Biased assessment of gestational age at birth when obstetric gestation is known

Luigi Gagliardi, Carlo Brambilla, Raffaella Bruno, Stefano Martinelli, Vito Console

Abstract
The gestational age of 302 neonates whose obstetric gestational age was known was assessed at birth using the Dubowitz method; it was obtained from Dubowitz score both graphically from a nomogram and by calculation from the corresponding equation.

The values obtained graphically differed to a lesser extent from the obstetric gestational age than did the gestation derived algebraically. With infants small for gestational age (SGA), the difference between the methods was smaller and not significant. It is concluded that the concurrent knowledge of obstetric gestational age introduced a bias in the graphipal step; this did not happen in SGA infants probably because in these cases the available information is sometimes less certain.

These data demonstrate that even simple procedures are influenced by concurrent information; as a philosophical point about the interpretation of data in general, this study provides an empirical example of the 'theory-ladenness of facts' in medicine.

(Arch Dis Child 1993;68:32–4)

It has been suggested that 'any scientist should begin a scientific paper with the phrase: "Ladies and gentlemen, these are the opinions on which I base my facts"...'. In fact, many philosophers of science and scientists claim that all observations ('facts') are impregnated with the theory under which they were collected, that is, that facts separated from the hypotheses, beliefs, and expectations of the observer do not exist. A widely cited phrase summarising this concept is that of the 'theory-ladenness of facts'.

This problem has received little attention in clinical practice: clinicians sometimes admit that there is a subjective element in clinical assessment of signs, but seem to believe that this is a minor problem in medicine so that the hypothesis of its existence has rarely been tested (probably also because it is difficult to prove). A good example is the estimation of gestational age in neonates. The knowledge of gestational age is important for obstetricians and neonatologists, and it is routinely estimated both prenatally and at birth. The neonatologist is usually aware of the neonate's history, including the obstetric gestational age estimate. Despite a large body of studies on neonatal assessment of gestational age, the potential for bias of this situation has never been formally tested. Our hypothesis was that this information influenced the results of the neonatal assessment procedure, leading to a greater concordance between estimated and obstetric gestational age.

In our unit, gestational age is assessed at birth with the scale devised by Dubowitz et al. On this scale, 21 scored neurological and physical signs are recorded and added up. The final score is linearly related to gestational age. The authors provide two methods for transforming the score into gestational age: (i) algebraic using an equation (gestational age (weeks) = \(0.2642 \times \text{score} + 24.595\)) and (ii) graphic where a nomogram is entered on the abscissa (score), the straight line depicting the above mentioned equation is intersected, and the estimated gestational age is found on the ordinate. The graphic method is preferred in clinical practice and is shown in many textbooks because it is easier and quicker to use.

Of course, the two methods should yield the same results; some discrepancies could be expected as intersection of the line is usually done freehand, but there should not be any systematic difference (bias) between the methods. The presence of such a difference would indicate that the clinician is biased during the graphic procedure. To our knowledge, no study has been done to test this hypothesis.

Patients and methods
We gathered information about obstetric gestational age, and graphically estimated it, by chart review of 302 consecutive neonates admitted to our ward in 1989 and 1990 for whom the Dubowitz estimate and the obstetric presumed gestational age (based on dates and ultrasound) were recorded. The clinical examination and the graphic estimation were carried out by the clinician responsible for the infant’s care who was aware of the obstetric gestational age.

The nomogram used is an exact reproduction of that in the paper of Dubowitz et al.

Algebraic gestational age was calculated by computer using the reported score. The absolute difference of gestational age estimation (estimated gestational age – obstetric gestational age taken in absolute value) was then calculated for each neonate with both the algebraic and the graphic methods (‘algebraic error’ and ‘graphic error’ respectively). To test the null hypothesis of no bias these ‘errors’ were compared with a Wilcoxon matched rank sum test, a significant difference indicating a bias.

The number of times that the estimated gestational age differed by more than one and two weeks from the obstetric gestational age was
Absolute differences (‘errors’) between estimated and obstetric gestation. Results are median (10th, 90th centile) in weeks

<table>
<thead>
<tr>
<th></th>
<th>Algebraic error</th>
<th>Graphic error</th>
<th>Difference*</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole sample (n=302)</td>
<td>0.95 (0.18, 2.65)</td>
<td>0.9 (0.1, 2.5)</td>
<td>0.15 (−0.46, 0.82)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Infants who were SGA (n=79)</td>
<td>0.99 (2.3, 2.97)</td>
<td>1.0 (0.1, 3.2)</td>
<td>(−0.6, 0.88)</td>
<td>0.66</td>
</tr>
<tr>
<td>Infants who were not SGA (n=223)</td>
<td>0.99 (0.15, 2.51)</td>
<td>0.7 (0.1, 2.32)</td>
<td>(−0.46, 0.79)</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>

*Algebraic error - graphic error.

also computed for the two methods and compared by McNemar’s test.

These analyses were then separately performed in infants who were small for gestational age (SGA; this is an infant weighing <10th centile according to Lubchenco charts) and infants who were not SGA. We reasoned that, at least in some cases, in neonates who were SGA the obstetric determination of gestation tends to be less secure and the neonatologist is faced with conflicting information at birth (for example birth weight and obstetric gestational age). In these infants the neonatologist’s expectations are less strong than in infants who are not SGA; if the hypothesis of the theory-ladenness of facts is true, the bias also should be less strong in SGA infants.

All computations were done with the statistical package SPSS/PC+.

Results

The mean birth weight of the sample was 2134 g (range 610–4480) and the mean obstetric gestational age was 35 weeks (range 24–43); 200 neonates (66%) were <37 weeks gestational age and 79 (26%) were SGA.

The median error was 0.95 weeks with the graphic method and 0.9 weeks with the graphic method. The algebraic error was significantly greater than the graphic one (median 0.15 weeks; p<0.00001): the null hypothesis of no bias was therefore rejected (table).

The number of estimations differing by more than one week from the obstetric gestational age was higher for the algebraic rather than for the graphic method (47.7% v 34.3% respectively, p<0.00001). For differences exceeding two weeks, the corresponding figures were 21.9% v 13.1% (p<0.00001).

In infants who were SGA the difference between the methods was very small and not statistically significant (median 0.01 weeks, p=0.66), while in infants who were not SGA the algebraic error was significantly larger than the graphic one (median 0.2 weeks, p<0.00001), due to a smaller graphic error in this group (table).

Discussion

The purpose of this study was to investigate whether clinicians, when assessing the gestational age of a neonate, are influenced by their knowledge of obstetric gestational age. Despite a large body of studies on neonatal estimation of gestational age we could not find an evaluation of this problem.

The purpose was not to assess the accuracy of the Dubowitz method to estimate gestational age. For this reason, even though we could not quantify or improve the quality of obstetric gestational age, we do not think that this weakened our conclusions. Furthermore, a study conducted in northern Italy in several neonatal centres (including our own) showed that only 14% of the pregnancies had an uncertain gestational age.

The results of this study clearly indicate that, with the Dubowitz procedure, the study of obtaining the gestational age from the score using the graphic method introduced a systematic difference (a bias) as compared with the use of the algebraic equation. This bias tended to (misleadingly) increase the accuracy of the method itself.

The magnitude of this bias was on average small (median 0.15 weeks), but was sufficient to affect the performance of the scale as evidenced by the number of neonates whose estimated gestational age differed by more than one or two weeks from the obstetric gestational age: for both intervals, the graphic method results tended closer to obstetric gestational age than the true value (algebraic method) ones.

Of course, it is not the instrument itself (the nomogram), but rather its use, that is biased, and this is due to concurrent information about obstetric gestational age available to the clinician when he or she is performing a gestational age estimation. This interpretation is supported by the fact that when prior knowledge was less secure and included also conflicting information (for example gestational age and birth weight) as in SGA infants, the bias was smaller and not statistically significant. (Another explanation for this result in SGA infants is insufficient sample size: this possibility is unlikely as the power to detect a difference of 0.2 weeks, with α=0.05, was approximately 0.95.)

A similar influence of the knowledge of dates on the duration of pregnancy assessment has been reported in an obstetric setting by Gjørup et al. In their study, blind and non-blind estimates of the duration of amenorrhoea were compared in women seeking a legal abortion in early pregnancy. Non-blind estimates were in significantly closer agreement with stated duration of amenorrhoea than blind ones. In that study, however, blind and non-blind estimates were performed by different set of doctors, so that differences between observers were mixed with and could not be definitely distinguished between methods. Inter-rater differences are known to be large in neonatal gestational age assessment, so that in this study an intrarater (intrasubject) design was used to avoid this problem.

As a final point, we suspect that the bias described in this study is only an example of a generalised problem, and that the opening quotation from Marmot also applies to clinical medicine.

In clinical work, taking advantage of all available information is regarded as ‘good clinical practice’ and the source of bias to be avoided. Clinicians are generally unaware of the extent to which their judgment is influenced by concurrent information and tend
to believe that signs are objective. Our study empirically confirmed instead that concurrent expectations can influence the evaluation of clinical signs (that is, that facts are theory-laden). This probably represents an important problem in retrospective research, where data gathered from chart review (collected for clinical and not for research purposes) are taken at face value.