Peer reviewed article

Statistics in medical research

P. Sprent

Division of Mathematics, University of Dundee, Scotland

Summary

The role of statistics in medical research starts at the planning stage of a clinical trial or laboratory experiment to establish the design and size of an experiment that will ensure a good prospect of detecting effects of clinical or scientific interest. Statistics is again used during the analysis of data (sample data) to make inferences valid in a wider population.

In simple situations computation of simple quantities such as *P*-values, confidence intervals, standard deviations, standard errors or application of some standard parametric or nonparametric tests may suffice. Despite their wide use even these simple notions are sometimes misunderstood or misinterpreted by research workers in other disciplines who have only a limited knowledge of statistics.

More sophisticated research projects often need advanced statistical methods including the formulation and testing of mathematical models to make relevant inferences from observed data. Such advanced methods should only be applied with a clear understanding both of their purposes and the implication of any conclusions based upon their use.

Close collaboration between statisticians, whether professionals in that field or medical research workers with a sound statistical background, and other members of a research team is needed to ensure a seamless integration of the statistical elements into the reporting and discussion of research outcomes. Some suggestions are made as to how that collaboration is best achieved.

Key words: estimation; hypothesis testing; medical statistics; power; precision

Introduction

Use of statistical methods in medical research began more than 150 years ago. Florence Nightingale worked to improve the methods of constructing mortality tables. She was a fellow of the Royal Statistical Society and an honorary member of the American Statistical Association. John Snow applied simple statistical methods about the same time to support his theory that contaminated water was the source of a London cholera epidemic in 1854.

Statistics is now an integral part of most medical research projects. Major journals reporting research employ either a statistical adviser or use referees with statistical expertise when needed. These safeguards usually detect and sometimes make possible rectification of statistical deficiencies, but this is little more than a fire fighting exercise that can be avoided if authors seek statistical help both before the start of a research project and when interpreting data.

Both medical science and statistics have developed to a stage where there are a dwindling number of people with expertise in both areas, although most undergraduate medical courses include some training in elementary statistics and a few medical research workers are skilled users of quite sophisticated statistical techniques – usually ones relevant to their particular field of interest (eg, methods for analysing survival data). However, the statistical abilities of many doing clinical trials or laboratory experiments run to little more than carrying out simple procedures like a *t*-test, a Mann-Whitney test or a chi-squared test or perhaps a simple analysis of variance to produce the all-too-often-misunderstood *P*-values; they may also compute and quote standard deviations, standard errors, confidence intervals, correlation coefficients or fit a regression line.

In an appropriate context such calculations are useful, but they represent only the tip of a statistical iceberg. Sadly, inappropriate use of even simple methods sometimes occurs because the user has not understood the meaning or relevance of concepts like *P*-values, significance differences, confidence intervals. Useful as these tools often are at the preliminary stage of data analysis an armoury of more sophisticated statistical concepts must often be called upon to extract maximum relevant information from data.

Misuse or misunderstanding of even simple

This paper is an expanded version of a note distributed to participants on the course *Science Writing* organised by the Research Section of the Swiss Surgical Society at Grandhotel Giessbach, 3855 Brienz, Bern, Switzerland, 8–10 May 2003.

concepts is often accompanied by evidence that statistical advice has not been sought at the planning stage. A consequence may be that resources are wasted either because the trial or experiment is too small to have any hope of determining whether or not a potentially important treatment response occurs, or alternatively waste may result from carrying out a larger trial than is needed to resolve the point at issue. Statisticians can give useful guidance on optimum size and design.

I use a simple example in the next two sections to illustrate some of the strengths and limitations of simple statistical concepts and discuss some common misconceptions. The role of more advanced statistical concepts in medical research is considered briefly in the penultimate section and in the final section some comments are made on collaboration between statisticians and medical research workers.

Basic statistical concepts

Statistics has two roles in laboratory experiments and clinical trials

- The first is at the planning stage to ensure sound experimental design and optimal use of resources. Adequate statistical analysis of experimental data is usually only possible if the design is statistically sound.
- The second is in the analysis of results. Assertions based on experimental data need to be backed by a relevant statistical analysis. Even for pilot or for retrospective studies some statistical analysis is usually needed.

Papers are often submitted that give inadequate statistical analyses or present results of these in an inappropriate way. This may be due to poor collaboration between research teams and statisticians resulting in either (i) a statistician performing inadequate or even inappropriate analyses because he or she is not clear about the questions the research is designed to answer or (ii) the statistical content of the paper is potentially confusing or even misleading because the authors do not fully understand some of the statistical techniques being used.

The review by Murray [1] covers many aspects of the role of statistics in research methodology and useful guidelines for presenting statistical information are given by the same author in [2].

The technicalities of carrying out any but the simplest statistical analyses are mainly a matter for a trained statistician and computations may be catered for by good statistical software, but understanding the meaning of statistical aspects of reported results is important both for those conducting research and for readers of published reports.

Cynics sometimes claim statistics can prove anything. In fact, statistics can prove *nothing*. At a basic level it provides rational measures to reflect the degree of uncertainty associated with data-based assertions. At a more sophisticated level it provides indicators for how well data conform to some specified mathematical model, eg, tests goodness of fit to the model, and when appropriate provides estimates of certain constants or parameters in a model.

A simple hypothetical example explains some basic concepts. Suppose some standard surgical procedure gives rise to a post-operative complication in less than 10% of all cases and it is suggested that a procedural modification might reduce the incidence rate of complications. A simple clinical trial might be proposed in circumstances where a large number of cases may be expected over a reasonable time period, eg, about 1000 cases per year. Calling the standard procedure A and the modified procedure B, random allocation of patients to procedures is appropriate to avoid bias. Suppose that 914 patients are involved and the randomisation procedure results in 435 being subjected to procedure A and 479 to procedure B. Here, before the trial starts a statistician is assumed to have only a minor role of recommending randomisation, but as explained in the next section statisticians may also have to advise on the appropriate size (eg, number of units or cases) needed to ensure that useful inferences can be made. The statistician might also suggest possible alternative designs involving, say, matching of patients to avoid the potential influence on the complication rate of factors other than the procedure used.

Suppose that the hypothetical trial comparing procedures A and B gave the results in Table 1.

From these data we calculate the rate of complications for procedure A as a percentage, ie, $(36/435)\times 100 = 8.28\%$, and similarly that for procedure B is 5.01%. On this evidence, intuition suggests that the long-term rate is probably lower for procedure B. But is the observed decrease of more than 3% a reasonable reflection of what might be expected in all conceivable cases that are similar to the 914 recorded?

In statistical jargon the patients (units) in the trial form a *sample* and the totality of all conceivable 'similar' patients likely to be subjected to one or other of these procedures form a *population*. Using sample information to draw conclusions relevant to a population is described as making *statistical inferences*.

A random sample may broadly reflect the situation in the population, but there is always variation from sample to sample (referred to as *sampling variation*). For example, if the above study

Table 1
Numbers of patients exhibiting or not exhibiting a post-operative complication after two surgical procedures.

	Complication	No complication	Total
Procedure A	36	399	435
Procedure B	24	455	479
Total	60	854	914

Table 2
Numbers of patients exhibiting or not exhibiting a post-operative complication after two surgical procedures in a new sample.

	Complication	No complication	Total
Procedure A	31	374	405
Procedure B	23	387	410
Total	54	761	815

were followed up with another involving a further 815 patients we might get the results in Table 2.

The complication rates based on Table 2 for procedures A, B respectively are 7.65% and 5.61%, the difference now just over 2%. Such differences between the calculated percentages based on the data in Table 1 (first sample) and those in Table 2 (second sample) would, or should not, surprise us when we are dealing with human beings, whose characteristics as Everitt [3] has pointed out, are those succinctly described by Efron [4]: "there could be no worse experimental animals on earth than human beings: they complain, they go on vacations, they take things they are not supposed to take, they live incredibly complicated lives, and, sometimes, they do not take there medicine."

Suppose now we have only the information in Table 1. Given this, statistics cannot prove that the difference between complication rates for the two procedures will be just over 3% or any other value. It can give us indicators that reflect the plausibility of values that may be of interest. Two such indicators are a *P*-value and a 95% confidence interval, although we do not always get or want both of these, or we may instead be given or prefer some other closely related quantity.

There are some common misconceptions about the meaning of *P*-values and confidence intervals and also a tendency sometimes to attribute to their calculated values a degree of precision that is not justified. This lack of precision arises because statisticians usually have to make further assumptions that are only approximately true in order to calculate the relevant measures. These may include an assumption that observations are normally distributed, and that adequate randomisation procedures have been carried out, or that observations are independent of one another. Some or all of these assumptions may not be strictly true or even hopelessly false. Statisticians understand these matters and are aware of how and when breakdowns in assumptions may have consequences likely to be seriously misleading.

One may analyse the data in Table 1 to calculate a *P*-value and 95% confidence interval for the population rate difference between the two procedures, assuming that the proportions observed in samples are normally distributed. Alternatively, we might drop the normality assumption, because that assumption is known to be only approximately true when dealing with proportions (which is what the percentage rates here essentially are). How good the approximation will be depends on both the sample sizes and the actual percentages.

An analysis assuming normality gave $P = 0.047$ and a 95% confidence interval for the percentage rate difference (0.04%, 6.51%). An analysis without a normality assumption gives $P = 0.060$ and a 95% confidence interval (–0.4%, 6.81%).

To explain the implications of differences between the calculated values for the two analyses, one needs to be clear about what *P*-values and confidence intervals mean. The *P*-value is relevant to hypothesis testing, where a key concept is that of a *null hypothesis*. In this example the null hypothesis is that in the populations corresponding to each procedure the complication rate is the same. What the *P*-value tells us is the probability of getting a difference (in either direction) as great or greater than that observed, ie, of magnitude 8.28–5.01 = 3.27 or more when the null hypothesis of no population difference is true. It is not, as is sometimes mistakenly assumed, the probability that the null hypothesis is correct

A widely used convention is to reject the null hypothesis if the observed $P \leq 0.05$. The reasoning here is that if $P \leq 0.05$ we have in just one sample obtained a difference with the property that it or a larger one can be expected in at most about one sample in twenty (since 1/20 = 0.05) if the null hypothesis were true. It then seems more reasonable to accept a hypothesis that says the complication rates differ between procedures. This may not be true, but the odds are strongly in favour of it. In statistical jargon we say there is a *significant difference* or that we *reject the null hypothesis*. If $P > 0.05$ the convention is to continue to accept the null hypothesis. This does not mean the null hypothesis is true, but only that we do not feel we have enough evidence to be confident about rejecting it.

The 95% confidence interval is in essence an interval having the property that if we were to take lots of samples from our population and apply the rules for calculating that interval to the data from each sample, then 95% of all such calculated intervals would include the true fixed (but unknown) population value. The population value is unknown because if we knew it we wouldn't be wasting time estimating it!

P-values and confidence intervals are related. If we took as the null hypothesis *not* that the difference is zero but instead that it is some chosen numerical value lying anywhere within in the 95% confidence interval calculated for our particular sample, then the *P*-value calculated for that new null hypothesis would exceed 0.05. If we took as the null hypothesis any value for the difference lying outside the confidence interval, then the *P*-value would be less than 0.05. In particular. if the confidence interval does not include zero, when we take zero as the null hypothesis, we will find $P < 0.05$.

In the case of the first analysis in the numerical example above the confidence interval (0.04%, 6.51%) does not include zero so we reject the null hypothesis of zero difference. This is consistent with the observed P = 0.047. For the second analysis the confidence interval (–0.4%, 6.81%) includes zero so we do not reject the zero null hypothesis, consistent with the calculated $P = 0.060$ found in that analysis. Thus, in this example the convention of accepting $P = 0.05$ as a cut off mark for deciding whether to accept or reject the null hypothesis results in different decisions depending on whether or not we make an assumption that proportions are normally distributed! This contradiction is alarming only to anyone with little statistical experience. Trained statisticians know that statistical analyses are often sensitive to the assumptions made in conducting them, and they regard $P = 0.05$ as no more than a convenient and conventional yardstick for measuring the strength of evidence for or against the null hypothesis. If P is appreciably less than 0.05 when any reasonable statistical assumptions are made the evidence against the null hypothesis is strong and justifiable modification of assumptions is not likely to push a much lower P-value above 0.05. Similar arguments with obvious changes in wording apply to P-values appreciably greater than 0.05. It is only when analyses give P-values close to 0.05 (say, between approximately 0.04 and 0.06) that adding or dropping reasonable assumptions or approximations are likely to give P-values on opposite sides of $P = 0.05$. This means that if P is close to 0.05 some flexibility of interpretation may be appropriate.

In the above example in terms of confidence intervals in the first analysis zero falls just outside the 95% confidence interval, and in the second example only just inside the interval, suggesting that the analysis shows that zero has a fairly weak claim (ie, seems not very likely) to be the true difference.

The confidence interval is generally a more useful statistical concept than the P-value, although, specially when the P-value is much less than or much greater than 0.05 it is sometimes useful to have both. The explanation for the awe associated with $P = 0.05$ dates back to the pre-computer era when it was usually virtually impossible to calculate exact P-values, and one could only make a limited range of statements often specifically limited to whether $P \geq 0.10$, or lay somewhere between the discrete values $P = 0.10$, $P = 0.05$, $P = 0.01$ or else $P < 0.001$.

We have only discussed results for one analysis that assumed normality and one that did not. Taking the Table 1 data to another statistician or using different statistical software is likely to produce slightly different P-values and confidence intervals. For example, making normality assumptions but adding a correction which reflects the fact that for proportions based on counts the normal assumption is only approximate and tends to underestimate P, gives $P = 0.063$, close to the value given by the second analysis. Some computer software packages make this adjustment automatically, some allow it as an option, some ignore it. If in doubt ask a statistician, something it is always wise to do if it is not clearly explained what a computer program does, or if you do not understand the implications of what it does.

The above example shows how statistics can provide useful guidelines in a simple situation, while at the same time it warns of a lack of precision in the sense that all analyses rely to some extent on what assumptions can reasonably be made.

In more complex situations a further danger is that of over-analysis; calculating lots of P-values and doing separate analyses of parts only of the data, this last being particularly dangerous if the parts are chosen after one has seen the data, when it often leads to misleading claims. Some examples are given in [1], pp. 779–81.

From a clinician's viewpoint there is another consideration that outweighs any discussed so far. This is the difference between *statistical significance* and *practical importance*. This must be taken into account at the planning stage of any research project, for it is crucial to plan a laboratory experiment or clinical trial so that there is a good chance of detecting effects of practical importance if these exist, but to do so without wasting resources.

The size of a trial is vital to its prospects of detecting differences of some pre-specified magnitude. Studies with only a small number of units (eg, patients) will usually only detect large departures from a value specified in a null hypothesis. Some studies may be too small to detect any difference, no matter what plausible null hypothesis is specified. If a new treatment is used for only two patients where one dies and one survives the sample mortality rate is 50% but all one can say is that the population rate is neither 0% or 100%, but it could be anywhere in between. Large experiments enable us to make more precise estimates or tests. For example, if in 1000 cases there are 500 deaths, the sample mortality rate is still 50% but now a 95% confidence interval for the population mortality rate is 46.9% to 53.1%. For 10 000 cases with 50% mortality observed the 95% confidence interval reduces to 49.02 to 50.98%. The larger the sample size, the shorter the confidence interval and hence the smaller the differences from a null hypothesis that imply a significant difference.

Power

In practice clinicians are often not interested in extremely small differences. In the complications example in the previous section a surgical unit might only be interested in switching from procedure A to procedure B if there was strong evidence that the rate reduction for complications was at least 2%. Such a decision is not a statistical matter but one based primarily on factors such as good medical practice including ethical considerations, pressure on facilities and patients' welfare. If the complication caused little patient discomfort, only marginally prolonged hospital inpatient time, only slightly increased the workload on staff and procedure B cost more to carry out, or operating times were longer, it might be decided that a change from A to B for all cases would only be appropriate if at least a 2% long term reduction in complication rate were achieved.

In very broad terms the trial involving 914 patients discussed above showed, on the basis of a confidence interval, that there is merely a good chance that the reduction would be somewhere between about 0 and 6.5 per cent. It might be 2 per cent or more, but the trial was not big enough to establish with great confidence that there is that reduction – it could well be either greater or less than 2%. This is a somewhat indeterminate situation.

In an ideal world if the true reduction were at least 2% we would like our sample to indicate this, but we would have no wish to detect decreases appreciably less than 2%. We cannot achieve this ideal. What we can do, if we have relevant information about such factors as, in our example, the approximate rates of complication, the randomisation procedures used, etc., is to work out the sample sizes needed to give us, say, an 80% chance of detecting a difference of 2% if it really exists (and an even greater chance of detecting larger differences).

Statistical techniques for doing this are well established, although the actual calculations require a certain amount of expertise or availability of suitable computer programs. The 80% chance of detection (more usually expressed as the corresponding probability of 0.8) is called the *power* of the test.

In statistical jargon one is said to make an *error of the first kind* if the null hypothesis is rejected when it is in fact true. If we adhere to the decision to reject the null hypothesis when $P \leq 0.05$ then the probability of an error of the first kind is 0.05. If $P > 0.05$ we accept the null hypothesis, even though it may not be true. Accepting the null hypothesis when it is not true is called *an error of the second kind*, The probability that we make an error of the second kind depends upon how much the true difference departs from that stated in the null hypothesis. Errors of the second kind are closely related to the concept of power and often appear in statistical literature where power is discussed. The relationship is

Power = 1 – *probability of an error of the second kind*.

In practice for ethical, clinical or other practical reasons such as availability of patients, cost, or limited facilities, it may be difficult – even impossible – to carry out a clinical trial with power as high as 0.8. This is often the case when we are interested in low incidence numbers in large samples, as was the situation in the example based on Table 1.

Applying the appropriate power calculation to that situation shows we would need about 2650 patients allotted to each procedure to have a power 0.80 (ie, an 80% chance) for detecting a population 2% difference if it really existed. However, if the true difference were 3% the power with samples that size would be 0.99 (99%).

Compromises often have to be made because of resource limitations, etc. In the above example power calculations show there is a 75% chance of detecting a 3% difference in complication rates with samples of 1000 for each procedure. There would then still be a 38% chance of detecting a population difference of 2%

Power calculations are sensitive in an example like this to assumptions about the percentage of complications, etc. Pilot studies with relatively small samples often provide valuable information for estimating the power associated with experiments using larger samples. Power studies guide researchers towards an ideal size for a clinical trial and protect against waste of time and resources by carrying out trials that have little chance of detecting important differences because the trial is too small, or at the other extreme wasting resources by carrying out unnecessarily large experiments.

Brief comments on other simple aspects of statistics

In addition to those already introduced, most research workers meet statistical terms and concepts like normal distribution, binomial distribution, non-parametric methods, analysis of variance, long-tail distribution, exponential distribution, correlation, regression. Familiar tests include the t-test, chi-squared test, Wilcoxon signed-rank test, Wilcoxon rank-sum test and Mann-Whitney test (the last two are different forms of the same test), and perhaps others such as the log-rank test

if survival data is of interest. Terms like standard error, standard deviation, range. interquartile range are also freely used but sometimes cause confusion

The Normal distribution is central to much statistical analysis of measurement data and there are sound mathematical reasons why this is so – at least as an approximation – providing we can make certain assumptions. The normal distribution is characterised by its mean and standard deviation. These values for a population are generally unknown in advance, but estimates of the population mean and standard deviation can be made from a sample. For an independent sample from a normal distribution the sample mean and standard deviation (the latter usually with a conventional small modification) are estimates of their unknown population counterparts, estimates that improve as the sample size increases. For samples of 30 or more, if roughly half the observations are smaller than the mean, and the scatter around the mean both above and below it is roughly the same, and about two thirds of the observations lie within one standard deviation from the mean and nearly all of the observations are within 2 standard deviations of the mean, and none are more than about three standard deviations from the mean, then it is reasonable to assume normality. These are only rough guidelines. One must distinguish between the standard deviation and what is often loosely called the standard error but is more correctly described as the standard error of the mean. The latter is a measure of how accurately the sample mean estimates the population mean. It is in fact the standard deviation divided by the square root of n, the number of observations in the sample.

Avoid using a notation like 4.53 ± 1.72 without further explanation – in fact, it is better not to use the notation at all – because some people use this for the *mean ± standard error of the mean*, some for *mean ± standard deviation* and some for *mean ± 95% confidence interval for the population mean* [2]. If the ± notation is used at all the last use, with an explanation, should be the preferred one, but many writers use one of the other meanings without explanation. Incidentally, the assertion that the 95% confidence interval for the population mean is given by

sample mean $\pm$ 1.96 $\times$ *(standard error of the mean)*

is only strictly valid for samples of size about 30 or more from a normal distribution although a mathematical theorem called the *Central limit theorem* allows some relaxation of the normality requirement. For smaller samples the above formula underestimates the confidence interval even for a normal distribution.

If data indicate an obvious departure from normality, means and standard deviations become less appropriate. It is then usually better to use the *median* (the middle value when the data are arranged in ascending order) rather than the mean to measure centrality or location and the *interquartile range* rather than the standard deviation to measure spread. This range is the difference between a value having one quarter of the ordered observations below it *(first quartile)* and a value with one quarter above it *(third quartile)*. A simpler description is to use the median and to indicate spread by quoting the least and the greatest sample value. An increasingly popular data summary is the *five number summary*, consisting of

[least value, first quartile, median, third quartile, greatest value]

The binomial distribution is often, but not invariably, relevant when inferences about proportions based on counts are made. For large counts and proportions not too close to 0 or 1 (or the corresponding 0 and 100 per cent) the sample distribution of proportions approaches a normal distribution, an assumption made above for some of the calculations in the complication-rate example.

When normality assumptions are clearly unreasonable resort is often made to *nonparametric methods*. If data are a sample from a normal or near-normal distribution nonparametric methods are generally less efficient for hypothesis testing and estimation than are tests based on an assumption of normality, but they are often appreciably more efficient and informative if a normality assumption is not justified. These tests all require some assumptions (though less restrictive than that of normality) for validity and advice of a statistician should be sought. There are several introductory books on nonparametric methods [5, 6] designed for statistics students or for research workers who have attended a basic course of some 20 lectures in statistics.

Correlation is a well known concept when studying relationships between measurements when two or more quantities are measured on each of a number of units. *Correlation coefficients* are often calculated, but these are only a measure of one particular kind of association, values of the coefficient just slightly below 1 in magnitude implying that in a scatter diagram the points will lie reasonably close to a straight line (calculated as a *regression line*) or at least that there is a one-way trend in the relationship, ie, either a steadily increasing or steadily decreasing trend, and not, for instance, one that at first increases and then decreases. A weakness of the correlation coefficient is that a value near zero may imply no association or it may equally well suggest an association that on a scatter diagram would be associated with some curve other than a straight line. It is also important to remember that association need not imply cause and effect.

For medical research workers with only a basic knowledge of statistics there are several good books that discuss in detail, using real-data examples, statistical methods especially relevant in medical research. A widely used one is that by Altman [7]. Other popular ones, with which I am less familiar, but which are generally well regarded both by statisticians and clinicians, are those by Bland [8] and by Campbell and Machin [9]. Such books are invaluable if statistical help is needed but a suitably qualified statistician is not readily available.

More advanced statistics

Use of one or more of the concepts discussed above is important and may be all that is needed for assessing the validity of data-base assertions in simple situations. Their widespread use in the medical and surgical literature has been broadly welcomed by medical scientists and statisticians. However, modern developments towards more sophisticated trials and also in methods of data analysis mean that these complex trials often require more advanced statistical analyses.

For example, survival analysis studies are increasingly important. Data for survival times are usually not normally distributed. Typically, a few patients have a short survival time after diagnosis or treatment, perhaps only a matter of days, the median survival time may be, say, 10 months, but a few will survive 3, 4, 5, 6, 7 or more years. There are often complications with survival studies due to loss of contact with patients during the follow-up period, and for practical reasons follow-up may only be feasible for a limited period at the end of which some patients still survive. Non-parametric analysis, or sometimes analyses based on the exponential distribution are then appropriate. There is a vast literature on statistical analysis of survival data taking into account such factors as to whether patient drop-out may be treatment related, whether all patients actually receive the treatment intended at the start of the trial, what allowances should be made if the follow-up study is discontinued when some patients still survive, do factors such as sex, age, obesity, etc., affect survival rates.

Topics familiar to professional statisticians such as logistic regression and Poisson regression, both of which are special cases of what are called generalised linear models are playing an increasing role on clinical studies. Logistic regression is relevant to studies when only two possible outcomes are of interest, (eg, recovery or death; improvement or no improvement; side-effects present or absent) and where the probabilities of these outcomes may depend on a number of factors such as age, obesity, blood-pressure, sex, time between diagnosis and treatment, etc.

A range of generalised linear models and survival analysis are just two of many topics that are discussed with many illustrative examples of medical applications that highlight some of the complications by Everitt [3] in a book designed primarily for clinicians with a strong grounding in statistics

Collaboration

Most clinicians recognise the need for statistical expertise but this may not be readily available. The ideal situation is either one where a research team includes a medical scientist with advanced training in and understanding of statistical methods likely to be relevant, or else one where the research is being carried out in an institution with a statistical unit that includes specialists in applications of statistics in medicine. If either of these situations pertains there should be few problems of a statistical nature unless there are extraneous difficulties such as personality clashes.

In many cases the ideal situations just outlined do not hold. A research team may have to rely either on limited statistical expertise among its own members or seek advice from a professional 'general practitioner' statistician who has no specialised experience of or training in medical statistics. In these situations the statistical input may be satisfactory but at other times it will be less than adequate. The latter may happen if researchers with limited statistical expertise over-estimate their statistical abilities or if the general-practitioner statistician fails to understand fully the purpose and aims of the experiment or trial, or cannot fully appreciate the implications of any complications. Such problems often reflect communication difficulties between the parties.

The best advice one can give to reduce mistakes resulting from statistical inadequacies is that statisticians should make it their business to seek clarification on aims, objectives, experimental techniques and known complications and equally that researchers should seek full clarification of any statistical terminology or methodology they do not understand from whoever is dealing with the statistical aspects. Here explanation of principles is usually more important than technical detail of how things are calculated.

Modern statistical computer software provides programs for applying many advanced statistical techniques such as the many particular cases of generalised linear models and the even more complex generalised additive models, as well as coping with complications such as drop-outs in follow-up studies in survival data or unwanted correlations or the occasional aberrant observations. This is something of a mixed blessing, for such programs cannot replace a statistician, although they often allow researchers to carry out many of the routine and otherwise time-consuming computations in a matter of seconds when these would take hours, days weeks or even years if carried out by humans. With a few exceptions most such programs do not provide sufficient information on line or in hard-copy manuals about what the program is doing to ensure safe use unless the user has either a full un-

derstanding of what is involved or else has access to expert statistical advice.

Where any but simple statistics in involved an essential ingredient is close collaboration between those (whether professional statisticians or not) who are providing the statistical advice and the researchers to assure a virtually seamless blending of the statistical elements with other aspects of the reporting of the research. Human nature being what it is, it is sometimes easier to say this than to achieve it, but if you are not happy with the statistical advice or help you have received during a project look for some other source of statistical input in future. Most statisticians want to be helpful. Like arrogant medics, arrogant statisticians are a dying race.

Correspondence:
Professor P Sprent
32 Birkhill Avenue
Wormit
Newport-on-Tay
Fife, DD6 8PW
Scotland
E-Mail: psprent@aol.com

References

1 Murray GD. Statistical aspects of research methodology. Br J Surg 1991;78:777–81.
2 Murray GD. Statistical guidelines for the British Journal of Surgery. Br J Surg 1991;78:782–4.
3 Everitt B. Modern Medical Statistics. A Practical Guide. London: Arnold; 2003.
4 Efron B Forward: Limburg Compliance Symposium. Statistics in Medicine 1988;17:249–50.
5 Hollander M, Wolfe DA. Nonparametric Statistical Methods, 2nd edn. New York: Wiley. 1999.
6 Sprent P, Smeeton NC. Applied Nonparametric Statistical Methods, 3rd. edn. Boca Raton: Chapman & Hall/CRC; 2001.
7 Altman DG. Practical Statistics for Medical Research. Boca Raton: Chapman & Hall/CRC; 1991.
8 Bland M. An Introduction to Medical Statistics, 3rd. edn. Oxford: Oxford University Press; 2000.
9 Campbell MJ, Machin D. Medical Statistics. A Commonsense Approach. 3rd. edn. Chichester: Wiley; 1999.

Swiss
Medical Weekly

Official journal of
the Swiss Society of Infectious disease
the Swiss Society of Internal Medicine
the Swiss Respiratory Society

The many reasons why you should choose SMW to publish your research

What Swiss Medical Weekly has to offer:

- SMW's impact factor has been steadily rising, to the current 1.537
- Open access to the publication via the Internet, therefore wide audience and impact
- Rapid listing in Medline
- LinkOut-button from PubMed with link to the full text website http://www.smw.ch (direct link from each SMW record in PubMed)

- No-nonsense submission – you submit a single copy of your manuscript by e-mail attachment
- Peer review based on a broad spectrum of international academic referees
- Assistance of our professional statistician for every article with statistical analyses

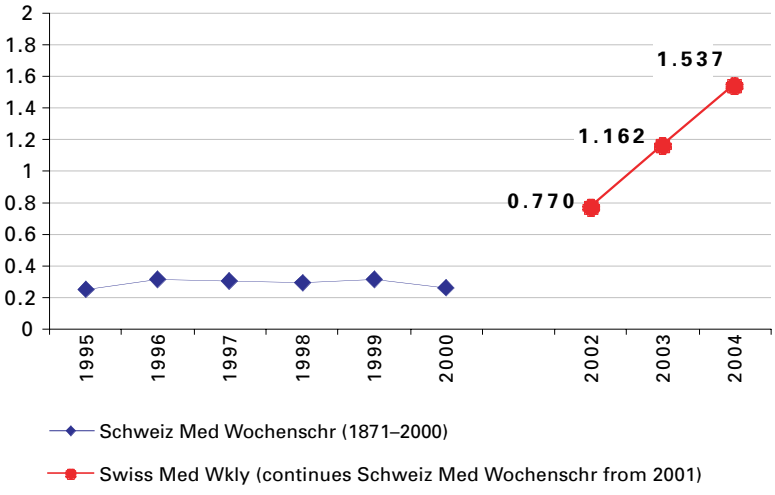
- Fast peer review, by e-mail exchange with the referees
- Prompt decisions based on weekly conferences of the Editorial Board
- Prompt notification on the status of your manuscript by e-mail
- Professional English copy editing
- No page charges and attractive colour offprints at no extra cost

We evaluate manuscripts of broad clinical interest from all specialities, including experimental medicine and clinical investigation.

We look forward to receiving your paper!

Guidelines for authors:
http://www.smw.ch/set_authors.html

Impact factor Swiss Medical Weekly



EMH FMH Schwabe
Editores Medicorum Helveticorum

All manuscripts should be sent in electronic form, to:

EMH Swiss Medical Publishers Ltd.
SMW Editorial Secretariat
Farnsburgerstrasse 8
CH-4132 Muttenz

Manuscripts: submission@smw.ch
Letters to the editor: letters@smw.ch
Editorial Board: red@smw.ch
Internet: http://www.smw.ch