Does Gestational Age Misclassification Explain the Difference in Birthweights for Australian Aborigines and Whites?

MICHAEL COORY

Coory M (Centre for Clinical Epidemiology and Biostatistics Faculty of Medicine and Health Science, University of Newcastle, David Maddison Clinical Sciences Building, Royal Newcastle Hospital, Newcastle, New South Wales, Australia, 2300. Does gestational age misclassification explain the difference in birthweight for Australian Aborigines and whites? International Journal of Epidemiology 1996; 25: 980–988.

Background. After 34 weeks gestation, summary measures of location for birthweight (e.g. means and centiles) increase more slowly for Australian Aborigines than for whites. A similar pattern has been observed for blacks in the US. This study tests whether the reported pattern is due to differential misclassification of gestational age.

Methods. Simulation was used to measure the potential effect of differential misclassification of gestational age. Reported gestational age data were obtained from Queensland Perinatal Data Collection (QPDC). Estimates of the true distributions of gestational age were obtained by assuming various (plausible) types of misclassification and applying these to the reported distributions. Previous studies and data from the QPDC were used to help specify the birthweight distributions used in the simulations.

Results. At full term, the parameters of the birthweight distributions were robust to gestational age misclassification. At preterm, the 10th centiles were robust to misclassification. In contrast, the 90th centiles were sensitive to even minor misclassification. Extreme types of misclassification were required to remove the divergence in median birthweights for Aborigines and whites.

Conclusions. Gestational age misclassification is an unlikely explanation for the reported divergence in average birthweights for Aborigines and whites. The results might help with the interpretation of other between-population comparisons.

Keywords: Aborigines, birthweight, gestational age, misclassification

One problem associated with using such databases is the misclassification of gestational age. Babies from economically disadvantaged populations such as Aborigines are believed to be more likely to have their gestational age misclassified than more affluent groups (e.g. whites). Therefore, the divergence in median birthweights could be due to differential misclassification of gestational age. Not surprisingly, Lancaster was critical of the Kliewer and Stanley study arguing that 'the results of such studies must be open to question unless the method of calculating gestational age is validated'. The aim of this present study was to quantify the effect of differential misclassification of gestational age on birthweight distributions using simulation.

METHODS
Estimation of the True Distributions of Gestational Age

Misclassification of gestational age implies that a certain proportion of babies will be classified to their...
correct gestational age, but the remainder will be misclassified into other categories of gestational age. This can be represented by a misclassification matrix.\(^6\)

\[
\begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1k} \\
P_{21} & P_{22} & \cdots & P_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
P_{k1} & P_{k2} & \cdots & P_{kk}
\end{bmatrix}
\]

The elements \(P_{ij}\) represent the probability that babies with true gestational age \(j\) (measured in completed weeks) are classified into gestational age category \(i\). For a particular true gestational age \(j\) the \(P_{ij}\) must sum to one (i.e. the column totals are all 1). The rows will not necessarily sum to 1. The diagonal elements represent the probability of correct classification. Perfect classification is represented by the identity matrix (diagonal elements all 1, off-diagonal elements all zero).

A study of inter-observer agreement for maturity scoring by Gagliardi \textit{et al.}\(^9\) was used as a starting point for the specification of the misclassification matrix. Based on the results of that study it was assumed that: (1) no misclassification occurred beyond 4 weeks either side of the true gestational age, (2) for a particular gestational age category, the probability of misclassification was equal in both directions, (3) there was a simple linear decrease in the probability of misclassification as the difference between the true and reported gestational age increased. Sensitivity analyses for each of these assumptions were carried out using four types of variations: (1) errors >4 weeks, (2) unequal probability of misclassification either side of the true gestational age, (3) uniform probability of misclassification whatever the size of the error, (4) exponential decreases in the probability of misclassification with the size of the error.

A misclassification matrix is the standard way of relating the true (but unknown) distribution of gestational age to a reported distribution\(^8\)

\[
t = M^{-1}r. \quad [1]
\]

\(M^{-1}\) denotes the inverse of the misclassification matrix. The vector \(r\) contains reported relative frequencies for each category of gestational age, and the vector \(t\) contains the estimated true relative frequencies. The elements of the vector \(r\) for Aborigines and whites were obtained from the Queensland Perinatal Data Collection (QPDC) using data for 1988–1992 (Figure 1). The QPDC is a whole population database similar to perinatal databases in each of the other Australian states. The gestational age is recorded in completed weeks. The best clinical estimate is used whether based on the date of the last menstrual period (LMP), ultrasound scanning, or maturity scoring by a clinician at birth. The method (or methods) used is (are) not recorded.
In this study, single week categories of gestational age from 28 to 41 weeks inclusive were used. Before 28 weeks, a composite 24–27 week category was formed because the comparatively small number of Aboriginal babies prevented the use of single week categories. The effects of gestational age misclassification on the post-term birthweight distributions were not considered because of the small number of postterm Aboriginal babies. Several misclassification probabilities (defined as the sum of the probabilities in column \( j \) except \( P_{jj} \)) were used: 0.05–0.35. To allow for differential misclassification, higher misclassification probabilities were used for Aborigines than whites. Only singleton livebirths were considered.

**Specification of the True Birthweight Distributions**

Initial simulations were run under the assumption that the true (but unknown) birthweight distributions are symmetrical and follow Normal curves at all gestational ages. Separate simulations were run for males and females as they have different mean birthweights. However, only the results for males are reported in this paper because it was the rate-of-increase in the birthweights, rather than their absolute values, that was important in determining the bias. Although the gestational age specific mean birthweights for females are about 150 g less than those for males, the rate-of-increase is similar. Hence, the results for males and females were similar.

Birthweights for whites and Aborigines were obtained using data from the QPDC for the period 1988–1992 (Figure 2). The observed medians were used as estimates of the centres of the true birthweight distributions as the median is less likely to be affected by skewness than the mean.

Misclassification of gestational age is unlikely to substantially affect the standard deviation of birthweight at 40 weeks because of the large numbers of babies compared with the numbers in the adjacent categories of gestational age (Figure 1). For this category, the standard deviation for birthweight observed in the QPDC was used. For the other categories, gestational age misclassification will tend to inflate the standard deviation of birthweight, so the true standard deviations would be lower than those observed. From the QPDC data, the coefficient of variation for birthweight at 40 weeks was 0.13. The standard deviation of birthweight increases with increasing gestational age, at about the same rate as the mean.\(^\text{10}\) Hence, for the initial simulations, the standard deviations of birthweight were specified to be 0.13 times the postulated means (Table 1).

Sensitivity analyses for each of the assumptions about the true birthweight distributions were carried out: (1) the preterm distributions were allowed to have a positive skewness of 1.0, (2) the rate-of-increase
Table 1  Birthweight standard deviations (SD) used in the simulations

<table>
<thead>
<tr>
<th>Gestational age</th>
<th>Aborigines</th>
<th>Whites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assumed SD based on CV = 0.13 (g)</td>
<td>SD calculated from QPDC(^b) (g)</td>
</tr>
<tr>
<td>24–27</td>
<td>137</td>
<td>183</td>
</tr>
<tr>
<td>28</td>
<td>168</td>
<td>243</td>
</tr>
<tr>
<td>29</td>
<td>176</td>
<td>254</td>
</tr>
<tr>
<td>30</td>
<td>202</td>
<td>272</td>
</tr>
<tr>
<td>31</td>
<td>211</td>
<td>307</td>
</tr>
<tr>
<td>32</td>
<td>235</td>
<td>342</td>
</tr>
<tr>
<td>33</td>
<td>271</td>
<td>368</td>
</tr>
<tr>
<td>34</td>
<td>303</td>
<td>392</td>
</tr>
<tr>
<td>35</td>
<td>318</td>
<td>421</td>
</tr>
<tr>
<td>36</td>
<td>347</td>
<td>444</td>
</tr>
<tr>
<td>37</td>
<td>389</td>
<td>452</td>
</tr>
<tr>
<td>38</td>
<td>412</td>
<td>453</td>
</tr>
<tr>
<td>39</td>
<td>433</td>
<td>456</td>
</tr>
<tr>
<td>40</td>
<td>465</td>
<td>465</td>
</tr>
<tr>
<td>41</td>
<td>468</td>
<td>480</td>
</tr>
</tbody>
</table>

\(^a\)Coefficient of variation.  
\(^b\)Queensland Perinatal Data Collection.

between birthweight means in adjacent categories of gestational age was increased by 10%, (3) standard deviations were allowed to be as large as those from the QPDC.

**Simulations**

FORTRAN programs that incorporated subroutines from the NAG library\(^1\) were used to generate the true distributions of birthweight. Where the true distributions were specified as symmetrical, the NAG subroutine GO5DDF was used to generate pseudo-random numbers from a Normal distribution with specified mean and standard deviation. True preterm birthweight distributions with a skewness of 1.0 were generated using pseudo-random numbers from a linearly transformed \(x^2\) distribution with 8 d.f. The NAG subroutine GO5DHF was used.

Mixtures of the true distributions were used to construct birthweight distributions contaminated by gestational age misclassification. The amount of contamination depended on two factors: (1) the true relative frequencies in the neighbouring categories of gestational age as estimated by equation [1], (2) type and probability of gestational age misclassification specified.

Because this work examined the effect of differential misclassification on differences in the birthweight distributions between whites and Aborigines, absolute rather than relative biases are reported (i.e. centile [or mean] from contaminated distribution minus centile [or mean] from true distribution) (Table 2). Biases were not calculated for the two end categories (24–27 weeks and 41 weeks) because they did not have enough neighbouring categories to permit construction of the contaminated birthweight distributions. Ten-thousand replications and 1000 births were used for each category of gestational age so that 95% confidence intervals (CI) for the mean and centiles had a half width of less than one gram.

**RESULTS**

**Estimation of the True Distributions of Gestational Age**

The estimated true distributions of gestational age had a smaller proportion of preterm (<37 weeks) and a larger proportion of full term (37–41 weeks) babies than the corresponding reported distributions (Figure 3). This result was robust to changes in the specification of the misclassification matrix. When misclassification probabilities >0.15 were used for the full-term categories, negative values were obtained for some elements of the vector \(t\) (specifically those corresponding to 35 and 36 weeks). Hence, misclassification probabilities for the full-term categories were not allowed to be >0.15 even when larger probabilities were used for the other categories.
Table 2 Absolute biases for different specifications of the true birthweight distributions

<table>
<thead>
<tr>
<th>Misclassification probability</th>
<th>Summary measures</th>
<th>Absolute bias (g)</th>
<th>Skewness of preterm distributions</th>
<th>Standard deviations same as QPDC^b</th>
<th>Difference between adjacent means increased by 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestational age = 30 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>10th centile</td>
<td>-17</td>
<td>-8</td>
<td>-17</td>
<td>-21</td>
</tr>
<tr>
<td>(all gestational age categories)</td>
<td>median</td>
<td>21</td>
<td>24</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>90th centile</td>
<td>137</td>
<td>162</td>
<td>134</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>52</td>
<td>51</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>0.35—preterm, 0.15—full term</td>
<td>10th centile</td>
<td>-17</td>
<td>-42</td>
<td>-44</td>
<td>-60</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>50</td>
<td>58</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>90th centile</td>
<td>226</td>
<td>358</td>
<td>349</td>
<td>396</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>67</td>
<td>67</td>
<td>68</td>
<td>52</td>
</tr>
<tr>
<td>Gestational age = 36 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>10th centile</td>
<td>84</td>
<td>85</td>
<td>98</td>
<td>87</td>
</tr>
<tr>
<td>(all gestational age categories)</td>
<td>median</td>
<td>199</td>
<td>224</td>
<td>213</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>90th centile</td>
<td>471</td>
<td>479</td>
<td>434</td>
<td>532</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>241</td>
<td>242</td>
<td>240</td>
<td>271</td>
</tr>
<tr>
<td>0.35—preterm, 0.15—full term</td>
<td>10th centile</td>
<td>63</td>
<td>65</td>
<td>81</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>188</td>
<td>217</td>
<td>203</td>
<td>205</td>
</tr>
<tr>
<td></td>
<td>90th centile</td>
<td>473</td>
<td>477</td>
<td>436</td>
<td>533</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>232</td>
<td>234</td>
<td>68</td>
<td>52</td>
</tr>
</tbody>
</table>

^a Coefficient of variation.  
^b Queensland Perinatal Data Collection.

Biases for Whites
For the preterm categories of gestational age, the 10th centiles were robust to gestational age misclassification. In contrast, the 90th centiles were sensitive to even low levels of misclassification (Figure 4). That is, misclassification mainly affects the upper part of the birthweight distributions. This was due to two factors: (1) large increases in relative frequencies in each successive category of gestational age (Figure 1), (2) comparatively large increases in the mean and standard deviation for each successive category of gestational age (Figure 2 and Table 1). Changes to the parameters of the postulated true birthweight distributions did not substantially change the biases. Table 2 shows the effect of the changes for two categories of gestational age: 30 and 36 weeks.

For the full-term categories the biases were small. Although the relative frequencies increase sharply from 37 to 40 weeks (Figure 1), the means and standard deviations are similar (Figure 2 and Table 1). Hence, at full term, the contaminated distributions were not much different from the true distributions.

Comparison of Biases for Whites and Aborigines
For a given misclassification probability, using the reported gestational age distribution for Aborigines (instead of the gestational age distribution for whites)—for the vector r—did not substantially change the biases. Similarly, using the medians of the reported birthweight distributions for Aborigines (instead of the medians for whites)—as the centres of the true birthweight distributions—did not substantially change the biases.

At full term the parameters of the birthweight distributions were robust to gestational age misclassification. Therefore, the Aboriginal/white differences in birthweight centiles observed at full term are not due to gestational age misclassification.

For mildly preterm babies (34-36) weeks the larger the misclassification probability the larger the positive
GESTATIONAL AGE MISCLASSIFICATION

bias (Figure 4). At these gestational ages the reported birthweight centiles for Aborigines are lighter than those for whites. Hence, the true difference must be even greater if a higher misclassification proportion is postulated for Aborigines than for whites.

In contrast, for the early categories of gestational age, the birthweight centiles for Aborigines are similar to those for whites. Differential misclassification of gestational age might mean that the true birthweight centiles for Aborigines are lighter than those for whites. If true, this would mean that Aborigines are lighter than whites at all gestational ages (i.e. the centiles do not diverge).

This issue was examined by running simulations where differential misclassification was specified for Aborigines and whites. The proportion misclassified for whites was set at 5%, and the proportion misclassified for Aborigines allowed to vary. Even with a large difference in the misclassification proportions, the difference in medians was much less than at full term (Figure 5). This suggests that the divergence in birthweight centiles for Aborigines and whites is a real phenomenon and not an artefact related to gestational age misclassification.

Simulations were also conducted with birthweight standard deviations for Aborigines as large as those observed in the QPDC. In addition, the rate-of-increase in the mean birthweights for Aborigines was increased by 10%, and the preterm distributions were allowed to have a positive skew of 1.0. These changes did not affect the conclusions.

Extreme types of misclassification were required to remove the divergence in median birthweights. For example, the divergence was removed under the following conditions: (1) a misclassification probability of 0.50 for Aborigines born before 34 weeks and all of the misclassification in the direction of earlier gestational ages, (2) a misclassification probability of 0.10 for whites born before 34 weeks and all of the misclassification in the direction of later gestational ages. Scenarios such as this seem unlikely.

DISCUSSION

This simulation study provides evidence against the hypothesis that the divergence in median birthweights for Aborigines and whites is due to the differential misclassification of gestational age. Moreover, the results might help with the interpretation of other studies. For example, for all categories of gestational age, the 10th centiles are unlikely to be substantially biased even in the presence of large levels of misclassification. In addition, at full term, between population comparisons of the mean (or median) birthweights are likely to be valid.
Misclassification probability = 0.05
(all gestational age categories)

Misclassification probability = 0.15
(all gestational age categories)

Misclassification probability = 0.25 preterm,
0.15 full-term

Misclassification probability = 0.35 preterm,
0.15 full-term

Figure 4 Absolute bias for means, medians, 10th and 90th centiles for whites
Some of the results quantify biases that have previously only been discussed in a qualitative way. For example, the much larger positive biases for the 90th centiles relative to the 10th centiles for the preterm categories of gestational age suggest that even low levels of misclassification can cause positive skewness (or even bimodality) in the birthweight distributions. This result is consistent with the widely held belief that the positive skewness and bimodality observed for the preterm birthweight distributions are artefacts resulting from errors in the measurement of gestational age.

Although gestational age measurements are known to be subject to error, good information about their accuracy in whole population databases is not available. Either ultrasound scanning or secure LMP dates would be suitable as the gold standard in a validation study, however, such an approach would mean that the group of babies whose gestational ages are subject to the most error would be excluded. Hence the generalizability of such a study to all records in a whole population database would be open to question. Maturity scoring would be carried out on a representative sample of babies in a whole population database but has been shown to be too inaccurate and imprecise to be used as the gold standard. Given these difficulties, one way of obtaining some insight into the effect of gestational age misclassification is via simulation.

To proceed with simulations, simplifying assumptions were made about the type of misclassification and the true distributions of birthweight. For example, an inter-observer agreement study of maturity scoring was used to specify the initial misclassification matrices. Admittedly, the gestational age measurements reported to whole population databases are usually obtained using methods other than (or in addition to) maturity scoring (i.e. LMP dates, or ultrasound). However, the type of misclassification specified was considered a worst case scenario since the other two methods are considered more precise than maturity scoring. The inter-observer agreement study was only used as a starting point because such studies can only provide estimates of the precision of the measurements, not bias. Later simulations allowed for bias (i.e. different probabilities of misclassification in either direction).

More extreme types of misclassification could have been considered. For example, it could be argued that when LMP dates are used there is likely to be a peak of misclassification at 4 weeks before the true gestational age. David emphasized this type of misclassification in a descriptive study of computerized birth records from North Carolina for the years 1975–1977. He claimed that misdating of 2% of births by one menstrual period would be sufficient to explain all the preterm births with unusual birthweights. However, such gross outliers can be easily identified and trimmed from the...
data set before undertaking the main analysis (as several researchers have done\textsuperscript{13}). In fact, David developed a technique for trimming such records. Modern methods of data quality control could even allow such errors to be detected at the data collection stage. Consequently, types of misclassification that produced records that were easily identifiable as outliers were not considered.

The misclassification probability varied between 0.05 and 0.35. Misclassification probabilities >0.15 might seem unlikely. However, clinicians traditionally mistrust gestational age estimates, even those based on secure LMP dates.\textsuperscript{15,16} Moreover, some authors have suggested that the proportion misclassified can be as large as 0.35.\textsuperscript{2}

Assumptions about the birthweight distributions were also varied. The preterm distributions were allowed to have a positive skew of 1.0. More severe skewness was considered unlikely since skewness for birthweight data from the QPDC was <1.0 for all the preterm categories of gestational age (the largest value was 0.7 at 27 weeks). Skewness for the full term birthweight distributions was not considered since studies have consistently found that these distributions are symmetrical.\textsuperscript{12} Simulations were also run allowing the standard deviations of the birthweight distributions to be as large as those for the QPDC data. These variations to the assumptions did not substantially change the biases.

Gestational age misclassification will attenuate the relationship between birthweight and gestational age. Therefore, the true rate-of-increase in the mean birthweights might be larger than those observed in the QPDC data. Simulations were run where the postulated rate-of-increase between means in adjacent categories of gestational age was increased by 10%. The effects of increases larger than 10% were not considered since they resulted in postulated mean birthweights that were implausibly small for the early preterm categories of gestational age and implausibly large for the full term categories.

Although significant misclassification of gestational age measurements in whole population databases is likely, the results of this study show that this cannot affect the divergence of Aboriginal and white birthweights. A genetic explanation for this phenomenon cannot be ruled out. However, the Australian Aboriginal population has a high prevalence of several risk factors for full term low birthweight. Examples include: maternal smoking, teenage pregnancy, short inter-pregnancy interval, and poor maternal nutrition.\textsuperscript{1} We should initially focus on these potentially modifiable risk factors, rather than immutable genetic ones.

REFERENCES


(Revised version received April 1996)