considered statistically significant.

Results

488,896 babies were included (254,843 boys and 234,053 girls). Infant characteristics are presented in Table 1. L, M, and S values were calculated for the two references (Supplementary Table), and the gestation- and sex-specific mean birth weight Z-scores, as well as the proportions of infants classified as SGA and LGA by the two references, were calculated (Table 2). The mean birth weight Z-scores shifted to negative at almost all gestational weeks in both boys and girls by both references. The shift was similar between the two references in late preterm and term neonates but more obvious by FOK2003 in neonates born <34 weeks' gestation. In neonates born <33 weeks' gestation, mean Z-scores calculated by INTERGROWTH21 were close to zero in both girls and boys (-0.21 to 0.05).

Compared to FOK2003, INTERGROWTH21 identified fewer boys and girls as SGA (boys: 7.8% vs. 9.3%, girls: 8.9% vs. 10%) or LGA (boys: 6.5% vs. 7.5%, girls: 6.6% vs 8.2%). Stratified by gestational age and sex, the ratio of the number of infants classified as SGA by INTERGROWTH21 to that by FOK2003 was, among neonates born <28 weeks' gestation, 0.549 (73/133) in boys and 0.831 (69/83) in girls, among neonates born at 28-39 weeks' gestation, 0.758 (13,624/17,982) in boys and 0.829 (14,467/17,459) in girls, and among neonates born >39 weeks' gestation, 1.11 (6,275/5,652) in boys and 1.072 (6,273/5,853) in girls.

13,723 (5.4%) boys and 12,984 (5.5%) girls had complications associated with SGA by our *a priori* criteria (Table 1). In infants identified as SGA by FOK2003, 5,141 (21.6%) boys and 5,482 (23.4%) girls had the relevant complications, accounting for 37.5% and 42.2%, respectively, of all the babies with the complications. In infants identified as SGA by INTERGROWTH21, 4,501 (22.5%) boys and 5,092 (24.5%) girls suffered from the relevant complications, accounting for 32.8% and 39.2%, respectively, of all the babies with the complications. In infants identified as appropriate for gestational age (AGA) by FOK2003, 7,347 (3.5%) boys and 6,565 (3.4%) girls had the above complications. In infants identified as AGA by INTERGROWTH21, 8,094 (3.7%) boys and 7,008 (3.5%) girls had the above complications. The diagnostic accuracy for these complications by the two references is presented in Table 3. Sensitivity was higher using FOK2003, while specificity was higher using INTERGROWTH21. The AUROC by FOK2003 was larger (0.674, 95% CI 0.670-0.677) than that by INTERGROWTH21 (0.658, 95% CI 0.655-0.661) (P<0.001), which was the case in both boys and girls (Table 3). Stratified by gestational age, the AUROC by FOK2003 was greater at every gestational week in boys except for 25 weeks than INTERGROWTH21. For

girls, the AUROC was similar between the two references in preterm neonates and better by FOK2003 in term neonates (Table 3). Applying INTERGROWTH21 instead of FOK2003, for the overall population, the NRI was -0.037 (95% CI: -0.045 to -0.03) and IDI was -0.0031 (95% CI: -0.0039 to -0.0024); for boys, the NRI was -0.05 (95% CI: -0.06 to -0.04) and IDI was -0.0042 (95% CI: -0.0053 to -0.0031); and for girls, the NRI was -0.025 (95% CI: -0.035 to -0.016) and IDI was -0.002 (95% CI: -0.003 to -0.001), all of which suggest worse performance by INTERGROWTH21 compared with FOK2003.

After excluding infants with a high risk of abnormal intrauterine growth, i.e., those born of multiple pregnancies, with congenital anomalies, with a history of IUGR or excessive in utero fetal growth, and born to mothers with medical conditions, e.g., chronic smoker, drug abuse, diabetes, hypertension, and chronic diseases, the gestation- and sex-specific mean birth weight Z-scores were larger and closer to zero (Figure 1). Using FOK2003, mean Z-scores were close to zero in neonates born at 28-41 weeks' gestation in both girls and boys. Using INTERGROWTH21, compared to FOK2003, the mean Z-scores seem closer to zero at 24-25 weeks' gestation in both boys (P=0.087) and girls (P=0.096), while further away from zero at 26-32 weeks' gestation for boys (P<0.001) and girls (P<0.001). Overall, the mean Z-scores by FOK2003 were closer to zero than by INTERGROWTH21 (-0.014 vs. -0.058, P<0.001), indicating a better fit (Figure 1).

Discussion

Preterm birth and SGA share many common causes and determinants. Infants with suboptimal growth in utero tend to be born preterm. Therefore, theoretically, babies born preterm should

have more than 10% of SGA and less than 10% of LGA, while those born at optimal gestation, i.e., 39, 40, or 41 weeks, will have approximately 10% of SGA or LGA. In our study, significantly more preterm neonates were classified as SGA by FOK2003 than by INTERGROWTH21. In preterm infants, for both boys and girls, the proportions of SGA classified by INTERGROWTH21 were close to or less than 10% at most of the gestational weeks, while those by FOK2003 were close to or less than 10% only at 24-25 weeks. Regarding the predictive power for SGA-related complications and the sex- and gestation-specific birth weight Z-score in low-risk infants, the local FOK2003 showed a better fit than INTERGROWTH21 further.

The ethnic differences in fetal growth are not negligible. However, despite a better fit of the local reference, when applying INTERGROWTH21 to low-risk infants born ≥33 weeks' gestation, the sex- and gestation-specific birth weight Z-scores were mostly around zero and similar to FOK2003, which may be due to the strict inclusion criteria and large sample sizes for constructing INTERGROWTH21 in this group of babies. In babies born <33 weeks' gestation, more high-risk infants, e.g., those born to mothers with pre-eclampsia, were included and the sample size per week was much smaller, leading to less robust centiles[23,24]. Nevertheless, the differences detected in our study regarding the proportions of infants classified as SGA and the AUROCs of predicting the selected complications associated with SGA are clinically small. Whether such birth weight differences may be due to non-modifiable factors such as genetics and ethnicity, or potentially modifiable factors such as nutrition and environmental constraints is unclear. Although infants of Indian ethnicity were usually reported to be smaller than Caucasian infants, in a study in an affluent region of India, among a selected population with a

quality residential environment and relatively higher parental education level as well as family income, the mean birth weight of healthy newborns was similar with those reported in healthy Caucasian infants (3.12kg vs. 3.3-3.5kg) as well as in our study[24,26,30]. From this point of view, if the impact of non-modifiable factors is minimal, the universal adoption of international standards would be of great importance to improve global neonatal health, especially for developing countries and low-resource settings. However, INTERGROWTH21 failed to provide reliable references for very preterm neonates, which therefore limits its application in developed regions.

Some researchers have suggested using customized birth weight centiles that adjusted for maternal ethnicity, height, weight, and parity instead of population centiles[26,31]. Although it is rational to consider maternal factors, it is unclear whether the mother's growth reflects her maximum genetic potential or has been influenced by medical factors. In a population-based study in Scotland, the customized references did not show any advantage in predicting adverse outcomes of infant death, low Apgar scores, and admission to neonatal unit[32].

Both FOK2003 and INTERGROWTH21 performed poorly in extremely preterm infants, which may be mainly due to the small sample sizes for centile construction at these gestational weeks, resulting in large standard errors for centile values. Cole recommended a sample size of 7000-25,000 per sex for constructing growth reference centile charts, which would derive a Z-score for standard error at the $10^{th}/90^{th}$ centile of 0.03 to 0.05[33]. Because of the very low birth rate, multicentre studies based on big data may be needed to recruit a sufficient sample size for these small babies. Additionally, as preterm babies are often born to pregnancies with complications that might influence birth weight, proper exclusion of high-risk infants may be necessary to

derive robust centiles. Otherwise, the optimal birth weight may be underestimated. Recent national and international studies have found lower extreme centiles, i.e., the 3rd, 10th, 90th, and 97th centiles, in more recently established birth weight references among preterm infants, which may be due to the advanced perinatal care allowing more severe infants to survive and a lack of proper exclusion of the high-risk infants[34-36].

Although it is a common way to compare growth references in the aspect of predictive power for adverse outcomes regarding SGA, there has been no consensus regarding which indicators are most appropriate[26,31]. Severe outcomes, such as neonatal death, may have significantly weakened associations with SGA in high-resource settings with skilled birth attendants and neonatologists compared to low-resource settings, especially for minor or moderate SGA (i.e., >3rd percentile and <10th percentile)[37]. Some preterm-related complications, e.g., necrotizing enterocolitis, respiratory distress syndrome, and intraventricular haemorrhage, are associated with SGA which was more commonly seen in moderate or late preterm neonates rather than very or extremely preterm neonates[5,10-12,16,38]. Other conditions, including NICU admissions and low Apgar scores, were more sensitive to preterm birth rather than SGA, thus the larger the proportion of more immature babies classified as SGA, the greater the predictive power for the complications would be for the corresponding reference[37,39]. Complications selected in our study were all reported to be strongly associated with SGA regardless of gestational age[6,10-15,40]. Further, our study is population-based with a very large sample size, thus we believe that the results were reliable and robust.

Conclusion

FOK2003, a local birth weight reference fitted better for ethnic Chinese babies than INTERGROWTH21, especially in infants born <33 weeks' gestation. Although the differences are clinically small, both references performed poorly in extremely preterm infants. Thus, a more robust birth weight chart based on a larger sample of appropriately selected infants, especially at the lower gestations is needed.

Acknowledgment:

References

1.   Modzelewski J, Pokropek A, Jakubiak-Proć M, Muzyka-Placzyńska K, Filipecka-Tyczka D, Kajdy A, et al. Large-for-gestational-age or macrosomia as a classifier for risk of adverse perinatal outcome: a retrospective cross-sectional study. J Matern-Fetal Neo M. 2021:1-8.

2.   Ye R, Pei L, Ren A, Zhang Y, Zheng X, Liu J-m. Birth Weight, Maternal Body Mass Index, and Early Childhood Growth: A Prospective Birth Cohort Study in China. Journal of Epidemiology. 2010;20(6):421-428.

3.   Fang F, Zhang Q-Y, Zhang J, Lei X-P, Luo Z-C, Cheng H-D. Risk factors for recurrent macrosomia and child outcomes. World Journal of Pediatrics. 2019;15(3):289-296.

4.   Mehta SH, Kruger M, Sokol RJ. Being too large for gestational age precedes childhood obesity in African Americans. American Journal of Obstetrics & Gynecology. 2011;204(3):e261-265.

5.   Aldana-Aguirre JC, Toye J, Shah PS, Yoon EW, Kumaran K. Patent ductus arteriosus and small for gestational age infants: Treatment approaches and outcomes. Early Human Development. 2019;131:10-14.

6.   Christensen RD, Yoder BA, Baer VL, Snow GL, Butler A. Early-Onset Neutropenia in Small-for-Gestational-Age Infants. Pediatrics. 2015;136(5):e1259-e1267.

7.   Whincup PH, Kaye SJ, Owen CG, Huxley R, Cook DG, Anazawa S, et al. Birth Weight and Risk of Type 2 Diabetes: A Systematic Review. JAMA. 2008;300(24):2886-2897.

8.   Huxley RR, Shiell AW, Law CM. The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. J

Hypertens. 2000;18(7):815-831.

9. Risnes K, Vatten L, Baker J, Jameson K, Sovio U, Kajantie E, et al. Birthweight and mortality in adulthood: a systematic review and meta-analysis. International Journal of Epidemiology. 2011;40(3):647-661.

10. Dogra S, Mukhopadhyay K, Narang A. Feed Intolerance and Necrotizing Enterocolitis in Preterm Small-for-Gestational Age Neonates with Normal Umbilical Artery Doppler Flow. Journal of Tropical Pediatrics. 2012;58(6):513-516.

11. Rani RRK, Antony BFC, Kuzhiyil AP. Neurodevelopmental Outcomes of Preterm Small for Gestational Age and Appropriate for Gestational Age Babies at One Year of Age. Journal Of Evolution Of Medical And Dental Sciences-JEMDS. 2020;9(43):3231-3236.

12. Mallick A, Venkatnarayan K, Thapar R, Tewari V, Shaw S. Morbidity Patterns of Late Preterm Babies Born Small for Gestation. The Indian Journal of Pediatrics. 2019;86(7):578-583.

13. Flavin MP, Osiovich H, Coughlin K, Sgro K, Ray J, Hu L, et al. Hypoglycemia in unmonitored full-term newborns—a surveillance study. Paediatrics & Child Health. 2018;23(8):509-514.

14. Fishel Bartal M, Chen H-Y, Blackwell SC, Chauhan SP, Sibai BM. Neonatal morbidity in late preterm small for gestational age neonates. The Journal of Maternal-Fetal & Neonatal Medicine. 2021;34(19):3208-3213.

15. Christensen RD, Baer VL, Henry E, Snow GL, Butler A, Sola-Visner MC. Thrombocytopenia in Small-for-Gestational-Age Infants. Pediatrics.

2015;136(2):e361-e370.

16. Mendez-Figueroa H, Truong V, Pedroza C, Chauhan S. Morbidity and Mortality in Small for Gestational Age Infants: A Secondary Analysis of Nine MFMU Network Studies. American Journal of Perinatology. 2017;34(4):323-332.

17. Albertsson-Wikland K, Karlberg J. Postnatal growth of children born small for gestational age. Acta Paediatrica. 1997;86(S423):193-195.

18. Haworth JC, Dilling L, Younoszai MK. Relation of Blood-Glucose to Haematocrit, Birthweight, and Other Body Measurements in Normal and Growth-Retarded Newborn Infants. The Lancet. 1967;290(7522):901-905.

19. Osifo OD, Numbere P-fP, Ovueni EM. Intra-abdominal solid visceral birth injury to neonates following normal spontaneous vaginal delivery. The American journal of case reports. 2011;12:19-22.

20. Beta J, Khan N, Khalil A, Fiolna M, Ramadan G, Akolekar R. Maternal and neonatal complications of fetal macrosomia: systematic review and meta-analysis. Ultrasound in Obstetrics & Gynecology. 2019;54(3):308-318.

21. Aslam HM, Saleem S, Afzal R, et al. "Risk factors of birth asphyxia". Italian Journal of Pediatrics. 2014;40(1):94.

22. Fok T, So H, Wong E, Ng P, Chang A, Lau J, et al. Updated gestational age specific birth weight, crown-heel length, and head circumference of Chinese newborns. Archives of Disease in Childhood-Fetal and Neonatal Edition. 2003;88(3):229-236.

23. Villar J, Giuliani F, Fenton TR, Ohuma EO, Ismail LC, Kennedy SH. INTERGROWTH-21st very preterm size at birth reference charts. Lancet.

2016;387(10021):844-845.

24.   Villar J, Ismail LC, Victora CG, Ohuma EO, Bertino E, Altman DG, et al. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. Lancet. 2014;384(9946):857-868.

25.   de Onis M, Onyango A, Borghi E, Siyam A, Blössner M, Lutter C. Worldwide implementation of the WHO Child Growth Standards. Public Health Nutr. 2012;15(9):1603-1610.

26.   Anderson NHMP, Sadler LCMMPH, McKinlay CJDMP, McCowan LMEMMD. INTERGROWTH-21st vs customized birthweight standards for identification of perinatal mortality and morbidity. Am J Obstet Gynecol. 2016;214(4):e501-509.

27.   Cole T. The LMS method for constructing normalized growth standards. European journal of clinical nutrition. 1990;44(1):45-60.

28.   Stomnaroska-Damcevski O, Petkovska E, Jancevska S, Danilovski D. Neonatal Hypoglycemia: A Continuing Debate in Definition and Management. Pril (Makedon Akad Nauk Umet Odd Med Nauki). 2015;36(3):91-97.

29.   Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008 Jan 30;27(2):157-72.

30.   Bhandari N, Bahl R, Taneja S, De Onis M, Bhan MK. Growth performance of affluent indian children is similar to that in developed countries. Bull World Health Organ. 2002;80(3):189-195.

31.   Groom KM, Poppe KK, North RA, McCowan LM. Small-for-gestational-age infants classified by customized or population birthweight centiles: impact of gestational age at delivery. Am J Obstet Gynecol. 2007;197(3):e231-239.

32.   Iliodromiti S, Mackay DF, Smith GC, 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. PLoS Medicine. 2017;14(1):e1002228

33.   Cole T. Sample size and sample composition for constructing growth reference centiles. Statistical methods in medical research. 2021;30(2):488-507.

34.   Boghossian NS, Geraci M, Edwards EM, Morrow KA, Horbar JD. Anthropometric Charts for Infants Born Between 22 and 29 Weeks Gestation. Pediatrics. 2016;138(6):e20161641.

35.   Norris T, Seaton S, Manktelow B, Baker PN, Kurinczuk JJ, Field D, et al. Updated birth weight centiles for England and Wales. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2018;103(6):F577-F582.

36.   Aris IM, Kleinman KP, Belfort MB, Kaimal A, Oken E. A 2017 US Reference for Singleton Birth Weight Percentiles Using Obstetric Estimates of Gestation. Pediatrics. 2019;144(1):e20190076.

37.   Nwabuobi C, Odibo L, Camisasca-Lopina H, Leavitt K, Tuuli M, Odibo AO. Comparing INTERGROWTH-21st Century and Hadlock growth standards to predict small for gestational age and short-term neonatal outcomes. J Matern Fetal Neonatal Med. 2020;33(11):1906-1912.

38. Giapros V, Drougia A, Krallis N, Theocharis P, Andronikou S. Morbidity and mortality patterns in small-for-gestational age infants born preterm. Journal of Maternal-Fetal and Neonatal Medicine. 2011;25(2):153-157.

39. Gardosi JO. Prematurity and fetal growth restriction. Early Hum Dev. 2005;81(1):43-49.

40. Mejri A, Dorval VG, Nuyt AM, Carceller A. Hypoglycemia in term newborns with a birth weight below the 10th percentile. Paediatrics & Child Health. 2010;15(5):271-275.

Table 1. Characteristics of the babies included in the study.

	Boys (n=254,843)	Girls (n=234,053)	P value
Maternal age, (median, IQR), years	31 (28-35)	31 (28-34)	0.132
Delivery mode			
Vaginal, % (n)	73.9 (188,266)	75.1 (175,801)	<0.001
CS, % (n)	26.1 (66,577)	24.9 (58,252)	
Preterm infants, % (n)	8.4 (21,311)	7.3 (17,088)	<0.001
Low birth weight, % (n)	7.7 (19,671)	9.2 (21,596)	<0.001
Low birth weight term neonates, % (n)	2.6 (6,615)	4.1 (9,710)	<0.001
Deaths in neonatal period, % (n)	0.1 (342)	0.1 (316)	0.969
Birth weight in term neonates, mean (sd), gram	3,253 (406)	3,144 (396)	<0.001
1-min Apgar score <4, % (n)	0.4 (1,072)	0.3 (814)	<0.001
Born to multiple pregnancy, % (n)	3.4 (8,748)	3.5 (8,305)	0.028
History of IUGR, % (n)	1.4 (3,681)	2.0 (4,678)	<0.001
Congenital anomaly, % (n)	5.3 (13,533)	4.7 (10,898)	<0.001
Adimission to NICU, % (n)	4.0 (10,201)	3.4 (7,854)	<0.001
Maternal diseases			
Hypertension, % (n)	5.8 (14,676)	6.1 (14,263)	<0.001
Diabetes, % (n)	7.5 (19,229)	7.3 (17,022)	<0.001
Chronic smoker, % (n)	0.1 (266)	0.1 (259)	0.531
Drug abuse, % (n)	0.1 (187)	0.1 (168)	0.877
Anemia during pregnancy, % (n)	5.1 (12,971)	5.3 (12,319)	0.006
Complications of SGA			
Hypoglycaemia, % (n)	2.2 (5,508)	2.0 (4,751)	0.001
Polycythemia, % (n)	1.7 (4,281)	1.6 (3,683)	0.003
Hypothermia, % (n)	0.1 (284)	0.1 (273)	0.620
Neutropenia, % (n)	0.2 (524)	0.1 (330)	<0.001
Thrombocytopenia, % (n)	0.6 (1,632)	0.5 (1,187)	<0.001

CS: cesarean section; IUGR: intrauterine growth restriction; NICU: neonatal intensive care unit; SGA: small for gestational age

Table 2. The sex- and gestation-specific mean birth weight Z-scores and proportions of SGA and LGA infants by the two birth weight references.

| | Boys | | | | | | Girls | | | | | |
| | FOK2003 | | | INTERGROWTH21 | | | FOK2003 | | | INTERGROWTH21 | | |
Gestation weeks	Mean Z score	SGA (%)	LGA (%)	Mean Z score	SGA (%)	LGA (%)	Mean Z score	SGA (%)	LGA (%)	Mean Z score	SGA (%)	LGA (%)
24	-0.13	9.4	n.p.	0.05	4.7	n.p.	-0.56	7.4	0	-0.19	6.6	n.p.
25	-0.40	12.6	n.p.	-0.03	6.8	2.6	-0.43	10.8	n.p.	-0.08	7.2	3.6
26	-0.44	17.3	n.p.	-0.06	8.7	3.1	-0.43	14.7	n.p.	-0.16	12.0	2.7
27	-0.53	20	n.p.	-0.16	11.9	n.p.	-0.36	14.1	2.4	-0.17	13.1	2.4
28	-0.46	18.5	3.3	-0.13	13.1	5.2	-0.33	17	4.0	-0.21	16.7	7.4
29	-0.40	17.3	3.6	-0.13	12.3	3.9	-0.32	15.5	3.5	-0.14	15.0	3.8
30	-0.35	17.3	4.2	-0.13	11.9	4.6	-0.35	14.7	4.0	-0.17	13.4	4.6
31	-0.41	18.5	4.3	-0.19	13.5	4.0	-0.35	13.9	3.5	-0.08	10.0	4.3
32	-0.28	14.0	4.0	-0.12	9.3	3.1	-0.39	15.9	3.8	-0.16	11.2	4.0
33	-0.36	16.0	4.4	-0.29	8.2	2.3	-0.46	17.9	3.6	-0.32	12.8	3.4
34	-0.33	15.8	4.5	-0.35	11.2	2.9	-0.43	17.7	5.4	-0.37	15.2	5.1
35	-0.32	15.5	5.3	-0.33	11.7	3.9	-0.44	18.4	4.8	-0.38	16.8	4.8
36	-0.33	16.3	5.1	-0.29	12.0	4.6	-0.44	20.1	5.4	-0.36	17.1	5.7
37	-0.24	14.5	6.2	-0.17	10.7	6.2	-0.31	16.6	6.3	-0.22	13.7	6.4
38	-0.05	9.0	7.7	-0.01	6.5	7.6	-0.09	10.2	8.0	-0.05	7.9	7.5
39	-0.04	7.9	7.4	-0.07	6.6	6.5	-0.05	8.6	7.9	-0.08	7.5	6.4
40	0.02	7.3	8.6	-0.10	7.7	6.5	0.03	7.6	9.5	-0.08	7.7	6.5
41	0.04	7.3	9.2	-0.13	9.2	6.5	0.12	7.1	11.2	-0.08	8.4	7.1
42	-0.32	17.1	6.3	-0.50	19.9	4.4	-0.18	15.5	9.1	-0.38	18.5	5.6

n.p. Data not published to maintain the confidentiality of individuals in groups with small numbers. SGA: small for gestational age; LGA: large for gestational age

Table 3. The diagnostic accuracy for predicting SGA-related complications of the two birth weight references

	Boys			Girls		
	FOK2003	INTERGROWTH 21	P value	FOK2003	INTERGROWTH 21	P value
Sensitivity	41.1%	35.7%	-	45.5%	42.1%	-
Specificity	91.7%	93.1%	-	91.1%	92.4%	-
Diagnostic odds ratio	7.7	7.6	-	8.6	8.8	-
Accuracy	89.0%	90.1%	-	88.6%	89.6%	-
F-measure	0.28	0.28	-	0.31	0.31	-
AUROC (95% CI)	0.664 (0.660-0.668)	0.644 (0.640-0.649)	<0.001	0.683 (0.679-0.688)	0.672 (0.668-0.677)	<0.001
AUROC (95% CI, GA<37 wk)	0.631 (0.624-0.639)	0.615 (0.608-0.622)	0.001	0.620 (0.612-0.628)	0.618 (0.610-0.626)	0.473
AUROC (95% CI, GA≥37 wk)	0.677 (0.671-0.683)	0.663 (0.657-0.668)	<0.001	0.705 (0.699-0.710)	0.693 (0.687-0.699)	0.004

SGA: small for gestational age; AUROC: area under the receiver operating characteristics curve; GA: gestational age

Diagnostic odds ratio=Positive likelihood ratio/Negative likelihood ratio;

Accuracy=($\sum$True positive+$\sum$True negative)/$\sum$Total population;

F-measure=2×Precision×Recall/(Precision+Recall) (Precision=True positive/(True positive+False positive); Recall= True positive/(True positive+False negative))
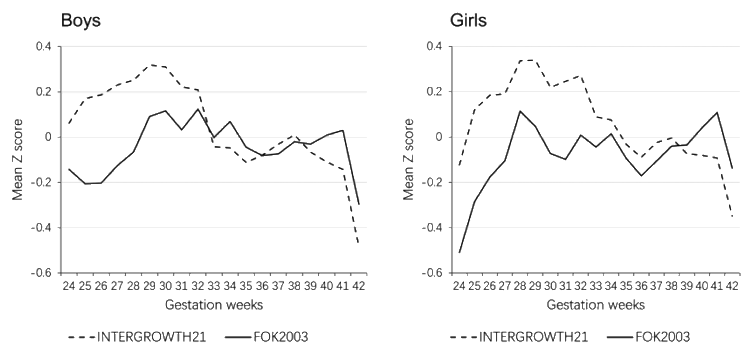
Figure 1. Gestation- and sex-specific mean Z-scores for low-risk infants.