ABSTRACT—Statistical malpractice is an insidious, and indeed prestige-laden and grant-rewarded, activity. Brilliantly clever, but fundamentally wrong-headed, number-crunchers are encouraged to devise inappropriate applications of mathematical methods to health problems. This species of misdirected zealot has so far been immune from criticism.

Science versus statistics

There is a worrying trend in academic medicine which equates statistics with science, and sophistication in quantitative procedures with research excellence [1]. The corollary of this trend is a tendency to look for the answers to medical problems from people with expertise in mathematical manipulation and information technology, rather than from people with an understanding of disease and its causes. The seeking of algorithms for scientific decision-making is an offence best described as statistical malpractice.

Epidemiology and cognate disciplines such as health economics are the main source of culprits, because statistical malpractice typically occurs when complex analytical techniques are combined with large data sets. The mystique of mathematics blended with the bewildering intricacies of big numbers makes a potent cocktail.

Malpractice must be distinguished from incompetence. There is a modest but thriving cottage industry whereby statisticians provide medical journals with articles describing how clinicians can avoid the pitfalls consequent upon the incompetent use of statistical methods, of which this is one.

The relationship between science and statistical analysis in medicine is quite simple: statistics is a tool of science which may or may not be useful for a given task. Indeed, as a general rule, the better the science, the less the need for complex analysis, and big databases are a sign not of rigour but of poor control. Basic scientists often quip that if statistics are needed, you should go back and do a better experiment. However, the ethical constraints of medical science mean that the ‘better’ experiments are often impossible, and we are stuck with statistics [2].

Science is an inextricable combination of the rational and the empirical: rational structures of entities and processes created in accordance with, and tested against, knowledge of the perceived world by means of iterative cycles of causal theorising together with purposive observation and experiment [3,4]. Statistics, on the other hand, is based upon axioms. Pure statistical analysis is therefore, in an important sense, a neutral activity, a set of summarising techniques that make only mathematical assumptions. Much of the prestige of the discipline derives from this objectivity: properly done, statistical techniques should not lead to false inferences.

Science is concerned with causes but statistics is concerned with correlations. While both science and statistics aim to reduce error to reveal underlying order, they have different emphases. The prime concern of science is with minimising systematic error (eliminating bias), while that of statistics is with minimising random error (maximising precision). Minimisation of bias is a matter of controlling and excluding extraneous causes from observation and experiment so that the cause of interest may be isolated and characterised. When systematic error is controlled, we can compare like with like. Maximisation of precision, on the other hand, is the attempt to reveal the true magnitude of a variable which is being obscured by the ‘noise’ consequent upon limitations of measurement.

The perils of averaging

The particular value of statistics to medical science lies in its ability to increase the precision of an estimate by dealing with random error. This does not necessarily require advances either in understanding or in techniques of measurement. Random errors will cancel out around the true value if enough instances are included. Averaging of repeated measures of the same instance can therefore increase precision, and the higher the number of repeated measures the greater the confidence that the average represents the true value. The problem is that truly random variation is seldom (or never) seen in biological systems. Therefore, averaging as a means of increasing precision is a somewhat hazardous procedure in biology (except where purely physical measurements are concerned) and the use of averaging needs to be justified whenever employed [5]. Averaging should be undertaken only when there is a qualitative similarity between the instances to be averaged; in other words, when it is reasonable on scientific grounds to assume that the instances will differ from one another only because of limitations of exactness in the measuring process [2].

Averaging of qualitatively different instances (where systematic differences have not been eliminated) leads to the production of non-existent artefacts akin to the 2.4 children of an ‘average’ family, or the ‘average’ quality of life enjoyed by a population containing one half dead people and the other half healthy individuals.
Such artefacts make sense only at the level of the group, and they arise because statistical analysis is not causal but correlative.

**Science by sleight-of-hand**

The root of most instances of statistical malpractice is the breaking of mathematical neutrality and the introduction of causal assumptions into analysis without justifying them on scientific grounds. This amounts to performing science by sleight-of-hand: the quickness of the statistics deceives the mind. The process is often accidental, the product of misunderstanding rather than of malice—as commonly happens when statistical adjustments or standardisation of populations are performed to remove the effects of confounding variables [6]. These are manoeuvres by which data sets are recalculated (eg by stratified or multivariate analysis) in an attempt to eliminate the consequences of uncontrolled ‘interfering’ variables which distort the causal relationship under study. For instance, standardising for age in a study of smoking and cancer may involve imbalances in the population age structure of smokers and non-smokers being ‘corrected’ using statistical manipulations.

There are, however, no statistical rules by which confounders can be identified, and the process of adjustment involves making quantitative causal assumptions based upon secondary analysis of the database in question. This may not matter much if the objective is simply to provide a summary statistic for the purpose of comparison, but is illegitimate as part of a scientific enquiry because it amounts to a tacit attempt to test two hypotheses using only a single set of observations.

Adjustment is therefore, implicitly, a way of modelling the magnitude of a causal process in order to subtract its effects from the data. However, modelling is not mathematically neutral and involves importing assumptions—an activity which requires to be justified for each case. Before statistical ‘correction’ is applied, the quantitative model describing the causal relationship between age, smoking and cancer should itself be replicated by testing against other independent sources of data. Yet this stage is typically omitted, presumably because adjustment is considered to be a purely mathematical, and therefore, neutral, procedure.

This question of modelling provides a useful distinction between statistics and science. A scientific model is built up from theories concerning causal assumptions, and has itself the character of a hypothesis with brings with it the need for replication. By contrast, statistical models are fundamentally post hoc and atheoretical, being derived from data sets by mathematical algorithms, such as constructing a ‘best fit’ curve. Examples of statistical models include ‘regression lines’, widely used in health economics for purposes such as generating equations for formula funding. The validity of such a model is dependent upon the sampled data set being a representative sample (or microcosm) of the ‘target’ data set to which it will be applied [3]. Generalisation is legitimate in so far as the sample is adequate in size and unbiased (ie the sample has the same relevant causal variables present in the same proportions). A statistical model does not, therefore, advance understanding, being of the nature of a summary of observations rather than a theory of how the world is: validity is derived wholly from the data fed into it.

**Practising science without a licence**

Statistical malpractice has an almost limitless potential for abuse, particularly in a context of the contemporary shift in emphasis away from curiosity-led science towards medical research driven by the imperatives of health policy and management. The difficulty is that statistical analysis is, apparently, applicable to any problem—just give it enough numbers and it will generate an answer. As a result, medicine has been deluged with more or less uninterpretable ‘answers’ generated by heavyweight statistics operating on big databases of dubious validity. Such numerical manipulations cannot, in principle, do the work of hypothesis testing. Yet this is precisely the use for which they are promoted. Meta-analysis is the most egregious example. The technique has been used, not just to provide a quantitative summary estimate of an effect—which is all it can legitimately do—but as an instrument for making scientific decisions. Thus, meta-analysis has been employed to determine whether or not a putative hazard is the cause of a disease and whether a medical intervention has a ‘significant’ therapeutic benefit (where ‘significant’ is taken to mean ‘real’). A tactic for narrowing confidence intervals has thus been confused with the expansion of scientific knowledge.

The hope is that precise qualitative information will enable human affairs to be managed so that health will be promoted even without knowledge of its causes [7, 8]. Statistical malpractice occurs, however, exactly because it has not dispensed with causation, but has merely concealed it under a cloak of mathematical neutrality. Medical research has become distorted by the quest for a level of accuracy in estimation which is attainable only at an inappropriate population level of analysis, and at the cost of ignoring the massive distortions of systematic error [4].

Statistical analysis has expanded beyond its legitimate realm of activity. The seductive offer of precision without the need for understanding is a snare to the incautious because exactitude is so often mistaken for explanation. Numerical technicians are now promoted to the status of general purpose experts on research methods, and the ultimate arbiters of the academic refereeing process. From the standpoint of medicine, this is a mistake: statistics must be subordinate to the requirements of science, because human life and
disease are consequences of a historical process of evolution which is not accessible to abstract mathematical analysis [9].

We are in danger of losing our grip on the profound insight that nature is mastered—and health promoted—not by force of numbers, but by the discovery of unity in hidden likeness and by learning how knowledge may be ordered such that it ‘commands more of the hidden potential of nature’ [3].

References
1. Altman DG. The scandal of poor medical research. Br Med J 1994;308:283-4.
2. Charlton BG. Mega-trials: methodological issues and implications for clinical practice. J R Coll Physicians Lond 1995;29:96-100.
3. Bronowski J. Science and human values (revised edn.) New York: Harper Colophon, 1975.
4. Charlton BG. The scope and nature of epidemiology. J Clin Epidemiol (in press).
5. Bernard C. An introduction to the study of experimental medicine. New York: Dover, 1865 (1957 edn).
6. Last JM. A dictionary of epidemiology. New York: Oxford University Press, 1988.
7. Rose G. The strategy of preventive medicine. Oxford: Oxford University Press, 1992.
8. Charlton BG. A critique of Geoffrey Rose’s ‘population strategy’ for preventive medicine. J Roy Soc Med 1995;88:607-10.
9. Crick F. What mad pursuit: a personal view of scientific discovery. London: Weidenfeld and Nicolson, 1989.

Address for correspondence: Dr Bruce G Charlton, Lecturer, Department of Psychology, University of Newcastle upon Tyne, NE1 7RU.

AUDIT MEASURES FOR GOOD PRACTICE IN Blood transfusion medicine

In the quality assurance of blood transfusion, little attention has been paid to the clinical interface of blood transfusion. Laboratory performance is monitored by the National External Quality Assessment Scheme (NEQAS) for Blood Group Serology and recommendations for good transfusion practice are made through the Blood Transfusion Task Force of the British Committee for Standards in Haematology (a subcommittee of the British Society for Haematology), but there is no systematic audit of the quality of the clinical transfusion process. To address this, the Research Unit of the Royal College of Physicians convened a workshop in 1992, to develop audit measures for the promotion of good practice in transfusion medicine. Membership of the workshop included representatives of the British Society of Haematology, the British Blood Transfusion Society and the Royal College of Pathologists. Nine audit proformas were produced and these have now been published by the Royal College of Physicains as a joint document from the four participatory bodies.

The audit proformas cover the following aspects of the clinical blood transfusion process:
- Institutional audit for blood transfusion practice
- Blood transfusion documentation
- Monitoring acute blood transfusion reactions
- Management of anaemic patients transfused with red blood cells
- Long-term blood transfusion therapy
- Management of transfusion in acute (non-surgical) haemorrhage
- Acute plasma volume replacement
- Management of patients presenting with an acute haemorrhagic diathesis
- Use of blood components in the newborn.

Price £7.50 (including p&p), £11.00 overseas. ISBN 1 86016 027 1

A4 report, 24 pages. Published NOVEMBER 1995

ROYAL COLLEGE OF PHYSICIANS OF LONDON