DRIP-FEED EDUCATION: STATISTICS NOTES IN THE BRITISH MEDICAL JOURNAL

J Martin Bland[1] and Douglas G Altman[2]
[1]Department of Health Sciences, University of York, United Kingdom
[2]Centre for Statistics in Medicine, University of Oxford, United Kingdom
mb55@york.ac.uk

Evidence-based medicine depends on decision makers, such as clinicians, being able to understand evidence. Statistical understanding is essential for this. Although most clinicians include some statistics in their training, it is often forgotten or confused by the time they want to apply it in practice. Statistics Notes, published in the British Medical Journal*, help to overcome this. In this paper we discuss how we came to write these short pieces, what they are like, and how they have been received and cited. We discuss the value of such series, both to medicine and to ourselves, and how the process is repeated in other journals. We recommend that similar series be undertaken in journals in other fields and suggest that laboratory research using animals would be a good focus for other authors to try.*

UNDERSTANDING EVIDENCE

The practice of evidence-based medicine, or evidence-based anything, depends on decision makers, such as clinicians, being able to understand evidence. They need to understand the subject area of evidence, including the medical specialty, the nature of a disease, the drugs or procedures being evaluated, and the outcomes that were assessed. They also need to understand the research methodology, for example the key methodological aspects of a clinical trial or a case control study. To do this, decision makers need to understand something of the disciplines involved in the conduct and reporting of these studies. These may include economics or psychology, but above all they includes statistics. For example, in the *British Medical Journal* and the *Lancet* over January 2010, 34 out of 38 research reports included statistical terms in the summary, terms such as confidence interval, P value, standardized mean difference, and hazard ratio. The four papers that did not include statistical terms were a historical case report, two economic evaluations, and a qualitative study.

Statistical understanding is important to all practicing clinicians, and especially so to those who themselves wish to add to the body of evidence by carrying out research. Although most clinicians, physicians, nurses, etc., received some statistics in their training, it is often forgotten or confused by the time they want to apply it in practice. They need another course. Yet they are very busy people, with many calls on their time. Nobody, apart from a few statisticians, sits down to read a statistics book, no matter how brilliantly written. A more feasible approach for these people is to take very small doses of statistics that can be absorbed in a few minutes. This is where Statistics Notes comes in.

STATISTICS NOTES BEGIN

In 1993, Doug Altman was asked by the *BMJ* to write a series of short pieces, around 600 words, to fill up spare parts of pages in the journal. Fillers, in other words. As they would not be peer reviewed, Doug thought that he should have a co-author, so that together they would have some chance of avoiding serious mistakes, at least. As we were good friends, had already written some successful papers, and had really enjoyed writing them together, he suggested me. He also wrote down an extremely long list of potential topics. So we planned a few notes and set to drafting them.

The first Statistics Note was published in 1994, entitled "Correlation, regression and repeated data" (Bland & Altman, 1994). The authors were Bland and Altman, because the first draft of this had been written by JMB. It was an issue in which he was particularly interested, and still is. We then each revised the text and went over it together.

The note began: "In clinical research we are often able to take several measurements on the same patient. The correct analysis of such data is more complex than if each patient were measured once. This is because the variability of measurements made on different subjects is usually much

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

greater than the variability between measurements on the same subject, and we must take both kinds of variability into account. Researchers sometimes put all the data together, as if they were one sample. Most statistics textbooks do not warn the researcher not to do this. It is so ingrained in statisticians that this is a bad idea that it never occurs to them that anyone would do it."

We then gave a simulation of two variables, *x* and *y*, generated from random numbers with no relation between *x* and *y* at all. Values of *x* and *y* were generated for each 'subject', then a further random number was added to make the individual 'observation'. We gave a table of these data and the graph shown in Figure 1.

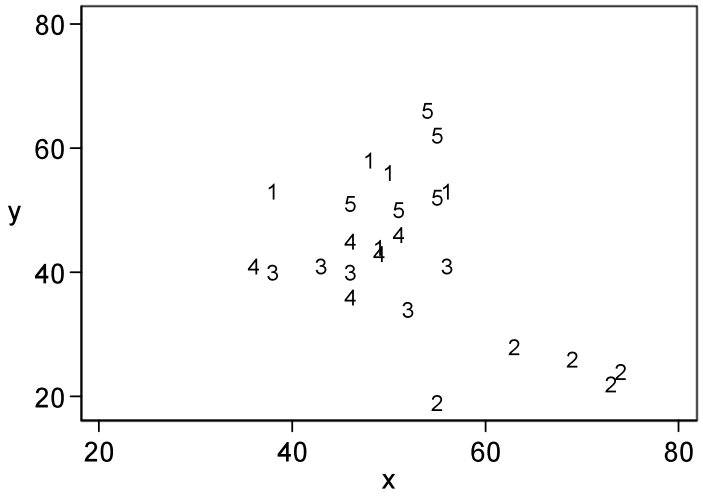

Figure 1. Graph showing simulated data for five pairs of observations on each of five subjects

For each subject separately, the correlation between *x* and *y* is not statistically significant. We have only five subjects and only five points. Using each subject's mean values, the correlation coefficient is $r = -0.67$, d.f. = 3, P = 0.22. However, if we put all 25 observations together we get $r = -0.47$, d.f. = 23, P = 0.02. Even though this correlation coefficient is smaller than that between the means, because it is based on 25 pairs of observations rather than five it becomes significant. The calculation is performed as if we have 25 subjects, and so the number of degrees of freedom for the significance test is increased incorrectly and a spurious significant difference is produced. The extreme case would occur if we had only two subjects, with repeated pairs of observations on each. We would have two separate clusters of points centred at the two subjects' means. We would get a high correlation coefficient, which would appear significant despite there being no relation whatsoever.

This Statistics Note was followed by eight more in 1994: "Regression towards the mean", "Diagnostic tests 1: sensitivity and specificity", "Diagnostic tests 2: predictive values", "Diagnostic tests 3: receiver operating characteristic plots", "One- and two-sided tests of significance", "Some examples of regression towards the mean", "Quartiles, quintiles, centiles, and other quantiles", and "Matching". Most of these were written from scratch, but "One- and two-sided tests of significance" was based on a section of *An Introduction to Medical Statistics* (Bland, 1995), then in draft. The two notes on regression towards the mean addressed a special interest of JMB. The three on diagnostic tests were a linked series from DGA. As of January 2010 we have published 55 Notes. A full list with references and links to the text can be found on martinbland.co.uk.

We published four more notes in 1995, and in 1996 there were 13, our peak year. In 1997 there were only 3 notes published, but in 1998 there were 8. This variability was partly because the *BMJ* waited until there was a need for a filler and partly because we had other things to do. After 1998 the numbers fell and remained low, as Figure 2 shows. We both became much more busy.

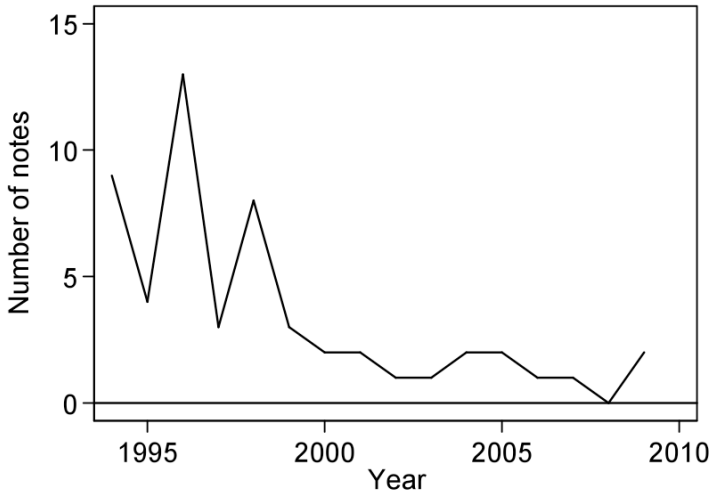

Figure 2. Published Statistics Notes by year

OTHER AUTHORS

The series attracted other statisticians. We decided that to keep editorial control over the series and to preserve some uniformity of style and, we fondly hoped, quality, we would include other authors only if one of us was an author too. The first guest author was John Matthews of the University of Newcastle. In 1996 he wrote three notes with Doug on the subject of interactions. The following year, Sally Kerry submitted an article on the design and analysis of cluster randomized trials to the *BMJ*. They did not want to publish this as it stood, but suggested that she contact me to discuss the possibility of including it in the Statistics Notes series. This was easy to do, as we were in adjoining departments at St. George's Hospital Medical School, and suited me as it was a topic in which I was already interested. Her original article formed the basis for a series of five notes on cluster randomized trials, written with me. Doug and I also wrote a note on units of analysis, which introduced the series. Doug later produced Statistics Notes with three other collaborators: Ken Schulz, Andrew Vickers, and Simon Day.

CITATIONS

To our considerable surprise, the Statistics Notes began to be cited. Some have been very highly cited; in January 2010 the highest was the Note on the Bonferroni correction, with 634 citations. The distribution of the number of citations achieved per note is shown in Figure 3.

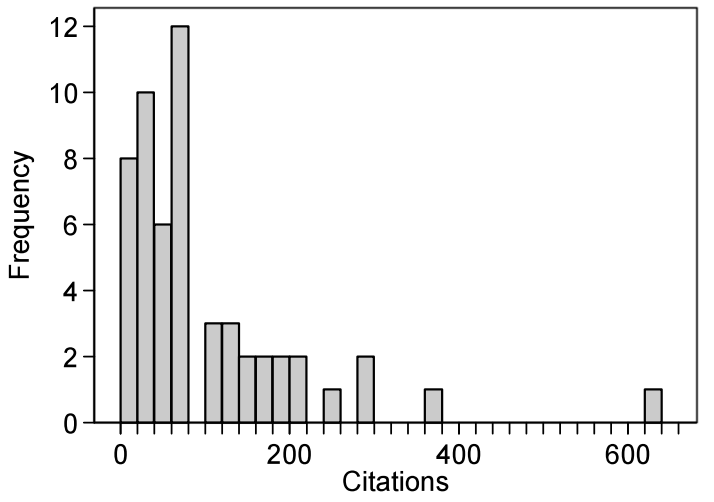

Figure 3. Distribution of citation numbers for Statistics Notes, January 2010

The mean number of citations per note was 98, median 67. I think that this is a very high number of citations for what are short educational pieces containing no original material. If Statistics Notes were an author, the Hirsch h index would be 37, i.e. the 37th most highly cited statistics note has 37 or more citations. Altogether, the 55 published Statistics Notes have more than 5,000 citations.

As would be expected, older notes tend to have more citations than recent ones, as Figure 4 shows.

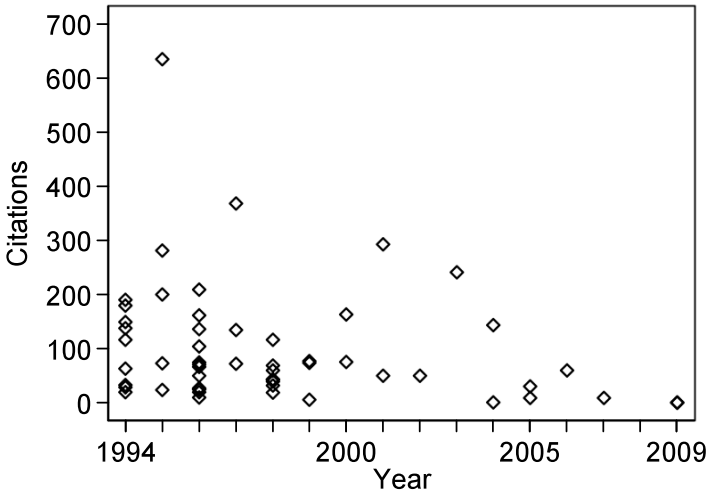

Figure 4. Number of citations of Statistics Notes, by year of publication

The large number of citations for these notes was completely unexpected. It provides a indication of how powerful an educational tool such series can be. Another illustration of this came one evening when JMB was having dinner with two other statisticians. One remarked that the Statistics Notes were a useful series and said that all new postgraduate students in his department were presented with a printed set. The other said that her department did this, too!

The most highly cited notes were on multiple significance tests - the Bonferroni method, Cronbach's alpha, analyzing controlled trials with baseline and follow up measurements, “absence of evidence is not evidence of absence”, interaction revisited: the difference between two estimates, measurement error and correlation coefficients, calculating correlation-coefficients with repeated observations—correlation within-subjects. All of these have been cited at least 200 times.

OTHER NOTES SERIES

The *BMJ* followed the Statistics Notes with other, shorter series. In January 1998 Bonnie Sibbald and Martin Roland began a series on Understanding Controlled Trials with ‘Why are randomized controlled trials important?’ (Sibbald & Roland, 1998). Economics Notes began in March 1998, with James Raftery’s ‘Economic evaluation: an introduction’ (Raftery, 1998).

Other journals have run similar series of statistical articles, and we are not the first. Perhaps the most well-known is that, begun in 1967 by Donald Mainland and continued from 1970 by Alvan Feinstein, in *Clinical Pharmacology & Therapeutics.* This predates us by many years and ran until 1981. There are others going on right now, I come across them occasionally.

Feinstein wrote most of his series, alone, but he did have at least one co-author. His piece with Kramer on “The Biostatistics Of Concordance” (Kramer & Feinstein, 1981) has been cited 722 times. Feinstein collected many of the series into a book, *Clinical Biostatistics* (Feinstein, 1977), published in 1977.

IS IT WORTH IT?

The Statistics Notes series was agreed with the *BMJ* soon after a Research Assessment, a process which is used in the UK to allocate research funds to universities, based on the assessed quality of their research. One of the elements of this Kafkaesque procedure was a count of each

department's publications over the four years covered by the assessment. Soon after this, a message came round JMB's institution telling academic staff that it was important that they kept a careful record of all publications. It might be thought quite fantastic that any academic would not do this, but we are more different from one another than we can easily imagine. Anyway, Statistics Notes would form a handy boost to publication numbers. Then the rules for the next Research Assessment came out, and they were not going to have a publication count any more. Apparently, it was encouraging trivial publishing. They required only four nominated publications from each researcher and academics were explicitly forbidden from disclosing the number of their publications in any way. And we had what looked like the best trivial publishing deal in history! Writing the notes would distract us from less trivial publications, but it seemed such a good idea that we went ahead anyway. Don't let research assessors tell you what to do!

The notes proved popular and have acquired a momentum of their own, together with an enormous number of citations. But do Statistics Notes work, i.e. do they help to improve readers' understanding of statistics and does this help improve the quality of medical research? Of course, we don't know, but we do know that the quality of medical research published in the major journals has improved greatly since we started our careers as medical statisticians. We are convinced that statisticians have had a large part to play in this improvement. We hope that we have helped in this process and our purely subjective belief is that Statistics Notes have been one small part of it

CONCLUSION

There is room for many more Statistics Notes from us and co-authors. We have made little impression on our list of potential topics. There is nothing original about our Statistics Notes, but this is surely the great thing about them. They are snippets of statistics offered in a place where those who might find them useful will see them and they can be repeated in other places for other audiences. Indeed, one of our Statistics Notes was republished verbatim in the *Australian Veterinary Journal* (Altman & Bland, 1996). So if you meet a journal editor who is looking for something to fill up the odd corner, why not offer a series of statistics notes? We would suggest finding a congenial colleague with whom to write them.

There is no shortage of world to conquer, either. Kilkenny et al. (2009) reviewed 271 papers reporting laboratory animal experiments. They reported that most of the papers surveyed did not use randomization or blinding and that almost one third of the papers that used statistical methods did not describe them or present their results adequately. So a few series of statistics notes in journals where those papers were published could do a lot of good.

REFERENCES

Altman, D. G., & Bland J. M. (1996). Absence of evidence is not evidence of absence. *Australian Veterinary Journal, 74*, 311.

Bland, J. M., & Altman, D. G. (1994). Correlation, regression and repeated data. *British Medical Journal 308*, 896.

Bland, M. (1995). *An Introduction to Medical Statistics* (Second Edition). Oxford: Oxford University Press.

Feinstein, A. R. (1977). *Clinical Biostatistics*. St. Louis: Mosby.

Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F. W., Cuthill, I. C., Fry, D., Hutton, J., & Altman, D. G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS ONE 4(11)*, e7824.

Kramer, M. S., & Feinstein, A. R. (1981). Clinical Biostatistics 54. The Biostatistics Of Concordance. *Clinical Pharmacology & Therapeutics, 29*, 111-3.

Raftery J. (1998) Economics notes: Economic evaluation: an introduction. *British Medical Journal, 316*, 1013-4.

Sibbald, B., & Roland, M. (1998). Why are randomised controlled trials important? *British Medical Journal 316*, 201.