Psychological Medicine, 1985, **15**, 471–476
Printed in Great Britain

EDITORIAL

The role of statistics in psychiatry[1]

'If for medical journals the 1960s and 1970s seem likely to be remembered as the era when the importance of ethics was emphasised, the last 20 years of this century promise to be that of statistics.' (Lock, 1982.)

1. INTRODUCTION

In undertaking to discuss the role of statistics in psychiatry I am naturally assuming that it has some role. Evidence for this assumption is very easily come by, just by glancing at recent issues of the major psychiatric journals. For example, of all the papers published in *Psychological Medicine* in 1980, 1981, 1982, and 1983 (excluding editorials), 81% contained statistical analyses. Similarly, White (1979), in a study of papers in the *British Journal of Psychiatry*, found that 83% presented numerical results. This would seem to suggest that statistics does not merely serve some role, but that it serves a very important role in psychiatry. Indeed, a comparison of these figures with corresponding ones for other basic sciences (such as psychology, biochemistry, etc.) might lead one to the conclusion that statistics is the most important basic science for psychiatry, at least as far as understanding the literature of the subject goes. Since clinicians read the literature in order to update their methods, research workers read it in order to learn of colleagues' results, and teachers read it in order to integrate the new material into their courses, the inevitable conclusion must be that statistics plays an absolutely vital role in psychiatry: all three of these groups must have some statistical knowledge in order to do their job effectively.

Having recognized the ubiquity of statistical methods in the psychiatric literature, a natural question is: *why* are they there? The answer, of course, lies in psychiatry itself. Psychiatry is characterized by its intrinsic complexity, its wealth of interacting factors and influences, and by the particular problems of defining and measuring the relevant concepts and indicators involved. To unravel the knots consequent upon these difficulties we must resort to sophisticated and powerful techniques. The straightforward methods of, say, classical physics will not suffice. We must, in fact, use the armamentarium of a specialist in research methodology – a modern statistician.

The next section examines the nature of this armamentarium; section 3 looks at the experts who wield its contents; and section 4 considers the relationship between statistics and medical ethics.

2. THE NATURE OF STATISTICS

Statistics has been likened to a telescope. The latter enables one to see further and to make clear objects which were diminished or obscured by distance. The former enables one to discern structure and relationships which were distorted by other factors or obscured by random variation. Statistics is thus a language for extracting patterns from quantitative (or, in fact, qualitative) observations of the subjects under study.

My choice of the word 'language' in the last sentence is deliberate. It is a common misconception to regard statistics as a mere vocabulary, a simple collection of techniques. But, in fact, the elements of statistics are bound together into a complex interlocking network. One technique or idea may be a special case of another. Two may be related by having the same abstract structure, but applied

[1] This paper is based on a talk given in the Advanced Psychiatry course, at the Institute of Psychiatry, University of London, September 1984. Address for correspondence: Dr D. J. Hand, Biometrics Unit, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF.

to different types of data. Another two might both be derived from the same mathematical basis but have evolved down completely different paths. When one begins to recognize the rich interconnections, the essential unity of the subject, the motivations and rationale behind the various statistical tools become much easier to comprehend. So, statistics is a language, not just a collection of isolated recipes, one of which is to be plucked out to match the presenting research question.

This misconception of statistics is sufficiently widespread to have been dignified with a name. It is called the 'cookbook fallacy'. One sometimes finds elementary textbooks written in this way. Such books typically reveal a mere superficial grasp of the concepts involved. Far worse than this, however, is the fact that they can have a pernicious and deleterious effect on the quality of any statistical work conducted by their readers. This is because, seeking to plug a ready-made solution into each research question, without communicating understanding of why the chosen tool is suitable, they risk encouraging the use of inappropriate techniques and can force the statistically inexpert research worker to mould his research questions to the statistical methods listed. And, as John Tukey put it: 'An approximate answer to the right question is worth a good deal more than an exact answer to an approximate problem.'

Apart from appreciating the overall unity of statistics there are other aspects of it which psychiatrists conducting research, who of necessity must come into contact with statisticians (convincing reasons for this assertion will be presented below), would benefit from understanding. A key issue is that statistics is a scientific discipline in its own right, and a very young one at that, all but its very early history lying in this century. This has many implications.

For example, as with any other science which has not become stale, it contains its share of controversies. There are, in fact, several different Schools of statistics, the proponents of which hold fundamentally different ideas about the nature of probability and statistical inference (see Barnett, 1982). This is where statistics and philosophy merge.

The youth of statistics means that, like any young thing, it is changing and advancing rapidly. A statistical tool learnt ten or twenty years ago might today be regarded as woefully inadequate, having been superseded by more sophisticated methods. This not only has implications for psychiatrists conducting research who wish to use statistics and to communicate with statisticians, but it also has implications for the (continuing) education of statistical consultants. We shall return to the subject of education in the next section. For the present, it is perhaps simply worth remarking that, in all sciences (not just statistics), it has been estimated that the half-life of the material one learns at University is five years. Thus, after 5 years one's knowledge is 50% outdated, after 10 years it is 75% outdated, and so on.

One factor causing change in most aspects of our modern lives is particularly influential in statistics. This, as the reader has probably already guessed, is the development of the electronic computer. It has affected all levels of statistical science.

Without computers, methods were limited to those for which the calculations could be carried out by hand within a reasonable time (meaning a matter of weeks at most). This focused methodological research attention on simpler methods (though statistics naturally had to develop simpler methods before more complicated or sophisticated ones) and on particular experimental research designs which permitted easier analysis, e.g. balanced orthogonal designs. The influence of this early developmental process is still evident in the way many basic statistical texts introduce their material, although nowadays there is no need for this. This is discussed further in Hand & Taylor (1985) and has brought us back to statistical education, which we shall discuss below.

It is vital that I should not give the wrong impression of the impact of computers on statistical practice. The primary impact is not that they force us to abandon simple methods and adopt complex ones, but rather that they free us from the need to worry about the numerical trivia. We can now concentrate on higher level questions of technique. The increased range of methods and the development of statistical programming languages (such as GENSTAT, GLIM, or S) means that we can choose or develop appropriate methods instead of being forcibly constrained to use something inappropriate. Moreover, we can try several methods – an analysis which might previously have taken a day could now take only a second.

Nothing, however, is an unqualified blessing to mankind. Complex statistical methods can now be applied with ease by the statistically inexperienced, using highly accessible packages like SPSS and SAS. This might sometimes be good, but can all too often be bad. It can result in a less than critical application of methods or in application of methods with a poor grasp of what the results mean or, indeed, a lack of understanding of whether the results are relevant at all; in the pre-computer era only someone who was fairly sure that he would understand the results would undertake a principle components analysis or a multivariate analysis of variance by hand. This has led several research teams to begin work on what seems to me must obviously be the next stage of statistical software: the statistical expert system (Hand, 1985*a*, *b*). These are systems which guide the research worker through an analysis, preventing him or her from misapplying the techniques. My own design for such a system has been moulded by my experience of acting as a consultant statistician to many research workers in psychiatry.

3. THE NATURE OF STATISTICIANS

Some research workers regard statisticians as obstructive and obscurantist and seek to do away with them by handling their statistics themselves. In this section we consider the justification for such opinions and the feasibility of doing one's own statistical work. We start with the latter point.

A common pattern for the modern statistician's training is as follows: a three-year bachelor's degree in mathematics (though there has recently been some debate in the statistical community about whether a mathematics degree is necessarily the ideal starting point for a consultant statistician), a one-year master's degree in statistics, or mathematical statistics, or medical statistics, etc., and then a three-year PhD project developing or refining a new statistical tool. Following this, an oft-quoted figure for working in a particular consulting environment – say, psychiatry, or industry, or agriculture, etc. – before attaining competence is three years. This means that training has taken a total of ten years. Bearing this in mind, it will come as no surprise to learn that consultant statisticians in private practice charge similar fees to doctors or lawyers. Evident from this will be the difficulty, for someone who has already trained to be a psychiatrist, of attaining comparable statistical expertise. Given the remarks made earlier about the rate of progress in scientific disciplines, it is clear that attempting to master both psychiatry and statistics would be futile.

Obviously, it should be acknowledged that, while attaining competence to the level of a statistical consultant may be infeasible, it might be practicable to develop sufficient expertise in the few techniques the research worker wishes to use. To some extent this might be true, but the dangers are great. There is the obvious risk of not understanding a technique as well as one thinks, and there is also the risk of finally utilizing a quaint old-fashioned method which has many well-known (to statisticians) weaknesses. In general, the research worker may attempt to answer research questions using inappropriate techniques or changing the questions so that the techniques fit. Again, we refer to the remark of Tukey, quoted above. In any case, this very notion of acquiring expertise in a few techniques hints at a cookbook approach.

If none of the above points seem very convincing, then one which is very difficult to argue against is the medical and psychiatric literature itself. In the next section we describe some studies of the level of statistical expertise evident in medical journals. A shocking picture is presented. The conclusion that must emerge is that it is not feasible to abandon contact with trained experts in the area; that is, with statisticians.

What about the view of statisticians as obstructive and obscurantist?

The image of consultant statisticians as obstructive presumably arises as a consequence of the nature of their job. Statisticians, as specialists in research methodology, must guide the research worker to questions which can be effectively answered, to methods which enable the research worker to answer those questions, and in interpreting the results of the analyses. From the point of view of the research worker who wishes to use a sample of students to make inferences about the differences between men and women in the population at large, being told that he cannot do so may appear obstructive, but the statistician is in fact just being honest.

The origins of the obscurantist image is obvious, and I think we all at times feel this about some scientific subgroup. Since statistics is a scientific discipline in its own right, it has its own vocabulary of technical terms and jargon. Much of this defies straightforward translation into simpler terminology, since it represents quite abstract and sophisticated concepts. A misrepresentation of someone who uses technical language as attempting deliberately to obscure things is commonplace. The resolution of the problem lies in a greater willingness of the collaborators, the client and the consultant statistician, to communicate. Communication is the key issue.

Having stressed the importance of communication, we should not be deceived into the belief that a few extra minutes of conversation will make the whole of statistics totally clear. Statistics is, after all, an intrinsically mathematical subject – with the implication that it requires long and hard effort. As Leonard Savage (1972) puts it:

> It cannot be too strongly emphasised that a long mathematical argument can be fully understood on first reading only when it is very elementary indeed, relative to the reader's mathematical knowledge. If one wants only the gist of it, he may read such material once only; but otherwise he must expect to read it at least once again. Serious reading of mathematics is best done sitting bolt upright on a hard chair at a desk. Pencil and paper are nearly indispensable.

So, the point is that if the statistician appears obscurantist, he is probably having some difficulty in formulating, in simple non-mathematical terms, highly abstract concepts which defy such reformulation.

While on the subject of communication between research worker and statistician it is perhaps worthwhile saying a little about the statistical education of the non-statistician. The difficulties associated with teaching service courses in statistics are well known, and have been the subject of much debate over the years. They typically and understandably arise because one is teaching students to whom the subject matter is of only secondary interest. Moreover, such courses must usually be squeezed into an already overcrowded syllabus, competing with other material which some might claim is more directly concerned with medical treatment. One feature about such courses which seems clear to me is that, although students should be encouraged to work through examples of the simpler techniques, the emphasis of such courses must lie on the ideas and reasons behind the techniques, rather than on the numerical manipulations. Mainland (1982) has put this distinction very nicely. He calls it *statistical thinking* rather than *statistical arithmetic*. The aim must be to develop a rudimentary critical faculty for statistical methods rather than a facility for arithmetic computation. This emphasis is not yet generally reflected in statistical texts (though some strive in that direction) but is surely increasingly desirable as the trivia of the arithmetic juggling is taken over by electronic machinery.

One further problem with teaching service courses in statistics at an undergraduate level is that years might pass before the student is called upon to use the material – by which time it will doubtless have been forgotten. There is clearly a need for regular courses in basic statistics, with the emphasis as described above, and which qualified psychiatrists undertaking research can take.

4. STATISTICS AND MEDICAL ETHICS

Basic ethical principles prevent us from avoidably subjecting patients to less efficacious treatments. Given two competing treatments (say, the traditional one and a newly proposed one), we conduct a clinical trial so that future cases can receive the better of the two treatments. It is obvious from this that we are ethically bound to ensure that our clinical trial is properly designed and analysed in order to reduce the risk of choosing the wrong treatment. Note that once a study has established some treatment as superior it might be a long time before an error is discovered. There is a natural tendency, involving funding as well as ethical considerations, to avoid repeating a study involving an 'established' poorer treatment.

There are, needless to say, many ways in which studies can be incorrectly designed. Since this is not a technical article, we shall not go into those ways here – the interested reader will find genuine examples in the references cited below or can examine texts on clinical trials. However, one simple way of producing an unethical study is perhaps worth mentioning briefly, since it has a number of implications and is undoubtedly the most common single source of unethical studies that I encounter. This is the question of sample size.

It is obvious that we should avoid too large a test sample: a trial necessarily involves subjecting *some* patients to the poorer treatment and we are ethically constrained to minimize this number. (There are also cost and time factors which work in this same direction.) However, and this is the kind of error I most often witness, we must also avoid too small a sample. Too small a study will have little, perhaps negligible, chance of permitting conclusions to be drawn in either direction. This means that the subjects' time has been wasted, they may have been put at needless risk, and research funds will have been wasted. About the only thing that can be said in favour of such studies is that they often result in research publications. I hope the reader will forgive me for being dogmatic when I state that this does not outweigh the unethical nature of such studies, for the point is that a conversation with a statistician before beginning the study will usually lead to a recognition of the uselessness of such studies. In general, as far as design goes, the word *before* is the key. Much futile suffering, anguish, and waste of resources could be avoided by adopting suitable designs. Typically, this will need to be by consultation with a statistician.

Inadequate design is one side of the coin. The other side is inadequate analysis of data. Several studies of the adequacy of data analysis in medical and psychiatric research have been conducted (see below) and all have found it wanting. A study which is inappropriately analysed is unethical for all of the reasons discussed above: it unnecessarily subjects patients to risk (the wrong conclusion may be drawn or the best use may not be made of the data), it wastes the time of both patient and clinician, and it wastes funds and other resources.

The discussion above has been oriented towards clinical trials but similar ethical considerations arise in other types of study. A poorly designed epidemiological study may fail to detect an important cause of a disease or may unfairly implicate some factor. As a result, people may be encouraged to change their lifestyles needlessly, which certainly inconveniences them and, in the worst case, if the study's conclusions were quite wrong, might increase their risk. The same applies in the analysis phase of epidemiological studies.

Design and analysis are followed by publication of research results. It hardly needs stating that it is unethical to publish incorrect conclusions if consultation with an expert – a statistician – could have avoided the errors.

As an aside, we should note that publishing a description of an inappropriate design or analysis is unethical, even if it does not lead to incorrect results. This is because such publication may encourage others to make the same mistake – with perhaps more serious consequences.

So far, my discussion has been in the abstract. Lest the reader imagine that I am conjuring up illusory spectres to deter potential research workers, consider the following:

(1) White (1979) conducted a survey of all 168 papers published in the *British Journal of Psychiatry* between July 1977 and June 1978. 139 papers contained numerical results and, of these, 63 contained statistical errors. Forty-seven papers contained major errors. White states 'at least one drew unsupportable conclusions. In many cases the errors were not considered to be severe, but they were often sufficient to raise doubts about some inferences.'

(2) Gore *et al.* (1977) studied the 77 reports appearing as papers and originals in 13 consecutive issues of the *British Medical Journal.* Sixty-two papers contained statistical analyses and, of these, 32 had statistical errors. In 18 cases fairly serious faults were discovered. The authors state: 'The summaries of five reports made some claim that was unsupportable on re-examination of the data.'

(3) Schor & Karten (1966) studied three issues of each of ten medical journals from the first three months of 1964. The ten 'contains those that are read more frequently and considered by many physicians to have excellent reputations'. Of the total number of 295 articles, 49·5% were case descriptions. Of the remaining 149, 27·5% were judged statistically acceptable, 67·8% were judged as in need of revision before publication, and 4·7% were judged unsalvageable – 'the problem posed by the investigator could not be solved by the kind of study described'. For the studies involving

statistics, none of the ten journals studied contained more than 40% of articles which were judged to be statistically acceptable. Two of the journals had no acceptable reports.

It should be acknowledged that the paper of Schor & Karten is now nearly twenty years old – so that one might hope that things have improved. The paper goes on to describe the impact that introducing a biostatistical review process had on the quality of one of the journals: the proportion of acceptable published papers rose from 26% to 74%. The authors attribute the shortfall from 100% to the failure of the editor to resubmit revisions for statistical review.

The nature of the errors found in the above is made clear by White (1979): 'errors are perpetrated with the simpler methods as well as with the more complicated ones'. And Altman (1982): 'The vast majority of statistical errors are unintentional, usually resulting from an inadequate understanding of statistical principles and methods.' Altman then goes on to say: 'Ignorance is not really an acceptable defence, since using statistical techniques (like using laboratory equipment or driving a car) requires adequate training (or supervision).'

In considering the figures given above, a further point which should be kept in mind is that the error rates are almost certainly optimistic. This is so, first, because many errors will be of a kind which cannot be detected by a perusal of the papers themselves. Schor & Karten describe the impact on authors of criticisms of the statistical contents of their papers. Some, the smallest group, 'revised their manuscripts in such a way that, although the design, the analysis and the conclusions remained unchanged (the conclusions were still invalid), it became more difficult for the reader, be he statistician or physician, to detect what was wrong with the study'. The second reason for the optimistic bias is that the journals chosen for study were of high prestige. Less reputable journals might well have more errors.

Altman (1981, 1982) proposes a number of recommendations aimed at tackling the various problems. Among these are that statisticians should serve on ethical committees and that all submissions with a statistical content should be refereed by a statistician. A slight problem here is that a statistician's primary interest is statistical methodology. He will probably already be refereeing material for statistical journals and an extra workload, of at best peripheral interest, can be unwelcome. It has been suggested that inducements should be offered.

5. CONCLUSION

It is clear that statistics serves a major role in modern psychiatry, and that awareness and understanding of statistical concepts is of increasing importance to all psychiatrists, but especially those who wish to advance the field by undertaking research themselves.

My aim has not been to deter clinicians who wish to undertake research. Quite the opposite, in fact. I would like to encourage them to do so, but I would very much like to encourage them to get it right. Good research is the way science advances. Poor research does a disservice to both its perpetrator and his science.

D. J. HAND

REFERENCES

Altman, D. (1981). Statistics and ethics in medical research: VIII. Improving the quality of statistics in medical journals. *British Medical Journal* **282**, 44–47.

Altman, D. (1982). Statistics in medical journals. *Statistics in Medicine* **1**, 59–71.

Barnett, V. (1982). *Comparative Statistical Inference*. John Wiley and Sons: Chichester.

Gore, S. M., Jones, I. G. & Rytter, E. C. (1977). Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *British Medical Journal* i, 85–87.

Hand, D. J. (1985*a*). Statistical expert systems: I. Design. *The Statistician* (in the press).

Hand, D. J. (1985*b*). Statistical expert systems: II. Necessary attributes. *Journal of Applied Statistics* (in the press).

Hand, D. J. & Taylor, C. C. (1985). *Practical Multivariate Analysis of Variance*. Chapman and Hall: London.

Lock, S. (1982). Preface. In *Statistics in Practice* (ed. S. M. Gore and D. G. Altman). British Medical Association: London.

Mainland, D. (1982). Medical statistics – thinking *vs*. arithmetic. *Journal of Chronic Diseases* **35**, 413–417.

Savage, L. (1972). *The Foundations of Statistics*. Dover Publications: New York.

Schor, S. & Karten, I. (1966). Statistical evaluation of medical journal manuscripts. *Journal of the American Medical Association* **195**, 1123–1128.

White, S. J. (1979). Statistical errors in papers in the *British Journal of Psychiatry*. *British Journal of Psychiatry* **135**, 336–342.

https://doi.org/10.1017/S0033291700031354 Published online by Cambridge University Press