Reasons for a policy change included concern over possible adverse effects of intramuscular vitamin K (71%), reports of vitamin K deficiency bleeding (49%), and problems over the use of an unlicensed oral preparation (27%).

Comment

Almost all infants born in the British Isles now receive vitamin K prophylaxis and the trend towards oral administration continues. Neverthless, 27% of nurseries cited concern about the unlicensed oral use of vitamin K preparations as a reason for policy changes. Professionals may feel trapped by a dilemma. Giving vitamin K by intramuscular injection is seen as undesirable and may be associated with an increased risk of cancer, yet the injection uses a licensed preparation and provides reliable protection against vitamin K deficiency bleeding. Conversely oral regimens avoid the trauma of injection and any potential risk of extremely high blood concentrations and have not been implicated in any cancer risk. However, the efficacy of multiple oral dose regimens is not well established. They are complicated to administer, and their use of unlicensed preparations may theoretically expose professionals to litigation in the event of failure of prophylaxis or of unforeseen adverse effects.

The data highlight the current confusion over the optimal dose of oral vitamin K. Formula fed infants, whose vitamin K intake is around 25 μg daily, rarely bleed from vitamin K deficiency. Hence it would be logical to suppose that a similar daily supplement given to breast fed infants would also be protective while avoiding unphysiological peak plasma concentrations.[2]

Whatever regimen is used, we suggest that parents should be given written information about vitamin K prophylaxis and deficiency bleeding early in pregnancy to allow time for deliberation. The recommendations of the maternity unit can then be stated, including endorsement of breast feeding, and signed consent requested.

1 Golding J, Greenwood R, Birmingham K, Mott M. Childhood cancer, intramuscular vitamin K and pethidine given during labour. *BMJ* 1992;305: 341-6.

2 Expert Committee. *Vitamin K prophylaxis in infancy.* London: British Paediatric Association, 1992.

3 Handel J, Tripp JH. Vitamin K prophylaxis against haemorrhagic disease of the newborn in the United Kingdom. *BMJ* 1991;303:1109.

4 McNinch AW, Tripp JH. Haemorrhagic disease of the newborn in the British Isles: two year prospective study. *BMJ* 1991;303:1105-9.

5 Chamberlain R, Chamberlain G, Howlett B, Claireaux A. *British births 1970.* Vol 1. *The first week of life.* London: Heinemann Medical, 1975.

(Accepted 18 January 1995)

Statistics Notes

Calculating correlation coefficients with repeated observations: Part 2—correlation between subjects

J Martin Bland, Douglas G Altman

This is the thirteenth in a series of occasional notes on medical statistics

In earlier *Statistics Notes*[1,2] we commented on the analysis of paired data where there is more than one observation per subject. It can be highly misleading to analyse such data by combining repeated observations from several subjects and then calculating the correlation coefficient as if the data were a simple sample.[1] The appropriate analysis depends on the question we wish to answer. If we want to know whether an increase in one variable within the individual is associated with an increase in the other we can calculate the correlation coefficient within subjects.[2] If we want to know whether subjects with high values of one variable also tend to have high values of the other we can use the correlation between the subject means, which we shall describe here.

Means of repeated measurements of intramural pH and $Paco_2$ for eight subjects[3]

Subject	pH	$Paco_2$	Number
1	6·49	4·04	4
2	7·05	5·37	4
3	7·36	4·83	9
4	7·33	5·31	5
5	7·31	4·40	8
6	7·32	4·92	6
7	6·91	6·60	3
8	7·12	4·78	8

The table shows the mean pH and $Paco_2$ for each of eight subjects, with the number of pairs of observations for each. The 47 pairs of measurements from which these means were calculated were given previously.[2] Here we are interested in whether the average pH for a subject is related to the subject's average $Paco_2$.

We can calculate the usual correlation coefficient for the mean pH and mean $Paco_2$. For the data in the table this gives $r=0·09$, $P=0·8$.

This analysis does not take into account the different numbers of measurements on each subject. Whether this matters depends on how different the numbers of observations are and whether the measurements within subjects vary much compared with the means between subjects. We can calculate a weighted correlation coefficient, using the number of observations as weights. Many computer programs will calculate this, but it is not difficult to do by hand.

We denote the mean pH and $Paco_2$ for subject i by $\bar{x}_i$ and $\bar{y}_i$, the number of observations for subject i by m_i, and the number of subjects by n. It is fairly obvious[4] that the weighted mean of the $\bar{x}_i$ is $\Sigma m_i\bar{x}_i/\Sigma m_i$. In the usual case, where there is one observation per subject, the m_i are all one and this formula gives the usual mean $\Sigma\bar{x}_i/n$.

An easy way to calculate the weighted correlation coefficient is to replace each individual observation by its subject mean. Thus the table would yield 47 pairs of observations, the first four of which would each be $pH=6·49$ and $Paco_2=4·04$, and so on. If we use the usual formula for the correlation coefficient on the expanded data we will get the weighted correlation coefficient. However, we must be careful when it comes to the P value. We have only 8 observations (n in general), not 47. We should ignore any P value printed by our computer program, and use a statistical table instead.

The actual formula for a weighted correlation coefficient is:

$$\frac{\Sigma m_i\bar{x}_i\bar{y}_i-\Sigma m_i\bar{x}_i\Sigma m_i\bar{y}_i/\Sigma m_i}{\sqrt{(\Sigma m_i\bar{x}_i^2-(\Sigma m_i\bar{x}_i)^2/\Sigma m_i)(\Sigma m_i\bar{y}_i-(\Sigma m_i\bar{y}_i)^2/\Sigma m_i)}}$$

where all summations are from $i=1$ to n. When all the m_i are equal they cancel out, giving the usual formula for a correlation coefficient.

For the data in the table the weighted correlation coefficient is $r=0·08$, $P=0·9$. There is no evidence that subjects with a high pH also have a high $Paco_2$. However, as we have already shown,[2] within the subject a rise in pH was associated with a fall in $Paco_2$.

1 Bland JM, Altman DG. Correlation, regression and repeated data. *BMJ* 1994;308:896.

2 Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 1—correlation within subjects. *BMJ* 1995;310:446.

3 Boyd O, Mackay CJ, Lamb G, Bland JM, Grounds RM, Bennett ED. Comparison of clinical information gained from routine blood-gas analysis and from gastric tonometry for intramural pH. *Lancet* 1993;341:142-6.

4 Armitage P, Berry G. *Statistical methods in medical research.* 3rd ed. Oxford: Blackwell, 1994:215.

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
J Martin Bland, *reader in medical statistics*

Medical Statistics Laboratory, Imperial Cancer Research Fund, PO Box 123, London WC2A 3PX
Douglas G Altman, *head*

Correspondence to: Dr Bland.

BMJ 1995;310:633