Seven caveats on the use of low birthweight and related indicators in health research

Marcelo L Urquia,1 Joel G Ray2

ABSTRACT
Birthweight and gestational age are the two most commonly used continuous variables in perinatal research. Dichotomous outcomes derived from these two variables include low birthweight, preterm birth and small for gestational age, each extensively used as perinatal and population health indicators within public health research and health surveillance systems. However, these dichotomous indicators have inherent limitations that need to be considered in the design, analysis and interpretation of epidemiological studies. In this report, we present seven caveats that may help researchers and users of epidemiological data avoid common (and not so common) pitfalls in the consideration of these indicators.

INTRODUCTION
Birthweight has been a traditional measure in perinatal research—both as an exposure and as an outcome variable. As pointed out by Wilcox, the popularity of birthweight stems from its measurement properties and its association with infant mortality. In most healthcare settings birthweight can be measured as a continuous variable, with high uniformity, accuracy and completeness, and is generally available within birth certificate and health-administrative databases. This allows for the monitoring of trends across time, geographical regions or different population groups.

For decades, low birthweight (LBW) was considered a useful indicator of the susceptibility of the ‘small baby’ to higher perinatal morbidity and mortality, and overall newborn health. However, there has also been a growing understanding of the limitations of using LBW as a marker of ill health. In the following, we present seven caveats about LBW and related indicators of perinatal health. Our intention is to help public health researchers, social epidemiologists and other users of perinatal data avoid some common pitfalls in the application and interpretation of the small baby and related indicators.

Caveat 1: LBW is not the optimal measure of a ‘small baby’
LBW—defined by a weight under 2500 g—has been strongly associated with infant mortality and other health outcomes later in life, yet, in the past decade, its relevance has been drawn into question. One concern is that the association between LBW and infant mortality may not be causal. A second concern is the realisation that impaired fetal growth and preterm birth (PTB) are different processes that may each lead to LBW. For this reason, in some settings such as Canada, LBW is no longer considered as informative, where it has been excluded from recent perinatal surveillance reports.

In recent years, the availability within health information system databases of an accurately dated gestational age at birth, vis-à-vis widespread access to early prenatal ultrasonographic dating, has made gestational age at birth a better predictor of adverse outcomes. Although birthweight and gestational age are highly correlated, each is independently associated with perinatal mortality. As term babies weighting less than 2500 g also have a higher mortality risk, their smallness was thought to originate from intrauterine growth restriction (IUGR). The notion of IUGR was introduced in the 1960s and was generally accepted by the 1970s, because it provided a measure of smallness, that is, relatively independent of gestational age. IUGR was operationalised as ‘small for gestational age’ (SGA), generally defined as the lightest 10% of births at each gestational age. By fixing the proportion of SGA babies at each gestational age, the influence of prematurity was removed, and SGA was exclusively considered to reflect impaired fetal growth.

As LBW may either reflect slow growth, short gestation or a combination of both, its interpretation should be complemented by PTB and SGA, whenever available.

Caveat 2: SGA is a proxy for IUGR, but cannot be validated against IUGR
Although SGA, when interpreted with measures of gestational age, was more informative than LBW alone, this ‘solution’ also presented its own problems. SGA is often used as a proxy for IUGR, and some believe that it is synonymous with IUGR. This practice somehow assumes that IUGR is a real and distinct entity manifested by SGA. The problem is that there is no ‘gold standard’ either to determine IUGR accurately, beyond size measures, or to validate SGA against that gold standard. Accordingly, SGA measures have been largely ‘validated’ by their association with greater adverse perinatal outcomes (i.e., using an outcome-based approach).

The term IUGR expresses the notion that a fetus, at a given point on its intrauterine growth trajectory, does not achieve its optimal expected size. This may be due to fetal or maternal diseases, or due to environmental influences, which may act in isolation or in combination. IUGR is a particular case of the more general notion of abnormal fetal growth (AFG), conceptually expressed in the following formula:

\[ AFG_{i \tau} = OFW_{i \tau} - EFW_{i \tau} \]

where \( OFW_{i \tau} \) is the observed weight at time \( \tau \), of the \( i \)th week gestation (when \( \tau \) date of birth, then
OFW (birthweight) and EFW, n is the expected (optimal) weight at time t, of the nth week gestation.

If AFG is approximately 0, then fetal growth is considered to be normal.

If AFG is less than 0, then the baby is lighter than expected.

If AFG is greater than 0, then the baby is heavier than expected.

This conceptual approach to AFG, although reasonable on paper, presents some practical problems. First, EFW, n is not truly known, and can only be indirectly estimated, with both systematic and random errors. Second, it does not tell how much AFG should depart from 0 to be considered truly abnormal, rather than arising due to random variation.

The current ‘solution’ common to all SGA definitions consists of using the mean birthweight (estimated in utero or observed at birth) as the EFW, and then choosing a cut point beyond which a deviation of the OFW from the mean birthweight is considered abnormal (ie, predisposing to pathological outcomes in the fetus/newborn).

Like its proxies—LBW and SGA—IUGR is not a distinct disease, but an intermediate state arising from multiple influences that can be difficult to tease apart. As such, the observed associations between different measures of IUGR and newborn morbidity and mortality may not be reflective of a causal effect of smallness or impaired growth. Rather, the association between LBW or SGA and perinatal mortality may be explained by shared upstream pathological processes that simultaneously lead to small size and death.

Similar observations apply to measures of the large baby, such as ‘large for gestational age’.

**Caveat 3: not all small babies result from a disease process, and not all babies affected by IUGR are small**

Differentiating physiological (constitutional) small babies from those who are pathologically small remains a challenge.

According to table 1, some fetuses may have IUGR due to different factors (eg, maternal placental vascular disease, chronic hypertension, maternal smoking or unknown causes), and they may, furthermore, be classified as SGA if their weight is below the conventional 10th percentile (situation 1). However, if the influence of placental disease or smoking impairs fetal growth, but not to the extent to which the actual fetal or newborn weight is below the 10th percentile, then SGA fails to reflect the state of IUGR properly (situation 2). Conversely, not all fetuses who are SGA are pathologically growth restricted; rather, some may be constitutionally small (situation 3). A sample of normal babies from healthy pregnancies shows a normal birthweight distribution and, inevitably, some babies fall within the left (lighter) tail of that curve. An example of ‘normal smallness’ is babies born to South Asian or East Asian parents, who are, on average, smaller than babies of white parents, without necessarily being at higher risk of adverse outcomes. Other constitutional differences in weight are given by the sex of the fetus (girls are smaller than boys), multiplicity (eg, twins are smaller than singletons), maternal parity (higher order babies are heavier than the first baby) and geographical altitude, rather than by maternal or fetal disease.16 Finally, it is presumed that most babies who neither experience IUGR nor fall below the 10th percentile are correctly classified as normal (situation 4). Researchers, however, have not been able to disentangle fully the two sources of smallness (ie, pathological vs physiological), so that misclassification is, to some extent, unavoidable. Distinguishing between pathological versus physiological smallness may have differential implications for the clinical management of these two types of conditions.

The actual magnitude of the potential misclassification (the prevalence of situations 2 and 3) is unknown because there is currently no concrete method for measuring IUGR. Therefore, for the remainder of the discussion in Caveat 3 we discuss how SGA measures have been developed and how they work.

Birthweight is a literal measurement of newborn weight, in grams, using a standardised weight scale. While a fetus remains in utero, estimation of its weight relies on a proxy measure, such as ultrasound-based fetal biometry, marked by quite pronounced intra and interobserver variability.17 There are several different formulae that can be used to estimate fetal weight by ultrasound, usually based on its abdominal circumference plus femoral length and/or head circumference.18–20

SGA measures vary by the method used to generate the birthweight curves on which they lie. Population-based ‘references’ use large samples of cross-sectional birth data, and include both healthy and unhealthy delivered newborns. Therein, by classifying as SGA the 10% of smallest babies in each gestational age stratum, the proportion of PTB babies among SGA babies is the same as that in the population. If the proportion of PTB in a typical population is approximately 6%, then this implies that most (94%) of the SGA signal comes from term infants. The adequacy of using population-derived SGA as a measure of fetal growth restriction has some limitations. First, SGA is not a measure of fetal growth, per se, but size at birth (ie, it is ex utero). Second, if babies born preterm are more likely to be growth restricted than those who remain in utero at the same gestational age,21–23 then fixing the proportion of the smallest 10% of births at each gestational age may underestimate IUGR among PTB infants. As more than 50% of LBW babies are born preterm, this underestimation may be substantial.

A second way that SGA measures are created is based on fetal growth ‘standards’ derived from ultrasound measurements taken on healthy unborn fetuses.21 24–26 By restricting to healthy fetuses, one assumes that departures from the normal values that they generate will, therefore, reflect pathological influences. This idea was further developed by Gardosi and colleagues,25 who proposed customised standards, tailored to individual profiles, by combining information on maternal and fetal physiological characteristics and the existing Hadlock equation for estimating fetal weight using ultrasonography. The use of Gardosi’s customised standard approach shows a stronger association between SGA and perinatal mortality than with population-based references.27–29 However, adding maternal characteristics such as age, parity and weight only marginally improves the identification of SGA infants,30 31 and the apparently better prediction of perinatal mortality has been argued to be an artefact of including a larger proportion of preterm infants.27 31

Compared to population-based birthweight references, fetal growth standards show larger fetuses at early gestational ages, such as 28 or 32 weeks, because a fetus that remains in utero is probably healthier than that born prematurely. Only at term do

<table>
<thead>
<tr>
<th>IUGR occurs</th>
<th>IUGR does not occur</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGA</td>
<td>Situation 1</td>
</tr>
<tr>
<td>Not SGA</td>
<td>Situation 2</td>
</tr>
</tbody>
</table>

IUGR, intrauterine growth restriction; SGA, small for gestational age.
the weights of population-based references and fetal growth standards achieve near agreement. When applied to actual populations, fetal growth standards tend to classify twice as many preterm babies as SGA (<10th percentile) than the population-based birthweight references restricted to newborns. Therefore, in the absence of additional information on direct pathological markers of IUGR and fetal/newborn wellness, it is unclear to what extent the better prediction of perinatal mortality based on customised standards reflects IUGR, or simply, a higher proportion of preterm babies.

**Caveat 4: population-based references and fetal growth standards share common limitations**

Despite the important aforementioned differences between population-based birthweight references and fetal growth standards, the two share some common limitations as a result of the statistical conceptualisation of normality, which classifies as ‘abnormal’ those babies whose weights fall at the tails of the curve. The various cut-off points used to define SGA—less than the 3rd, 5th or 10th percentiles or under two standard deviations—reflect different attempts to draw the line of abnormality in the continuum of the birthweight distribution. As mentioned above, one common source of misclassification is that some small babies are small simply because of biological variability, rather than from a disease process. More importantly, a second aspect of misclassification is the fetus who experiences growth restriction, but whose weight remains above the lower tail cut point of, say, the 10th percentile (situation 2 in table 1). This may be a common problem, because such fetuses are grouped with the other 90% of births that are at or above the 10th centile. For example, maternal smoking is typically associated with a modest birthweight reduction of 150–250 g. This means that a term newborn whose expected birthweight is 3500 g, but due to maternal smoking has a weight of 3500 g, may still not be below the 10th percentile birthweight; as such, its state of IUGR is not realised by the SGA criterion. IUGR due to smoking would only be detected by standards based on healthy pregnancies if the expected birthweight of the same mother in the absence of smoking would be approximately 150–250 g higher than the 10th percentile, which at 40 weeks of gestation would be around the 25th percentile. In other words, the ability of SGA measures to detect IUGR decreases as the expected birthweight increases.

**Caveat 5: reducing the prevalence of SGA may be an intangible public health goal**

Those who use SGA as a risk factor, as an outcome or as a covariate, need to consider that, by definition, the proportion of SGA in the population is a function of the weight cut points. However, over time, in many countries the birthweight distribution of babies has shifted rightward. For example, Kramer et al developed birthweight percentile curves based on Canadian singleton births from 1994 to 1996, providing cut-off values to classify SGA births at the 10th percentile. However, one decade later, birthweight increased, and what was once the 10th centile weight value became the 8th centile weight value. Accordingly, does this mean that the proportion of ‘SGA’ births fell from 10% to 8%? and that IUGR, as a pathological process, also declined? Or, did babies get physiologically bigger, thereby shifting the whole weight distribution to the right? Will these shifts continue in the next couple of decades? These issues raise concerns about the meaning of monitoring time trends in SGA or about the need to update the reference cut points periodically. As measures of SGA are ill-suited to distinguish between constitutionally and pathologically small babies, interpreting their trends is problematical. Using unambiguous indicators of newborn health or wellbeing is probably the preferred way to monitor population trends.

**Caveat 6: PTB is a better indicator than LBW, but it too has some limitations**

At first glance, PTB seems to be a better indicator of newborn health and disparity than LBW, because its meaning is restricted to the length of gestation that determines fetal size. However, early delivery may be triggered by some of the unmeasured pathological factors that also lead to LBW and neonatal morbidity and mortality, thus raising the possibility of confounding. Confounding due to unmeasured factors may be substantial, because at least one in four cases of PTB are of unknown aetiology. Moreover, secular increases in PTB have occurred in Canada, Scotland and the USA. Part of this change may be attributed to more multiple births, but PTB rates have even risen among singletons. Demographic explanatory variables for the latter include older maternal age and delayed childbearing, while obstetric interventions, such as caesarean delivery and induction of labour, may also have affected these trends. Given that PTB is strongly associated with infant mortality, it is somewhat paradoxical that PTB rates have risen while infant mortality has declined. This may be partly explained by improvements in neonatal care in high-income countries and the tendency to deliver fetuses at a late preterm period, in lieu of early maternal or fetal compromise, thereby reducing the risk of serious neonatal morbidity and death. Because of these trends, PTB is now a more ambiguous measure of infant risk, and should be used with greater care, as might be the case in the study of social inequalities and perinatal health.

**Caveat 7: uncrirical use of birthweight or PTB measures may lead to misleading conclusions in the assessment of health inequalities between population groups**

The following two examples illustrate Caveat 7:

**Example 1: ethnic inequalities**

The label ‘small baby’ can be difficult to interpret when analysing health disparities by race, ethnicity and/or immigration status. In particular, PTB rates do not vary appreciably by maternal country of birth, but LBW and SGA does. In Canada, South Asians show a two to three times higher odds of being LBW/SGA than immigrants from industrialised countries, but almost the same proportion of PTB. Simply concluding that South Asians have a higher risk of IUGR may be misleading, because differences in birthweight may simply be a reflection of the constitutionally smaller body size among South Asians compared to the mostly white population. Also, the application of a Canadian reference birthweight chart that defines SGA at less than the 10th percentile results in 21% of South Asian newborns being categorised as SGA; whereas, using a South Asian-specific reference chart produces the expected SGA proportion of 10%. Are we to believe that this 11% net difference is a reflection of a constitutional (physiological) or a pathological process? While the evidence to answer this is inconclusive, one study showed that infants born to Chinese and South Asian mothers living in British Columbia, Canada, actually experience lower mortality across gestational ages than the Canadian-born population, and that this is so despite having higher rates of SGA by the Canadian-born standard. The latter suggests that the use of a single reference may artefactually overestimate SGA rates among some ethnic groups. Such
misclassification may also lead to unnecessary stigmatisation and parental stress, expensive over-investigation of the individual infant, or the implementation of unnecessary public health initiatives.15

Example 2: socioeconomic disparities
Comparison of LBW rates between groups may not accurately reflect disparities if their defining components (fetal growth restriction and length of gestation) behave differently between comparison groups. If socioeconomic status (SES) gradients in SGA and PTB are in the same direction, then the gradient for LBW will also follow the same direction. However, if the SES gradients in SGA and PTB are in opposite directions, those for LBW may be attenuated or cancelled out, thereby underestimating the degree of inequalities in the population. Attenuated and even reversed SES gradients in LBW and PTB have been observed in Chile and Brazil.49 50 Although there is no direct evidence to explain this phenomenon, one suspected reason may be the differential access to elective caesarean delivery, which is more common among women of higher SES.51 52 As caesarean section is often associated with PTB, it delivery, which is more common among women of higher SES.

These examples show that uncritical interpretation of LBW, SGA and PTB may lead to spurious conclusions that can ultimately misinform health policy.

CONCLUSIONS
Currently used measures of the small fetus and newborn are not definite pathological entities unto themselves, but are indicators in evolution. To date, these measures have relied heavily on birthweight criteria. The challenge is to move beyond birthweight and develop measures that incorporate information about the pathological and physiological determinants of birthweight and gestational age within solid theoretical frameworks. Until then, health researchers and those who generate and interpret perinatal data need to familiarise themselves with the assumptions and limitations inherent in existing measures. This is particularly important within low and middle-income countries, where birthweight may be the only informative indicator that is widely collected and available within administrative data sources.

Contributors MU led the writing of the manuscript. JGR contributed to, revised and approved all versions.

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES


