Drip-Feed Education: Statistics Notes in the *British Medical Journal*

J Martin Bland and Douglas G Altman

Dept. of Health Sciences
University of York
York YO10 5DD

and

Centre for Statistics in Medicine
University of Oxford
Wolfson College Annexe
Linton Road
Oxford OX2 6UD

mb55@york.ac.uk

Martin Bland was invited to speak in a session on evidence-based medicine at the Eighth International Conference on Teaching Statistics, held in Slovenia in July 2010. He suggested to Doug Altman that they collaborate on a talk on their long-running series of Statistics Notes in the *British Medical Journal*. This is an expanded version of the talk.

Understanding evidence

The practice of evidence-based medicine, or evidence-based anything, depends on decision makers being able to understand evidence. To do this, they need to understand the subject area of the evidence, such as the medical specialty, the nature of a disease being considered, the drugs or procedures being evaluated, and the outcomes that were assessed. They also need to understand the research methodology, such as the key methodological aspects of a clinical trial or a case control study. They need to understand something of the disciplines involved in the conduct and reporting of these studies. Above all, these disciplines include statistics. For example, in the January 2010 issues of the *British Medical Journal* and the *Lancet*, 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. So statistics was of central importance in understanding the evidence reported in these research papers.

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. We doubt that anybody, apart from a few statisticians, sits down to read a statistics book, including ours, no matter how brilliantly written. What people can do 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 *British Medical Journal*, which we shall from here on write as *BMJ* for short, to write a series of short pieces, around 600 words, to fill up spare parts of pages in the journal. Fillers, in other words. Some of these would

expand on a method used in a paper in the same issue, some would stand alone. 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, most important, had really enjoyed writing them together, he suggested Martin. He also wrote down an extremely long list of potential topics. So we planned a few notes and set to drafting them. We then each revised the text and went over it together.

The first Statistics Note was published in 1994, entitled "Correlation, regression and repeated data". [1] The authors were Bland and Altman, because the first draft of this had been written by Martin. It was an issue in which he was particularly interested, and still is.

To give an idea of what these Statistics Notes are like, we shall quote some of it. 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 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." Dr. Fisher held forth on the same subject in a recent issue; [2] of course, we agree with her or him.

We then gave a simulation of two variables, denoted with startling originality as 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.

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 analysis is performed as if we have 25 independent observations, 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 could get a high correlation coefficient, which would appear significant despite there being no relation whatsoever.
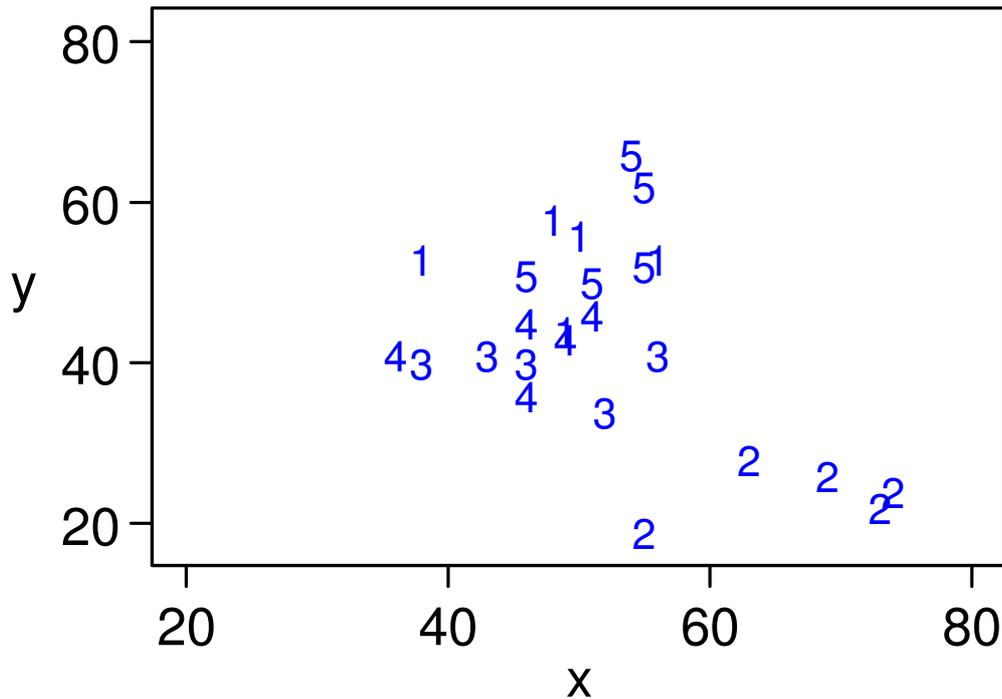
Figure 1. Graph showing simulated data for five pairs of observations on each of five subjects

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 Martin's *An Introduction to Medical Statistics, 2nd edition* [3], then in draft. The two notes on regression towards the mean addressed a special interest of Martin. The three on diagnostic tests were a linked series from Doug, based on text from his then recent book.. [4]

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 but more because we had other things to do. After 1998 we both became much more busy and the numbers fell and remained low, as Figure 2 shows. 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.
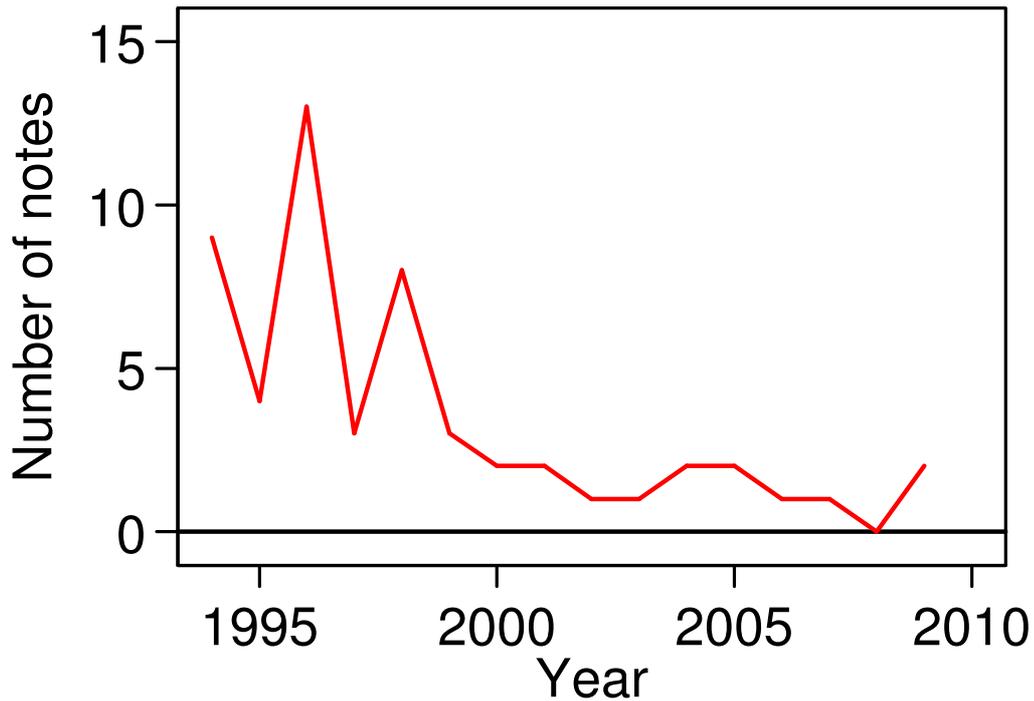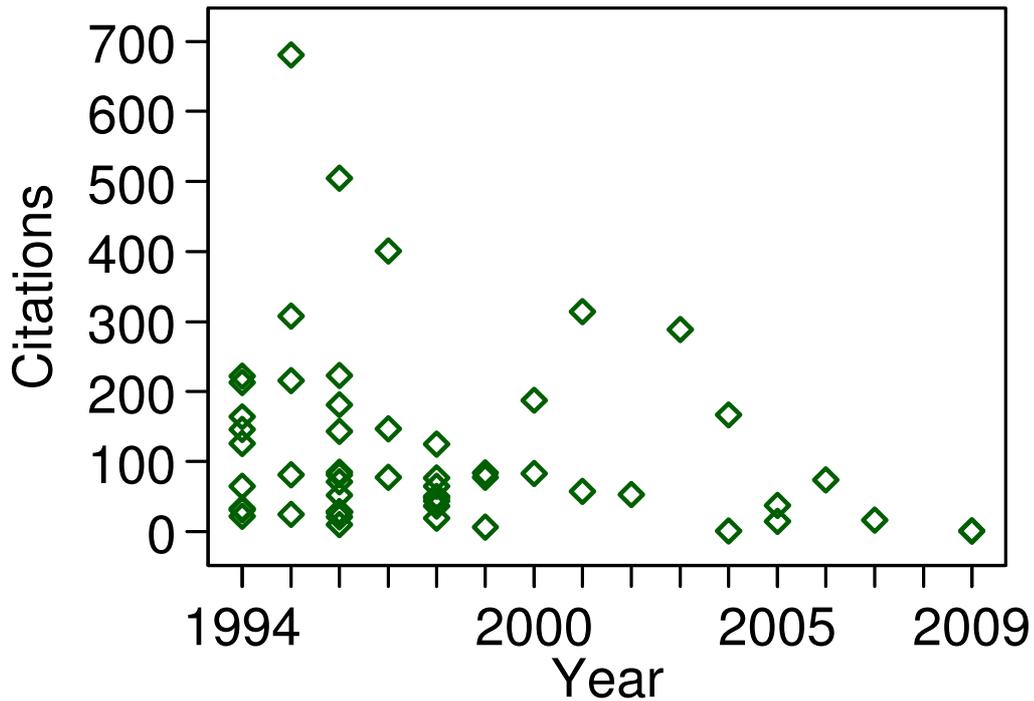
Figure 2. Published Statistics Notes by year

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. Like so many good ideas, these began with a chat at an RSS meeting. 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 Martin to discuss the possibility of including it in the Statistics Notes series. This was easy to do, as they were in adjoining departments at St. George's Hospital Medical School, and suited Martin as it was a topic in which he was already interested. Her original article formed the basis for a series of five notes on cluster randomized trials, written with Martin. Doug and Martin 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. We had hoped people might read them, but expected readers to throw them away afterwards. Some notes have been very highly cited; in July 2010 the highest was the note on the Bonferroni correction, with 681 citations. The number of citations achieved per note, by year of publication, is shown in Figure 3.

Figure 3. Number of citations of Statistics Notes, by year of publication

The mean number of citations per note is 115, median 74. We think that this is a very high number of citations for short educational pieces containing no original material. If Statistics Notes were an author, its 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 6,337 citations. As might be expected, older notes tend to have more citations than recent ones, as Figure 3 shows.

The most highly cited notes were on multiple significance tests - the Bonferroni method, Cronbach's alpha, measurement error proportional to the mean, 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, diagnostic-tests — predictive values. All of these have been cited at least 200 times by July 2010.

The note about Cronbach's alpha was asked for by the *BMJ*. It was the only one which was published to coincide with a paper which used it. Apart from this one, linking statistics notes to other research papers had proved far too difficult to coordinate.

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 Martin 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!

Other series of statistical articles

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?" [5] Economics Notes began in March 1998, with James Raftery's "Economic evaluation: an introduction". [6]. These series were both relatively short, not being intended to continue indefinitely.

Other journals have run similar series of statistical articles, and we are not the first. [7] Perhaps the most well-known is that, begun in 1967 by Donald Mainland and continued from 1970 by Alvan Feinstein, in *Clinical Pharmacology & Therapeutics*. That series predates us by many years and ran until 1981. Also those articles were quite long. 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" [8] has been cited 722 times. Feinstein collected many of the series into a book, *Clinical Biostatistics*, [9] published in 1977. We might do the same, eventually.
The *BMJ* itself had also published many articles about statistics before Statistics Notes: There was a series by Swinscow, which became the book *Statistics at Square One*, [10], and a series by Sheila Gore (now Bird) which addressed some of the topics we later covered in Statistics Notes. The *BMJ* had also published a series on statistics and medical ethics by Doug himself, and the series by Gore and Altman were combined in another book. [11]. Austin Bradford Hill's classic 1937 textbook *Principles of Medical Statistics* [12] was based on a series in the *Lancet*. There have been other similar series and some are going on right now; we come across them occasionally.

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 Martin'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 researchers to increase their publication count by splitting papers in several shorter papers and other acts of what was called "trivial publishing". Who would have thought that an incentive to publish more papers would result in more publications, even though these might contain only the same amount of research? The research assessors now 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! As an employee of a charity (at that time), Doug could treat university research assessment with lordly disdain. Martin wondered whether it was worth continuing, as writing the notes might distract him from less trivial publication. But he likes seeing his name in print and it seemed such a good idea that he put such thoughts from his mind. In the words of Davy Crocket, "Be always sure you are right, then go ahead." 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. There is a rumour that citations are to be used in the next Research Assessment, so if this happens we may be proved right to have persevered. But a

much more important question is whether Statistics Notes work. Do they help to improve readers' understanding of statistics and does this help to 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 noticeably since we started our careers as medical statisticians, even if there remains much scope for further improvement. 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

Conclusions

There is room for many more educational articles such as 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.* [13] 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. A recent review of 271 papers reporting laboratory animal experiments 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 the methods used or present their results adequately. [14] So a few series of statistics notes in journals where those papers were published could do a lot of good.

References

1. Bland, JM, Altman, DG. (1994) Correlation, regression and repeated data. *British Medical Journal*, **308**, 896.

2. Dr. Fisher's Casebook: A Misleading Assumption, *Significance*, **7(2)**, 121.

4   Altman, DG. (1991) *Practical statistics for medical research*. London: Chapman and Hall.

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

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

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

7   Johnson AL, Altman DG. Statistical articles in medical journals. In: Armitage P, Colton T (eds) *Encyclopedia of biostatistics*. 2nd edn. Chichester: John Wiley, 2005: 3151-62

.8. Kramer, MS, Feinstein, AR. (1981) Clinical Biostatistics 54. The Biostatistics Of Concordance. *Clinical Pharmacology & Therapeutics,* **29**, 111-3.

9. Feinstein, AR. (1977) *Clinical Biostatistics*. St. Louis: Mosby.

10. Swinscow, TDV. (1976) *Statistics at square one*. London, British Medical Association

11.   Gore, SM and Altman, DG. (1982) *Statistics in Practice*. London: British Medical Association.

12   Hill, AB. (1937) *Principles of Medical Statistics*. London: The Lancet.

13.   Altman, DG, Bland JM. (1996) Absence of evidence is not evidence of absence. *Australian Veterinary Journal*, **74**, 311.

14.   Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, MFW., Cuthill, IC., Fry, D., Hutton, J., Altman, DG. (2009) Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS ONE,* **4(11**), e7824.

Martin Bland is Professor of Health Statistics at the University of York, Doug Altman is Professor of Statistics in Medicine and Director of the Centre for Statistics in Medicine at the University of Oxford. They met in 1972 while working at St. Thomas's Hospital Medical School and, despite the divergence of their careers, have been friends and co-authors ever since.