When is birthweight at term abnormally low? A systematic review and meta-analysis of the association and predictive ability of current birthweight standards for neonatal outcomes

GL Malin, a RK Morris, b,c R Riley, d MJ Teune, e KS Khan f

a School of Medicine, the University of Nottingham, Nottingham, UK b Department of Obstetrics, The School of Clinical and Experimental Medicine, College of Medical & Dental Sciences, University of Birmingham, Birmingham, UK c Department of Obstetrics and Gynaecology, Birmingham Women’s Foundation NHS Trust, Birmingham, UK d School of Health and Population Sciences, University of Birmingham, Birmingham, UK e Department of Obstetrics & Gynaecology, Academic Medical Centre, Amsterdam, The Netherlands f Women’s Health Research Unit, The Blizzard Institute, Barts and The London School of Medicine, Queen Mary, University of London, London, UK

Correspondence: Dr G Malin, School of Medicine, D Floor, Queen’s Medical Centre, Derby Road, Nottingham NG7 2UH, UK.
Email gemma.malin@nottingham.ac.uk

Accepted 20 September 2013. Published Online 8 January 2014.

Background Intrauterine growth restriction is a cause of neonatal morbidity and mortality. A variety of definitions of low birthweight are used in clinical practice, with a lack of consensus regarding which definitions best predict adverse outcomes.

Objectives To evaluate the relationship between birthweight standards and neonatal outcome in term-born infants (at ≥ 37 weeks of gestation).

Search strategy MEDLINE (1966–January 2011), EMBASE (1980–January 2011), and the Cochrane Library (2011:1) and MEDION were included in our search.

Selection criteria Studies comprising live term-born infants (gestation ≥ 37 completed weeks), with weight or other anthropometric measurements recorded at birth along with neonatal outcomes.

Data collection and analysis Data were extracted to populate 2 x 2 tables relating birthweight standard with outcome, and meta-analysis was performed where possible.

Main results Twenty-nine studies including 21,034,114 neonates were selected. Absolute birthweight was strongly associated with mortality, with birthweight <1.5 kg giving the largest association (OR 48.6, 95% CI 28.62–82.53). When using centile charts, regardless of threshold, the summary odds ratios were significant but closer to 1 than when using absolute birthweight. For all tests, summary predictive ability comprised high specificity and positive likelihood ratio for neonatal death, but low sensitivity and a negative likelihood ratio close to 1.

Author’s conclusions Absolute birthweight is a prognostic factor for neonatal mortality. The indirect evidence suggests that centile charts or other definitions of low birthweight are not as strongly associated with mortality as the absolute birthweight. Further research is required to improve predictive accuracy.

Keywords Fetal growth restriction, low birthweight, neonatal morbidity, neonatal mortality, systematic review.

Introduction Intrauterine growth restriction remains a significant problem in current obstetric and neonatal practice, and is a significant cause of perinatal mortality and morbidity. 1,2 Statistically ‘normal’ birthweight is defined as being within a range around the central tendency (e.g. centile ranges). This simple approach has many deficiencies. Clinically, infants who are of low birthweight may belong to one of four groups. There are those that suffer intrauterine growth restriction, whereby the fetus does not achieve their growth potential because of environmental factors, such as placental insufficiency or maternal health status. 3 Others may have a structural or chromosomal abnormality that affects their growth. 3 Another group of infants who have low birthweight are those that are constitutionally small. These

DOI: 10.1111/1471-0528.12517
babies reach their growth potential; they are not subject to a pathological process. Low birthweight also refers to babies who are normally grown but are born prematurely. Prematurity is independently associated with increased mortality and long-term morbidity. A number of methods have been used to attempt to identify infants who are most at risk of adverse outcomes, including neonatal morbidity and mortality. These include: population-based centile charts, with the most commonly used threshold being the tenth centile; customised charts, where the mother’s body mass index (BMI) and ethnicity are used to calculate individualised growth centiles; and ponderal index, which takes into account neonatal weight and length. The published associations between each standard for defining growth restriction and adverse outcome vary, and there is no current consensus regarding the best method. Within current UK practice a variety of different population and customised centile charts are used antenatally, with a different growth chart used for the postnatal period, and with the absolute birthweight (<2.5 kg) often being used to determine the need for increased care or observation in the neonatal period.

The aim of this systematic review was to re-examine the association between measures of low birthweight, including absolute birthweight and other anthropometric measurements, such as ponderal index, with adverse neonatal outcomes. We attempted to avoid the confounding influence of prematurity and to determine which definition of growth restriction has the strongest prognostic association with, and is the best predictor for, subsequent morbidity and mortality.

In this article, the term ‘prognostic’ refers to the strength of association between a birthweight test and the odds of an adverse outcome, as measured by an odds ratio. The term ‘predictive’ refers to the ability of a test to discriminate between babies who will and babies who will not experience an adverse outcome, as measured by sensitivity, specificity, and positive and negative likelihood ratios. A test may have strong prognostic ability, but not necessarily good predictive ability, and so it is important to consider both.

Methods

A protocol-driven systematic review was performed using widely recommended methods for reviews (Appendix S1), and is reported according to the MOOSE (meta-analysis of observational studies in epidemiology) guidance. This study was performed as part of a larger systematic review to determine the association of birthweight standards with outcomes throughout life, and therefore the search strategies and study selection process refer to the studies included for the overall project. The articles relating to outcomes in the neonatal period are reported in this article, whereas those relating to childhood and adult outcomes are reported separately.

Sources

We searched MEDLINE (1966–January 2011), EMBASE (1980–January 2011), the Cochrane Library (2011:1), and MEDION for relevant published articles. In order to identify ‘grey’ literature, OpenGrey and Web of Science were also searched for relevant citations. In MEDLINE the search consisted of a combination of medical subject headings (MeSHs; e.g. infant, small for gestational age, fetal growth retardation), keywords (e.g. intrauterine growth retardation, low birthweight), and word variants using the Boolean operator ‘OR’ for capturing citations of the relevant text. These were combined using ‘AND’ with a combination of MeSHs (e.g. human development, infant mortality, diabetes mellitus), keywords (e.g. developmental delay, handicap, cardiovascular disease), and word variants to capture relevant outcomes. The search was restricted to human studies, but no language restrictions were applied. The MEDLINE search strategy (Appendix S2) was adapted for use in other electronic databases. Hand searching of recent major journals was also performed. The search was performed by two investigators: R.K.M. and G.M. A comprehensive database collating all citations was constructed using REFERENCE MANAGER 12.0.

Study selection and data extraction

Initially, the database was scrutinised by two reviewers (R.K.M. or G.M., partly in duplicate), and full articles of all citations that were likely to meet the predefined selection criteria were obtained. Articles in languages other than English were translated. Final inclusion or exclusion decisions were made after examination by two reviewers (G.M. and R.K.M.) in accordance with the most recent guidance, and with strict adherence to the following criteria:

- **Population:** Live-born infants who have had weight or other anthropometric measurements recorded at birth and were born at term (gestation ≥ 37 completed weeks).
- **Index test:** Any measure of weight or growth at birth, including: absolute birthweight (thresholds <2.5 kg, <2.0 kg, <1.5 kg); population or customised centile charts (thresholds <10th centile, <5th centile, <3rd centile); ponderal index or other growth ratios.
- **Outcome:** Any measure of compromise of neonatal, childhood, or adult wellbeing, such as: mortality; neonatal morbidity, including hypoxic ischemic encephalopathy; childhood or adult motor disability; childhood or adult disease, including diabetes mellitus, cardiovascular disease, and hypertension.
- **Study design:** Observational studies that allowed the generation of a 2 × 2 table (true positives, false positives, false positives, false negatives).
negatives, and true negatives) to compute an estimate of the association between test result and outcomes. Studies with five or fewer individuals were excluded on account of unreliability.

All articles were carefully examined to identify duplications in population. Where this was identified, the most recent and complete versions of the work were selected. There was no language restriction in study selection. The reference lists of selected studies and review articles were scrutinised and additional relevant articles were obtained. Information was extracted from the selected articles in duplicate (G.L.M. and R.K.M.) using a data collection sheet. Data were extracted on study characteristics (including the threshold values used), quality, and results, and were entered onto an Excel spreadsheet. Data were used to construct 2 × 2 tables of the association between the measure of growth at birth using the threshold reported in the article and the postnatal outcome for each individual. If results for multiple thresholds were reported, we sought to construct a separate 2 × 2 table for each threshold. In studies where data were felt to be relevant but 2 × 2 tables could not be constructed, or the outcome or population reported in the article did not meet the specific inclusion criteria, the authors were contacted. The study was not included unless the specific data could be provided. Difficulties in data extraction were resolved by seeking input from a third reviewer (K.S.K.). From the overall data set, the subset of studies reporting neonatal adverse outcomes was selected for inclusion in this report.

Study quality assessment
All articles meeting the selection criteria were assessed for methodological quality, defined as confidence that the study design, conduct, and analysis minimised any bias in the estimation of an association. We assessed quality using the complete STARD and QUADAS checklists. These are validated for the reporting and methodological quality of diagnostic test accuracy studies, and we selected the quality elements that were felt to be most relevant for this review on prognostic tests and associations.\(^ {14,15}\) We did not assign a quality score, as this has been shown to give flawed results.\(^ {16}\) We considered cohort study design to be superior to case-control design. A study was rated high quality if it had at least four of the following items: an adequate description of the population; an adequate description of the test (definition of low birthweight) and the outcome measure; consecutive recruitment; prospective recruitment; >90% completions of follow-up; appropriate outcome measurement; blinding of the investigators performing the outcome measure, and a statement regarding the use of intervention between the index test and outcome. A study was deemed to be of medium quality when three criteria were met and low quality if two or less were adhered to.

Data synthesis for prognostic association
The 2 × 2 tables were used to compute odds ratios (ORs) and 95% confidence intervals (95% CIs) for each index test–outcome pair, and the results were pooled for each index test (considering each definition and threshold of growth as a separate test) using meta-analysis. The OR was selected as the summary statistic, as it represents the effect of the exposure on the odds in an unbiased fashion and enables the results of both case–control and cohort studies to be included.\(^ {17}\) It is frequently used to demonstrate an epidemiologic association,\(^ {17}\) and here it provides a measure of a test’s prognostic ability.

With clinical and statistical heterogeneity expected between studies, a random-effects model was used throughout to account for this, which synthesises the logarithmic odds ratio estimates for each test and weights each study by the inverse of the variance within the study plus between-study variance. This method provides a summary estimate of the average prognostic effect of a test. As the prognostic ability of a test may vary from this average from setting to setting, after each random-effects meta-analysis, if \( I^2 > 0\% \) we also estimated a prediction interval (EPI). This reveals the potential prognostic association if the test is applied in a single setting similar to one of the studies from our analysis.\(^ {18}\) EPI was calculated where three or more studies were included in the meta-analysis.

We plotted summary OR data in forest plots and assessed the between-study heterogeneity in the prognostic association for each test by estimating \( I^2 \) (the level of variability in prognostic effects arising from between-study heterogeneity)\(^ {19}\) and \( \tau^2 \) (the among-study variance of the true prognostic effect).\(^ {20}\) Where the number of studies reporting a given birthweight standard and outcome allowed, we performed subgroup analysis to examine the effect of potential confounding factors. Singleton or multiple birth status, ethnicity, exclusion of congenital anomalies, birth of the study population during or after 1990 (because of recent advances in antenatal and neonatal care), and study quality were considered to be important factors that may influence the strength of the association between low birthweight and adverse outcome.

In each study, when a table contained cells with a value of 0, 0.5 was added to all cells to allow the calculation of log ORs and their variances for meta-analysis.\(^ {21}\) Meta-analyses were performed where two or more studies reported the same index test and outcome measure. The primary outcomes were considered to be neonatal mortality and a composite measure of neonatal neurological morbidity and non-neurological morbidity. A composite outcome measure for morbidity was employed to maximise the number of events that could be included in the analysis and avoid the need to select a single morbidity as a pri-
mary outcome measure; however, a hazard of composite outcome measures is the assumption that the significance of the result applies to all components. To address this issue, we analysed the component outcomes as subgroups. When the composite outcome measure was used, care was taken to ensure that each individual was only counted once in each analysis, particularly where studies reported multiple outcomes for a single population. Where multiple outcomes were reported, attempts were made to select the outcome that was most consistent with the other studies: for example, in the neonatal non-neurological morbidity analysis, hypoglycaemia was the most commonly reported outcome and therefore this was selected primarily, followed by other conditions. To explore for the presence of funnel plot asymmetry (small study effects), and thus potential publication bias, the Peters test was performed in each meta-analysis containing at least ten studies.

For the purposes of our meta-analyses, we used data where birthweight had been dichotomised around a threshold specified in the primary studies. In order to compare the effect of birthweight when it was analysed as a continuous variable, we examined all of the included studies where logistic regression analysis had been performed with birthweight included as a continuous variable, and qualitatively summarised the findings.

**Data synthesis for predictive ability**

Where there was a strong and statistically significant prognostic association between a test and an outcome measure (defined by an OR > 5, with a 95% CI > 1), we went on to calculate sensitivity, specificity, and likelihood ratios, again using data from the $2 \times 2$ tables and synthesising predictive measures using a bivariate random-effects meta-analysis model. This allowed us to examine the predictive ability of the test: that is, whether the test can accurately discriminate between those who do and those who do not have a poor outcome (as measured by sensitivity and specificity), and how much a positive or negative test result modifies the odds of a poor outcome (as measured by the positive and negative likelihood ratios).

All analyses were performed in **Stata 10.0** (StataCorp, College Station, TX, USA) using the metan, metandi, and metabias commands. Plots were generated using **Stata**-**Direct**.

**Results**

As shown in Figure 1, after an initial search of 36,956 citations, we included 92 primary articles in the overall systematic review, of which 29 contained data relating birthweight standards to neonatal outcomes. Five of these were included after contact with authors who provided data or information. In total, data were available for 21,034,114 neonates. Details of the studies included are given in Table S1; a list of excluded studies is available from the authors upon request. A total of 145 further articles were felt to contain potentially relevant data, but the authors could not be contacted, could not supply data to create $2 \times 2$ tables, or upon clarification regarding the population the study was excluded. If a study included infants of $<37$ weeks of gestation, it was only included if separate data regarding term infants was given or the authors provided this. A number of studies contained duplicate populations: where there was duplication of the test and outcome measure the least complete study was excluded from the review. If the population was the same but the measure of growth restriction or adverse outcome differed, then both studies were included, but care was taken not to include multiple studies reporting from the same population within a single meta-analysis, or within the overall count of the number of individuals included in the review.

The majority of studies used population growth chart below the tenth percentile ($n = 17$) or birthweight under 2.5 kg ($n = 9$) as the index test that defined fetal growth restriction. A wide variety of neonatal outcome measures, including mortality and morbidity (e.g. seizures, hypothermia, hypoglycaemia, respiratory distress), were reported. For comparison, we grouped outcomes according to mortality, neurological morbidity, and non-neurological morbidity.

**Prognostic association with neonatal mortality**

A forest plot of the summary meta-analysis odds ratios and 95% confidence intervals for each measure of fetal growth restriction in relation to neonatal mortality is given in Figure 2. A birthweight below 1.5 kg showed the strongest association with neonatal mortality (OR 48.6, 95% CI 28.62–82.53), with no between-study heterogeneity in this effect. Raising the birthweight threshold to 2.0, 2.5, or 2.9 kg gradually reduced the association and increased the heterogeneity, but the summary effect estimate remained highly significant at each threshold. Population centile charts were also strongly associated with neonatal mortality, but generally showed a weaker association at all thresholds than absolute birthweight, because the summary ORs were closer to 1 (Figure 2).

**Prognostic association with neonatal morbidity**

The association between measures of fetal growth restriction and neonatal morbidity are given in Figure 3. The analysis was subdivided into reported neurological morbidity (including seizures, hypoxic ischaemic encephalopathy, intraventricular haemorrhage) and non-neurological morbidity (including hypoglycaemia, respiratory distress syndrome, cardiac failure), according to the definitions given
in the primary studies. A birthweight below 2.0 kg was most strongly associated with neurological morbidity (OR 17.34, 95% CI 5.63–53.70); however, this was based on a single study of 770 neonates. There was a significant association between weight below the third, fifth, and tenth centiles and neurological morbidity. A birthweight below the tenth centile according to a customised growth chart and a ponderal index of ≤ 2.25 did not show a significant association with this outcome. For non-neurological morbidity, birthweights below the third, fifth, or tenth centiles on population chart and birthweights more than 2SD below the population mean showed significant association with this outcome, with summary odds ratios of a similar magnitude. Subgroup analysis for individual morbidities was only possible for birthweight below the tenth centile on the population chart and neonatal hypoglycaemia (any threshold, three studies, OR 3.72, 95% CI 0.85–16.19),9,37,41 and seizures (two studies, OR 2.35, 95% CI 1.58–3.49).41,47

**Quality assessment**

The results for the quality assessment are presented in Table 1. The majority of the studies included were of cohort design (97%), and most were retrospective studies (73%). Most studies were of high or moderate quality according to our pre-specified criteria. Studies often failed to adequately describe the test or outcome in a way that would make them reproducible, and very few studies

**Figure 1.** Study selection process for systematic review of the prognostic and predictive ability of current birthweight standards for short- and long-term outcomes.
described any interventions that were performed between the time of the birthweight measurement and the outcome test. Where possible a subgroup analysis using only high-quality studies was performed, and the results are presented in Table 2.

Subgroup analyses of prognostic association
The results for subgroup analyses to address potential confounding factors of the association between birthweight and adverse neonatal outcome, within the meta-analysis groups for each birthweight standard, are presented in Table 2. No subgroup analyses were possible for neonatal morbidity according to these criteria. Too few studies reported ethnicity in enough detail to permit subgroup analysis. Limiting to a singleton population slightly weakened the association between birthweight below 1.5 kg and neonatal death, but did not affect the association between birthweight below 2.5 kg and the same outcome.

Birthweight as a continuous variable
None of the included studies that considered neonatal outcomes examined birthweight as a continuous variable via logistic regression analysis, so it is not possible to comment on this further.

Direct comparison of prognostic association for absolute versus population centiles
Only one study directly compared absolute birthweight and centile on population chart in the same population. For neonatal mortality, a birthweight below 2.9 kg had an odds ratio of 2.64 (95% CI 1.45–4.82) and a birthweight below the tenth centile on the population chart had an odds ratio of 5.31 (95% CI 2.85–9.89) for the same outcome.

Publication bias for prognostic association results
To examine funnel-plot asymmetry (small study effects), and thus the potential for publication bias, the Peters test was applied to the only meta-analysis containing ten or more studies (birthweight below tenth centile and neonatal mortality). There was no significant evidence of small study effects in this group ($P$ = 0.996).

Predictive ability of standards of low birth weight to predict neonatal death
The outcome that had the strongest prognostic association overall with low birth weight was neonatal death. For birthweight tests with a large (OR $> 5$) and statistically significant prognostic association with this outcome, their predictive ability for individual babies was summarised by

---

**Table 2. Definition of definitions, No. of Studies, No. of Individuals, Year of birth, Odds ratio (95% CI)**

<table>
<thead>
<tr>
<th>Definition</th>
<th>No. of Studies</th>
<th>No. of Individuals</th>
<th>Year of birth</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth weight &lt; 1.5 kg</td>
<td>4</td>
<td>1,441,363</td>
<td>1961-2005</td>
<td>48.6 (28.62, 82.53)</td>
</tr>
<tr>
<td>Birth weight &lt; 2.0 kg</td>
<td>3</td>
<td>539,124</td>
<td>1958-1998</td>
<td>19.12 (4.54, 80.59)</td>
</tr>
<tr>
<td>Birth weight &lt; 2.5 kg</td>
<td>8</td>
<td>1,458,983</td>
<td>1957-2005</td>
<td>8.46 (6.25, 11.48)</td>
</tr>
<tr>
<td>Birth weight &lt; 2.9 kg</td>
<td>1</td>
<td>9201</td>
<td>1981-1983</td>
<td>2.64 (1.45, 4.82)</td>
</tr>
<tr>
<td>Population chart ≤ 3rd centile</td>
<td>1</td>
<td>82,361</td>
<td>1988-1996</td>
<td>8.01 (3.78, 17.0)</td>
</tr>
<tr>
<td>Population chart ≤ 5th centile</td>
<td>1</td>
<td>82,361</td>
<td>1988-1996</td>
<td>6.23 (3.08, 12.61)</td>
</tr>
<tr>
<td>Population chart &lt; 10th centile</td>
<td>13</td>
<td>19,444,776</td>
<td>1981-2006</td>
<td>4.11 (3.70, 4.56)</td>
</tr>
<tr>
<td>Ponderal Index (weight / (100 x length/cm)²) ≤ 0.75</td>
<td>1</td>
<td>5,305</td>
<td>1980-1986</td>
<td>3.57 (2.21, 5.79)</td>
</tr>
<tr>
<td>Fetal growth ratioa ≤ 0.75</td>
<td>1</td>
<td>5,305</td>
<td>1980-1986</td>
<td>7.40 (0.38, 144.11)</td>
</tr>
<tr>
<td>Fetal growth ratioa ≤ 0.80b</td>
<td>1</td>
<td>5,305</td>
<td>1980-1986</td>
<td>33.58 (3.03, 371.43)</td>
</tr>
<tr>
<td>Birth weight &lt; mean – 1SD</td>
<td>2</td>
<td>70,398</td>
<td>1989-1993</td>
<td>5.86 (4.19, 8.32)</td>
</tr>
<tr>
<td>Birth weight &lt; mean – 2SD</td>
<td>1</td>
<td>15,596</td>
<td>Unreported</td>
<td>11.98 (7.85, 18.27)</td>
</tr>
</tbody>
</table>

---

**Figure 2. Forest plot of odds ratios for the association between birthweight standards and neonatal mortality.**
using meta-analysis to calculate summary sensitivity, specificity, and likelihood ratios (Table 3). These measures reveal the discriminative ability of each test and how test results modify a baby’s odds of having a neonatal death. For each test the specificities and positive likelihood ratios were high, but the sensitivity and negative likelihood ratios were generally poor (Table 3). This can be explained by the fact that although a higher proportion of deaths occurred within the low birthweight group, because this group represents a small fraction of the overall population, a large absolute number of deaths still occurred within the normal weight groups, and therefore sensitivity is low and the ‘false negative’ numbers are high, giving a poor negative likelihood ratio (close to 1). For example, the highest positive likelihood ratio was for birthweights below 1.5 kg, indicating that any baby under this weight multiplied their pre-test odds of neonatal death by 49.1 (95% CI 27.3–88.5); however, the negative likelihood ratio was only 1.01 (1.00–1.01), indicating that the odds of death barely change after a negative test result. Thus, although a birthweight below 1.5 kg substantially increases the odds of a poor outcome, a birthweight above 1.5 kg does not increase the odds of a good outcome.

**Discussion**

**Main findings**

Low birthweight showed a strong, consistent association with neonatal mortality. The relationship was highest at lower thresholds and gradually decreased (but remained strong) as the threshold increased. The absolute birthweight seemed to be more strongly related to this outcome than centiles on population weight charts, especially for thresholds of 1.5 and 2.0 kg. Restricting the analysis to singletons, year of birth since 1990, or by country of origin did not change the magnitude of the association. Other definitions of fetal growth restriction were based on single studies and showed mixed results, but none appeared to be more strongly associated with neonatal mortality than the absolute birthweight. The results for neonatal morbidity were mixed, but no single definition of growth restriction appeared to be consistently more strongly associated with adverse outcomes than others. All of the birthweight and population chart thresholds assessed for predictive ability showed a high specificity and positive likelihood ratio for neonatal death, and thus babies who test positive are at a substantially higher risk of neonatal mortality. However, each test generally had a low sensitivity.
and negative likelihood ratio close to 1, and thus a negative test result does not improve the odds that a baby will not have a neonatal death.

**Strengths and limitations**

This review provides the best available evidence, at the time of writing, regarding the association between different measures of fetal growth restriction and adverse outcomes. No other review has attempted to compare different definitions of growth restriction to inform clinical practice. The strength of our review and the validity of our inferences lie in the methodology used. We have complied with existing guidelines for the reporting of systematic reviews of diagnostic and observational studies. We have used the most up-to-date techniques for performing and interpreting meta-analysis. An extensive literature search was performed in relevant databases with no language restrictions applied. Every effort was made to obtain the most complete data set possible through contact with authors and experts in the field. Peters test showed that there was no evidence of small study bias within our largest meta-analysis; other groups were too small to assess. We also considered both the prognostic association of birthweight tests with outcome (as summarised by an odds ratio) and their predictive ability (as summarised by sensitivity, specificity, and likelihood ratio).

There are several limitations to our review. Different numbers of studies contributed to each analysis, and there were few direct comparisons. Indeed, in the only study that compared absolute birthweight and centile chart in the same population, the association for birthweight below the tenth centile was observed to be stronger than the association with absolute birthweight below 2.9 kg, for this outcome. There was a lack of data in some analyses, e.g. customised centile charts and ponderal index in relation to adverse outcome, but as every effort was made to acquire both published and unpublished data we do not feel that anything further could be done to address this. Although

---

**Table 1. Methodological quality of studies included in systematic review of birthweight standards for neonatal outcomes**

<table>
<thead>
<tr>
<th>Quality item</th>
<th>Number (%) of studies n = 29</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cohort study design</td>
<td>28  (97)</td>
</tr>
<tr>
<td>Population adequately described</td>
<td>28  (97)</td>
</tr>
<tr>
<td>Consecutive recruitment</td>
<td>22  (76)</td>
</tr>
<tr>
<td>Prospective recruitment</td>
<td>6   (21)</td>
</tr>
<tr>
<td>Appropriate outcome measure</td>
<td>29  (100)</td>
</tr>
<tr>
<td>Outcome measure blinded</td>
<td>0</td>
</tr>
<tr>
<td>&gt;90% of individuals had outcome measure</td>
<td>26  (90)</td>
</tr>
<tr>
<td>Index test and outcome measure described</td>
<td>14  (48.5)</td>
</tr>
<tr>
<td>Intervention between index test and outcome</td>
<td>1   (3)</td>
</tr>
<tr>
<td>Quality classification</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>24  (83)</td>
</tr>
<tr>
<td>Medium</td>
<td>4   (14)</td>
</tr>
<tr>
<td>Low</td>
<td>1   (3)</td>
</tr>
</tbody>
</table>

---

**Table 2. Subgroup analysis according to birthweight standard and neonatal mortality, where possible, for study quality, year of birth of study population, location of study, and singleton population**

<table>
<thead>
<tr>
<th>Birthweight standard</th>
<th>Number of studies</th>
<th>Subgroup</th>
<th>OR (95% CI)</th>
<th>Estimated prediction interval (EPI)</th>
<th>I², t²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neonatal death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birthweight &lt;1.5 kg</td>
<td>31,46,48</td>
<td>High-quality studies</td>
<td>53.29 (30.08-94.39)</td>
<td>–</td>
<td>I² = 0, t² = 0</td>
</tr>
<tr>
<td>Birthweight &lt;1.5 kg</td>
<td>23,32</td>
<td>Singletons</td>
<td>41.85 (16.53-105.94)</td>
<td>–</td>
<td>I² = 0, t² = 0</td>
</tr>
<tr>
<td>Birthweight &lt;2.5 kg</td>
<td>43,32,43,52</td>
<td>Singletons</td>
<td>8.39 (4.90-14.36)</td>
<td>0.86-81.36</td>
<td>I² = 81, t² = 0.20</td>
</tr>
<tr>
<td>Birthweight &lt;2.5 kg</td>
<td>53,43,45,46,48</td>
<td>High-quality studies</td>
<td>8.15 (5.76-11.54)</td>
<td>2.40-27.66</td>
<td>I² = 80, t² = 0.12</td>
</tr>
<tr>
<td>Birthweight &lt;2.5 kg</td>
<td>46,48</td>
<td>Year of birth ≥ 1990</td>
<td>9.74 (5.31-17.86)</td>
<td>–</td>
<td>I² = 91, t² = 0.17</td>
</tr>
<tr>
<td>Population chart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10th centile</td>
<td>6,34,36,38,40,47</td>
<td>Singletons</td>
<td>4.03 (3.88-4.18)</td>
<td>–</td>
<td>I² = 0, t² = 0</td>
</tr>
<tr>
<td>Population chart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10th centile</td>
<td>8,33,34,36,40,50,51</td>
<td>Year of birth ≥ 1990</td>
<td>4.23 (3.73-4.81)</td>
<td>3.23-5.55</td>
<td>I² = 31, t² = 0.01</td>
</tr>
<tr>
<td>Population chart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10th centile</td>
<td>4,73,36,47,50</td>
<td>Congenital anomalies excluded</td>
<td>4.01 (3.86-4.16)</td>
<td>–</td>
<td>I² = 0, t² = 0</td>
</tr>
<tr>
<td>Population chart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10th centile</td>
<td>6,34,36,47,50,51</td>
<td>Studies in USA/Europe</td>
<td>4.04 (3.89-4.19)</td>
<td>–</td>
<td>I² = 0, t² = 0</td>
</tr>
</tbody>
</table>
every effort was made to control for potential confounding factors through subgroup analysis, because of the quality and reporting of the primary studies this was not always possible. We strictly limited our review to infants born at 37 weeks of gestation or later to avoid the confounding effect of preterm birth; however, the method of estimating gestation in the primary studies was often inaccurate. Very few studies used ultrasound measurement of crown–rump length at 10–13 weeks of gestation, which is the most accurate method\(^6^9\); the majority used the mother’s last menstrual period and some used a clinical examination of the newborn, which are less reliable and may have resulted in preterm infants being included inadvertently. We also recognise that within the group of ‘term’ infants there is a continuing spectrum of gestational age and birthweight, and the risks are not equal, i.e. a baby at 37 weeks of gestation will have a higher risk of adverse outcome than a baby at 40 weeks of gestation, irrespective of birthweight. However, as the majority of studies did not report outcomes according to gestation and birthweight, we could not examine this issue further with the current data. Current clinical practice tends to group infants of 37 weeks of gestation and over together in the way that they are managed, so we feel that the approach in this review remains valid.

As a result of poor reporting in the primary studies, our ability to perform subgroup analysis according to ethnicity was limited. It is known that Afro-Caribbean and Asian populations have smaller babies, and therefore it is likely that the same thresholds would not give the same results in all ethnic backgrounds.\(^6^0\) We did not analyse according to social class: again this was not possible with the information available. We limited the population to singletons where possible, and found that this did not significantly affect the results. We also recognised that the year of birth may be an important factor in neonatal outcome, particularly mortality, because of advances in neonatal care, and therefore performed an analysis limited to studies where the population was born in or after 1990. This did not significantly alter the odds ratios for either birthweight below the tenth centile on the population chart or birthweight below 2.5 kg, the only groups for which this analysis was possible. Customised charts may perform best in subgroups, such as women who are obese, and this type of analysis was not possible.\(^6^1\)

Comparing different standards of birthweight through analyses using different populations may not give a true result; however, no studies reported more than two standards in the same population, and only one study compared absolute birthweight and population centile charts, thereby limiting our ability to deal with this issue.

We attempted to consider all clinically important outcomes within this review; however, one important adverse outcome of fetal growth restriction that has been omitted is stillbirth. This exclusion was made because the remit of the project was to look at parameters of weight at birth and subsequent adverse outcome, rather than tests performed in the antenatal period. We also felt that there was too much potential for confounding to examine the association between birthweight and stillbirth, given that stillbirth may occur days or weeks prior to delivery, and therefore lead to the inclusion of premature infants in the analysis.

### Interpretation

There is a vast literature exploring the relationship between fetal growth restriction and adverse outcomes, using different methodologies to do so. The aim of our review was to consider the association and prediction of different thresholds of birthweight or centile charts, and we therefore excluded studies where 2 × 2 tables could not be obtained from the original article or authors could not provide this. We therefore could not make a complete assessment of the association of birthweight as a continuous variable with adverse health outcomes. In order to address this we considered whether the studies included in the review had examined the association between a continuous birthweight measure and adverse outcomes via logistic regression analysis; however, no studies relating to neonatal outcomes had performed this analysis. We did not identify any other systematic reviews attempting to compare different standards of low birthweight with neonatal outcomes.

---

**Table 3. Results for the predictive ability (sensitivity, specificity, and likelihood ratios) of different birthweight standards for neonatal mortality**

<table>
<thead>
<tr>
<th>Birthweight standard</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive likelihood ratio (95% CI)</th>
<th>Negative likelihood ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birthweight &lt;1.5 kg(^{31,32,46,48})</td>
<td>0.008 (0.004–0.146)</td>
<td>0.99 (0.99–1.00)</td>
<td>49.1 (27.3–88.5)</td>
<td>1.01 (1.00–1.01)</td>
</tr>
<tr>
<td>Birthweight &lt;2.0 kg(^{31,45,52})</td>
<td>0.05 (0.03–0.07)</td>
<td>0.99 (0.99–1.00)</td>
<td>13.3 (2.27–78.28)</td>
<td>0.94 (0.85–1.02)</td>
</tr>
<tr>
<td>Birthweight &lt;2.5 kg(^{31,32,43,45,46,48,49,52})</td>
<td>0.31 (0.19–0.47)</td>
<td>0.94 (0.88–0.97)</td>
<td>5.27 (3.57–7.76)</td>
<td>1.37 (1.15–1.62)</td>
</tr>
<tr>
<td>Population chart &lt;3rd centile(^{57})</td>
<td>0.24 (0.12–0.41)</td>
<td>0.96 (0.96–0.96)</td>
<td>6.31 (3.57–11.14)</td>
<td>0.79 (0.66–0.94)</td>
</tr>
<tr>
<td>Fetal growth ratio &lt;0.80(^{44})</td>
<td>0.67 (0.09–0.99)</td>
<td>0.94 (0.93–0.95)</td>
<td>11.9 (3.87–32.52)</td>
<td>0.36 (0.07–1.75)</td>
</tr>
<tr>
<td>Birthweight &lt; mean – 2 SD(^{39})</td>
<td>0.13 (0.09–0.19)</td>
<td>0.99 (0.99–0.99)</td>
<td>10.53 (7.25–15.28)</td>
<td>0.88 (0.83–0.92)</td>
</tr>
</tbody>
</table>
recommendations. The authors have no competing interests to declare.

conclusion

Our meta-analysis confirms that birthweight has a strong prognostic association with neonatal mortality, with low birthweight substantially increasing the risk of a poor outcome. However, although specificity and positive likelihood ratios were excellent, sensitivity was usually <0.5 and negative likelihood ratios were close to 1. This means that, compared with the pre-test risk of neonatal death (prevalence), babies with a low birthweight (test positive) are at a substantially increased risk, but the risk for those with a normal birthweight (test negative) does not change.

recommendations

Future research is necessary to establish whether there is a birthweight standard that can accurately predict adverse neonatal outcomes. Initially, it is important to compare the different standards across the same population to enable an unbiased comparison, and to further explore the standards that were less frequently reported and therefore could not be included in the meta-analysis within our review, such as ponderal index and customised centile charts. This could be performed through an individual patient data (IPD) meta-analysis, where multiple definitions of fetal growth restriction could be compared across the same population, and factors such as ethnicity more adequately assessed.

Important factors to consider in any future IPD analysis are that of the accurate estimation of gestational age (i.e. pregnancies dated by first-trimester ultrasound scan only) and comparing outcome by week of gestational age rather than grouping all term infants together. Another option would be to perform further analysis on the large Scandinavian birth registries, which record a variety of birth anthropometry that can be linked to health outcomes.

Finally, it is likely that more accurate risk predictions could be made using birthweight as a continuous variable, rather than dichotomising it using a threshold, as is currently the general practice. The use of measures of functional growth rather than weight alone, such as body composition or metabolic parameters, may help to differentiate between infants who are small because of growth restriction, and therefore might be at higher risk of adverse outcome, and those who are constitutionally small.

Conclusion

Birthweight tests are strongly associated with neonatal mortality and morbidity, especially at lower absolute birthweight thresholds, and babies that test positive (i.e. abnormal growth) are at a substantially increased risk of neonatal mortality; however, babies who test negative (i.e. normal growth) do not have a decreased risk of neonatal mortality. Further research is required to identify the optimum definition of low birthweight that helps best predict the risk of adverse outcomes, and this may require using birthweight as a continuous variable, developing prognostic models that also contain other factors, and using individual patient data meta-analysis.

Details of ethics approval

As this systematic review only included studies previously published in peer-reviewed journals or presented in major scientific meetings, ethics approval was not sought.

Funding

G.M. was funded by the Mary Crosse Fellowship, Birmingham Women’s Hospital. K.M. is a National Institute for Health Research (NIHR) Clinical Lecturer.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. List of studies included in our systematic review of birthweight standards and neonatal outcomes.

Appendix S1. Protocol for systematic review and meta-analysis of the prognostic and predictive ability of current birthweight standards for short- and long-term outcomes.

Appendix S2. Medline search strategy for systematic review and meta-analysis of the prognostic and predictive ability of current birthweight standards for short- and long-term outcomes.

References

Higgins JT. Heterogeneity in meta-analysis should be expected and
expected.

22

23

Higgins JP, Thompson SG, Deeks JJ. Measuring inconsistency in

24

Deeks J. Systematic reviews of evaluations of diagnostic and

25

Harbord RM, Harris RJ, Sterne JAC, Steichen T. METABIAS: Stata
module to test for small-study effects in meta-analysis, 2009. Boston
College Department of Economics, Statistical Software Components
S404901.

26

Harries RJ, Bradburn M, Deeks J, Harbord RM, Altman D, Steichen T,
et al. METAN: Stata module for fixed and random effects
meta-analysis, 2008. Statistical Software Components S456798,
Boston College Department of Economics, revised 2009.

27

Harbord RM. METANDI: Stata module to perform meta-analysis of
diagnostic accuracy, 2008. Statistical Software Components S456932,
Boston College Department of Economics.

28

Customised birthweight standards accurately predict perinatal

29

Minakami H, Izumi A, Sato I. Gestational age-specific normal birth
weight for Japanese twins - Risk of early neonatal death in
small-for-gestational-age and large-for-gestational-age twins. J

30

Minor VK, Divon MY. Fetal growth restriction at term: myth or

31

Gardosi J. Intrauterine growth restriction: new standards for

32

Whalther FJ, Ramaekers LH, Walther FJ, Ramaekers LH. The ponderal
index as a measure of the nutritional status at birth and its relation to

33

Botero DL. Intrauterine growth retardation and long-term effects on

34

Pepe MS, Janes H, Longton G, Leisenring W, Newcombe R. Limitations
of the odds ratio in gauging the performance of a diagnostic,

35

Akers J, Aguilar-Ibañez R, Baba-Akbari Sari A, Beynon S, Booth A.
Systematic Reviews: CRD’s Guide for Undertaking Reviews in
Health Care. York, UK: NHS Centre for Reviews and Dissemination,
University of York, 2009

36

Stroup DF, Berlin JA, Morton SC, Olin K. Meta-analysis of

37

The development of QUADAS: a tool for the quality assessment of

38

Levin M. Towards complete and accurate reporting of studies of
diagnostic accuracy, 2008. Statistical Software Components
S404901.

39

Towards complete and accurate reporting of studies of
diagnostic accuracy, 2008. Statistical Software Components
S404901.

40

Towards complete and accurate reporting of studies of
diagnostic accuracy, 2008. Statistical Software Components
S404901.

41

Towards complete and accurate reporting of studies of
diagnostic accuracy, 2008. Statistical Software Components
S404901.

42

Towards complete and accurate reporting of studies of
diagnostic accuracy, 2008. Statistical Software Components
S404901.

43

Towards complete and accurate reporting of studies of
diagnostic accuracy, 2008. Statistical Software Components
S404901.