Monitoring postnatal growth of preterm infants: present and future1–3

Francesca Giuliani,4 Leila Cheikh Ismail,5,6 Enrico Bertino,4 Zulfiqar A Bhutta,7,8 Eric O Ohuma,5,9 Ilaria Rovelli,4 Agustin Conde-Agudelo,10,11 José Villar,5,6,12 and Stephen H Kennedy5,6,12*

4Department of Public Health and Pediatrics, University of Turin, Turin, Italy; 5Nuffield Department of Obstetrics & Gynaecology, and 6Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, United Kingdom; 7Center of Excellence in Women & Child Health, The Aga Khan University, Karachi, Pakistan; 8Center for Global Child Health, Hospital for Sick Children, Toronto, Canada; 9Centre for Statistics in Medicine, University of Oxford Botnar Research Centre, Oxford, United Kingdom; 10Perinatology Research Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD; 11National Institutes of Health/Department of Health and Human Services, Detroit, MI

ABSTRACT
Background: There is no consensus with regard to which charts are most suitable for monitoring the postnatal growth of preterm infants.
Objective: We aimed to assess the strategies used to develop existing postnatal growth charts for preterm infants and their methodologic quality.
Design: A systematic review of observational longitudinal studies, whose primary objective was to create postnatal growth charts for preterm infants, was conducted. Thirty-eight items distributed in 3 methodologic domains (“study design,” “statistical methods,” and “reporting methods”) were assessed in each study. Each item was scored as a “low” or “high” risk of bias. Two reviewers independently selected the studies, assessed the risk of bias, and extracted data. A total quality score [(number of “low risk” of bias marks/total number of items assessed) × 100%] was calculated for each study and for all included studies were computed.
Results: Sixty-one studies met the inclusion criteria. Twenty-seven (44.3%) of the 61 studies scored ≥50%, of which 10 scored >60% and only 1 scored >66%. The median (range, IQR) quality score for the 61 included studies was 47% (26–75%, 34–56%). The scores for the domains study design, statistical methods, and reporting methods were 44% (19–67%, 33–52%), 25% (0–88%, 13–38%), and 33% (0–100%, 0–33%), respectively. The most common shortcomings were observed in items related to anthropometric measures (the main variable of interest), gestational age estimation, follow-up duration, reporting of postnatal care and morbidity, assessment of outliers, covariates, and chart presentation.
Conclusions: The overall methodologic quality of existing longitudinal studies was fair to low. To overcome these problems, the Preterm Postnatal Follow-up Study, 1 of the 3 main components of The International Fetal and Newborn Growth Consortium for the 21st Century Project, was designed to construct preterm postnatal growth standards from a prospective cohort of “healthy” pregnancies and preterm newborns without evidence of fetal growth restriction. Am J Clin Nutr doi: 10.3945/ajcn.114.106310.

Keywords: preterm birth, syndrome, phenotypes, perinatal outcomes, growth charts, postnatal growth, systematic review

INTRODUCTION
Preterm birth is the leading cause of perinatal mortality worldwide and the second largest cause of deaths in children <5 y of age (1, 2). Those preterm infants who survive are at risk of a range of health problems in later life, such as high blood pressure (3, 4) and impaired neurodevelopment (5), especially if born extremely preterm and/or with low birth weight (LBW)13 (6). Therefore, ensuring that postnatal growth is as healthy as possible is critical to improving survival and long-term outcomes in preterm infants. This requires having robust standards to monitor their growth (7), which is problematic in preterm infants given the lack of consensus regarding the most suitable charts to use (8).

Here, we discuss conceptual issues related to the strategies presently used for monitoring the postnatal growth of preterm infants and present the results of a systematic review that assesses the quality of longitudinal studies (as opposed to cross-sectional studies, which are the basis for most charts currently in use in clinical practice) using a set of predefined quality criteria for study design, and statistical and reporting methods. On the basis of this extensive knowledge base, we outline in brief the protocol and principal findings of the Preterm Postnatal Follow-up Study (PPFS), which was designed specifically to correct the

1Presented at the meeting “Evaluating the Evidence to Support Guidelines for the Nutritional Care of Preterm Infants: The Pre-B Project” held at the USDA/Agricultural Research Service Children’s Nutrition Research Center, Baylor College of Medicine, Houston, TX, 31 July–1 August 2014.
2Supported by the INTERGROWTH-21st grant 9038 from the Bill & Melinda Gates Foundation to the University of Oxford.
3Supplemental Tables 1 and 2 are available from the “Online Supporting Material” link in the online posting of the article and from the same link in the online table of contents at http://ajcn.nutrition.org.
12These authors contributed equally to this work.
*To whom correspondence should be addressed. E-mail: stephen.kennedy@obs-gyn.ox.ac.uk.
13Abbreviations used: BL, body length; BW, birth weight; FGR, fetal growth restriction; HC, head circumference; INTERGROWTH-21st, The International Fetal and Newborn Growth Consortium for the 21st Century; LBW, low birth weight; LMP, last menstrual period; PPFS, Preterm Postnatal Follow-up Study. doi: 10.3945/ajcn.114.106310.
shortcomings of current growth-monitoring strategies for these high-risk infants so as to construct preterm postnatal growth charts by using longitudinal measurements derived from a cohort of “healthy” preterm newborns.

In reviews (7, 10) we highlighted the current approaches to monitoring postnatal growth in preterm infants using 1) fetal growth charts based on ultrasound estimations of fetal weight, 2) birth weight (BW)-for-gestational-age charts, 3) postnatal longitudinal growth charts for preterm infants, 4) prescriptive growth standards for infants born at term (37–41 wk), or 5) a combination of numbers 2 and 4. All have their own problems, which should be considered.

At present, the American Academy of Pediatrics recommends that postnatal growth for preterm infants should approximate the intrauterine growth of fetuses of the same gestational age (11). However, this strategy is still being debated and may be inadequate in the early neonatal period (12), because it is evident that the intra- and extrauterine environments are markedly different. For example, the latter is associated with increased energy expenditure and nutrient losses affecting postnatal growth (13), which has been described in preterm infants with BWs between 501 and 1500 g (14).

Moreover, if estimated fetal weight is to be used, it needs to be calculated at each gestational age to construct ultrasound-based charts of intrauterine growth, but the methods available for estimating fetal weight are known to generate large errors. They combine the lack of precision of ultrasound measurements with the inaccuracy of the algorithms used to convert one-dimensional ultrasound traits into a three-dimensional neonatal trait (i.e., weight). Consequently, fetal growth curves derived from ultrasound measurements may not be appropriate either for the longitudinal or cross-sectional evaluation of preterm infant growth, especially with regard to weight (15).

Alternatively, BW-for-gestational-age “neonatal charts” are often used up to term. However, we believe that they should not be used to monitor postnatal growth because they are derived from cross-sectional data collected at birth and therefore cannot provide “growth” patterns; rather, they only describe size at birth in pregnancies that had complications before delivery. In addition, it is evident that if preterm infants are going to reach the same fetal growth rates, they should have (after the documented postnatal weight lost) considerably higher growth velocity than fetuses in utero. The long-term effects of such a forced velocity need to be evaluated: for example, does rapid weight increment increase metabolic risk, leading to diabetes, hypertension, or dyslipidemia in adulthood (8)? Finally, it is assumed that, by term, these infants can be graduated to the WHO Child Growth Standards, although they will be considered already undernourished.

Instead, we believe that the charts should be developed from longitudinal postnatal anthropometric measurements (which is precisely what they should be monitoring), obtained specifically from preterm infants, because the charts of term infants are prone to have the preterm growth variables plotted at the lowest centiles. The key question that remains, however, when constructing such charts is which preterm population to select—that is, is it possible to develop standards for monitoring the growth of preterm infants that are similar to those we have already for infants and children?

Although numerous postnatal growth charts for preterm infants have been reported in the literature, the methodologic quality of the studies from which they were derived has not been critically appraised and the clinical implications of these limitations are often ignored. We review these charts below.

METHODS

The study followed a prospective protocol and is reported with the use of the MOOSE (Meta-analysis Of Observational Studies in Epidemiology) guidelines for systematic reviews of observational studies (16). We searched MEDLINE (http://www.ncbi.nlm.nih.gov/pubmed), EMBASE (https://www.elsevier.com), CINAHL (https://health.ebsco.com/products/the-cinahl-database), and LILACS (http://lilacs.bvsalud.org/en) (all from inception to 15 April 2014) and Google Scholar using a combination of MeSH or key word terms related to preterm infant (“preterm infant,” “premature infant,” “infant, premature,” “infant, extremely premature,” “infant, low birth weight,” “infant, very low birth weight,” “infant, newborn”) and growth charts (“growth charts,” “growth curves,” “anthropometric charts,” “intrauterine growth charts,” “neonatal growth charts,” “weight growth,” “growth velocity,” “postnatal growth,” “catch-up growth,” “postnatal growth failure”). To identify additional publications, we searched bibliographies of the retrieved articles. No language restrictions were imposed.

We included observational longitudinal studies whose primary objective was to create postnatal growth charts for preterm infants. Studies were excluded from the systematic review if the primary aim was anything other than the construction of postnatal growth charts (e.g., comparisons between different populations), if their aim was to reanalyze previously published charts, or if they used cross-sectional data. All published studies deemed suitable were retrieved and reviewed independently by 2 investigators (FG and IR) to determine inclusion. Disagreements were resolved through consensus or consultation with a third reviewer (LCI).

The methodologic quality assessment of each included study was performed by using a prespecified list of criteria developed by FG, EB, and JV, and agreed upon by all of the authors. This is a modified and adapted version of the criteria used in our previous evaluations of fetal size (17) and neonatal anthropometric charts (18). The quality criteria were divided into 3 domains (domain 1: “study design”; domain 2: “statistical methods”; and domain 3: “reporting methods”) deemed to be important for the quality of the studies aimed at creating postnatal growth charts for preterm infants (Table 1). Twenty-four criteria were assessed in the 3 domains (16 in study design, 6 in statistical methods, and 2 in reporting methods). However, some criteria consisted of >1 item. In total, 38 items were assessed in each included study (27 in study design, 8 in statistical methods, and 3 in reporting of results) by 3 investigators [2 neonatologists (FG and IR) and 1 medical statistician (EOO)] working independently. Each item was scored as either a “low” or “high” risk of bias. If there was insufficient information available to make a judgment about some items, then they were scored as “not evaluable.” Disagreements were resolved either by consensus or consultation with another reviewer (LCI).

Data were extracted independently from each article by 2 investigators (FG and IR) by using a standardized and pilot-tested data collection form. Information was extracted on the first author’s name, geographic location of the study, year of publication, sample size, and the 38 prespecified methodologic items. Disagreements with regard to extracted data were resolved by
<table>
<thead>
<tr>
<th>Domain</th>
<th>Low risk of bias</th>
<th>High risk of bias</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Study design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1. Aim of the study</td>
<td>To construct postnatal growth charts for preterm infants. The aim in the abstract and in the text was the same.</td>
<td>Not clearly defined. Different aim in the abstract and in the text.</td>
<td>Better and more reliable charts will be obtained if the primary aim of the study was to create them.</td>
</tr>
<tr>
<td>1.2. Design</td>
<td>Clearly described and longitudinal.</td>
<td>Unreported or unclear.</td>
<td>Longitudinal approach required. References and standards are different tools.</td>
</tr>
<tr>
<td>1.3. Distinction between reference and standard growth chart</td>
<td>The authors clearly stated whether the chart was a reference or a standard or this information could be discerned from the Methods section.</td>
<td>Not clearly defined.</td>
<td></td>
</tr>
<tr>
<td>1.4. Definition of target population</td>
<td>Clear definition of target population (e.g., geographical area, race-ethnicity, single or multiple pregnancy, population characteristics, among others).</td>
<td>Not clearly defined.</td>
<td>It determines to which population the chart can be appropriately applied.</td>
</tr>
<tr>
<td>1.5. Sample selection</td>
<td>The sample was part of the target population and the inclusion/exclusion criteria were clearly reported. Data were collected prospectively enrolling the neonates consecutively.</td>
<td>Inclusion/exclusion criteria were not clearly reported. The study was not planned or the data were not collected prospectively (e.g., data set collected for a different study or hospital or national registries).</td>
<td>Planning the study, selecting the sample within the target population, listing inclusion/exclusion criteria, and collecting data prospectively will lead to better and more reliable charts.</td>
</tr>
<tr>
<td>1.6. Selection of preterm neonates</td>
<td>Only selected preterm infants according to GA.</td>
<td>Selection of “preterm” infants according to birth weight instead of GA.</td>
<td>Low birth weight represents a heterogeneous group of conditions. Selection of newborns according to birth weight will lead to an overrepresentation of SGA or IUGR infants, which should be studied separately.</td>
</tr>
<tr>
<td>1.7 Sample size</td>
<td>Clear report of the number of: 1.7.1. Neonates at each GA at birth 1.7.2. Losses to follow-up 1.7.3. Visits attended by each infant</td>
<td>No report or unclear report of the number of neonates at each GA at birth, losses to follow-up, or visits attended by each infant.</td>
<td>The precision of the estimates increases with the number of observations.</td>
</tr>
<tr>
<td>1.8. Number of measurements and time elapsed between 2 measurements</td>
<td>1.8.1. Measurements were taken at birth or within first 24 h. 1.8.2. Measurement follow-up schedule after birth: measurements were taken before discharge if the infant was discharged before 15 d of life. Subsequently, the measurements were taken at least every 2 wk for the first 2 mo, then at least monthly until 6 mo of age, and at least every 3 mo up to 2 y.</td>
<td>Measurements were taken after first 24 h of life or the intervals between the measurements were longer than those set in criterion 1.8.2. or were unreported</td>
<td>The frequency of anthropometric measurements should take into account the occurrence of periods of fast growth (e.g., the first 2 mo).</td>
</tr>
<tr>
<td>1.9. Follow-up duration</td>
<td>At least 2 y.</td>
<td>Less than 2 y.</td>
<td>Follow-up should be long enough during infancy to allow interface with child growth charts.</td>
</tr>
<tr>
<td>1.10. Infant nutrition</td>
<td>Clear description of actual feeding practices.</td>
<td>Not clearly described.</td>
<td>There are differences in the nutritional practices of preterm newborns among neonatal units. The resulting growth charts are influenced by this aspect, which should be described.</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 1 (Continued)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Low risk of bias</th>
<th>High risk of bias</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.11. Postnatal care</td>
<td>Clear description of postnatal care. If the study is multicenter, the basic level of neonatal care should be similar among centers.</td>
<td>Not clearly described. If the study was multicenter, the basic level of neonatal care was very different among centers or eventual differences were not considered.</td>
<td>Postnatal care has an influence on growth. Routine practices should therefore be described.</td>
</tr>
<tr>
<td>1.12. Postnatal morbidities</td>
<td>Clear report of the occurrence of major morbidities known to affect postnatal growth.</td>
<td>No report or unclear report of major morbidities known to affect postnatal growth.</td>
<td>The report of postnatal morbidities that can affect postnatal growth is useful in interpreting the charts.</td>
</tr>
<tr>
<td>1.13.1. Based on LMP and confirmed by early ultrasound assessment</td>
<td>1.13.2. Based on a difference of &lt;2 wk between ultrasound and LMP assessments</td>
<td>1.13.3 Exclusion of neonates with unreliable GA</td>
<td></td>
</tr>
<tr>
<td>1.14. Postnatal age evaluation</td>
<td>Clear report of the postnatal age used (corrected or chronological).</td>
<td>Not reported or unclear.</td>
<td>Postnatal age used should be clear to the users.</td>
</tr>
<tr>
<td>1.15. Anthropometric evaluation</td>
<td>1.15.1. Anthropometric traits were measured using standardized instruments.</td>
<td>Standardized instruments or protocols were not used or reported. The instruments were not periodically calibrated. The measures were taken by only one operator. The operators were not trained/standardized.</td>
<td>Poor reliability of measurements increases dispersion.</td>
</tr>
<tr>
<td>1.15.2. Instruments were calibrated at least fortnightly.</td>
<td>1.15.3. Anthropometric traits were measured by using standardized protocols.</td>
<td>1.15.4. Measures were taken by at least 2 different operators.</td>
<td></td>
</tr>
<tr>
<td>1.15.5. Operators were trained.</td>
<td>1.15.6. Operators were standardized.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.16. Characteristics of the mothers</td>
<td>Report of baseline characteristics and pregnancy complications of the mothers.</td>
<td>Not reported or unclear.</td>
<td>Useful information when interpreting growth curves.</td>
</tr>
<tr>
<td>2. Statistical methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1. Assessment of outliers</td>
<td>2.1.1. Method used to detect the outliers was appropriate, well described, and justified.</td>
<td>The method used was not appropriate (e.g., low or high percentiles), not reported, or not justified. Outliers were not evaluated or not corrected or excluded.</td>
<td>The presence of outliers (observations markedly different in value from the others of the sample) can affect the estimates.</td>
</tr>
<tr>
<td>2.2. Covariates</td>
<td>Growth charts were presented according to: 2.2.1 Sex.</td>
<td>Specific charts according to sex and GA were not produced.</td>
<td>Sex exerts an effect on neonatal size of preterm infants, which is more relevant for higher GA.</td>
</tr>
<tr>
<td>2.2.2 GA at birth (mandatory)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3. Statistical models</td>
<td>The model used to construct the charts was clearly reported and identified.</td>
<td>Not reported, unclear, or model not appropriate.</td>
<td>It is crucial to use an adequate statistical model for creating reliable charts.</td>
</tr>
<tr>
<td>2.4. Assessment of goodness-of-fit of the models</td>
<td>Method used for assessment of goodness-of-fit was reported.</td>
<td>Not reported or not done.</td>
<td>It allows the reader to evaluate the adequacy of the model used to trace charts.</td>
</tr>
<tr>
<td>2.5. Lack of precision of the estimates</td>
<td>SEs (or confidence limits) of outer percentiles were reported, and they were small.</td>
<td>Not done, not reported, or nonsmall SEs.</td>
<td>It is a measure of the reliability of charts.</td>
</tr>
<tr>
<td>2.6. Smoothing</td>
<td>Smoothed centiles</td>
<td>Raw centiles.</td>
<td>Smoothing reduces the fluctuations observed in raw centiles due to sampling variability.</td>
</tr>
</tbody>
</table>
### Results

The searches produced 943 citations, of which 116 (12%) were considered to be potentially eligible (Figure 1). A total of 827 studies were excluded, mainly because they assessed the effect of interventions on growth (21%), focused only on pregnancies or fetuses (19%), had a cross-sectional design or assessed only term infants (18%), or evaluated the association between growth and clinical and/or neurological outcomes (18%). References for excluded studies can be obtained from the authors. Sixty-one (53%) studies fulfilled the inclusion criteria (14, 19–78), of which 50 (82%) were published in English, 5 (8%) in Portuguese, 4 (7%) in Spanish, and 2 (3%) in French.

The main characteristics, assessment of methodologic quality and total quality score for each included study, are shown in Supplemental Tables 1 and 2. Twenty-one (34%) studies were conducted in North America (19 in the United States and 2 in Canada), 17 (28%) in Europe, 12 (20%) in Latin America, 9 (15%) in Asia, and 1 (2%) each in Africa and Australia. The earliest study was published in 1948 and the latest in 2014. Thirty-seven (61%) studies were published before 1980 and 24 (39%) in or after 2000. Some of the included studies reported results related to the same sample (22–24, 27, 28) or the same research program (36, 37, 52, 53, 72). Data were collected prospectively in 45 (74%) studies, retrospectively in 15 (25%), and were unclear in 1 (2%) study. Seventeen (28%) studies reported on weight only, and 44 (72%) reported on weight and body length (BL) and/or head circumference (HC). Table 2 depicts the risk of bias for each of the methodologic criteria assessed in the 3 domains.

### Study design–related methodologic criteria

Forty-eight (79%) studies created reference charts and 13 (21%) standard charts. In approximately half the studies, the authors failed to define clearly the target population and/or report if their objective was to create reference or standard charts. In addition, LBW or very-LBW infants were used as a proxy for preterm infants in almost 40% of the studies. The sample size used to draw charts ranged from 17 (29) to 4973 (39) (median: 126). Eighteen (30%) studies had a sample size ≥100. Data on measurements were collected prospectively and explicitly for research purposes in approximately two-thirds of the studies. In 25 (41%) studies, the time elapsed between 2 measurements was longer than that currently recommended in most follow-up protocols after birth (79, 80). The duration of follow-up varied from 30 d (68) to 11 y (47) (median: 1 y). Twenty-eight (46%) studies reported a follow-up until discharge from the neonatal intensive care unit, and 25 (41%) studies reported a follow-up until at least 1 y. Only 12 (20%) studies reported a follow-up duration ≥2 y (4 related to the same research program and 2 to the same sample). The routine practices for newborn nutrition and postnatal care and postnatal morbidities were described in 52%, 11%, and 36% of the studies, respectively. Only 13 (21%) studies estimated gestational age through the use of both the last menstrual period (LMP) and an early ultrasound scan. Moreover, only 4 (7%) studies clearly excluded neonates with unreliable gestational ages. Several studies (35–37, 40, 48, 52–56, 58, 64–66, 69, 72, 73, 76, 77) used methods of clinical assessment for determining gestational age at birth, such as Ballard (81), Dubowitz (82), or Capurro (83) scores. There was a low risk of bias in postnatal age evaluation. In general, the anthropometric evaluation was subjected to a high risk of bias: standarized instruments were used in 31 (51%) studies, their calibration was reported in 16 (26%), the measurement techniques and protocols were clearly described in 26 (43%), and the measurements were taken more than once in only 10 (16%) studies. Few studies (41, 58, 64, 70, 71, 78) reported on the specific techniques used to calibrate instruments and perform measurements (84–88). The baseline characteristics and pregnancy complications of the mothers were reported in only one-quarter of the studies.

---

**TABLE 1** (Continued)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Low risk of bias</th>
<th>High risk of bias</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Reporting methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1. Characteristics of the study population</td>
<td>Baseline characteristics (pregnancy and neonatal morbidity and mortality) were presented in tables or in text.</td>
<td>Not presented in tables or described in the text.</td>
<td>It determines if the population studied belongs to the target population.</td>
</tr>
<tr>
<td>3.2. Charts presentation</td>
<td>3.2.1. Values of (at least) 10th, 50th, and 90th percentiles were not reported or not computable. z Scores were not presented or not computable.</td>
<td></td>
<td>Assessment of neonatal size by using graphical methods has a lower precision than the use of numerical values. The possibility to express the size as a z score increases the reliability of size evaluation. It also allows comparison between different subjects or populations.</td>
</tr>
</tbody>
</table>

1GA, gestational age; IUGR, intrauterine growth-restricted; LMP, last menstrual period; SGA, small for gestational age.

---

**3.2. Charts presentation**

- **Figures 1 and 2.**
- **Values of 10th, 50th, or 90th percentiles were not reported or not computable.**
- **Z Scores were not presented or not computable.**

- **Baseline characteristics (pregnancy and neonatal morbidity and mortality) were presented in tables or in text.**

- **Not presented in tables or described in the text.**

The possible time to express the size as a z score increases the reliability of size evaluation. It also allows comparison between different subjects or populations.
Statistical methods–related methodologic criteria

The highest risk of bias was found in the field “assessment of outliers.” In fact, only 5 (8%) studies clearly stated how outliers were detected and only one corrected or excluded them in a clear way. Most (92%) studies reported and clearly identified the model used to construct the charts. Assessment of the goodness-of-fit of the proposed equation and presentation of growth charts according to sex were performed in only one-third of the studies. Eighteen (30%) studies presented the charts by gestational age, either by each week or by gestational age groups, whereas 27 (44%) used BW categories as a covariate. Smoothed centiles and SEs or confidence limits of extreme centiles were reported in only 20–25% of studies.

Reporting methods–related methodologic criteria

Fifty-nine percent of studies presented the study population characteristics in tables or text. Only 13 (21%) studies reported at least the 10th, 50th, and 90th centiles, or parameters for allowing their computation. In addition, only 15% of the studies presented the charts with $z$ scores or in a format that allowed them to be computed. The majority of the studies presented their results as mean values (33 studies) or means ± SDs (18 studies). More than 20 y ago, centiles were produced from the data of the Infant Health and Development Program for the following: 1) weight, BL, and HC by BW (36, 37) and sex (52); 2) weight-for-length in 3-cm intervals (53); and 3) HC-for-BL by BW and sex (72). At the same time, 5th, 50th, and 95th centiles for weight, BL, and HC by sex were published, but only until 42 wk of corrected age (78). A few studies presented centiles up to 3–4 mo of postnatal age (43, 58, 73).

We attempted to compare the mean and/or centiles of the available charts, but it was not possible because most studies described only the growth pattern of a cohort of preterm infants without presenting a true chart; only 4 (7%) studies that presented a chart had a reliable assessment of gestational age and in those 4 studies the results were presented in a different way, so we could not compare the curves.

The overall quality scores for all of the included studies and for each of the 3 domains are summarized in Figure 2. The median (range, IQR) total quality score for the 61 included studies was 47% (26–75%, 34–56%). Twenty-seven (44%) studies scored ≥50%, of which 10 (16%) scored >60% with only 1 (2%) study scoring >66%. The median (range, IQR) quality score was 44% (19–67%, 33–52%) for the domain study design, 25% (0–88%, 13–38%) for statistical methods, and 33% (0–100%, 0–33%) for reporting methods.

The future: prescriptive postnatal growth standards for preterm infants

In light of the findings of the systematic review, and the well-recognized need for a robust tool to monitor preterm postnatal growth and to facilitate comparisons across interventional studies...
We aimed to develop longitudinal prescriptive standards, as opposed to reference charts, to monitor the growth of infants born preterm (7, 90). Prescriptive standards describe how fetuses and newborns should grow when nutritional, environmental, and health constraints on growth are minimal, as opposed to reference charts, which describe how fetuses and newborns have grown at a particular time and/or place.

We developed the preterm standards in the context of the Fetal Growth Longitudinal Study within The International Fetal and Newborn Growth Consortium for the 21st Century.

<table>
<thead>
<tr>
<th>Domain</th>
<th>High risk of bias, n (%)</th>
<th>Low risk of bias, n (%)</th>
<th>Risk of bias not evaluable, n (%)</th>
<th>Reason for assigning score not evaluable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Study design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1. Aim of the study</td>
<td>30 (49)</td>
<td>31 (51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2. Design</td>
<td>0 (0)</td>
<td>61 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3. Distinction between reference and standard</td>
<td>32 (52)</td>
<td>29 (48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4. Definition of target population</td>
<td>33 (54)</td>
<td>28 (46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5. Sample selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5.1. Preplanned study</td>
<td>18 (30)</td>
<td>43 (70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5.2. Enrollment in accordance with target population</td>
<td>11 (18)</td>
<td>48 (79)</td>
<td>2 (3)</td>
<td>Target population not defined</td>
</tr>
<tr>
<td>1.6. Selection of preterm neonates</td>
<td>24 (39)</td>
<td>37 (61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.7. Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.7.1. Number of neonates at each GA at birth</td>
<td>45 (74)</td>
<td>16 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.7.2. Number of losses to follow-up</td>
<td>24 (39)</td>
<td>37 (61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.7.3. Number of visits attended by each infant</td>
<td>6 (10)</td>
<td>55 (90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8. Number of measurements and time elapsed between 2 measurements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8.1. Measurements taken at birth or within first 24 h</td>
<td>42 (69)</td>
<td>19 (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8.2. Measurement follow-up schedule after birth</td>
<td>25 (41)</td>
<td>36 (59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.9. Follow-up duration</td>
<td>48 (79)</td>
<td>13 (21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.10. Infant nutrition</td>
<td>29 (48)</td>
<td>32 (52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.11. Postnatal care</td>
<td>54 (89)</td>
<td>7 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.12 Postnatal morbidities</td>
<td>39 (64)</td>
<td>22 (36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.13. GA estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.13.1. Based on LMP and confirmed by early ultrasound assessment</td>
<td>48 (79)</td>
<td>13 (21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.13.2. Based on a difference of &lt;2 wk between LMP and ultrasound assessments</td>
<td>21 (34)</td>
<td>11 (18)</td>
<td>29 (48)</td>
<td>Ultrasound or LMP assessment not recorded</td>
</tr>
<tr>
<td>1.13.3. Exclusion of neonates with unreliable GA</td>
<td>6 (10)</td>
<td>4 (6)</td>
<td>51 (84)</td>
<td>Reliability of GA not evaluated</td>
</tr>
<tr>
<td>1.14. Postnatal age evaluation</td>
<td>0 (0)</td>
<td>61 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.15. Anthropometric evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.15.1. Use of standardized instruments</td>
<td>30 (49)</td>
<td>31 (51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.15.2. Calibration of instruments</td>
<td>29 (48)</td>
<td>16 (26)</td>
<td>16 (26)</td>
<td>No information on instruments</td>
</tr>
<tr>
<td>1.15.3. Use of standardized protocols</td>
<td>35 (57)</td>
<td>26 (43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.15.4. Measures taken by at least 2 operators</td>
<td>49 (80)</td>
<td>12 (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.15.5. Trained operators</td>
<td>45 (74)</td>
<td>16 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.15.6. Standardized operators</td>
<td>14 (23)</td>
<td>3 (5)</td>
<td>44 (72)</td>
<td>No information on training</td>
</tr>
<tr>
<td>1.16. Characteristics of the mothers</td>
<td>46 (75)</td>
<td>15 (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Statistical methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1. Assessment of outliers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1.1. Method used to detect the outliers</td>
<td>56 (92)</td>
<td>5 (8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1.2. Correction or exclusion of outliers</td>
<td>4 (6)</td>
<td>1 (2)</td>
<td>56 (92)</td>
<td>Outliers not identified</td>
</tr>
<tr>
<td>2.2. Covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.1. Sex</td>
<td>42 (69)</td>
<td>19 (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.2. GA at birth</td>
<td>30 (49)</td>
<td>31 (51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3. Statistical models</td>
<td>5 (8)</td>
<td>56 (92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4. Assessment of goodness-of-fit of the models</td>
<td>0 (0)</td>
<td>19 (31)</td>
<td>42 (69)</td>
<td>Raw centiles</td>
</tr>
<tr>
<td>2.5. Lack of precision of the estimates</td>
<td>2 (3)</td>
<td>12 (20)</td>
<td>47 (77)</td>
<td></td>
</tr>
<tr>
<td>2.6. Smoothing</td>
<td>2 (3)</td>
<td>16 (26)</td>
<td>43 (71)</td>
<td>No centiles</td>
</tr>
<tr>
<td>3. Reporting methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1. Characteristics of the study population</td>
<td>25 (41)</td>
<td>36 (59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2. Charts presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2.1. Values of (at least) 10th, 50th, and 90th centiles reported or computable</td>
<td>48 (79)</td>
<td>13 (21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2.2. z Scores reported or computable</td>
<td>52 (85)</td>
<td>9 (15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GA, gestational age; LMP, last menstrual period.
a prospective, standardized, multiethnic, population-based project, conducted between 2008 and 2013 in 8 urban areas around the world (80). The project’s overall aim was to study growth, health, nutrition, and neurodevelopment in the same cohort from the first trimester of pregnancy to 2 y of age by using the same conceptual approach as the WHO Multicentre Growth Reference Study (91), so as to produce international standards for fetal growth and the postnatal growth of preterm infants, which would complement the existing WHO Child Growth Standards (92).

In the Fetal Growth Longitudinal Study, we recruited 4607 women who initiated antenatal care before 14 wk with reliable menstrual dates and a confirmatory ultrasound dating scan (93); met the entry criteria of optimal health, nutrition, education, and socioeconomic status (80); and were not exposed during pregnancy to environmental hazards (94). We aimed to study all preterm births ($\geq 26+0$ but $<37+0$ wk) without ultrasound evidence of fetal growth restriction (FGR) in this cohort. They were eligible for inclusion in PPFS. The preterm birth rate of 4.9% was not surprising given the healthy, low-risk status of the mothers.

In brief, we examined this cohort every 2 wk during the first 8 wk and then every 4 wk until 8 postnatal months to obtain the following:
1) anthropometric measurements (weight, BL, and HC),
2) a clinical evaluation,
3) morbidity data, and
4) food intake. The study generated international standards according to postmenstrual age and sex for postnatal weight, length, and HC (10) that exhibit different growth patterns compared with the published INTERGROWTH-21st Newborn Size Standards (95) and overlap with the WHO Child Growth Standards (92) by 64 wk of postmenstrual age. We believe these standards are the most robust tool available to monitor the postnatal growth of preterm infants born after 33 wk, who represent the majority of the preterm population.

The conceptual approach, as well as its implications and limitations, has been discussed in detail (7, 9). This strategy provided a population in PPFS, with prospectively evaluated ultrasound evidence of optimal fetal growth, that was conceptually as close as possible (according to the degree of preterm infant maturation) to the prescriptive approach used to construct the WHO Child Growth Standards (92).

DISCUSSION

To the best of our knowledge, this is the first systematic review to evaluate the methodologic quality of studies that aimed to create postnatal growth charts for preterm infants on the basis of longitudinal data. It follows similar reviews from our group that assess the quality of studies used to create fetal size (17) and neonatal anthropometric charts (18). The reliability and robustness of our results are corroborated by the following: the use of the most rigorous methodology for performing a systematic review of observational studies; a comprehensive literature search without language or publication date restrictions that identified studies published from 1948 to the present day in English, Portuguese, Spanish, and French; the inclusion of a relatively large number of studies; strict assessment of the methodologic quality of included studies through the use of a predetermined set of criteria in 3 domains; and a quantitative and qualitative summary of the evidence.

The results of the systematic review show that the methodologic quality of the included studies was fair to low, with very few that were of high quality: more than half had a quality score $\leq 50\%$, and only one scored $>66\%$. Most charts included in the review had the following methodologic weaknesses:

1) An unreliable estimation of gestational age, which is a major flaw given that accurate determination of gestational age is essential for the postnatal assessment of a preterm neonate.
2) A lack of standardization in performing the anthropometric measurements.
3) Small sample sizes (almost one-third of the studies included $\leq 100$ infants, with 7 including $<50$ infants).
4) Short follow-up, often only until discharge from the neonatal intensive care unit (half of the included studies), which meant that it was not possible to determine when in postnatal life the preterm infants’ growth became
comparable to that of term infants, as assessed by the
WHO Child Growth Standards (91).

5) The use, in almost half the studies, of BW (especially very
or extremely LBW) as a proxy for prematurity, either in
the sample selection or presentation of the results, which
produced charts by BW group for a mixed population of
preterm, term small-for-gestational-age, and FGR infants.
This selection bias led to a lack of data for larger infants
(e.g., late or moderately preterm infants with a BW
>1500 g) in almost half the studies.

6) Shortcomings in the assessment of outliers, creation of
charts according to covariates, reporting of the study pop-
ulation characteristics, and presentation of the charts.

Some caution, however, is needed when interpreting the results
of our review. First, we were unable to assess several meth-
oodologic items comprehensively, such as the following: the
exclusion of neonates with an unreliable gestational age, preg-
nancy dating where the discrepancy between the gestational age
estimates based on the LMP and an early ultrasound scan was
<2 wk, calibration of the measurement instruments, standardi-
zation of operators, correction or exclusion of outliers, as-
essment of goodness-of-fit models, lack of precision of the
estimates, and smoothing. It was not clear whether failure to
meet these criteria was due to imperfections in the execution of
the study or incomplete reporting. Second, although we used
a list of methodologic criteria to judge quality, which was
modified and adapted from previous reviews, the choice was, to
some extent, arbitrary. Third, the reviewers who performed the
data extraction and methodologic assessment were not blinded
to the authors/institutions of included studies, although we doubt
masking would have altered our main conclusions. Finally, the
older included studies were scored by using quality criteria,
which have only been established in recent years.

It is also important to note that several postnatal growth charts
for preterm infants, currently widely used in clinical practice,
were not included in the present review because they were
created with the use of cross-sectional data and/or term-born
infant data (96–100). Their main characteristics are shown in
Figure 3. For example, in 1971, Gairdner and Pearson (100)
published charts of weight, length, and HC from 28 wk until
the first 2 y of life according to sex. The charts were derived
from several sources including longitudinal follow-up of term
infants (88, 101) and both cross-sectional and longitudinal data
on preterm infants born from 28 to 40 wk (25, 102). In 1976,
Babson and Benda (97) published charts from 26 wk to 10 y of
age, which are a combination of cross-sectional preterm infant
measurements (103) and longitudinal data on term infants (104,
105). Almost 30 y later, Fenton (97, 98) published an updated
version of the Babson and Benda charts from 22 wk on the
basis of a meta-analysis of studies that combined cross-sectional
neonatal BW charts (106–108) with post-term CDC charts
(109). In 2013, the Fenton growth charts were revised (99) with
the use of updated cross-sectional studies for the preterm period
(106, 110–114) and harmonized with the WHO Child Growth
Standards (115). Finally, there are the UK-WHO Neonatal and
Infant Close Monitoring Charts (116), which were designed
for plotting the growth of very preterm infants and those with
significant early health problems from 23 wk to 2 y corrected
age. The new charts amalgamated recalculated UK 1990
birth data (117) with the WHO Child Growth Standards after
term (115). However, the resulting charts (23–42 wk, 2 wk to
6 mo corrected age post-term, and 6 mo to 2 y corrected age
post-term), as remarked by the authors, do not describe how
preterm infants grow after birth because they show birth mea-
urements of infants born at different gestational ages until
42 wk and then longitudinal data on healthy, breastfed term
infants.

FIGURE 3 Main characteristics of postnatal growth charts for preterm infants excluded from the systematic review. BL, body length; GA, gestational
age; HC, head circumference; LBW, low birth weight; NICM, neonatal infant close monitoring; W, weight.
Although the American Academy of Pediatrics recommends that the goal for a preterm infant should be to achieve a postnatal growth rate which approximates that of a “normal” fetus of the same gestational age (118), postnatal growth patterns in preterm infants are markedly different from those of term infants. Preterm neonates almost always show a postnatal cumulative nutritional deficit as well as extraterine growth restriction (119, 120). Hence, all of the charts above are valuable tools but inadequate to describe the actual growth of infants born preterm, because most of the charts for the period <40 wk were derived from cross-sectional BW data. In addition, preterm birth cannot be considered a normal event, and many other variables can affect the postnatal growth of preterm infants (121).

To overcome these methodologic and conceptual weaknesses, we designed the PPFs component of the INTERGROWTH-21st Project (80) to ensure the following:

1. Gestational age was accurately assessed by combining LMP with early ultrasound assessment to date each pregnancy (122) and by excluding women if there was >1 wk discrepancy between the gestational age estimates based on the LMP and ultrasound scan (80).

2. Anthropometric measurements were taken by using identical standardized instruments and techniques (123).

3. We had a reasonable sample size by enrolling >200 preterm infants from a cohort of >4000 women followed from the first trimester of pregnancy representing, to the best of our knowledge, the largest cohort of “healthy” preterm infants ever studied for the specific purpose of creating postnatal growth standards.

4. We monitored fetal growth by serial ultrasound to exclude preterm cases with evidence of FGR.

5. The follow-up period was extended to 8 mo to avoid the so-called right-edge effect in constructing the growth standards (125).

6. We avoided using BW as a proxy for prematurity.

7. Well-described analytic, statistical, and reporting methods were used.

The recently published PPFs standards (9) complement the set of international standards produced by the INTERGROWTH-21st Project, which include standards for crown-rump length in the first trimester of pregnancy (93), fetal growth by ultrasound up to term (125), and newborn size (95) that match the WHO standards for term infants conceptually and methodologically (92), enabling growth and development to be monitored from the first trimester of pregnancy to 5 y of age, irrespective of location or ethnicity.

Full acknowledgment of all those who contributed to the development of the INTERGROWTH-21st Project protocol appears at www.intergrowth21.org.uk.

The authors’ responsibilities were as follows—FG: conceptualized and designed the study, acted as first reviewer, interpreted the data, and wrote the initial draft of the manuscript; LCI, EB, JV, and SHK: conceptualized the study and revised the draft critically for important intellectual content; ZAB and AC-A: revised the draft critically for important intellectual content; EOO: carried out the analyses and revised the draft critically for important intellectual content; IR: acted as second reviewer and interpreted the data; all authors: read and approved the final version of the manuscript. None of the authors declared a conflict of interest.

REFERENCES


84. Patri AM, Sepúlveda BH, Valenzuela CP, Cortés OA. Anthropometry of Chilean children from 0 to 6 years of age. Santiago (Chile): Editorial Andres Bello; 1984 (in Spanish).


