

- ³ Report of the Joint Working Party on Lead Pollution around Gravelly Hill. *Lead pollution in Birmingham*. Department of the Environment Central Unit on Environmental Pollution, Pollution Paper No 14. London: HMSO, 1978.
- ⁴ Archer A, Giltrow JP, Waldron WA. Blood lead concentrations in pre-school children in Birmingham. *Proc R Soc Med* 1980;**73**:328-32.
- ⁵ Bryce-Smith D, Stephens R. *Lead or health*. A review of contemporary lead pollution and a commentary on the HM Government Working Party report *Lead and health*. London: Conservation Society, 1980.
- ⁶ World Health Organisation. *Environmental health criteria 3. Lead*. Geneva: WHO, 1977.
- ⁷ Chamberlain AC, Heard MJ, Little P, Newton D, Wells AC, Wiffen RD. *Investigations into lead from motor vehicles*. AERE-R9198. Harwell: United Kingdom Atomic Energy Authority, 1978.
- ⁸ Jost D, Sartorius R. Improved ambient air quality due to lead in petrol regulation. *Atmospheric Environment* 1979;**13**:1463-5.
- ⁹ Lawther PJ, Commins BT, Ellison J McK, Biles B. More observations on airborne lead. In: *Environmental health aspects of lead*. Proceedings of international symposium, Amsterdam, 1972.
- ¹⁰ Jones RD, Commins BT, Cernik AA. Blood lead and carboxyhaemoglobin in London taxi drivers. *Lancet* 1972;ii:302-3.
- ¹¹ Moore MR, Meredith PA, Campbell BC, Goldberg A, Pocock SJ. Contribution of lead in drinking water to blood lead. *Lancet* 1977;iii:661-2.
- ¹² Working Party on the Monitoring of Foodstuffs for Heavy Metals, Fifth Report. *Survey of lead in food: first supplementary report*. London: HMSO, 1975.
- ¹³ Hirao Y, Patterson CC. Lead aerosol pollution in the High Sierra overrides natural mechanisms which exclude lead from a food chain. *Science* 1974;**184**:989-92.
- ¹⁴ Tjell JC, Hovinand MF, Mosek H. Atmospheric lead pollution of grass grown in a background area in Denmark. *Nature* 1979;**280**:425-6.
- ¹⁵ Rabinowitz M. Plant uptake of soil and atmospheric lead in Southern California. *Chemosphere* 1972;**4**:175-80.
- ¹⁶ Interdepartmental Working Group on Heavy Metals. *Lead in the environment and its significance to man*. Department of the Environment Central Unit on Environmental Pollution, Pollution Paper No 2. London: HMSO, 1974.
- ¹⁷ Anonymous. Children and lead: some remaining doubts. *Arch Dis Child* 1980;**55**:497-9.
- ¹⁸ Department of the Environment. *Lead in drinking water: a survey in Great Britain, 1975-76*. Department of the Environment Central Unit on Environmental Pollution, Pollution Paper No 12. London: HMSO, 1977.
- ¹⁹ Thomas HE, Elwood PC, Welsby E, St Leger AS. Relationship of blood lead in women and children to domestic water lead. *Nature* 1979;**282**:712-3.
- ²⁰ Moore MR, Goldberg A, Meredith PA, Lees R, Low RA, Pocock SJ. The contribution of drinking water lead to maternal blood lead concentrations. *Clin Chim Acta* 1979;**95**:129-33.
- ²¹ Moore MR, Hughes MA, Goldberg DH. Lead absorption in man from dietary sources. The effect of cooking upon lead concentrations of certain food and beverages. *Int Arch Occup Environ Health* 1979;**44**:81-90.
- ²² Sartor F, Rondia D. Blood lead levels and age: a study in two male urban populations not occupationally exposed to lead. *Arch Env Health* 1980;**35**:110-6.
- ²³ Beattie AD, Dagg JH, Goldberg A, Wang I, Ronald J. Lead poisoning in rural Scotland. *Br Med J* 1972;ii:488-91.
- ²⁴ Crawford MD, Gardner MJ, Sedgwick PA. Infant mortality and hardness of water supplies. *Lancet* 1972;ii:988-92.
- ²⁵ Pocock SJ. Factors influencing household water lead: a national survey. *Arch Env Health* 1980;**35**:45-51.
- ²⁶ Walter SP, Yankel AJ, von Lindern IH. Age-specific risk factors for lead absorption in children. *Arch Env Health* 1980;**35**:53-8.
- ²⁷ Statutory Instruments 1980 No 1248. *Health and safety: The Control of Lead at Work Regulations 1980*. London: HMSO, 1980. (With approved code of practice.)
- ²⁸ Duggan MJ. Lead in urban dust: an assessment. *Water, Air, and Soil Pollution* (in press).
- ²⁹ Barltrop D, Strehlow CD, Thornton I, Webb JS. Absorption of lead from dust and soil. *Postgrad Med* 1975;**51**:801-4.
- ³⁰ Charney E, Sayre J, Coulter M. Increased lead absorption in inner city children: where does the lead come from? *Pediatrics* 1980;**65**:226-31.
- ³¹ Day JP, Hart M, Robinson MS. Lead in urban street dust. *Nature* 1975;**253**:343-5.
- ³² Barltrop D. Lead poisoning in childhood. *Postgrad Med J* 1968;**44**:537-48.

Medicine and Mathematics

Statistics and ethics in medical research

VIII—Improving the quality of statistics in medical journals

DOUGLAS G ALTMAN

Publication of a paper implies that the work is both sound and worth while. As I pointed out in my first article, it bestows both respectability and credibility on the work—a “seal of approval.” Once a paper has been published the results may influence both medical practice and further research by other scientists, and if the subject is of general interest the “mass media” may report the findings.

The ultimate responsibility for the general standard of published research rests with the medical journals. Perhaps unwillingly, the journals have the role of guardians of quality. This is particularly important with regard to statistical methods, which the majority of readers of medical papers are not able to judge for themselves and so must take on trust. The system of appraisal by independent referees is not ideal, but it is probably

the most practical method of quality control. Referees are usually selected, however, for their expertise in the relevant medical topic; their ability to assess the statistical aspects is left somewhat to chance. The result is that the statistical methods used in many research papers do not receive adequate scrutiny, with the consequences described in the previous articles.

The poor quality of statistics in published papers has been a cause of concern for many years, and is not confined to medical research. In 1964 Yates and Healy¹ wrote: “It is depressing to find how much good biological work is in danger of being wasted through incompetent and misleading analysis of numerical results.” Concern should be particularly great in the medical field because of the ethical implications, but the medical journals have generally been slow to appreciate that the statistical aspects can be fundamental to the validity of research.

Statistics in medical papers

Probably as a reflection of widespread unease, there have been several reviews of the quality of statistics in published papers

Division of Computing and Statistics, Clinical Research Centre, Harrow, Middx HA1 3UJ

DOUGLAS G ALTMAN, bsc, medical statistician (member of scientific staff)

over the past 15 years.²⁻⁶ These views are not strictly comparable because they looked at different statistical aspects in different journals at different times. Nevertheless, they all found many statistical errors or important errors of omission—in 72%, 49%, 52%, 45%, and 44% of papers studied, respectively. Further, a review of papers in five general medical journals found that 20% of the statistical procedures used were unidentified.⁷

It is impossible to assess the seriousness of many of the errors found. For example, an invalid analysis *may* give the same answer as an appropriate one, omission of information about randomisation does not necessarily mean that subjects were not allocated to treatments at random, and so on. It is, though, a measure of the disturbingly high prevalence of bad statistics that the reviewers of 62 papers in the *BMJ*⁴ thought that it was “some comfort that only five papers drew a false conclusion.”

Reviews of statistical procedures have sometimes been accompanied by editorials⁸⁻⁹ reinforcing the suggestions made in most of the papers that the standards of teaching should be improved and that there should be greater participation by statisticians in medical research. Such articles, however, stop short of the obvious suggestion that many of the papers should not have been published, at least as they stood, since any errors detected after publication could equally well have been detected at the refereeing stage.

Not all journals are equally culpable. The number of journals that use statisticians as referees, and sometimes also as members of editorial boards, has gradually increased, and several journals have publicly recognised the need to improve their statistical reviewing.¹⁰⁻¹² As Rennie¹⁰ says: “Our goal is the publication of data that are correctly observed and properly analysed.” Such sentiments should be endorsed by all medical journals.

Raising statistical standards

Later I shall examine in some detail what the journals can do to improve standards. It is, however, important to realise that there are other aspects to the problem, which can broadly be summarised by the question: “Why is the standard of statistics so low in papers *submitted* for publication?”

TEACHING OF STATISTICS

The recent widespread move to include statistics in the syllabus for medical students and other science undergraduates is a welcome development. Such teaching is likely to be most beneficial when it gets away from a rigid method-orientated approach and concentrates more on general concepts. For medical students it may be more successful when not taught as an isolated subject, but closely related to another course such as epidemiology.¹³

Statistics is not an easy subject, however. A short introductory course is not sufficient to equip qualified doctors or scientists to carry out their own statistical analyses adequately, both because of the necessarily limited scope of such courses and also because several years may elapse before they need to use the knowledge. Thus although there is room for improvement in undergraduate teaching, it is unlikely to have much effect on the quality of statistics in medical research.

Of greater value in this respect would be postgraduate courses in statistics for those who had previously had an introductory course, and aimed particularly at those intending to do research. Such courses should try to give a greater understanding of statistical concepts: to help researchers to understand properly the simpler statistical methods (including when not to use them), to appreciate the principles of more advanced methods, and to know when to seek expert help. If such courses exist they are rare.

Similar comments apply to textbooks, where there is a wide

gap between the elementary¹⁴ and the comprehensive.¹⁵ Simple textbooks are usually much too strongly method-orientated to give a good grasp of the underlying principles behind much of statistics.

INVOLVEMENT OF STATISTICIANS

In general, the larger a project the more likely it is that a statistician will be directly concerned. Yet a survey¹⁶ of 211 cancer treatment studies in progress in 1978 showed that in only 47% was a statistician fully concerned (in design, data collection, and analysis). There was some involvement in a further 44%, but in 9% there was none. Unfortunately, not all medical researchers have direct access to a statistician, but large collaborative studies usually need considerable statistical advice,¹⁷ preferably with a statistician as an active participant. Even for small studies statistical advice before the research begins may be very valuable, especially in helping to match the design to the objectives of the study, and also to give the statistician a greater understanding of the research. Yet, despite common pleas for early involvement, most consultancy concerns the analysis of data that have already been collected. A bigger problem, though, is that many projects are carried out without the benefit of any statistical advice at all. Increased involvement of statisticians in medical research would clearly improve the overall standard of statistics, but this requires greater availability of medical statisticians than at present.

Successful consultancy relies on the ability of both researcher and statistician to understand each other's language, which is not always easy. Sprent¹⁸ has suggested that “Interdisciplinary communication is probably the most pressing problem in the pursuit of knowledge.” The difficulties from the statistician's viewpoint have been discussed so often that a 1977 bibliography¹⁹ gave nearly 40 references. One aspect not often mentioned is that statisticians receive little or no preparation for consultancy work, either with respect to the sort of practical statistical problems that arise, or the role of consultant. This is a definite shortcoming in the education of statisticians, especially important because of their influence on the conduct of medical research.

ETHICAL COMMITTEES

Ethical committees have the opportunity to review many protocols for intended research on human subjects, and have the important sanction of withholding their approval. In view of the ease with which research can be rendered unethical by statistical mismanagement (as discussed in previous articles) it should be an automatic part of the review by ethical committees to look formally at the experimental design, and preferably also at the intended form of analysis. May²⁰ has written: “A poorly designed or poorly conceived experiment is unethical by definition and should not be permitted. Further it is the responsibility of the review committee to ensure that the conception and design meet the accepted canons of scientific method because we are dealing with experimentation which may not be for the individual subject's direct benefit.” We can share his surprise that statisticians are not universally represented on ethical committees.

WHY PUBLISH ?

One reason for the relentless production of low quality papers (not only with respect to the statistics) is the pressure on many individuals to publish as much as possible, with quantity being much more important than quality. At present it is known that other papers with poor statistics are being published, so a scientist may well think that there is no incentive (or need) to do better. But if journals were more careful about what they published we might advance to a state where fewer papers of a

higher standard were produced. This might also help to stem the counter-productive flow of new journals.

Role of the medical journals

There is general agreement among the medical journals in their attitude towards publishing the results of unethical research. Such research may have yielded valuable findings, but, as one editor wrote²¹: "publication in a reputable journal automatically implies that the editor and his reviewers condone the experimentation." In effect, papers describing unethical research are treated as "inadmissible evidence." For papers that may be deemed unethical because of their incorrect use of statistical methods, however, the attitudes of the journals vary enormously. Surely the same sort of argument as above should be extended, with publication similarly implying editorial approval of the data analysis and interpretation of results. It is illogical to refuse (quite rightly) to publish possibly useful findings of unethical research and yet be prepared to publish papers in which the results are invalidated by incorrect use of statistical methods.

One of the more obvious dangers of publishing questionable papers is that the conclusions may be quoted uncritically in the national press (since journalists are not usually qualified to criticise). Any ensuing critical letters will not receive similar publicity.

STATISTICAL REVIEW OF PAPERS

Since the reviews of published papers²⁻⁶ have found errors in about half of the papers examined, it is obvious that statistical review *before* publication ought to be highly effective. In 1964 the *Journal of the American Medical Association* raised the proportion of published papers considered statistically acceptable from one-third to three-quarters when it introduced a comprehensive statistical reviewing procedure.²²

Some of the following suggestions about ways in which journals can raise the quality of statistics in published papers have been made before,^{6 11 22} most notably in two recent papers.^{23 24} The most important recommendations are:

Statisticians should help referee

Journals should recruit statistically experienced people as referees, preferably with representation on editorial boards. Statistical review should be a formal procedure and not based on a casual inquiry to the nearest available statistician to "check that everything is all right." This is particularly important for specialist journals, where some depth of knowledge of the subject is often necessary.

All papers using any statistical procedure should be refereed by a statistician

Any paper in which inferences are drawn from the data presented should be seen by a statistician, whatever the level of statistical content. Indeed, the papers that cause the most trouble are usually those using only simple statistical methods "... where formal statistical review had seemed unwarranted,"^{21 0} rather than those with more complicated analyses. Short reports should not be exempt but should get higher priority. To reduce the work load the statistical assessment could be carried out only when a paper is likely to prove otherwise acceptable.

Revised papers should be returned to the same referee for reappraisal

A statistical refereeing system cannot work well without this condition. Failure to do this was the main reason why only 75% of published papers were completely acceptable even after the introduction of such a scheme.²²

Journals using a statistical refereeing system should state clearly what their policy is

This may help to discourage the submission of poor papers, and it would be valuable information for readers to know whether or not a journal uses such a system.

There should be statistical guidelines for contributors

All journals have instructions for contributors; very few mention statistics, and these rarely say much. It would obviously be undesirable for each journal to have different guidelines, but some agreement on this could be achieved in the same way as it has been on formats for references, perhaps in collaboration with the statistical societies. Some suggestions are given below.

All research papers should include a separate section on statistical methods

This should include information on relevant aspects of design, data collection, and analysis. Particularly important (if relevant) are the treatment allocation policy, response rate (and how non-responders were dealt with), and clear descriptions of analyses. Unusual methods of analysis should be given a specific reference (not a whole textbook!) with the reason for their use. This is a very important section of a paper, and should not be shortened at the expense of essential information.

Journals should give priority to well-executed and well-documented studies

Editorial boards should carefully consider the quality of study design, performance, analysis, and presentation of results when evaluating manuscripts. Standards should not be relaxed just because a paper is topical or interesting. Also, journals should not reject statistically valid papers purely because the findings were negative. (Obviously, this does not extend to those studies, discussed in the third article, that are too small to detect important differences.) As Bradford Hill said 25 years ago: "A negative result may be dull but often it is no less important than the positive; and in view of that importance it must, surely, be established by adequate publication of the evidence."²⁵

Less important but still desirable additional features are:

Authors should be encouraged to supply additional information (especially on methodology) to help the referees but not for publication

One of the problems when assessing papers is lack of information necessary for proper statistical assessment; this is the main reason for the fifth recommendation above. The extra information could be a more detailed account of the design, a fuller description of the methods used and the results, and copies of other related papers.

Authors should be encouraged to include the raw data in their papers

Obviously this is only practicable for small studies, but could be eased by using "miniprint" tables.

Journals should employ editorial staff with some understanding of statistics

This is perhaps less important if a comprehensive statistical refereeing system is adopted but is still highly desirable, especially in the event of disagreement between authors and referees.

For all journals to implement a comprehensive statistical refereeing system might well require many more medical statisticians than are currently available. It is much more likely, however, that there will be a continued steady increase in the use of statistical referees by journals, which should not cause major problems. Even the appointment by a journal of a single statistician can be enormously successful in raising the quality of statistics in published papers.

GUIDELINES FOR STATISTICAL REFEREES

Apart from checking on the validity of the statistical methods used, referees should ensure that there is adequate explanation and justification of what was done. It is also particularly important that the conclusions are reasonable, and that the summary is a fair reflection of the content.

The referee's report should be able to be understood by the authors, who may have only minimal statistical training.

GUIDELINES FOR CONTRIBUTORS

What sort of statistical guidelines should journals provide? Clearly these should not include advice on how to carry out research, although they might include discussion of the merits of different types of design. Such guidelines would not be a set of rules, but rather advice. The main emphasis should be on how best to describe clearly what procedures were used and what inferences were drawn.

Comprehensive guidelines would be of great benefit; these could perhaps be produced by a working party including representatives of medical journals and statistical societies. The following general suggestions relate to some of the more important aspects; they cannot be taken as comprehensive.

Design—This should be described clearly with, if relevant, information on treatment allocation, sample selection, if and how randomisation was used, whether or not the study was "blind" in any way, how sample size was determined (power), etc.

Data collection—Surveys should have response rates specified, and the representativeness of the sample and the possible effects of non-response should be discussed.

Analysis—The use of unusual forms of analysis should be justified, preferably with a reference, but all analyses should be very clearly described. It may be necessary to demonstrate the validity of the assumptions for some analyses (*t* tests, regression, etc).

Presentation of results—The results presented should be those most relevant to the question asked. Thus analysis of paired data should be accompanied by information—for instance, mean and standard deviation—about the within-person differences. Significance levels should not be given in place of quantitative results.

Interpretation of results—Special care should be taken to distinguish between statistical significance and clinical significance. Confidence intervals may greatly aid interpretation, especially where results are not statistically significant.

CONCLUSIONS

Reviews of published papers²⁻⁶ have all found unacceptably high proportions of papers with statistical errors. Some journals may feel that their policy of publishing letters criticising individual papers is an adequate safeguard. To take this attitude is to fail to appreciate the responsibility of the journals, both for ethical and scientific reasons, to avoid publishing sub-standard papers. In any case letters to journals usually produce a reply from the authors repeating their incorrect claims. Further, most papers are never read by anyone with the statistical knowledge to detect the flaws. If the credibility of published research is to

be raised it is essential that more journals introduce comprehensive statistical review procedures.

Summary

In these articles I have concentrated very much on one aspect of research. This is not meant to imply that statistics is of overriding importance, but rather that it is an area where much improvement is both highly desirable and possible.

By emphasising the ethical implications of carrying out research and publishing papers with incorrect statistics, I have argued that this is not just a matter for the individual researcher. There needs to be a wider appreciation of the importance of correct statistical thinking, and a great improvement in the standard of published research so that the sorts of errors discussed become very much the exception rather than commonplace. In the long term improved teaching and the greater involvement of statisticians will help; in the short term it is essential to have higher standards for published papers.

I am especially grateful to Martin Bland, Ted Coles, Stewart Mann, Charles Rossiter, and Patrick Royston for their perceptive criticism of earlier drafts of these articles. I must also thank Nicola Wilson Smith for the large amount of typing she has done.

This is the eighth in a series of eight articles. No reprints will be available from the author.

References

- Yates F, Healy MJR. How should we reform the teaching of statistics? *Journal of the Royal Statistical Society A* 1964;127:199-210.
- Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966;195:1123-8.
- Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J* 1970;iii:637-40.
- Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J* 1977;i:85-7.
- White SJ. Statistical errors in papers in the *British Journal of Psychiatry*. *Br J Psychiatry* 1979;135:336-42.
- Glantz SA. Biostatistics: how to detect, correct, and prevent errors in the medical literature. *Circulation* 1980;61:1-7.
- Feinstein AR. Clinical biostatistics. XXV A survey of the statistical procedures in general medical journals. *Clin Pharmacol Ther* 1974;15:97-107.
- Anonymous. A pillar of medicine. *JAMA* 1966;195:1145.
- Anonymous. Statistical errors. *Br Med J* 1977;i:66.
- Rennie D. Vive la différence ($p < 0.05$). *N Engl J Med* 1978;299:828-9.
- Shuster JJ, Binion J, Moxley J, et al. Statistical review process. Recommended procedures for biomedical research articles. *JAMA* 1976;235:534-5.
- Rosen MR, Hoffman BF. Statistics, biomedical scientists, and *Circulation Research*. *Circ Res* 1978;42:739.
- Clarke M, Clayton DG, Donaldson LJ. Teaching epidemiology and statistics to medical students—the Leicester experience. *Int J Epidemiol* 1980;9:179-85.
- Swinscow TDV. *Statistics at square one*. London: BMA, 1976.
- Armitage P. *Statistical methods in medical research*. Oxford: Blackwell, 1971.
- Tate HC, Rawlinson JB, Freedman LS. Randomised comparative studies in the treatment of cancer in the United Kingdom: room for improvement? *Lancet* 1979;ii:623-5.
- Breslow N. Perspectives on the statistician's role in cooperative clinical research. *Cancer* 1978;41:326-32.
- Sprent P. Some problems of statistical consultancy (with discussion). *Journal of the Royal Statistical Society A* 1970;133:139-64.
- Woodward WA, Schucany WR. Bibliography for statistical consulting. *Biometrics* 1977;33:564-5.
- May WW. The composition and function of ethical committees. *J Med Ethics* 1975;1:23-9.
- Woodford FP. Ethical experimentation and the editor. *N Engl J Med* 1972;286:892.
- Schor S. Statistical reviewing program for medical manuscripts. *American Statistician* 1967;21:28-31.
- O'Fallon JR, Dubey SD, Salsburg DS, Edmonson JH, Soffer A, Colton T. Should there be statistical guidelines for medical research papers? *Biometrics* 1978;34:687-95.
- Mosteller F, Gilbert JP, McPeck B. Reporting standards and research strategies for controlled trials. Agenda for the editor. *Controlled Clinical Trials* 1980;1:37-58.
- Hill AB. Contribution to the discussion of a paper by D J Finney. *Journal of the Royal Statistical Society A* 1956;119:19-20.