A longitudinal study of weight gain in pregnancy in Malawi: unconditional and conditional standards

Jiajun Xu, Mari Luntamo, Teija Kulmala, Per Ashorn, and Yin Bun Cheung

ABSTRACT
Background: To monitor weight gain during pregnancy and assess its relation with perinatal health outcomes, both unconditional (cross-sectional) and conditional (longitudinal) standards of maternal weight are needed.

Objective: This study aimed to develop and validate unconditional and conditional maternal weight standards for use in Malawi, Africa.

Design: Longitudinal data were drawn from an antenatal care intervention study conducted in Malawi. Participants were selected for this analysis if they had a healthy profile defined by body mass index and infectious disease measures and delivered healthy singletons defined by birth weight, gestational age, and neonatal survival status. A total of 1733 measurements from 358 women were randomly split to form development and validation samples.

Results: Unconditional and conditional standards were developed and validated. An electronic spreadsheet implements the calculations. Weight gain during pregnancy was substantially slower in this cohort than the US Institute of Medicine recommendation. The percentiles increased linearly; therefore, the use of the conditional standards is robust to inaccuracies in gestational age estimates.

Conclusion: The standards can facilitate researchers and clinicians to examine maternal weight and weight gain and estimate their associations with pregnancy outcomes in Malawi. This trial was registered at www.clinicaltrials.gov as NCT00131235.


INTRODUCTION

Monitoring of weight and weight gain during pregnancy is a routine practice in antenatal care. Inadequate and excessive gestational weight gains are both risk factors of adverse pregnancy and birth outcomes, such as low birth weight or macrosomia (1–4). Research studies of total gestational weight gain in relation to perinatal health outcomes are problematic because it is confounded by less time to gain weight for preterm deliveries (5). The true degree of association between gestational weight gain and health outcomes is therefore unclear. The use of z scores and percentiles derived from a reference would help to understand the relation.

Maternal weight gain reference values are relatively well established for white women (5–9). Reference values have also been developed in several South American countries (10–12) and China (13). There is a scarcity of information for monitoring weight and weight gain during pregnancy in African women. Using data collected from 1978 to 1980 in a rural area in Kenya, Jansen et al (14) estimated that the mean weight gain between 3 mo and term was 5.8 kg. In the late 1980s, Theron and Thompson (15) developed a centile chart for gestational weight gain in an urban population in Western Cape, South Africa. In the mid-1990s, a Malawian cohort study estimated a mean gestational weight gain of 259 g/wk (16). Studies in multiethnic societies have suggested that African and white women may have different patterns in weight gain during pregnancy (7, 17). The availability of locally relevant references or standards would thus be useful.

Monitoring of weight requires unconditional references, also known as cross-sectional references. Monitoring of weight gain requires conditional references, also known as longitudinal references (18–20). In child health research and practice, it has been suggested that growth monitoring with the use of conditional references is more sensitive than is the use of unconditional references in identifying children with health problems (19, 21, 22). In the monitoring of maternal anthropometric measures, weight (in kg) at 5 or 7 mo of pregnancy had a larger OR than pregnancy weight gain (in kg) in the prediction of low birth weight (23). The study also suggested that locally derived cutoff points were needed for efficient use of the indicator (23). However, it is known that analyses using z scores and the original metrics may give different results and offer different perspectives; stunted children may catch-up in terms of height-for-age z scores but not centimeter is an example (18). Whether the predictive ability of

1From the Department of Statistics and Actuarial Sciences, University of Hong Kong, PR China (JX); the Department of International Health, University of Tampere School of Medicine, Tampere, Finland (ML, TK, PA, and YBC); the Sexual and Reproductive Health Unit, National Institute for Health and Welfare, Helsinki, Finland (TK); the Department of Paediatrics, Tampere University Hospital, Tampere, Finland (PA); and the Center for Quantitative Medicine, Duke–National University of Singapore Graduate Medical School, Singapore (YBC).
2YBC was supported by the Singapore Ministry of Health’s National Medical Research Council under its Clinician Scientist Award. ML was supported by a grant from the Finnish Cultural Foundation. LAIS was supported by grants 79787 and 207010 from the Academy of Finland, the Foundation for Pediatric Research in Finland, and the Medical Research Fund of Tampere University Hospital. Azithromycin and its placebo were provided free of charge by Pfizer Inc, which also provided funding for the polymerase chain reaction testing for sexually transmitted infections.
3Address correspondence to YB Cheung, Center for Quantitative Medicine, Duke–NUS Graduate Medical School, Singapore 169857. E-mail: yinbun.cheung@duke-nus.edu.sg.

Received September 3, 2013. Accepted for publication November 1, 2013. doi: 10.3945/ajcn.113.074120.
weight and weight gain should be compared in terms of absolute values or z scores is unclear.

A reference describes what happens in the general population. A standard describes what happens in a “healthy” population. A standard is more suitable for monitoring and health care. The WHO is explicit that its new child growth charts is a set of standards, not references (24). This study aimed to develop unconditional and conditional maternal weight standards for use in Malawi.

SUBJECTS AND METHODS

Study design

This work was based on data from the Lungwena Antenatal Intervention Study (LAIS) conducted in Malawi. Details of the trial were reported recently (25, 26). Briefly, Lungwena is an area in southern Malawi where a government health center served a rural area (~100 km²) with ~17,000 people. More than 95% of pregnant women attended the antenatal care clinic. The literacy rate was low, and the main occupations were subsistence farming and fishing. Malaria was common.

LAIS recruited pregnant women who came to the health center during December 2003–October 2006 for antenatal care and whose ultrasonography-assessed gestational age was 14–25 completed weeks at that time. It randomly assigned the consented participants to 1 of 3 groups: standard antenatal care with 2 doses of sulfadoxine-pyrimethamine (SP) during pregnancy for preventive treatment of malaria, with monthly SP, or with monthly SP plus 2 doses of azithromycin. Two doses of SP was standard care at that time. Sample size was calculated for the comparison between the randomized groups (25). The trial was performed according to Good Clinical Practice guidelines and the ethical standards of Helsinki Declaration. It was approved by the College of Medicine Research and Ethics Committee, Malawi and the Ethical Committee of Pirkannmaa Hospital District, Finland.

Gestational age at enrollment was determined by measuring the fetal biparietal diameter and femur length with an ultrasound imager. Follow-up visits were scheduled at 4-wk intervals until 36 completed gestation weeks and weekly thereafter. The weight of mothers with light clothes was measured by using electronic standing scales at enrollment and subsequent visits. The result was recorded in kilogram to the nearest 0.1 kg. The scales were calibrated with standard weights once a week.

Participants

For the current purpose of developing standards, we selected a “healthy” group of participants whose pregnancy completed in healthy birth outcomes. The control group of the randomized trial was excluded from this analysis because 2 doses of SP are now known to be suboptimal and is no longer the recommended standard of care (27, 28). Currently there is no consensus, and definite evidence about the efficacy of azithromycin together with SP in preventing preterm deliveries, but there is strong evidence about its safety (25, 29). Therefore, the groups with monthly SP and monthly SP plus azithromycin were pooled in this analysis.

Furthermore, the following participants were excluded: women who were HIV positive at enrollment, women with malaria parasitemia at enrollment or midpregnancy, women with a BMI (in kg/m²) <18.5 at enrollment (median 20 wk of gestation), women whose pregnancies ended in still births or death of the infant within 1 mo of birth, deliveries earlier than 37.0 wk or later than 42.0 wk, and birth weight below the first percentile (2.3 kg for both sexes) or above the 99th percentile of the WHO 2006 growth standards. Although it is more common to use <2.5 kg as an exclusion criterion, the WHO 2006 standards were based on a healthy cohort. It shows that, in a healthy cohort, infants may have a birth weight <2.5 kg. Therefore, we used the first percentile of the WHO growth standards.

Statistical methods

For development of standards with the use of repeated-measurement data, it is important that the statistical parameters do not gravitate toward women with more measurements. To make an analysis data set homogeneous in number of measurements per participant, we excluded women who had <4 measurements from the analysis. Four measurements were the mode of the distribution; only 12% of women had <4 measurements. Furthermore, among the women with >4 measurements, we randomly selected 4 measurements for the development of standards and the remaining for the validation.

To allow for nonnormal distribution, we used the parametric modeling approach of Royston and Wright (30) to develop the unconditional standards. The detrended quantile-quantile plot, also known as worm plot (31), was used for model diagnostics. We fitted the normal, exponential-normal, and modulus-exponential-normal models and found that the exponential-normal model provided good fit. Briefly, let \( Y(t) \) be maternal weight at time \( t \) and \( M(t), S(t), \) and \( G(t) \) be, respectively, the median, SD, and skewness parameters of maternal weight as functions of the measurement time \( t \) to be estimated from the data. The standardized variable

\[
z_t = \frac{\exp\left( G(t) \left( \frac{Y(t) - M(t)}{S(t)} \right) \right) - 1}{G(t)}
\]

follows a standard normal distribution. Conversely, percentiles are obtained by plugging the appropriate standard normal value (eg, \( z = -1.28 \) for the 10th percentile) into

\[
Y(t) = M(t) + \frac{S(t)}{G(t)} \ln(z \times G(t) + 1)
\]

Wright and Royston (32) also developed a Stata macro “xriml” to implement this transformation and model fitting by maximum likelihood.

The conditional standards provide z scores and percentiles for current maternal weight given the previous weight and measurement time (19, 20). Expected current weight given observed previous weight is obtained from linear regression of current maternal weight z scores (\( z_m \)) on previous maternal weight z score.
It has been shown that the expected value for \( z_2 \) and its SD are (19) as follows:

\[
E(z_2|z_1) = \rho \times z_1 \tag{3}
\]

\[
SD = \sqrt{1 - \rho^2} \tag{4}
\]

where \( \rho \) is the correlation coefficient between \( z_1 \) and \( z_2 \). We used the Markov correlation structure to estimate the correlation. The Markov correlation structure is chosen because it is biologically plausible and succinct. It depends on only one parameter, \( \alpha \). The correlation coefficient between 2 \( z \) scores at different time points within the same person can be derived as follows:

\[
corr(z_{ij}, z_{ik}) = \alpha^{|t_{ij} - t_{ik}|} \tag{5}
\]

where \( i \) denotes the woman, \( i \) and \( k \) represent 2 points of weight measurement in completed weeks of gestation, and the absolute value of \( t_{ij} - t_{ik} \) represents the gap between the 2 time points. The further apart they are, the weaker the correlation. Estimation of the parameter \( \alpha \) was implemented by the Stata macro “xtqls” (33). We also examined the stability of the estimated correlation structure in different ranges of gestational age. Conditional weight gain percentiles and \( z \) scores are obtained by applying the above formula. We provide an electronic spreadsheet for the calculation (see Supplemental Table 1 under “Supplemental data” in the online issue). Statistical inference for regression coefficients were conducted by bootstrapping method to allow for multiple observations per person. In the validation, the Shapiro-Wilk normality test was applied to the \( z \) scores in each gestational age interval (34, 35). The worm plot was also used to assess normality. The Pearson’s correlation coefficient between conditional \( z \) scores in an interval and the corresponding \( z \) scores at the beginning of the interval was estimated to assess the performance of the conditional standards (proper performance should give approximately zero correlation). Furthermore, logistic regression was used to check that the proportions below the 10th percentile or above the 90th percentile were independent of gestational age. All statistical analyses were performed in Stata version 12.

### RESULTS

#### Development and validation data sets

There were 880 women with singleton pregnancies in the 2 intervention groups. A total of 471 women were excluded because of the selection criteria for defining a “healthy” cohort. Examination of individual trajectories suggested that 24 measurements from 21 women were implausible values and they were excluded. Furthermore, 50 women who had <4 measurements were also excluded. Following the procedure described in the previous section, the development sample consisted of 1432 measurements from 358 women. Each contributed 4 measurements. The validation sample consisted of 301 measurements from 203 women. The number of measurements per woman ranged from 1 to 5 in the validation sample.

#### Unconditional standards

The earliest measurement record was at 14 wk, and the latest record was at 41 wk. Each woman had a measurement every 4 or 5 wk. The fitted \( M(t), S(t) \), and \( G(t) \) were as follows:

\[
M(t) = 48.491 + 0.245t \tag{6}
\]

\[
S(t) = 4.838 + 0.039t \tag{7}
\]

\[
G(t) = -0.157 \tag{8}
\]

All the coefficients in the 3 functions were statistically significant (each \( P < 0.01 \)). The percentiles can be calculated by using Equation 2 accordingly. For example, the third percentile \((z_{0.03} = -1.881)\) for \( t = 35 \) completed weeks of gestation is as follows:

\[
y_{35,0.03} = \left\{ \frac{48.491 + 0.245 \times 35}{4.838 + 0.039 \times 35} \right\} + 0.157 \times 0.157 \times (-1.881) + 1 = 46.844 \tag{9}
\]

The unconditional standard chart for maternal weight is shown in Figure 1. The percentiles increased linearly. The vertical distance between percentiles increased slightly over time.

#### Conditional standards

When the Markov structure was applied to the \( z \) scores from the development data set, the estimated correlation coefficient parameter \( \alpha \) was 0.9949. When Equation 5 was used, the correlation between, eg, 30 wk and 40 wk would be 0.9949 raised to the 10th power, giving 0.9502. To check the robustness of the estimate, the data set was partitioned into 2 parts—gestational age <30 and \( \geq 30 \) wk—and the parameter \( \alpha \) was separately estimated. The estimated values were 0.9944 and 0.9951, respectively. The correlation was stable during pregnancy. With the correlation between 2 measurements, weight percentiles and \( z \) scores conditional on the earlier weight are readily

![FIGURE 1. Unconditional chart for maternal weight in Malawi.](image-url)
obtainable by using the formula aforesaid. The electronic spreadsheet we provide facilitates this. A hypothetical example is shown in Figure 2. If a woman weighed 59.6 kg at 25 wk, the unconditional $z$ score was 0.802 at 25 wk. We can calculate the conditional $z$ scores corresponding to the median and 10th and 90th percentiles at her next visit (29 wk) by using Equations 3–5:

$$z_{29, 0.5} = 0.802 \times 0.9949^{29-25} = 0.786$$

(10)

$$z_{29, 0.1} = 0.802 \times 0.9949^{29-25} + z_{0.1} \times \sqrt{1 - (0.9949^{29-25})^2} = 0.529$$

(11)

$$z_{29, 0.9} = 0.802 \times 0.9949^{29-25} + z_{0.9} \times \sqrt{1 - (0.9949^{29-25})^2} = 1.043$$

(12)

where $z_{0.1} = -1.282$ and $z_{0.9} = 1.282$ are the standard normal values for the 10th and 90th percentiles. Then the conditional $z$ scores are transformed back to body weight to form the conditional percentiles with the use of Equation 2, i.e.,

$$Y_{29, 0.5} = 48.491 + 0.245 \times 29 + 4.838 + 0.039 \times 29 \ln[-0.157 \times 0.786 + 1] = 60.60$$

(13)

for the median. The median conditional weight gain expected for this participant was therefore 60.6–59.6 = 1 kg. The other percentiles were obtained similarly. If this woman’s weight dropped by 1 kg to 58.6 kg at 29 wk, her unconditional $z$ scores would be 0.484 at 29 wk and her weight would still be above the median of the unconditional standard. However, the use of her known previous weight and the use of Equations 3 and 4 showed her conditional $z$ score to be

$$\frac{0.484 - 0.802 \times 0.9949^{29-25}}{\sqrt{1 - (0.9949^{29-25})^2}} = -1.509$$

(14)

at 29 wk, which fell below the 10th percentile in the conditional standard. The correlation between initial weight $z$ scores at each pair of consecutive visits and the conditional weight gain $z$ scores was close to zero ($r = -0.06$), as expected for valid conditional standards.

**Validation**

In the validation data set, 25 (8.3%) and 36 (12.0%) observations were below the 10th percentile and above the 90th percentile of the unconditional standards, respectively. The 95% CIs were 4.2% and 12.4% for the classification below the 10th percentile and 7.1% and 16.9% for the classification above the 90th percentile. Furthermore, logistic regression showed no association between gestational age and classification below the 10th percentile or above the 90th percentile (each $P > 0.05$). The percentiles classified the validation observations appropriately.

The worm plot of the $z$ scores when the unconditional standard was applied to the validation data set is shown in Figure 3. The worms were around the horizontal line, and most data points were within the 95% CI, except for some deviation among the observations with the earliest gestation weeks (14–18 wk; upper-left corner of Figure 3). The U shape in that panel suggested positive skewness. The skewness in the $z$ scores in this subgroup was 1.20, and the test for normality gave $P = 0.019$. However, the departure was largely driven by one influential observation, at 3.05 $z$ scores. Without this value, the skewness was 0.53 and the test for normality gave $P = 0.534$. Among the other 8 panels in the worm plot, the skewness ranged from $-0.76$ to 0.64 (median: $-0.02$), and each test for normality gave $P > 0.05$.

The validation sample had 301 observations from 203 women, giving 98 intervals. Conditional weight gain $z$ scores were calculable for the 98 intervals. The correlation between the conditional weight gain $z$ scores and the initial weight $z$ scores were close to zero ($r = -0.05$), as expected for valid conditional standards. The worm plot of the conditional $z$ scores is shown in Figure 4. Because the sample size here was relatively small, only 4 age intervals were used in the worm plot. Again, the worms were flat and most data points were within the 95% CI. Normality test in each panel did not suggest departure from normal distribution (each $P > 0.20$).
DISCUSSION

We have developed standards on gestational weight gain in a low-income African setting. In contrast with some previous studies, we provided not only unconditional (cross-sectional) but also conditional (longitudinal) standards. We have provided not only percentiles but also parameters for calculating \( z \) scores. It has been suggested that conditional standards are more sensitive in child growth monitoring (19, 21, 22). It is known that categorization leads to loss of precision; therefore, \( z \) scores are preferred to percentiles in research. We also provided an electronic spreadsheet for easy application of the standards. Furthermore, we have not only developed the standards but also provided model diagnostic and validation results. They showed the proper performance of the standards.

A random-effects model is one of the appropriate approaches for developing growth standards with the use of longitudinal data. However, it does not always fit the data well, and some users do not provide model diagnostics results (36). Our research initially began with the use of the random-effects model proposed by Royston (20) (details not shown). However, model diagnostic results showed that the random-effects approach did not fit this data set satisfactorily. We decided to switch to an appropriate alternative. The combined use of cross-sectional data analysis methods to generate \( z \) scores and then modeling the relation of the \( z \) scores between ages is another appropriate approach for developing growth standards by using longitudinal data (19, 37).

A concern about step one in this 2-step approach is that participants with a small number of observations may be underrepresented (20). The use of 4 observations for the development sample not only addressed this concern but also spared a set of observations for model validation, which is an issue that some studies in this area have omitted to look at (36). This number of observations per participant is similar to that of the INTERGROWTH 21st Fetal Growth Longitudinal Study (38).

The conditional standard takes weight at the first visit or previous visit into account. A useful feature of the standards is that the percentiles increase linearly. As a result, accurate gestational age information is not needed for using the conditional standards to monitor weight gain. This is particularly useful in a low-income setting, where an accurate estimate of gestational age may not be available. Only the number of weeks between 2 visits needs to be reliable. Because the SD increases slightly over gestational age, despite the linearity in the percentiles and accurate entry of number of weeks between visits, misspecification of the initial gestational age would lead to some inaccuracy in the weight gain \( z \) scores. However, experimentation with the electronic spreadsheet shows that varying the initial gestational age by 4 wk rarely gave a discrepancy that exceeds 0.01 \( z \) scores. Nevertheless, the use of conditional standards in clinical practice warrants caution. Conditioning a woman’s current weight on the previous weight that was the outcome of poor early weight gain may be normalizing a poor weight gain trajectory. Conditional standards only show one specific aspect of the picture. It may be more useful in identifying acute but not chronic growth faltering. The joint use of unconditional and conditional standards would tell a more comprehensive picture.

Formal comparison of the current standards with previous standards is impossible because of various differences in methods. For example, Hutcheon et al (5) showed weight gain calculated by subtracting self-reported prepregnancy weight from measured weight at antenatal clinics, whereas our standards are based only on measured weight. Kulmala et al (16) plotted the mean &plus; 1 SD curves, implicitly assuming normal distribution, whereas our standards handled nonnormal distribution by using the exponential-normal model. Nevertheless, it is clear that the average weight gain in this sample, which had healthy birth outcomes and a healthy profile at first antenatal visit, is only roughly half of that recommended by the US Institute of Medicine (39) and is close to the 10th percentile in the urban population in Western Cape, South Africa (15). The use of locally inappropriate standards may lead to a high false-positive rate.

In a low-income setting, clinical assessments that are simple, portable, and inexpensive are important. Monitoring of maternal weight gain is a common practice, but its predictive value has been debated; consequently, the practice is variable even within the same country (40, 41). Two issues should be considered. First, it is important to develop and use diagnostic tools appropriately. Some previous criticisms about monitoring gestational weight gain are essentially about the inappropriate development and use of references, such as monitoring unconditional but not conditional values (41). Without the proper standards available, the predictive value of maternal weight gain would remain unknown (5). Second, it may be that no single diagnostic measure can sufficiently predict health outcomes. However, each of these measures can contribute to some extent and possibly be pooled to form a composite measure.

In a low-income setting, it is common that women start antenatal care after the first trimester. A limitation of LAIS was that it recruited the participants when gestational age was 14–25 wk and did not have information on prepregnancy weight. Therefore, we did not examine weight gain according to categories of prepregnancy weight or BMI. For the purpose of identifying a healthy cohort for developing growth standards as opposed to growth references—one of the inclusion criteria was BMI at enrollment &gt;18.5. The analysis sample did not include women with pathologically low weight at enrollment. Overweight is uncommon in Malawi. In the current sample, only 3 women had BMI at enrollment &gt;27. They were kept in the analysis sample because we could not definitely determine whether they were overweight before pregnancy.

Another limitation is that the women were recruited from only one area in Malawi. Nevertheless, the analysis sample was chosen on medical grounds to represent a healthy standard. It was not
chosen to be representative of pregnancies in the population. Only participants who had a healthy profile at baseline, were offered currently recommended care, and had healthy birth outcomes were included in this development of standards. Clinically, this selected sample is a proper target for comparison. Furthermore, birth outcomes in the study area are similar to some other countries in southern Africa. For example, the incidence of low birth weight was 15.1% in an earlier Malawian study (42) and 16.2% in its neighboring country, Mozambique (43). In the control group of LAIS, the incidence of low birth weight was 15.7% (26). Lungwena, where the analysis sample for LAIS was drawn, is fairly typical of low-income settings in southern Africa. Additional research to examine the generalizability and applicability of the proposed standards will be useful.

Kenneth Maleta contributed to the design and conduct of LAIS. The authors’ responsibilities were as follows—PA and TK: designed LAIS; PA, TK, and ML: conducted the study; YBC: designed this secondary analysis; and YBC and JX: co-analyzed the data and cowrote the first draft of the manuscript. All authors helped interpret the findings and critically revise the manuscript. All authors approved the submission. All authors declare that they had no conflict of interest. The sponsors had no role in the study design, data collection, data analysis, interpretation of data, or preparation of the manuscript or in the decision to submit for publication.

REFERENCES


