POPULATION STUDY ARTICLE

Comparison of the INTERGROWTH-21st standard and a new reference for head circumference at birth among newborns in Southern China

Wan-Qing Xiao[1,2,3], Li-Fang Zhang[1,2], Jian-Rong He[1,2], Song-Ying Shen[1], Anna L. Funk[4], Jin-Hua Lu[1,2], Xue-Ling Wei[1,2], Jia Yu[2], Li Yang[2], Fang Li[5], Hui-Min Xia[1,6] and Xiu Qiu[1,2,3]

BACKGROUND: Previous studies proposed that there were racial or ethnic disparities in fetal growth, challenging the use of international standards in specific populations. This study was to evaluate the validity of applying the INTERGROWTH-21st standard to a Chinese population for identifying abnormal head circumference (HC), in comparison with a newly generated local reference.
METHODS: There were 24,257 singletons delivered by low-risk mothers in four perinatal health-care centers in Southern China. New HC reference was constructed and comparison in distribution of HC categories was performed between the INTERGROWTH-21st standard and new reference after applying these two tools in study population. Logistic regression was used to examine the association between abnormal HC and adverse neonatal outcomes.
RESULTS: There were 4.40% of the newborns identified with microcephaly (HC > 2 standard deviation below the mean) using the INTERGROWTH-21st standard, comparing to the proportion of 2.83% using new reference. The newborns identified with microcephaly only by the INTERGROWTH-21st standard were not at a higher risk of adverse neonatal outcome, compared with those identified as non-microcephaly by both tools (OR 0.73, 95% CI 0.47–1.13).
CONCLUSION: The new HC reference may be more appropriate for newborn assessment in Chinese populations than the INTERGROWTH-21st standard.

INTRODUCTION

Head circumference (HC) at birth is a proxy for fetal brain development[1] and may be important to predict adverse perinatal and long-term outcomes. Small HC is associated with increased risk of infant mortality[2] and delayed neurodevelopment of children in later life.[3] Meanwhile, a large head size in fetuses or newborns is associated with adverse labor outcomes, such as operative delivery and unplanned cesarean.[4–8] Measuring HC in infants is a simple, quick, and economical method for determining head size and overall brain volume.[9–11] Based on an age- and sex-adjusted reference curve, clinicians can identify newborns with abnormal HC who may suffer from microcephaly (>2 standard deviation (SD) or 3 SD below the mean) and have a high risk of neurodevelopmental disorders.[11,12]

There are several published references for HC at birth in the United States, Canada, Finland, India, Indonesia, and Turkey.[9,13–17] Significant differences among these curves can be found in the same centile for the same gestational age and sex, which suggests that racial or ethnic disparities, as well as differences in exposures to prenatal risk factors, across different populations may affect the fetal development in utero. These references usually describe fetal growth in different populations, while a "standard" inclines toward optimal growth based on an ideal low-risk population in a prescriptive sense and has an implication to render normal pregnancies abnormal.[18–20] In 2014, the INTERGROWTH-21st project provided an international standard to assess newborn HC based on healthy and well-nourished women.[21] However, racial or ethnic disparities in fetal growth challenged the use of international standards in specific populations, and little evidence is provided about whether this global standard for newborn head size is suitable for a Chinese population.

The existing Chinese national reference for HC at birth was generated using data from newborns born in 15 cities between 1986 and 1987,[22] without consideration of infants' sex. This existing reference may have become out of date because of the fact that rapid urbanization and economic growth over the past decades in China has brought enormous improvements in prenatal health care including nutritional status among pregnant women. For example, large-scale physical growth surveys in China showed a significant increase of 0.4 cm in average HC of newborns between 1985 and 2005.[23] Few studies in China have established population-specific HC references at birth,[24,25] but even these

[1]Division of Birth Cohort Study, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, Guangdong, China; [2]Department of Woman and Child Health Care, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, Guangdong, China; [3]Department of Obstetrics and Gynecology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, Guangdong, China; [4]Emerging Disease Epidemiology Unit, Institut Pasteur, Paris, France; [5]Department of Obstetrics and Gynecology, Guangzhou Eighth People's Hospital, Guangzhou Medical University, Guangzhou, Guangdong, China and [6]Department of Neonatal Surgery, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, Guangdong, China
Correspondence: Xiu Qiu (qxiu0161@163.com, xiu.qiu@bigcs.org)
These authors contributed equally: Wan-Qing Xiao, Li-Fang Zhang

Comparison of the INTERGROWTH-21st standard and a new reference for head...
W.-Q. Xiao et al.

530

have not necessarily considered preterm birth or produced a detailed HC reference chart.

The objectives of the present study were to evaluate the validity of applying the INTERGROWTH-21st standard to a Southern Chinese population for identifying abnormal HC and compare it with a newly generated local HC reference by examining the association between abnormal HC identified by the two tools and adverse neonatal outcomes.

METHODS

Study population
The HC surveillance for newborns was conducted in four perinatal health care centers: Guangzhou Women and Children's Medical Center (GWCMC), which comprised two campuses, Guangzhou Huadu Women and Children Health Care Hospital and Guangzhou Liwan Women and Children Health Care Hospital, located in the central, north and west areas of Guangzhou, China, respectively. We included all the singletons born in these centers between 24^{+0} and 41^{+6} weeks of gestation from February 10, 2017 to May 31, 2018. Exclusion criteria for ensuring the low-risk status of newborns and their mothers included maternal age ≥35 years or <18 years, height <150 cm, body mass index (BMI) ≥28 kg/m² or <16 kg/m², history of cardiovascular disease, drug abuse, two previous pregnancies ending in miscarriage, history of stillbirth, macrosomia, preterm birth, or congenital malformations in previous pregnancies. We also excluded newborns whose mother used assisted reproductive technology or had the following complications in the index pregnancy: gestational hypertension, gestational diabetes mellitus, intrahepatic cholestasis of pregnancy, heart disease, liver disease, kidney disease, type 2 diabetes mellitus, chronic hypertension, thyroid disorders, systemic lupus erythematosus, tuberculosis, cancer, psychiatric disorders, idiopathic thrombocytopenic purpura, hematologic disorder, severe anemia, and sexual transmitted disease including gonorrhea, syphilis, genital warts or herpes, genital chlamydia trachomatis or mycoplasma infection, and acquired immunodeficiency syndrome. Stillbirths and infants with congenital abnormalities were further excluded. In addition, since at least 50 observations for each gestational week group were required to establish new HC curves,[21] newborns at early gestational ages without enough observations were also excluded.

Collected HC data were entered into the Guangzhou Perinatal Health Care and Delivery Surveillance System (GPHCDSS), which covers >99% of deliveries in Guangzhou as previously described.[26,27] The GPHCDSS collects information on maternal demographic characteristics, medical history, and maternal complications during pregnancy, as well as neonatal information including gestational age at birth and sex. Gestational age was based on the ultrasound examination in the first or early second trimester as a routine practice, which was rigorously followed in the four study centers. In rare cases, gestational age was calculated by the date of last menstrual period when ultrasound examination was not available. Data of 47,369 singletons were collected from the four centers during the study period. After excluding 22,168 newborns who were not at the low-risk status, 103 stillbirths, 456 with congenital abnormalities, 227 with missing or implausible data (HC values with residual >+3.89 SD or <−3.89 SD by robust regression with iteratively reweighted least square procedure) on HC measurements, and 158 born earlier than 33 weeks of gestation because of insufficient observations in these groups, 24,257 singleton live births were left in our final analysis (Supplemental Fig. S1 (online)). The study was approved by the Guangzhou Women and Children's Medical Center Ethics Approval Board (No. 2016111865-2).

HC measurement
HC of newborns was measured by midwives immediately after delivery, using the "seca 212" head tape-measure band, which is made of non-stretch Teflon. All the midwives in the four centers were trained with standardized measurement procedures before the conduction of the study. For each newborn, the midwife should measure the HC twice by wrapping the band around the broadest part of the forehead above the eyebrow, above the ears, and around the most prominent part of the back of the head. HC measurements were read to the nearest millimeter and the average value of the two measurements was then recorded. The allowable difference between the two measurements was no more than 4 mm, otherwise the charge midwife would retake the HC with the same process.
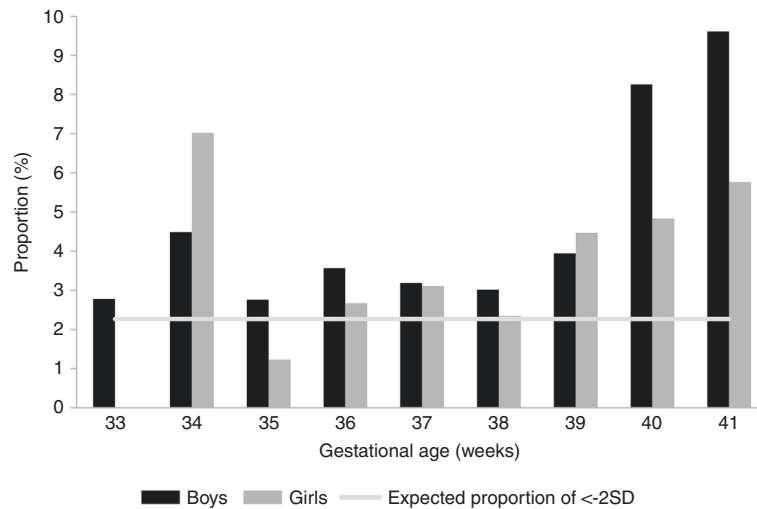
Composite adverse neonatal outcome
A composite of neonatal outcomes was defined as one or more of the following conditions: low Apgar score (<7) at 5 min after birth, neonatal asphyxia, and admission to neonatal intensive care unit (NICU). Data of neonatal outcomes were extracted from electronic medical records of GWCMC, while medical records from the other two centers were inaccessible.

Statistical analysis
We used robust regression with iteratively reweighted least square procedure to identify implausible HC values before the construction of the HC reference to avoid biased gestational age-specific HC distributions, either by misclassification of gestational age or invalid HC measurements. We considered HC values with residual >+3.89 SD or <−3.89 SD as outliers (n = 34). When applying the INTERGROWTH-21st standard to the local population, we compared the proportions of newborns identified with HC > 2 SD below the mean at different gestational ages with expected proportion in a Gaussian (normal) distribution. We used generalized additive models for location, scale, and shape framework (GAMLSS) to construct a gestational age-specific HC reference and smoothed it with cubic spline. Three methods were explored in the GAMLSS to take account of skewness, platykurtosis, and leptokurtosis: (1) a LMS (lambda, mu, and sigma) method, which assumes a Box–Cox transformation; (2) a LMST (lambda, mu, sigma, Box–Cox t distribution) method, which assumes a shifted and scaled (truncated) t distribution, and (3) a LMSP (lambda, mu, sigma, Box–Cox power exponential distribution) method, which assumes a Box–Cox power exponential transformation. We selected the best model based on model fit assessment from the Bayes Information Criterion. When estimating the lambda (L), mu (M), and sigma (S) parameters, the link function for the L and S was "log," while the skewness and kurtosis values were constant. All models were fitted for boys and girls separately.

We generated figures and tables for the 3rd, 10th, 25th, 50th, 75th, 90th, and 97th centiles, and SDs were obtained after smoothing for boys and girls separately. We compared the new −3 SD, −2 SD, mean, 2 SD, and 3 SD curves with the INTERGROWTH-21st standard visually. We also compared the classification of HC measurements at birth according to the −3 SD, −2 SD, 2 SD, and 3 SD cut-offs by the two tools.

Logistic regression was used to estimate the odds ratios (ORs) and 95% confidence intervals (CIs) for the associations between abnormal head size and composite adverse neonatal outcome. Abnormal HC, including microcephaly and macrocephaly, were defined as HC > 2 SD below the mean and above the mean, respectively, using the INTERGROWTH-21st standard or the new reference. Newborns were categorized into four or three groups jointly based on the two tools: (1) for microcephaly, newborns were firstly categorized as two groups: non-micro (not microcephaly by either of the two tools) and micro-either (microcephaly by either of the two tools), while the latter group was then classified into: micro-IG only (<−2 SD by the INTERGROWTH-21st standard alone), micro-NR only (<−2 SD by the new reference alone), and micro-both (<−2 SD by both tools); (2) for macrocephaly, newborns were first categorized as two groups: non-macro (not macrocephaly by either of the two tools) and macro-

Comparison of the INTERGROWTH-21st standard and a new reference for head...
W.-Q. Xiao et al.

531

Fig. 1 Proportions of newborns identified with microcephaly by the INTERGROWTH-21st standard at different gestational ages. Data in table were presented as the number of newborns with head circumference (HC) < −2 SD/total number of newborns. Expected proportion of HC < −2 SD was based on the assumption that the HC had a normal distribution (2.28%)

	Total	33 Weeks	34 Weeks	35 Weeks	36 Weeks
All the newborns	1067/24,257	1/70	8/146	5/226	18/561
Boys	616/12,919	1/36	4/89	4/145	12/336
Girls	451/11,338	0/34	4/57	1/81	6/225
	37 Weeks	38 Weeks	39 Weeks	40 Weeks	41 Weeks
All the newborns	61/1932	164/6011	362/8627	369/5639	79/1045
Boys	36/1128	102/3376	178/4512	232/2808	47/489
Girls	25/804	62/2635	184/4115	137/2831	32/556

either (macrocephaly by either of the two tools), while the latter group was then classified into: macro-NR only (>2 SD by the new reference alone) and macro-both (>2 SD by both tools). Considering the difference in gestational age distribution among the microcephaly groups, we conducted stratified analyses by gestation age based on the intersection of the two tools (≥38 vs <38 weeks of gestation). Statistical significance was considered at the level of $p < 0.05$. All analyses were performed using R version 3.2.4 (R, Inc) and SAS version 9.3 (SAS Institute, Cary, NC).

RESULTS
Application of the INTERGROWTH-21st standard to the local population
Figure 1 shows the proportion of microcephaly after applying the INTERGROWTH-21st standard to the local population at different gestational ages. There were 1067 (4.40%) newborns classified as HC < −2 SD according to the INTERGROWTH-21st standard, and when stratified by neonatal sex, 616 (4.77%) of the boys and 451 (3.98%) of the girls were identified as HC < −2 SD, respectively. The proportions of newborns identified with HC < −2 SD were higher at most gestational ages than expected proportion (2.28%) in a Gaussian (normal) distribution, especially among boys delivered at 40 and 41 weeks of gestation. The inconsistency between the observed and expected proportions suggested that there might be a need to construct a new HC reference in the local population.

Construction of new HC curves and comparison with the INTERGROWTH-21st standard
Smoothed gestational age-specific HC curves for boys and girls at birth in the new reference are presented in Fig. 2. The 3rd, 10th, 25th, 50th, 75th, 90th, and 97th percentiles; the means; and SDs of HC; as well as the values of L, M, and S parameters, at each gestational week from 33 to 41 in smoothed curves are shown in Table 1. The average absolute differences between observed and smoothed centiles were 0.12 cm for boys and 0.14 cm for girls. The HC measurements at all percentiles for each gestational age were higher for boys than for girls, with differences ranging from 0.1 to 0.5 cm.

The comparison of new HC curves with the INTERGROWTH-21st standard is presented in Fig. 3. For boys, the HC cut-offs of −2 SD and −3 SD after 38 weeks of gestation were higher in the INTERGROWTH-21st standard than in the new reference with a maximal difference of 0.8 cm, while an opposite pattern was found before 38 weeks of gestation. The HC values in the 2 SD and 3 SD curves at all gestational weeks for boys were lower in the new reference than in the INTERGROWTH-21st standard with a maximal difference of 0.9 cm. For girls, similar differences were observed between the new reference and the INTERGROWTH-21st standard.

Table 2 shows the classification of HC at birth according to the new reference and the INTERGROWTH-21st standard. There were 688 (2.83%) and 40 (0.16%) newborns classified as HC < −2 SD and <−3 SD, respectively, according to the new reference, comparing to the numbers of 1067 (4.40%) and 65 (0.27%) when using the INTERGROWTH-21st standard. Among the newborns delivered at or later than 38 weeks of gestation, 417 were classified as HC <−2 SD by the INTERGROWTH-21st standard only, while 38 were classified as HC < −2 SD by the new reference only among those delivered earlier than 38 weeks of gestation. There were 498 (2.05%) newborns classified as HC >2 SD according to the new reference, comparing to the number of 271 (1.11%) when using the INTERGROWTH-21st standard.
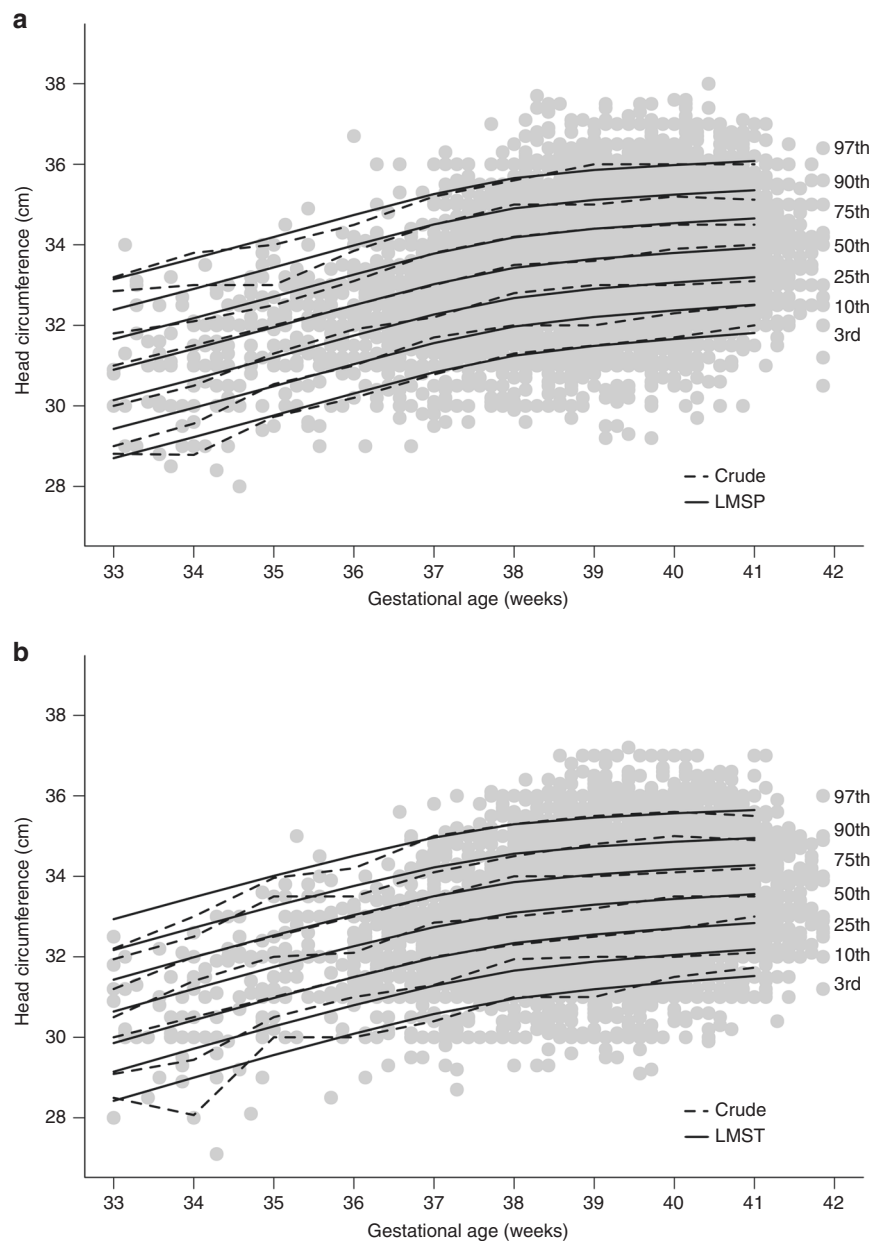
Comparison of the INTERGROWTH-21st standard and a new reference for head...
W.-Q. Xiao et al.

532

Fig. 2 Crude and smoothed curves for head circumference (HC) at birth of boys (**a**) and girls (**b**) in the new reference. Different curves represent the 3rd, 10th, 25th, 50th, 75th, 90th, and 97th percentiles according to gestational age. Gray circles represent actual observations. LMST (lambda, mu, sigma, Box–Cox t distribution) method, which assumes a shifted and scaled (truncated) t distribution. LMSP (lambda, mu, sigma, Box–Cox power exponential distribution) method, which assumes a Box–Cox power exponential transformation

HC categories and adverse neonatal outcomes
The associations between microcephaly defined by the two tools and composite adverse neonatal outcome are shown in Table 3. Compared to the non-micro group, newborns in the micro-either group had a higher risk of composite adverse neonatal outcome (OR 1.56, 95% CI 1.26–1.91), while the ORs for the micro-NR only and micro-both groups were 7.53 (95% CI 2.97–19.12) and 2.00 (95% CI 1.57–2.56), respectively. No differences in the risk of composite adverse neonatal outcome were found between newborns in the micro-IG only group and in the non-micro group (OR 0.73, 95% CI 0.47–1.13). Stratified analysis showed that the preterm micro-NR only group (<38 weeks of gestation) had a higher risk of composite adverse outcome (OR 2.87, 95% CI 1.12–7.33), compared to the preterm non-micro group, whereas the term micro-IG only group (≥38 weeks of gestation) did not have such a higher risk compared to the term non-micro group

(OR 0.88, 95% CI 0.57–1.36). Both the preterm and the term micro-both groups had higher risk of composite adverse outcome compared to the non-micro group, with ORs of 2.22 (95% CI 1.34–3.68) and 1.84 (95% CI 1.38–2.46), respectively (Table 3). We did not find significant associations between macrocephaly identified by the two tools and composite adverse neonatal outcome (Supplemental Table S1 (online)).

DISCUSSION
In the present study, we found that 4.40% of the newborns were identified with HC > 2 SD below the mean by the INTERGROWTH-21st standard in a low-risk Chinese population, and the inconsistency with corresponding proportion in a normal distribution highlighted the need for new local HC reference. An up-to-date local sex-specific HC reference was then constructed for newborns

Comparison of the INTERGROWTH-21st standard and a new reference for head...
W.-Q. Xiao et al.

533

Table 1. Head circumference (in cm) of singleton boys and girls at birth according to gestational age in the new reference

GA	n	Centiles							Mean	SD	Parameters		
		3rd	10th	25th	50th	75th	90th	97th			L	M	S
Boys													
33 weeks	36	28.7	29.4	30.1	30.9	31.7	32.4	33.1	31.0	1.4	0.677	30.898	0.038
34 weeks	89	29.2	30.0	30.7	31.4	32.2	32.9	33.7	31.3	1.4	0.677	31.416	0.037
35 weeks	145	29.8	30.5	31.2	32.0	32.7	33.4	34.2	31.9	1.1	0.677	31.953	0.036
36 weeks	336	30.3	31.0	31.7	32.5	33.3	34.0	34.7	32.5	1.1	0.677	32.500	0.036
37 weeks	1128	30.8	31.6	32.3	33.0	33.8	34.5	35.3	33.0	1.1	0.677	33.023	0.035
38 weeks	3376	31.2	32.0	32.7	33.4	34.2	34.9	35.7	33.5	1.1	0.677	33.428	0.035
39 weeks	4512	31.5	32.2	32.9	33.7	34.4	35.1	35.9	33.7	1.2	0.677	33.651	0.034
40 weeks	2808	31.7	32.4	33.1	33.8	34.5	35.2	36.0	33.8	1.2	0.677	33.797	0.034
41 weeks	489	31.8	32.5	33.2	33.9	34.7	35.4	36.1	33.9	1.0	0.677	33.923	0.033
Girls													
33 weeks	34	28.4	29.1	29.9	30.6	31.4	32.2	32.9	30.5	1.1	0.538	30.640	0.038
34 weeks	57	29.0	29.7	30.4	31.2	32.0	32.7	33.5	31.2	1.3	0.538	31.202	0.037
35 weeks	81	29.6	30.3	31.0	31.7	32.5	33.3	34.0	31.9	1.1	0.538	31.749	0.036
36 weeks	225	30.1	30.8	31.5	32.3	33.0	33.8	34.5	32.2	1.1	0.538	32.263	0.035
37 weeks	804	30.6	31.3	32.0	32.7	33.5	34.2	35.0	32.7	1.2	0.538	32.737	0.034
38 weeks	2635	31.0	31.7	32.3	33.1	33.9	34.6	35.3	33.1	1.1	0.538	33.095	0.033
39 weeks	4115	31.2	31.9	32.6	33.3	34.0	34.7	35.5	33.3	1.1	0.538	33.295	0.033
40 weeks	2831	31.4	32.0	32.7	33.4	34.2	34.9	35.6	33.4	1.1	0.538	33.438	0.032
41 weeks	556	31.5	32.2	32.8	33.6	34.3	35.0	35.6	33.6	1.0	0.538	33.555	0.031

GA gestational age, SD standard deviation

from 33 to 41 weeks of gestation. When comparing our new HC curves to the INTERGROWTH-21st standard, we found significant differences in the HC cut-offs of −2 SD and −3 SD at both earlier and later gestational age groups. The proportions of newborns with HC > 2 SD and > 3 SD below the mean by the INTERGROWTH-21st standard were over 1.5 times as high as those by the new reference. The INTERGROWTH-21st standard identified many more newborns with HC > 2 SD below the mean compared with the new reference among those delivered at or later than 38 weeks of gestation; however, the risk of composite adverse neonatal outcome for these "microcephaly" newborns were similar to their counterparts in the non-micro group.

The INTERGROWTH-21st standard was developed among healthy pregnant women assuming that babies from different ethnic backgrounds would follow the same standard if they were under ideal conditions.[21,28] Conversely, wide variation in fetal growth has been found even when the mothers were in equal optimized conditions.[18] A fetal growth study showed highly significant differences in fetal HC measurements by ultrasound exam at the same gestational age among Asian, White, Hispanic, and Black women who were at low-risk status.[29] Our findings about the validity of applying the INTERGROWTH-21st standard to the local population might add some new support to the disparities in fetal head growth among different populations.

There were similarities and differences between the features of our newly generated HC reference and the INTERGROWTH-21st standard. The HC values at all percentiles for each gestational age were higher for boys than for girls, which was in line with the INTERGROWTH-21st standard.[21] The head size difference according to sex was also observed in standards based on other populations.[9,17] However, differences between boys and girls in the new curves (≤0.5 cm) were smaller than those from other studies including the INTERGROWTH-21st standard.[13,21,30] In our new curves, we found a flattened HC growth after 38 weeks of gestation, which was consistent with previously published HC growth curves[9,13–15] in contrast to the INTERGROWTH-21st standard. This result was supported by previous studies that fetal growth including the HC growth would slow down after 38 weeks of gestation in normal pregnancies,[31,32] which might be attributed to the reduction of utero-placental function in late pregnancy.[33]

When comparing our new reference to the INTERGROWTH-21st standard, we found significant differences in the −2 SD and the −3 SD curves at both earlier and later gestational ages. Specifically, the cut-offs of −2 SD and −3 SD in the INTERGROWTH-21st standard tended to be higher than those in the new reference among newborns born after 38 weeks of gestation, which accounted for >60% of all the newborns in our population. When using the INTERGROWTH-21st standard as a screening tool for microcephaly, the proportions of newborns with HC <−2 SD or <−3 SD would each increase up to 1.5-folds, while the microcephalic newborns identified by the INTERGROWTH-21st standard only were not at greater risk of composite adverse neonatal outcome compared to those in the non-micro group. Few studies have reported the associations between microcephaly and neonatal mortality and morbidity including NICU admission and hospitalization.[2] Further studies are needed to provide stronger evidence for these associations and improve individualized care for newborns with abnormal head measurements.

Consistent with our findings, a study indicated that fetal smallness would be over-diagnosed in the Chinese population when using the INTERGROWTH-21st fetal growth standard, especially for HC measures.[34] The differences in the performance of our new reference and the INTERGROWTH-21st standard may be partly attributed to the genetic variations in fetal size among different races or ethnicities, under the premise of similar maternal characteristics in this study with the INTERGROWTH-21st study. An international standard is valuable for the comparison across different countries and for the growth evaluation in populations without suitable local standards.[35]
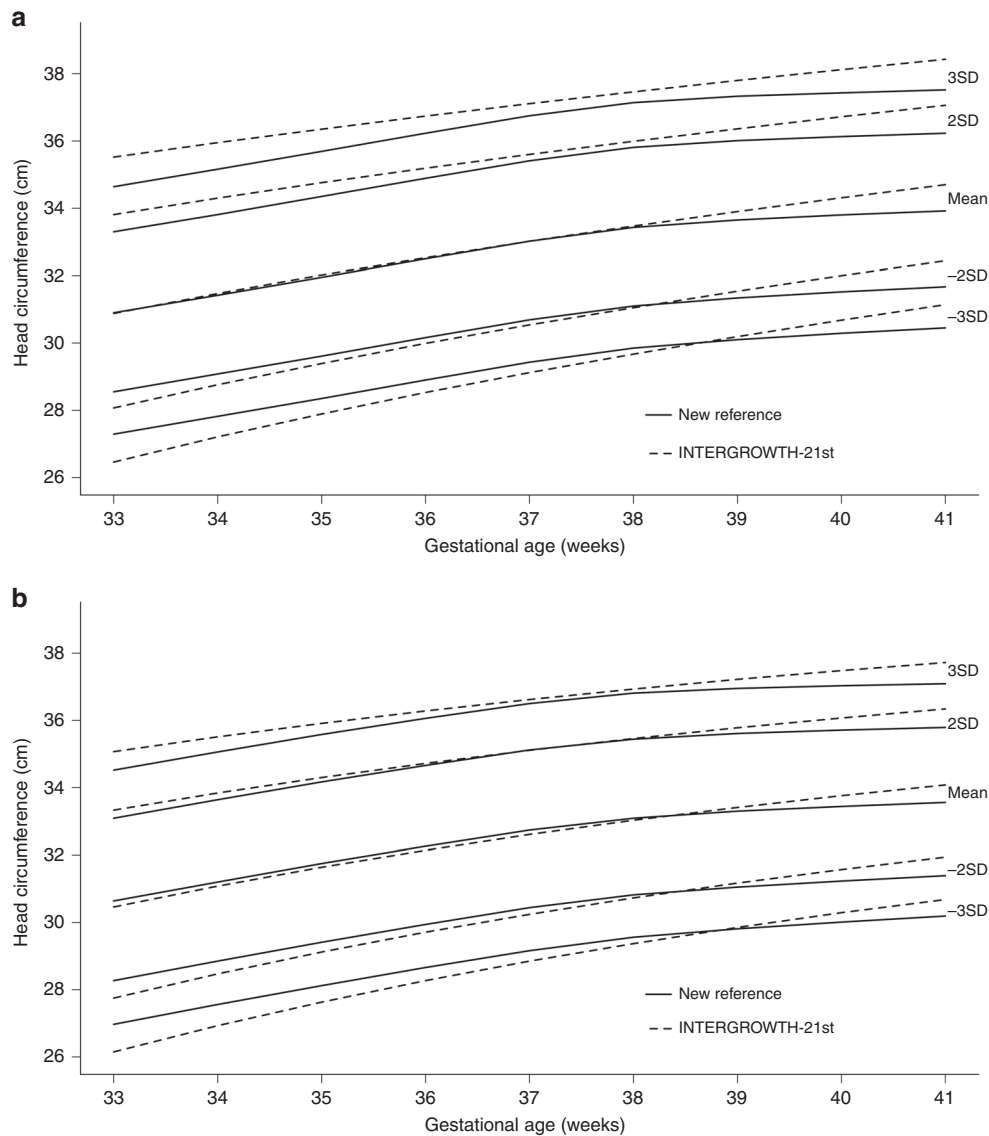
Comparison of the INTERGROWTH-21st standard and a new reference for head...
W.-Q. Xiao et al.

534

Fig. 3 Comparison of head circumference curves at birth of boys (**a**) and girls (**b**) between the INTERGROWTH-21st standard and the new reference. The new reference is represented as a solid line and the INTERGROWTH-21st standard is represented as a dashed line

However, it needs to be validated and cut-offs for identifying suboptimal infant growth should be tailored before it is applied to the local population.[36]

There are several strengths of the present study. First, previously published population-based HC references were usually generated using routine registry data, which had relative low quality. By contrast, the measurement of HC in this study was conducted according to standardized procedures, which might decrease measurement errors. Second, the analysis was conducted among mothers who were at low-risk status according to strict exclusion criteria so that our results could be compared with other growth references/standards that used similar inclusion and exclusion criteria including the INTERGROWTH-21st standard. Third, data on maternal and neonatal information were collected through the GPHCDSS system or electronic medical records of the study center, and the accuracy of data with minimal missing values enabled a robust analysis. One of the limitations of the present study is that we have a limited number of newborns for early gestational ages and therefore were unable to construct a HC curve for those born before 33 weeks. Second, the sample size in earlier gestational weeks was small, which might lead to a

relatively large difference between the observed and smoothed curves of the 3rd and the 10th centile compared to later gestational weeks. Third, similar to the INTERGROWTH-21st standard and other commonly used HC references, our new reference did not exclude newborns with extensive head moulding, which might influence the measurement of their HCs. Previous studies have shown that head moulding may lead to a larger or smaller HC at birth than it would be after the effect of moulding has subsided.[37,38] Therefore, including newborns with extensive head moulding may result in underestimating or overestimating the extreme percentiles of reference curves. Fourth, data of neonatal outcomes from two centers of our study sites were not available that reduced the sample size for examining the association between HC and neonatal outcomes. Last, the neonatal outcomes we used (low Apgar score after birth, neonatal asphyxia, and admission to NICU) were cross-sectional (measured at birth) and gestational age dependent, which might not reflect the prognosis of abnormal HC well. Unfortunately, we could not examine the associations between the HC measurements and neurodevelopment outcomes, which are most relevant to abnormal HC.

Comparison of the INTERGROWTH-21st standard and a new reference for head...
W.-Q. Xiao et al.

535

Table 2. Classification of head circumference at birth according to the INTERGROWTH-21st standard and the new reference

	New reference	INTERGROWTH-21st					Total
		>3 SD	≤3 SD and >2 SD	≤2 SD and ≥−2 SD	<−2 SD and ≥−3 SD	<−3 SD	
All the newborns	>3 SD	1 (0.00)	24 (0.10)	0 (0.00)	0 (0.00)	0 (0.00)	25 (0.10)
	≤3 SD and >2 SD	0 (0.00)	246 (1.01)	227 (0.94)	0 (0.00)	0 (0.00)	473 (1.95)
	≤2 SD and ≥−2 SD	0 (0.00)	0 (0.00)	22,654 (93.39)	417 (1.72)	0 (0.00)	23071 (95.11)
	<−2 SD and ≥−3 SD	0 (0.00)	0 (0.00)	38 (0.16)	578 (2.38)	32 (0.13)	648 (2.67)
	<−3 SD	0 (0.00)	0 (0.00)	0 (0.00)	7 (0.03)	33 (0.14)	40 (0.16)
	Total	1 (0.00)	270 (1.11)	22,919 (94.48)	1002 (4.13)	65 (0.27)	24257 (100.00)
Term (≥38 weeks)	>3 SD	1 (0.00)	23 (0.11)	0 (0.00)	0 (0.00)	0 (0.00)	24 (0.11)
	≤3 SD and >2 SD	0 (0.00)	217 (1.02)	218 (1.02)	0 (0.00)	0 (0.00)	435 (2.04)
	≤2 SD and ≥−2 SD	0 (0.00)	0 (0.00)	19,889 (93.28)	417 (1.96)	0 (0.00)	20306 (95.23)
	<−2 SD and ≥−3 SD	0 (0.00)	0 (0.00)	0 (0.00)	493 (2.31)	32 (0.15)	525 (2.46)
	<−3 SD	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.00)	31 (0.15)	32 (0.15)
	Total	1 (0.00)	240 (1.13)	20,107 (94.30)	911 (4.27)	63 (0.30)	21322 (100.00)
Preterm (<38 weeks)	>3 SD	0 (0.00)	1 (0.03)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.03)
	≤3 SD and >2 SD	0 (0.00)	29 (0.99)	9 (0.31)	0 (0.00)	0 (0.00)	38 (1.29)
	≤2 SD and ≥−2 SD	0 (0.00)	0 (0.00)	2765 (94.21)	0 (0.00)	0 (0.00)	2765 (94.21)
	<−2 SD and ≥−3 SD	0 (0.00)	0 (0.00)	38 (1.29)	85 (2.90)	0 (0.00)	123 (4.19)
	<−3 SD	0 (0.00)	0 (0.00)	0 (0.00)	6 (0.20)	2 (0.07)	8 (0.27)
	Total	0 (0.00)	30 (1.02)	2812 (95.81)	91 (3.10)	2 (0.07)	2935 (100.00)

Data were presented as n (%)
SD standard deviation

Table 3. Associations between microcephaly identified by the INTERGROWTH-21st standard and the new reference and composite adverse neonatal outcome

	n (%)	OR	95% CI
All the newborns[a]	13751 (100.00)		
Two groups			
Non-micro	12,961 (94.25)	Ref.	
Micro-either	790 (5.75)	**1.56**	**(1.26–1.91)**
Four groups			
Non-micro	12,961 (94.25)	Ref.	
Micro-IG only	305 (2.22)	0.73	(0.47–1.13)
Micro-NR only	18 (0.13)	**7.53**	**(2.97–19.12)**
Micro-both	467 (3.40)	**2.00**	**(1.57–2.56)**
Term (≥38 weeks)	12,280 (100.00)		
Two groups			
Non-micro	11,576 (94.27)	Ref.	
Micro-either	704 (5.73)	**1.41**	**(1.10–1.79)**
Four groups			
Non-micro	11,576 (94.27)	Ref.	
Micro-IG only	305 (2.48)	0.88	(0.57–1.36)
Micro-NR only	—	—	—
Micro-both	399 (3.25)	**1.84**	**(1.38–2.46)**
Preterm (<38 weeks)	1471 (100.00)		
Two groups			
Non-micro	1385 (94.15)	Ref.	
Micro-either	86 (5.85)	**2.35**	**(1.49–3.68)**
Four groups			
Non-micro	1385 (94.15)	Ref.	
Micro-IG only	—	—	—
Micro-NR only	18 (1.22)	**2.87**	**(1.12–7.33)**
Micro-both	68 (4.62)	**2.22**	**(1.34–3.68)**

Composite adverse neonatal outcome included neonatal asphyxia, low Apgar score, and NICU admission
Bold indicates statistical significance (p < 0.05)
Non-micro not microcephaly by either of the two tools, Micro-either microcephaly by either of the two tools, Micro-IG only <−2 SD by the INTERGROWTH-21st standard alone, Micro-NR only <−2 SD by the new reference alone, Micro-both, <−2 SD by both tools, OR odds ratio, CI confidence interval
[a]Data were analyzed among newborns delivered in GWCMC

CONCLUSIONS

The newly generated sex-specific reference for HC at birth may be more appropriate for application in local populations compared to the INTERGROWTH-21st standard and would decrease the cost of screening and diagnosis for microcephaly, especially among the newborns delivered at or after 38 weeks of gestation. Further studies are needed to examine the ability of our new reference to predict neurodevelopment in childhood.

AUTHOR CONTRIBUTIONS

X.Q. and H.-M.X. conceived the study. X.Q., H.-M.X., W.-Q.X., L.-F.Z., J.-R.H., S.-Y.S. and A. L.F. designed and directed the study. W.-Q.X. and L.-F.Z. drafted the manuscript. W.-Q. X., L.-F.Z., S.-Y.S. and J.-H.L. cleaned and analyzed the data. W.-Q.X., X.-L.W., J.Y. and L. Y. were involved in data acquisition and data management. X.Q., J.-R.H. and F.L. revised the manuscript. All authors have read and approved the final version of the manuscript for publication.

ADDITIONAL INFORMATION

The online version of this article (https://doi.org/10.1038/s41390-019-0446-0) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Melamed, N. et al. Sonographic estimation of fetal head circumference: how accurate are we? Ultrasound Obstet. Gynecol. **37**, 65–71 (2011).

Comparison of the INTERGROWTH-21st standard and a new reference for head...
W.-Q. Xiao et al.

536

2. Auger, N. et al. Congenital microcephaly in Quebec: baseline prevalence, risk factors and outcomes in a large cohort of neonates. *Arch. Dis. Child Fetal Neonatal Ed.* **103**, F167–F172 (2018).

3. Millichap, J. G. Head circumference and neurocognitive outcomes. *Pediatr. Neurol. Briefs* **29**, 54 (2015).

4. Rabei, N. H. et al. Intrapartum fetal head circumference and estimated fetal weight as predictors of operative delivery. *Int. J. Gynaecol. Obstet.* **137**, 34–39 (2017).

5. de Vries, B. et al. Is neonatal head circumference related to caesarean section for failure to progress? *Aust. N. Z. J. Obstet. Gynaecol.* **56**, 571–577 (2016).

6. Lipschuetz, M. et al. A large head circumference is more strongly associated with unplanned cesarean or instrumental delivery and neonatal complications than high birthweight. *Am. J. Obstet. Gynecol.* **213**, 833.e1–12 (2015).

7. Mujugira, A. et al. Fetal head circumference, operative delivery, and fetal outcomes: a multi-ethnic population-based cohort study. *BMC Pregnancy Childbirth* **13**, 106 (2013).

8. Elvander, C., Hogberg, U. & Ekeus, C. The influence of fetal head circumference on labor outcome: a population-based register study. *Acta Obstet. Gynecol. Scand.* **91**, 470–475 (2012).

9. Barbier, A. et al. New reference curves for head circumference at birth, by gestational age. *Pediatrics* **131**, e1158–e1167 (2013).

10. Sullivan, J. C. et al. Reliability of self, parental, and researcher measurements of head circumference. *Mol. Autism* **5**, 2 (2014).

11. Harris, S. R. Measuring head circumference: update on infant microcephaly. *Can. Fam. Physician* **61**, 680–684 (2015).

12. Garcia-Alix, A. et al. [Ability of neonatal head circumference to predict long-term neurodevelopmental outcome]. *Rev. Neurol.* **39**, 548–554 (2004).

13. Olsen, I. E. et al. New intrauterine growth curves based on United States data. *Pediatrics* **125**, e214–e224 (2010).

14. Sankilampi, U. et al. New population-based references for birth weight, length, and head circumference in singletons and twins from 23 to 43 gestation weeks. *Ann. Med.* **45**, 446–454 (2013).

15. Kandraju, H. et al. Gestational age-specific centile charts for anthropometry at birth for South Indian infants. *Indian Pediatr.* **49**, 199–202 (2012).

16. Haksari, E. L. et al. Reference curves of birth weight, length, and head circumference for gestational ages in Yogyakarta, Indonesia. *BMC Pediatr.* **16**, 188 (2016).

17. Kurtoglu, S. et al. Body weight, length and head circumference at birth in a cohort of Turkish newborns. *J. Clin. Res. Pediatr. Endocrinol.* **4**, 132–139 (2012).

18. Kiserud, T. et al. The World Health Organization fetal growth charts: concept, findings, interpretation, and application. *Am. J. Obstet. Gynecol.* **218**, S619–S629 (2018).

19. Lee, A. C. et al. Estimates of burden and consequences of infants born small for gestational age in low and middle income countries with INTERGROWTH-21(st) standard: analysis of CHERG datasets. *BMJ* **358**, j3677 (2017).

20. Ananth, C. V., Brandt, J. S. & Vintzileos, A. M. Standard vs population reference curves in obstetrics: which one should we use? *Am. J. Obstet. Gynecol.* **220**, 293–296 (2019).

21. Villar, J. 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* **384**, 857–868 (2014).

22. Research Cooperation Group on the physical development of newborns in 15 cities of China. The status of physical development in neonates of different gestational ages in China. *J. Clin. Pediatr.* **9**, 72–77 (1991).

23. Zhang, Y. Q. & Li, H. Changes in weight, length, head circumference, and ponderal index at birth of healthy term newborns in nine cities in China during the period of rapid social development 1985-2005. *Econ. Hum. Biol.* **19**, 45–50 (2015).

24. Huang, X. Y. et al. [Intrauterine growth curves for body weight, body length, head circumference, chest circumference, and crown-rump length in 16 887 neonates with a gestational age of 27-42 weeks in Shenzhen, China]. *Zhongguo Dang Dai Er Ke Za Zhi* **19**, 877–886 (2017).

25. Jia, Z. X. et al. Percentile curves for gestational age-specific head circumference of newborns in Beijing. *Chin. J. Child Health Care* **23**, 580–583 (2015).

26. He, J. R. et al. A new birthweight reference in Guangzhou, southern China, and its comparison with the global reference. *Arch. Dis. Child* **99**, 1091–1097 (2014).

27. He, J. R. et al. Birth weight changes in a major city under rapid socioeconomic transition in China. *Sci. Rep.* **7**, 1031 (2017).

28. Kiserud, T. & Piaggio, G. The World Health Organization fetal growth charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLoS Med.* **14**, e1002220 (2017).

29. Buck Louis, G. M. et al. Racial/ethnic standards for fetal growth: the NICHD fetal growth studies. *Am. J. Obstet. Gynecol.* **213**, 449.e1–449.e41 (2015).

30. Thomas, P. et al. A new look at intrauterine growth and the impact of race, altitude, and gender. *Pediatrics* **106**, E21 (2000).

31. Deter, R. L. et al. Fetal growth cessation in late pregnancy: its impact on predicted size parameters used to classify small for gestational age neonates. *J. Matern. Fetal Neonatal Med.* **28**, 755–765 (2015).

32. Deter, R. L., Hill, R. M. & Tennyson, L. M. Predicting the birth characteristics of normal fetuses 14 weeks before delivery. *J. Clin. Ultrasound* **17**, 89–93 (1989).

33. Yamada, T. et al. Growth of Japanese fetuses--analysis from the viewpoint of population and biological growth curves. *Acta Paediatr. Jpn.* **29**, 229–232 (1987).

34. Cheng, Y. et al. Impact of replacing Chinese ethnicity-specific fetal biometry charts with the INTERGROWTH-21(st) standard. *BJOG* **123**(Suppl 3), 48–55 (2016).

35. Juliusson, P. B. et al. Growth of Belgian and Norwegian children compared to the WHO growth standards: prevalence below -2 and above +2SD and the effect of breastfeeding. *Arch. Dis. Child* **96**, 916–921 (2011).

36. Anderson, N. H. et al. INTERGROWTH-21st vs customized birthweight standards for identification of perinatal mortality and morbidity. *Am. J. Obstet. Gynecol.* **214**, 509.e1–7 (2016).

37. Simic Klaric, A., Tomic Rajic, M. & Tesari Crnkovic, H. Timing of head circumference measurement in newborns. *Clin. Pediatr. (Phila.)* **53**, 456–459 (2014).

38. Souza, S. W., Ross, J. & Milner, R. D. Alterations in head shape of newborn infants after caesarean section or vaginal delivery. *Arch. Dis. Child* **51**, 624–627 (1976).