Evaluation of the Sheffield system for identifying children at risk from unexpected death in infancy

Sir,

The conclusions drawn by Oakley et al. (Archives, 1978, 53, 649) are so at variance with the data they present that we must comment. Briefly, the re-analysis below shows that the birth data from Birmingham and Newcastle are remarkably similar and are consistent with the two sets of Sheffield birth data (Protestos et al., 1973; Carpenter et al., 1977). Also the hypothesis that the birth scoring system (Carpenter et al., 1977) would be equally effective in Birmingham, Newcastle, and Sheffield is not rejected (P = 0.14). Furthermore, although our birth scoring system performed relatively poorly on the Birmingham data compared with Sheffield, its level of performance there is the same as that of the scoring system studied by Alberman and Goldstein (1970) when it was shown that maximum benefit is gained by the differential allocation of resources to a high-risk group.

Table 3 of Oakley et al. (1978) appears to show that the birth data from Birmingham differ from those collected in Newcastle, and the text indicates that this is how the authors view their data. Their mistake arises from the method of analysis. For example, invasive investigations in pregnancy are reported as significantly associated with the Birmingham cases but not with Newcastle ones. What is not apparent is that the difference between Birmingham cases and controls is smaller than the difference between cases and controls in Newcastle in this respect, as shown in Table C1.

Table C1 Percentage differences between cases of unexpected death and live controls in Birmingham and Newcastle

<table>
<thead>
<tr>
<th>Variable</th>
<th>City</th>
<th>Cases (%)</th>
<th>Controls (%)</th>
<th>Percentage difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasive investigations during pregnancy</td>
<td>Birmingham</td>
<td>32-5</td>
<td>26-5</td>
<td>6-0</td>
</tr>
<tr>
<td></td>
<td>Newcastle</td>
<td>19-6</td>
<td>12-5</td>
<td>7-1</td>
</tr>
<tr>
<td>Sex</td>
<td>Birmingham</td>
<td>51-8</td>
<td>57-8</td>
<td>-6-0</td>
</tr>
<tr>
<td></td>
<td>Newcastle</td>
<td>69-6</td>
<td>46-4</td>
<td>23-2</td>
</tr>
</tbody>
</table>

Re-analysis of risks associated with factors. The question then arises, is there any evidence that their data from Birmingham and Newcastle differ in regard to the risks of unexpected death associated with the 18 factors they list? There are several approximate statistical methods for examining this question. We have used the exact method of Zelen (1971) and have applied it to the data presented in Table 3 of the paper by Oakley et al. The only significant difference (P = 0.025) between the two sets of data is that relating to sex, Table C1. This is mainly because the proportion of male cases in Birmingham is unusually low (51.8%) while that in Newcastle is exceptionally high (69.6%) (Carpenter, 1972). At least one P value as small as this may occur by chance in 18 significance tests.

The overall impression is that the two data sets are remarkably similar. The average P value for the 18 tests of homogeneity is 0.72 which is substantially larger than the expected value of 0.5. If the 18 comparisons are regarded as independent, the two sets of data appear significantly similar (P < 0.00003), but the variables are all interrelated which probably accounts for the unexpected similarity. For 5 factors: birthweight, twin birth, urinary tract infection during pregnancy, gestation < 37 weeks, and admission to special care baby unit, the relative risks cannot be more similar, given the number of infants with the factor in each set of data.

Having found that the data from Birmingham and Newcastle are essentially similar, are they compatible with the findings of the two surveys of birth data in Sheffield reported by Protestos et al. (1973) and Carpenter et al. (1977)? Our analysis is confined to 7 of the 8 factors in the birth scoring system described by Carpenter et al. (1977). We cannot include the 2nd stage of labour because the data are grouped differently. For each factor, we tested whether the 4 estimates of relative risk, derived from the 4 data sets, could be regarded as homogeneous. In no case was the hypothesis of homogeneity rejected. The smallest P value was 0.085 for mother's blood group. This is hardly surprising because the two Sheffield data sets are significantly different in respect of mother's blood group (Carpenter et al., 1977). Combining the 7 significance tests also provided no suggestion of heterogeneity.

Consequently the best information available is provided by combining all the data. The results of doing so are presented in Table C2 which shows that the differences are significant for each of the 7 variables. Differences are similar to those previously reported by Carpenter et al. (1977).

Table C2 Combined birth data from Birmingham, Newcastle, and two Sheffield surveys. (334 cases of unexpected infant death and 389 live controls)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cases (%)</th>
<th>Controls (%)</th>
<th>Percentage difference P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother's age &lt; 24</td>
<td>60-8</td>
<td>43-8</td>
<td>17-0</td>
</tr>
<tr>
<td>Birthweight &lt; 3000 g</td>
<td>45-3</td>
<td>27-1</td>
<td>18-2</td>
</tr>
<tr>
<td>Intention to breast feed</td>
<td>20-5</td>
<td>31-4</td>
<td>-10-9</td>
</tr>
<tr>
<td>Primiparous</td>
<td>22-7</td>
<td>37-0</td>
<td>-14-3</td>
</tr>
<tr>
<td>Twins</td>
<td>5-7</td>
<td>0-5</td>
<td>5-2</td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td>12-4</td>
<td>7-3</td>
<td>5-1</td>
</tr>
<tr>
<td>Blood group A</td>
<td>35-4</td>
<td>46-3</td>
<td>-10-9</td>
</tr>
</tbody>
</table>

*Combining the four data sets by Cochran's test (Armitage, 1971) gives mean % differences which are up to 1-2% smaller and which are slightly less significant. But exact tests (Zelen, 1971) of the combined relative risks are more significant than shown.

Birth scoring. Using stepwise discriminant analysis, Oakley et al. (1978) obtained discriminant functions for Birmingham and Newcastle which had only 3 variables in common, only one of which was included in our birth scoring system—i.e. mother's age. The reader is left with the impression of a mass of heterogeneous data. Stepwise analyses are very arbitrary in their selection of variables and heterogeneity can be established only by formulating...
and testing the appropriate hypothesis. Oakley et al. give no hint that this has been done, nor have they tested whether their data conflict with our scoring system. All their analysis shows is that if we had started with one of their sets of data we might be using a scoring system which superficially would look very different from our present one.

Oakley et al. report the results of applying our second scoring system to their data. Testing the hypothesis that infants with scores $>500$ are at the same risk in Birmingham and Newcastle we find $P = 0.13$. When we compare their results with our evaluation (Table 9 of Carpenter et al., 1977) we again find no significant differences ($P = 0.14$). Taking all these data together, 54% of cases had scores $>500$ compared with 30% of controls. It is now clear that the choice of 500 as the critical score is too low if the high risk group is only to include about 16% of controls. Raising the critical score also reduces the proportion of cases in the high risk group. In any case there is no good reason for merely considering high and low risk groups as risk rises exponentially with score.

The principal criticism of Oakley et al. is that 'the degree of discrimination that the Sheffield "at birth" score achieved in either Birmingham or Newcastle during the study period is not one that would be of use in a prospective prevention programme, . . .'. Sensitivity and specificity may be combined in the Youden index, $J$ (Armitage, 1971) which is simply the difference between the percentage of cases and controls in the high risk group. For our prospective study, $J = 38\%$; for the Newcastle data, $J = 34\%$; for the Birmingham data, $J = 13\%$. Alberman and Goldstein (1970) made a statistical evaluation of the risk registers and showed that maximum benefit is gained by the differential allocation of resources to a high risk group. The value of $J$ for the risk register they discussed is 13%, which is identical with that observed in Birmingham.

We have never said that the scoring system developed for identifying high-risk infants in Sheffield would work in other cities. The wide range of postperinatal mortality rates indicates that differences between cities are likely, but so far the reported differences may be attributed to random variation. We believe a scoring system along the lines we have described provides a practical method of improving the effectiveness of primary care.

References


R. G. Carpenter AND Angela Gardner
Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT

J. L. Emery
Children's Hospital, Western Bank, Sheffield S10 2TH

Drs Oakley and Stanton comment:

Our co-author Mrs J. Tavare is now living in the USA, and we have had insufficient time to obtain her reply to the comments of Dr Carpenter and his colleagues on the statistical methods used in our paper.

There are always many ways of analysing a particular data set. Dr Carpenter and his colleagues are right in observing that our paper was primarily concerned with the practical application of the Sheffield scoring system to prospective prevention programmes, although we confirmed that the system discriminated at a statistically significant level between children who die, and the living control population.

A system which in Birmingham identified fewer than half the children who subsequently died while scoring one-third of the normal population at risk may be statistically valid, but we believe it has no place in primary care where we are dealing with the lives of infants and not statistics. Using the present system could well result in a positive maldistribution of primary care services by creating a false sense of security, and directing attention away from potentially vulnerable children who do not score at risk.

It is possible that future work in the Department of Health and Social Security's Multicentre Postneonatal Study will produce an effective at-risk system. Meanwhile, we are still of the opinion that scoring systems should only be used as research tools to indicate groups of infants that require particular study to investigate the aetiological factors of the sudden infant death syndrome.

J. R. Oakley
Children's Hospital, Western Bank, Sheffield S10 2TH

A. N. Stanton
Princess Mary Maternity Hospital, Great North Road, Newcastle upon Tyne NE2 3BD