MEDICAL PRACTICE

Medicine and Mathematics

Statistics and ethics in medical research

Misuse of statistics is unethical

DOUGLAS G ALTMAN

> "Some people hate the very name of statistics but I find them full of beauty and interest. Whenever they are not brutalised, but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary. They are the only tools by which an opening can be cut through the formidable thicket of difficulties that bars the path of those who pursue the Science of man."
>
> FRANCIS GALTON[1]

In 1949 a divorce case was heard in which the sole evidence of adultery was that a baby was born almost 50 weeks after the husband had gone abroad on military service. To quote Barnett[2]: "The appeal judges agreed that the limit of credibility had to be drawn somewhere, but on medical evidence 349 (days), whilst improbable, was scientifically possible." So the appeal failed.

If we look at the distribution of length of gestation[3] (fig 1), which the judges apparently did not do, I think that most people would feel that the husband was hard done by. Even if we take reports of extremely long pregnancies as accurate, it is clear that, although "scientifically possible," a pregnancy lasting 349 days is an extremely unlikely occurrence. For those who believe as I do that a pregnancy of 51 weeks* exceeds the bounds of credibility, suppose it had been only 48 weeks, or 45?

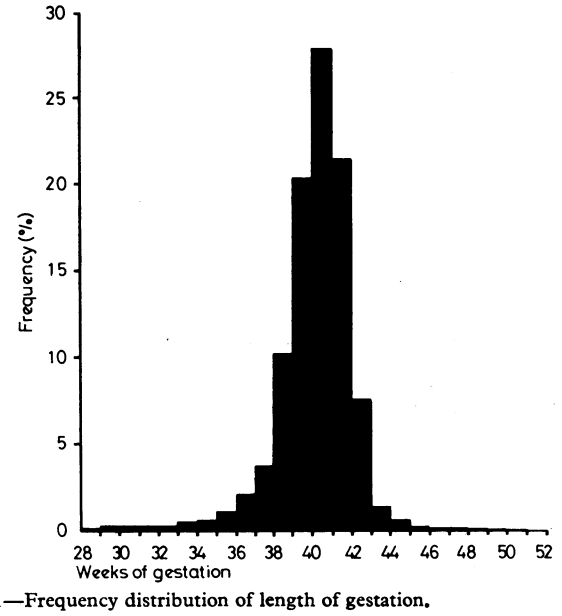

FIG 1—Frequency distribution of length of gestation.

If this case were heard now, where would *you* draw the line on the basis of fig 1?

This case illustrates a failure to use statistical methods when they ought to have been used, a fairly common occurrence. Saying that an event is possible is quite different from saying that it has a probability of, say, one in 100 000. Although not an example from medical research, this case concerned essentially the same difficulty as in many more frequently encountered problems, such as defining hypertension or obesity. Everything varies; it is in trying to draw lines between good

*Using the standard convention of counting in completed weeks from the first day of the last menstrual period and assuming conception to have occurred 14 days later.

Division of Computing and Statistics, Clinical Research Centre, Harrow, Middx HA1 3UJ
DOUGLAS G ALTMAN, BSC, medical statistician (member of scientific staff)

and bad, high and low, likely and unlikely, and so on, that many problems arise. Although statistics cannot answer a given question, they can often shed considerable light on the problem.

Statistics and medical ethics

So what is the relation between statistics and medical ethics? It is well appreciated that ethical considerations may affect the design of an experiment. Perhaps the most obvious examples are clinical trials—we cannot, for example, carry out controlled trials of cigarette smoking. The purpose of this series of articles is to discuss in some detail a different and much neglected aspect of the relation—how the statistical aspects affect the ethics.

Stated simply, it is unethical to carry out bad scientific experiments.[4] Statistical methods are one aspect of this. However praiseworthy a study may be from other points of view, if the statistical aspects are substandard then the research will be unethical. There are two principal reasons for this.

Firstly, the most obvious way in which a study may be deemed unethical, whether on statistical or other grounds, is the misuse of patients (or animals) and other resources. As May[5] has said: "... one of the most serious ethical problems in clinical research is that of placing subjects at risk of injury, discomfort, or inconvenience in experiments where there are too few subjects for valid results, too many subjects for the point to be established, or an improperly designed random or double-blind procedure."

Secondly, however, statistics affects the ethics in a much more specific way: it is unethical to publish results that are incorrect or misleading. Errors in the use of statistics may occur at all stages of an investigation, and one error can be sufficient to render the whole exercise useless. A study may have been perfectly conceived and executed, but if it is analysed incorrectly then the consequences may be as serious as for a study that was fundamentally unsound throughout.

There are many ways in which the statistical content of research may be deficient. In a fascinating and somewhat frightening recent paper, Sackett[6] identified 56 possible biases that may arise in "analytic research," over two-thirds of which related to aspects of study design and execution. Figure 2 shows

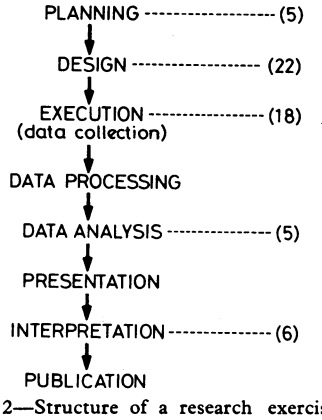

FIG 2—Structure of a research exercise. Explanation of numbers is given in text.

how these possible biases are distributed over the stages of a research exercise. In general this distribution also reflects very well the relative seriousness of statistical errors at each stage, and indicates where there is greatest need for statistical expertise. Errors in the analysis or interpretation of results can usually be rectified if detected in time—that is, before publication—but deficiencies in the design are nearly always irremediable. The end point of the process is usually publication. Problems may well arise when this is considered to be the most important aspect of the whole exercise, a not uncommon occurrence.

Publication

Once published, a piece of research achieves both respectability and credibility so that it is important for journals to make strenuous efforts to detect substandard research. In recent years there have been several good studies of the quality of statistics in papers in medical journals to support the idea that there is much room for improvement. For example, Schor and Karten[7] reported that, of 149 papers reporting analytical studies in several journals, only 28% were judged acceptable, 67% were deemed deficient but could be improved, and 5% were totally unsalvageable.

The editor of the journal wrote as follows:

> "The study is an indirect argument for greater knowledge and appreciation of statistics by the medical author, for a reiteration on his part that the biostatistician is not a worrisome censor, but a valuable ally, and that biostatistics, far from being an unrelated mathematical science, is a discipline essential to modern medicine—a pillar in its edifice."[8]

More recent studies[9–11] have shown that there are still far too many papers being published in which the statistical analyses are incorrect. Conflicting results from similar studies can often be attributed to varying degrees of statistical competence.[12–14]

The ethical implications of publishing research containing incorrect or unfounded results or conclusions are little affected by the nature of the errors made, and are indeed much the same as the consequences of publishing spurious results. The cost in time and energy in trying to reproduce such results can be enormous.[15] Alternatively, the results may rest unchallenged for many years. Suppose a randomised controlled trial is carried out in which a conclusion is reached that the new treatment is significantly better than the previous standard treatment. The publication of such a finding may well affect patient care, and it may then be considered to be unethical to carry out further trials as one group would be denied the new treatment that was "known" to be better. Clearly, both of these consequences of publication will hold whether or not the conclusions were justified unless any deficiencies are very obvious (and many that Sackett[6] lists would not be) or if there is considerable protest. A solitary critical letter, perhaps from a statistician, hidden away on the correspondence page is unlikely to be sufficient. Similar consequences apply in the opposite case where a treatment is incorrectly found to be ineffective.

Summary

The ethical implications of statistically substandard research may be summarised as follows:

(1) the misuse of patients by exposing them to unjustified risk and inconvenience;

(2) the misuse of resources, including the researchers' time, which could be better employed on more valuable activities; and

(3) the consequences of publishing misleading results, which may include the carrying out of unnecessary further work.

These are specific and highly undesirable outcomes. Failure to guard against these is surely as unethical as using experimental methods that offend against moral principles, such as failing to obtain fully informed consent from subjects. Surprisingly, this aspect seems to have been totally ignored by books on medical ethics.

All stages of research shown in fig 2 are vulnerable to statistical mismanagement. As an example consider one aspect of planning a study: "reading up published reports." If published papers are accepted uncritically you might be trying to verify someone else's spurious results. Remember too that authors will tend to

refer to other published work that supports their arguments and ignore papers that do not.

The next few articles will illustrate some ways in which errors at different stages of a study can compromise the ethical status of the research, and discuss some ways in which they may be avoided. These will serve only as examples, since it is impossible to be comprehensive. In the final article I will consider the role of the medical journals in this context.

References

1 Galton F. *Natural inheritance.* London: Macmillan, 1889.
2 Barnett V. The study of outliers: purpose and model. *Applied Statistics* 1978;**27**:242-50.
3 Chamberlain R. Birth weight and length of gestation. In: *British births 1970.* Vol 1. *The first week of life.* National Birthday Trust Fund and Royal College of Obstetricians and Gynaecologists. London: Heinemann, 1975.
4 Denham MJ, Foster A, Tyrrell DAJ. Work of a district ethical committee. *Br Med J* 1979;ii:1042-5.
5 May WW. The composition and function of ethical committees. *J Med Ethics* 1975;**1**:23-9.
6 Sackett DL. Bias in analytic research. *J Chronic Dis* 1979;**32**:51-63.
7 Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966;**195**:1123-8.
8 Anonymous. A pillar of medicine. *JAMA* 1966;**195**:1145.
9 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.
10 Feinstein AR. A survey of the statistical procedures in general medical journals. *Clin Pharmacol Ther* 1974;**15**:97-107.
11 Ambroz A, Chalmers TC, Smith H, Schroeder B, Freiman JA, Shareck EP. Deficiencies of randomized control trials. *Clin Res* 1978;**26**:280A.
12 Gifford RH, Feinstein AR. A critique of methodology in studies of anticoagulant therapy for acute myocardial infarction. *N Engl J Med* 1969;**280**:351-7.
13 Peto R. Clinical trial methodology. *Biomedicine* (special issue) 1978;**28**: 24-36.
14 Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. *Am J Med* 1979;**66**:556-64.
15 Muller M. Why scientists don't cheat. *New Scientist* 1977;**74**:522-3.

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

Process and Outcome

An audit of antenatal care: the value of the first antenatal visit

P K CHNG, MARION H HALL, I MacGILLIVRAY

Summary and conclusions

A critical analysis of the events recorded at the first antenatal visits in a city where all pregnant women are seen by specialist obstetricians for booking for antenatal care and confinement showed that many women attended too late for optimal care. The selection of women for their risk of complications was not very effective, partly because of failure to take account of information that was available, but mainly because many obstetric complications cannot be predicted, except by classifying large proportions of pregnant women as high risk. Even with the greatest care, inappropriate bookings are made at the first visit, and reappraisal of booking for continuing care and confinement is necessary during pregnancy.

Introduction

The first antenatal visit provides an opportunity to review the medical and obstetric history of the pregnant woman, make a physical examination, perform appropriate investigations, arrange suitable antenatal care for the rest of the pregnancy, and book the confinement in a setting with the facilities and professional expertise likely to be necessary. Advice on diet, drug consumption, and other health matters may be given, and problems discussed.

It might be thought that if the first antenatal visit screening were performed by specialist obstetricians in a teaching hospital with an accepted unit policy then subsequent obstetric difficulties in those selected as abnormal could be reduced to a minimum, and women selected as normal could anticipate a problem-free pregnancy and confinement. We report our attempts to discover whether this was so in a city where general practitioners refer all pregnant women to specialist obstetricians.

Department of Obstetrics and Gynaecology, University of Aberdeen, Aberdeen AB9 2ZD

P K CHNG, MB, MRCOG, registrar
MARION H HALL, MD, MRCOG, consultant
I MacGILLIVRAY, MD, FRCOG, regius professor

Methods of study and population

This study was a retrospective analysis of the case records from the city district of Aberdeen of all 2186 women who delivered in 1975. Details of maternal characteristics, past obstetric and medical history, abnormal findings and investigations at all antenatal visits (both to hospital and general practitioner), and complications of the entire pregnancy were obtained by an experienced obstetric registrar (PKC) and punched on to computer tape for analysis.

Initially 2186 women (including 23 sets of twins) were included in the study. We excluded 11 because they had no antenatal care at all and 268 because their general practitioner records were not available. Thus the complete analysis was made for 1907 women.

The options open to the obstetricians arranging antenatal care were (*a*) *hospital care*, where women would be seen usually by an obstetrician of registrar grade or above but occasionally by senior house officers in training or (*b*) *combined care*, where women would attend hospital only at booking, at 34, and at 40 weeks' gestation, and attend her general practitioner for all other visits. The options for booking for