International Journal of Epidemiology
© International Epidemiological Association 1988
Vol. 17, No. 2
Printed in Great Britain

LEADING ARTICLE

Misleading Statistics: Errors in Textbooks, Software and Manuals

J MARTIN BLAND* AND DOUGLAS G ALTMAN†

'The trouble of people is not that they don't know, but that they know so much that ain't so.'[1]

Statistics is an essential tool in medicine in general and in medical research in particular. This is shown by the large number of articles on statistical topics carried by medical journals, by the demand for consultancy from medical statisticians and by the number of textbooks in medical statistics, especially introductory texts, now on the market. Yet, despite this importance and wealth of advice, the statistical component of much medical research is very badly done.[2-4]

Statistics is a difficult subject. Few people, including statisticians, find it easy at first acquaintance. Modern medical education usually contains only a short introduction to statistics, and very little about scientific thinking and research methods in general. Thus the medical researcher has at best acquired only a limited knowledge of statistical methods from formal education. There are few postgraduate courses available and many clinicians do not have the time to attend them anyway. Researchers are reduced to learning the subject from books or to relying on the statistical software to do their thinking for them.

Readers of the *IJE* are likely to be asked for their statistical help and advice by other medical researchers, or asked to recommend suitable books and microcomputer software. Unfortunately, many texts in statistics aimed at doctors are bad books, filled with errors, and conveying no understanding on which to build.[5] Statistical software, too, may hold traps for the unwary and the associated manuals may be highly misleading.

STATISTICAL TEXTBOOKS

Many introductory textbooks in medical statistics contain fundamental conceptual errors. For example, in explaining significance tests, it is often wrongly stated that the P value is the probability that the null hypothesis is true. There are incorrect ideas which seem to be passed on from book to book without any justification, such as that the standard deviation should only be calculated from data which come from a Normal Distribution. Inappropriate statistical methods may be used, such as two sample t tests for paired data or contingency chi-squared tests for examining trend in ordered contingency tables.[6]

Because authors do not have much experience of the analysis of data outside their own work, they often do not have real data to illustrate statistical methods. They therefore invent them, resulting in completely unrealistic examples.

The fundamental importance of good research design is also absent from many introductory textbooks. What are we to make of a book which includes a chapter on clinical trials, but does not mention random allocation anywhere?

Sometimes authors appear to be making it up as they go along. In a recent book, the author applied both linear regression and correlation techniques to some fictitious looking data relating blood pressure to body weight in newborn babies. These were used to test the null hypotheses that the slope of the regression line and the correlation coefficient were each zero. These two tests are identical and should give identical P values. The author found a much larger P value, and hence a much less significant result, for the slope of the regression line than for the correlation coefficient. This was explained by saying that regression provides a less sensitive test than does correlation! In fact the author had made an arithmetical error in the calculations and then used a spurious argument to explain the discrepancy.

As methods and formula are often given without justification, it is extremely difficult for the reader to spot errors. We know three different books which give incorrectly the formula for the chi-squared test for a two-by-two table. The Catch 22 is that it is unlikely that

* Department of Clinical Epidemiology and Social Medicine, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK.
† Medical Statistics Laboratory, Imperial Cancer Research Fund, PO Box No. 123, Lincoln's Inn Fields, London WC2A 3PX, UK.

anyone with the knowledge to detect such errors would be reading the book.

SOFTWARE MANUALS

We have often come across people trying to learn statistics from that magnificent tome, the SPSS manual. Computing and statistics are so linked that researchers often use statistical software and associated manuals as important sources of information. The proliferation of micros has led to the development of many simple statistical packages and these are widely used. People tend to like to use an untried program on the friendly little machine they know rather than tried and tested packages on mainframes. This can soon lead them into the mire.

A colleague was recently misled by chi-squared tests calculated on his Macintosh. The program offered procedures 'Chi-square (one group)' and 'Contingency table'. Wanting chi-squared tests for two-by-two tables he erroneously picked 'Chi-square (one group)' and got a lot of highly significant differences where there were none. This illustrates the unfriendliness of the program, but the true horror was to be seen in the manual. This gave an example for 'Chi-square (one group)' in which observed mean cholesterol levels for men and women of 171.1 and 180.7 were compared to expected means of 190 and 180. Why these should be expected was not stated. The program was then used to compare the observed and 'expected' means as if they were frequencies! To carry out this completely spurious calculation, the user has to choose 'Contingency table' from the next menu!

While some of the better manuals contain a lot of information, they are not a very suitable source for learning about statistics. The main function of a program manual is to help the user to get the program to do what is required. It is unlikely that a manual could both do this and teach statistics at the same time.

SOFTWARE

The accuracy of some simple computer programs is also questionable, especially where statistical functions have been added to existing, non-statistical programs. Some spreadsheet and word processing packages now incorporate statistical calculations. We know two major programs which calculate the standard deviation using n rather than $n-1$ as the divisor of the sum of squares. Even cheap pocket calculators give the choice of σ_n and σ_{n-1}. Some programs also use the inaccurate formula $\Sigma x_i^2 - (\Sigma x_i)^2/n$ for the sum of squares about the mean rather than the more accurate $\Sigma(x_i - \bar{x})^2$. The former is acceptable when using a calculator, but not as an algorithm for a computer program, where the potential errors in subtracting one large number from another are hidden.

Some of the well established programs can be misleading at times, too. In the default for t tests, one program uses unpooled variances with fractional degrees of freedom, which is an approximation, rather than the correct pooled variance form. The latter must be obtained using a subcommand.

AUTHORSHIP

It is a curious phenomenon of statistics, which we have not encountered elsewhere, that the authors of introductory texts in the subject often, indeed predominantly, are neither qualified nor expert in the subject. They are not statisticians. Furthermore, this is presented as a positive advantage. The author is claimed to be able to communicate the subject to the medical audience not because he or she is an expert in statistics but because of a common medical background. Statistical software, too, is often not written by statisticians, but by computer programmers or statistics users. Programmers know how to make the computer perform calculations in the fastest, most user-friendly way, but they do not necessarily know which calculations should be done. Too often, the result is misleading and inaccurate statistical programs which the typical user has no means, or notion, of checking. After all, if it comes out of a computer, it must be true.

Now we are not trying to set up a statisticians' trade union and draw up lines of demarcation. It does not matter who writes a textbook, manual or program if it is correct. Nor do we wish to imply that all statistical works written by non-specialists are bad. Some are excellent, but too many are full of errors. We must also admit that an author with statistical training does not guarantee a good book. However, we do not think that any statistician could make the sort of errors referred to above. A statistician might write an incomprehensible book about statistics, or a cumbersome program, but would be unlikely to write a dangerous one.

PUBLICATION AND REVIEW

Any textbook in statistics should be refereed before publication by an experienced statistician, especially if the authors are not themselves expert in the field. Despite the moral responsibility of publishers to try to publish books which are substantially correct, it seems clear that many publishers of introductory statistics books do not have them refereed by anyone competent in statistics. Statistical software and manuals, too, should be assessed by a statistician, not only to test the accuracy of calculations, but also their appropriateness.

Books are published, full of statistical errors, and sent for review to journals in the health and related fields. There they are reviewed, often by a member of the speciality of the journal, not by a statistician. These reviewers are unlikely to spot the errors and misunderstandings with which the books abound, because they themselves do not have the knowledge of statistics required. They are thus judging the book on whether it is readable, attractively laid out, short and cheap, not whether it is a good book from which to learn statistics. We examined eight reviews of one of the elementary medical statistics texts mentioned above. All four written by statisticians were highly critical, rightly so as the book abounds with errors, but the four by non-statisticians praised the book as useful and readable. We suggest that medical journals should publish more reviews of statistical text books and that all such books should be reviewed by statisticians. Double reviews by a statistician and a clinician could be especially revealing.

Statistical software is reviewed in medical journals in much the same way. Comments relate to the ease of use, compatibility with other software and facilities offered. Only in a statistical journal are the accuracy and correctness of the program likely to be assessed.

Bad textbooks and bad software are dangerous. Bad statistics makes bad research, bad research may lead to bad medicine, and bad medicine may cost lives. If, as a result of errors in books, manuals and programs, bad research is done, it is at best a waste of effort and at worst a hazard to patients.

Medical researchers are often advised to consult a statistician before embarking on a study. Clearly, we must also advise that they consult a statistician before buying a book or computer program! Furthermore, medical statisticians and epidemiologists need to read a book very carefully before recommending it.

REFERENCES

[1] Josh Billings' *Encyclopaedia of wit and wisdom*, 1874.

[2] Glantz S A. Biostatistics: how to detect, correct and prevent errors in the medical literature. *Circulation* 1980; **61**: 1–7.

[3] Tyson J E, Furzan J A, Reisch J S, Mize S G. An evaluation of the quality of therapeutic studies in perinatal medicine. *J Pediatr* 1983; **102**: 10–13.

[4] Avram M J, Shanks C A, Dykes M H M, Ronai A K, Stiers W M. Statistical methods in anesthesia articles: an evaluation of two American journals during two six-month periods. *Anesth Analg* 1985; **64**: 607–611.

[5] Bland J M, Altman D G. Caveat doctor: a grim tale of medical statistics textbooks. *Brit Med J* 1987; **295**: 979.

[6] Day S J. Errors in text books in analysing contingency tables. In Press, 1988.