Finnish Centre for Radiation and Nuclear Safety and Finnish Cancer Registry, Helsinki, Finland
Anssi Auvinen, *senior scientist*
Timo Hakulinen, *professor*
Kristina Servomaa, *laboratory head*
Tapio Rytömaa, *research director*

Extramural Programs Branch and Radiation Epidemiology Branch, National Cancer Institute, Bethesda, MD, USA
G Iris Obrams, *chief*
John D Boice Jr, *chief*

Correspondence to:
Dr Bigbee.

matched for age. The data, summarised in table 1, show a tendency for increased frequency of variants with loss of the glycophorin A allele in all three groups of workers compared with the controls, although none of the differences is significant. The pooled results, with an average increase of frequency of $\sim 1\text{-}2\times10^{-6}$ corresponding to radiation doses of ~4-8 cGy,[4] indicate that these workers' average exposure was unlikely to greatly exceed 10-20 cGy, the approximate minimum radiation dose detectable by our assay.

Comment

We undertook this biodosimetry study to ascertain whether many Chernobyl cleanup workers received substantial radiation exposures that were either undocumented or inaccurately recorded. Our initial biodosimetry data strongly suggest that this is unlikely. It also seems that there is not a large subset of these workers who received doses substantially above the average physical doses. Thus the estimates of physical doses, while perhaps incomplete and imprecise, cannot be rejected as inadequately characterising the workers' exposures. Our results support the use of these estimates to assess possible health hazards and as the basis of power calculations for epidemiological studies of populations of Chernobyl cleanup workers. To strengthen this conclusion, we are now performing fluorescence in situ hybridisation (FISH) based chromosomal translocation analysis in the peripheral blood lymphocytes of these workers. We are also studying the incidence of leukaemia and prevalence of thyroid cancer and are constructing assessments of exposure using combined physical dosimetry records, extensive questionnaire data, and biological dosimetry methods.

Funding: This work was supported by Contract No N01-CP-50520 from the Radiation Epidemiology Branch of the US National Cancer Institute.

Conflict of interest: None.

1 United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). *Sources, effects and risk of ionizing radiation. 1988 Report to the general assembly, with annexes (E.88.IX.7)*. New York, NY: United Nations, 1988.
2 Jensen RH, Bigbee WL. Direct immunofluorescence labeling provides an improved method for the glycophorin A somatic cell mutation assay. *Cytometry* (in press).
3 Langlois RG, Akiyama M, Kusunoki Y, Bigbee WL, Grant SG, DuPont BR, *et al*. Analysis of somatic mutations at the glycophorin A locus in atomic bomb survivors: a comparative study of assay methods. *Radiat Res* 1993;**136**:111-7.
4 Jensen RH, Langlois RG, Bigbee WL, Grant SG, Moore D II, Pilinskaya M, *et al*. Elevated frequency of glycophorin A mutations in erythrocytes from Chernobyl accident victims. *Radiat Res* 1995;**141**:129-35.
5 Straume T, Langlois RG, Lucas J, Jensen RH, Bigbee WL, Ramalho AT, *et al*. Novel biodosimetry methods applied to victims of the Goiânia accident. *Health Phys* 1991;**60**:71-6.

(Accepted 9 April 1996)

Statistics Notes

Transformations, means, and confidence intervals

J Martin Bland, Douglas G Altman

This is the 18th in a series of occasional notes on medical statistics

When we use transformed data in analyses,[1] this affects the final estimates that we obtain. Figure 1 shows some serum triglyceride measurements, which have a skewed distribution. A logarithmic transformation is often useful for data which have positive skewness like this, and here the approximation to a normal distribution is greatly improved. For the untransformed data the mean is 0.51 mmol/l and the standard deviation 0.22 mmol/l. The mean of the $\log_{10}$ transformed data is –0.33 and the standard deviation is 0.17. If we take the mean on the transformed scale and back transform by taking the antilog, we get $10^{-0.33}$=0.47 mmol/l. We call the value estimated in this way the geometric mean. The geometric mean will be less than the mean of the raw data.

When triglyceride is measured in mmol/l the log of a single observation is the log of a measurement in mmol/l. The average of n such transformed measurements is also the log of a number in mmol/l, so the antilog is back in the original units, mmol/l.

The antilog of the standard deviation, however, is not measured in mmol/l. Calculation of the standard deviation of the log transformed data requires taking the difference between each log observation and the log geometric mean. The difference between the log of two numbers is the log of their ratio.[2] As a ratio is a dimensionless pure number, the units in which serum triglyceride was measured would not matter; the standard deviation on the log scale would be the same. As a result, we cannot transform the standard deviation back to the original scale.

If we want to use the standard deviation or standard error it is easiest to do all calculations on the transformed scale and transform back, if necessary, at the end. For example, the 95% confidence interval for the mean on the log scale is –0.35 to –0.31. To get back to the original scale we antilog the confidence limits on the log scale to give a 95% confidence interval for the geometric mean on the natural scale (0.47) of 0.45 to 0.49 mmol/l. For comparison, the 95% confidence interval for the arithmetic mean using the raw, untransformed data is 0.48 to 0.54 mmol/l. These limits are wider than those for the geometric mean. This is because with highly skewed data the extreme observations have a large influence on the arithmetic mean, making it more prone to sampling error. Lessening this influence is one advantage of using transformed data.

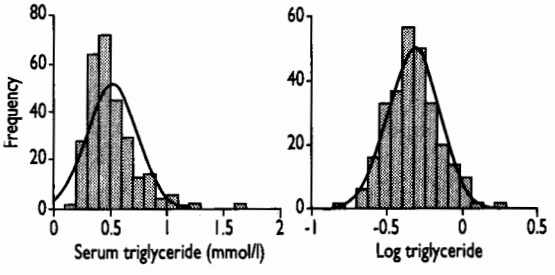

Fig 1—*Serum triglyceride and $\log_{10}$ serum triglyceride concentrations in cord blood for 282 babies, with best fitting normal distribution*

If we use another transformation, such as the reciprocal or the square root,[1] the same principle applies. We carry out all calculations on the transformed scale and transform back once we have calculated the confidence interval. This works for the sample mean and its confidence interval. Things become more complicated if we look at the difference between two means. We shall look at this in another Statistics Note.

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

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

Correspondence to:
Professor Bland.

BMJ 1996;**312**:1079

1 Bland JM, Altman DG. Transforming data. *BMJ* 1996;**312**:770.
2 Bland JM, Altman DG. Logarithms. *BMJ* 1996;**312**:700.