Detecting skewness from summary information

Douglas G Altman, J Martin Bland

As we have noted before, many statistical methods of analysis assume that the data have a normal distribution.¹ When the data do not they can often be transformed to make them more normal.² Readers of published papers may wish to be reassured that the authors have carried out an appropriate analysis. When authors present data in the form of a histogram or scatter diagram then readers can see at a glance whether the distributional assumption is met. If, however, only summary statistics are presented—as is often the case—this is much more difficult. If the summary statistics include the range of the data then some idea of the distribution may be gained. For example, a range from 7 to 41 around a mean of 15 suggests that the data have positive skewness. However, as the range is based on the two most extreme (and hence atypical) values this inference is not reliable. Similar asymmetry affecting the lower and upper quartiles³ would be much more convincing evidence of a skewed distribution. Usually, however, the only summary statistics presented are the mean and either the standard deviation or standard error. Such information cannot show that the data are near to a normal distribution, but they can sometimes show that they are not.

There are two useful tricks. The normal distribution extends beyond two standard deviations either side of the mean. It follows that for measurements which must be positive (like most of those encountered in medicine) if the mean is smaller than twice the standard deviation the data are likely to be skewed. Table 1 shows urinary cotinine levels related to number of cigarettes smoked daily. Clearly the data must be highly skewed, as the mean is smaller than the standard deviation in each group. This aspect of the data was not apparent in the original paper, which gave just the means and standard errors. (We added the standard deviations, derived simply as standard error $\times \sqrt{n}$.) As a conse-

Table 1—Urinary cotinine excretion ($\mu\text{g}/\text{mg}$ creatinine) related to number of cigarettes smoked daily⁴

Cigarettes smoked per day	No in group	Mean	SE	SD
1-9	25	0.31	0.08	0.40
10-19	57	0.42	0.10	0.75
20-29	99	0.87	0.19	1.89
30-39	38	1.03	0.25	1.54
>40	28	1.56	0.57	3.02
Unspecified	25	0.56	0.16	0.80

quence, the use of *t* tests was not easily seen to be incorrect.

The second indicator of skewness can be used when, as in table 1, there are data for several groups of individuals. As we have noted,² deviations from the normal distribution and a relation between the standard deviation and mean across groups often go together. If the standard deviation increases as the mean increases then this is a good indication that the data are positively skewed, and specifically that a log transformation may be needed.² There is a clear relation between mean and standard deviation for the cotinine data. As we have noted, log transformation often removes skewness and makes the standard deviations more similar.

In this example we can detect skewness from summary statistics, but we cannot tell what the effect of log transformation would have been. That requires the raw data.

1 Altman DG, Bland JM. The normal distribution. *BMJ* 1995;310:298.

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

3 Altman DG, Bland JM. Quartiles, quintiles, centiles and other quantiles. *BMJ* 1994;309:996.

4 Matsukura S, Taminato T, Kitano N, Seino Y, Hamada H, Uchihashi M, et al. Effects of environmental tobacco smoke on urinary cotinine excretion in nonsmokers. *N Engl J Med* 1984;311:828-32.

ICRF 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;313:1200

Lesson of the Week

Fetal blood after late abortion in RhD negative women should be RhD typed to ensure appropriate anti-D immunoglobulin dosage

Importance of accurate assessment of fetomaternal haemorrhage after late abortions

Imelda Bromilow, J K M Duguid, S Walkinshaw

Anti-D immunoglobulin is routinely administered to RhD negative women after an abortion. Before 20 weeks of gestation 25 IU of anti-D immunoglobulin should be given and after 20 weeks of gestation 500 IU. The Kleihauer stain test should be routinely performed to assess the adequacy of this dose.¹ RhD typing of the aborted fetus is not usually performed.

We report a case in which large doses of anti-D immunoglobulin were given inappropriately because the Kleihauer test detects only fetal haemoglobin; further testing with flow cytometry showed that the abortus was RhD negative.

Case report

A pregnant woman underwent a late termination of pregnancy for a lethal fetal malformation after 20 weeks of gestation. The procedure was carried out using mifepristone pessaries. There were no clinical problems: maternal blood loss was not excessive and the placenta was delivered normally. Anti-D immunoglobulin (500 IU) was given, and Kleihauer testing indicated that a fetal bleeding of about 100 ml had occurred. Additional anti-D immunoglobulin was given to a total dose of 10 500 U. Further Kleihauer testing 24 hours later showed that fetal cells were still evident in the maternal circulation.

Correspondence to: Dr Duguid.

BMJ 1996;313:1200-1

**National Blood Service,
Mersey and North Wales
Centre, Liverpool L7 8TW**
Imelda Bromilow, *clinical
scientist*
J K M Duguid, *consultant
haematologist*

**Liverpool Women's
Hospital, Liverpool
L8 7SS**
S Walkinshaw, *consultant
haematologist*

Flow cytometry was then performed to assess more accurately the size of the fetomaternal haemorrhage. Flow cytometry used in this way detects RhD positive cells in a RhD negative circulation and gives a more objective assessment than the Kleihauer test for calculating the number of fetal cells after fetomaternal haemorrhage.² Results for the original sample showed that there were no RhD positive cells in the maternal circulation, strongly suggesting that the fetus was RhD negative (fig 1).

Discussion

This case highlights several points. Babies born to RhD negative mothers are RhD typed. If the baby is RhD negative then anti-D immunoglobulin is not required. RhD typing of a fetus (abortus) after a late (>20 weeks) termination of pregnancy is not, however, routine. After 20 weeks the fetus has a well established

circulation, and large fetomaternal bleeds well in excess of 4 ml can occur. Monitoring of the size of fetomaternal haemorrhage should therefore be routinely undertaken using a Kleihauer test to ensure that an adequate dose of anti-D immunoglobulin has been administered. If this indicates that a large (>4 ml) fetomaternal haemorrhage has occurred the maternal sample should be tested by flow cytometry. This technique measures the number of RhD positive cells among RhD negative cells and is therefore an alternative way of RhD typing fetal blood after a fetomaternal haemorrhage.

Anti-D immunoglobulin specifically eliminates RhD positive fetal cells; it has no effect on RhD negative fetal cells. A Kleihauer result may remain strongly positive despite administration of large amounts of anti-D immunoglobulin because it detects fetal haemoglobin. This is what happened in our case. It shows the importance of RhD typing the fetus of a late termination of pregnancy, and if this cannot be done, the importance of the use of flow cytometry in conjunction with Kleihauer testing in cases with evidence of a clinically important volume of fetomaternal haemorrhage.

If this regimen were to be routinely adopted the unnecessary use of human anti-D immunoglobulin would be reduced, resulting in a more effective use of all resources and an improvement in the quality of patient care.

Funding: None.

Conflict of interest: None.

1 National Blood Transfusion Service Immunoglobulin Working Party. Recommendations for the use of anti-D immunoglobulin. *Prescribers' Journal* 1991; 31:137-45.

2 Duguid JKM, Bromilow IM. Value of Kleihauer testing after administration of anti-D Immunoglobulin. *BMJ* 1994; 309:240.

(Accepted 10 September 1996)

Fig 1—Results of flow cytometry using anti-D fluorescein isothiocyanate. There are no RhD positive cells in maternal sample

WHEN I USE A WORD . . .

Ae, ae, ae!

A young man came in 12 hours after taking 100 paracetamol tablets. We gave him acetylcysteine. He had no major symptoms or signs, but there were a few petechiae over the shoulders and his platelet count was $9 \times 10^9/l$. I searched Medline: thrombocytopenia in paracetamol overdose does occur and is a poor prognostic sign (*Am J Hematol* 1994;45:258-9). Of course, I didn't look for just <paracetamol> AND <thrombocytopenia>, but for <paracetamol OR acetaminophen> AND <thrombocytopenia OR thrombocytopaenia>. It is always important, when searching, to remember that drugs may have alternative names—but how often do you need to remember that a word may be misspelled?

Of 5169 papers on thrombocytopenia published in 1990-5, 55 (1.1%) used the wrong spelling -paenia; the figure for the previous 24 years was 155 (0.8%) of 10 445 papers ($\chi^2 = 3.33$; $P=0.07$). So, authors and editors are getting it wrong, and perhaps more often now than before. In contrast, the incorrect spelling "gentamycin" occurred in 10.5% of 7893 papers in 1966-89 and in only 8.7% of 2618 papers in 1990-5 ($P=0.006$); things may be bad for gentamicin, but my two minutes' worth of computerised research suggests that they are at least getting better.

Spelling wasn't always standardised. Old English spelling was consistent, but the conjunction of Anglo-Saxon and Norman-French produced inconsistencies that lasted until the printing press and dictionaries gradually forced regularity, if not always rationality. Samuel Johnson,

in his dictionary of 1755, used the etymologically incorrect variant -our for many words that should have ended -or; in some cases this has survived (colour, honour), in others not (errour, inferiour). When Noah Webster published his first dictionary in 1806 he introduced some spellings of his own, partly to return to the original etymology but also because he wanted to make American orthography distinct from English—for example, he removed the *u* from -our (color, honor) and (against the etymology) changed -re to -er (center, theater).

But Webster did not institute the change to *e* in words that are spelled with *ae* in English. The digraph *ae* was originally a representation of the Old English letter *aesc* (pronounced "ash"); it disappeared in the 13th century, but re-emerged in the 16th century to represent the Greek *ai* or the Latin *ae*. It then persisted in Greek and Latin names and technical vocabulary, but tended to disappear in words that entered general use ([a]ether, d[a]emon, pr[a]eface, for example). This trend happened more readily in the United States. In Webster's *New International Dictionary* (1911) anaemia, for example, is the main form of the headword, with anemia a secondary variant, but in the *New Third International* (1961) the reverse is the case.

So when you see "thrombocytopaenia" you know that the writer wrongly thinks that -penia (Greek penia, poverty) is the Americanised form of -paenia. To put it another way, the spelling is spaccious.—JEFF ARONSON is a clinical pharmacologist in Oxford