diagnosis of hypovitaminosis D osteopathy. An alternative approach in clinically suspected disease would be to store serum samples and to evaluate clinical improvement after a test dose of 10 000 IU of vitamin D plus extra calcium. Apart from being cheaper and more practical, this approach may even prove to be the best treatment.

We conclude that hypovitaminosis D osteopathy in immigrant women may remain undiagnosed for an unacceptably long time. A higher degree of suspicion in this patient group should shorten the delay and may reduce the number of undue investigations.

1 Parfitt AM. Osteomalacia and related disorders. In: Avioli LV, Krane SM, eds. *Metabolic bone disease.* Philadelphia: Saunders, 1990:329-60.
2 Preece MA, McIntosh WB, Tomlinson S, Ford JA, Dunnigan MG, O'Riordan JLH. Vitamin D deficiency among Asian immigrants to Britain. *Lancet* 1973;i:907-10.
3 Peach H, Compston JE, Vedi S. Value of the history in diagnosis of histological osteomalacia among Asians presenting to the NHS. *Lancet* 1983;ii:1347-9.
4 Peach H, Compston JE, Vedi S, Horton LWL. Value of plasma calcium, phosphate, and alkaline phosphatase measurements in the diagnosis of histological osteomalacia. *J Clin Pathol* 1982;35:625-30.
5 Krall EA, Sayhoun N, Tannenbaum S, Dallal GE, Dawson-Hughes B. Effect of vitamin D intake on seasonal variations in parathyroid hormone secretion in postmenopausal women. *N Engl J Med* 1989;321:1777-83.
6 Thallasinos NC, Wicht S, Joplin GF. Secondary hyperparathyroidism in osteomalacia. *BMJ* 1970;i:76-9.

Statistics Notes

Presentation of numerical data

Douglas G Altman, J Martin Bland

The purpose of a scientific paper is to communicate, and within the paper this applies especially to the presentation of data.

Continuous data, such as serum cholesterol concentration or triceps skinfold thickness, can be summarised numerically either in the text or in tables or plotted in a graph. When numbers are given there is the problem of how precisely to specify them. As far as possible the numerical precision used should be consistent throughout a paper and especially within a table. In general, summary statistics such as means should not be given to more than one extra decimal place over the raw data. The same usually applies to measures of variability or uncertainty such as the standard deviation or standard error, though greater precision may be warranted for these quantities as they are often used in further calculations. Similar comments apply to the results of regression analyses, where spurious precision should be avoided. For example, the regression equation[1]

birth weight= $-3\cdot0983527+0\cdot142088\times$ chest circumf $+0\cdot158039\times$ midarm circumf,

purports to predict birth weight to 1/1 000 000 g.

Categorical data, such as disease group or presence or absence of symptoms, can be summarised as frequencies and percentages. It can be confusing to give percentages alone, as the denominator may be unclear. Also, giving frequencies allows percentages to be given as integers, such as 22%, rather than more precisely. Percentages to one decimal place may sometimes be reasonable, but not in small samples; greater precision is unwarranted. Such data rarely need to be shown graphically.

Test statistics, such as values of t or χ^2, and correlation coefficients should be given to no more than two decimal places. Confidence intervals are better presented as, say, "12·4 to 52·9" because the format "12·4-52·9" is confusing when one or both numbers are negative. P values should be given to one or two significant figures. P values are always greater than zero. Because computer output is often to a fixed number of decimal places P=0·0000 really means P<0·00005—such values should be converted to P<0·0001. P values always used to be quoted as P<0·05, P<0·01, and so on because results were compared with tabulated values of statistical distributions. Now that most P values are produced by computer they should be given more exactly, even for non-significant results—for example, P=0·2. Values such as P=0·0027 can be rounded up to P=0·003, but

not in general to P<0·01 or P<0·05. In particular, the use of P<0·05 (or, even worse, P=NS) may conceal important information: there is minimal difference between P=0·06 and P=0·04. In tables, however, it may be necessary to use symbols to denote degrees of significance; a common system is to use *, **, and *** to mean P<0·05, 0·01, and 0·001 respectively. Mosteller gives a more extensive discussion of numerical presentation.[2]
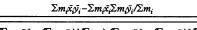
The choice between using a table or figure is not easy, nor is it easy to offer much general guidance. Tables are suitable for displaying information about a large number of variables at once, and graphs are good for showing multiple observations on individuals or groups, but between these cases lie a wide range of situations where the best format is not obvious. One point to consider when contemplating using a figure is the amount of numerical information contained. A figure that displays only two means with their standard errors or confidence intervals is a waste of space as a figure; either more information should be added, such as the raw data (a really useful feature of a figure), or the summary values should be put in the text.

In tables information about different variables or quantities is easier to assimilate if the columns (rather than the rows) contain like information, such as means or standard deviations. Interpretation of tables showing data for individuals (or perhaps for many groups) is aided by having the data ordered by one of the variables—for example, by the baseline value of the measurement of interest or by some important prognostic characteristic.

1 Bhargava SK, Ramji S, Kumar A, Mohan MAN, Marwah J, Sachdev HPS. Mid-arm and chest circumferences at birth as predictors of low birth weight and neonatal mortality in the community. *BMJ* 1985;291:1617-9.
2 Mosteller F. Writing about numbers. In: Bailar JC, Mosteller F. eds. *Medical uses of statistics.* 2nd ed. Boston: NEJM Books, 1992:375-89.

Correction

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

An authors' error occurred in this Statistics Note (11 March 1995, p 633). The formula for a weighted correlation coefficient had a superior 2 missing in the denominator. The full formula should have read:

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

IRCF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, PO Box 777, Oxford OX3 7LF
Douglas G Altman, *head*

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

Correspondence to:
Mr Altman.

BMJ 1996;312:572