Statistics in medicine: some considerations from a clinician’s point of view

Medical research is growing at an impressive rate, and an increasing number of papers are submitted to biomedical journals each year. As only a fraction of those submitted will finally be published, criteria for accepting medical research papers are becoming more and more stringent. An essential part of many of these papers is the statistical analysis whose final output is often a p-value. It is widely felt that biomedical journals are biased to accept preferentially papers reporting positive results. Accordingly, to obtain a p-value <0.05 seems to be a major goal for some investigators, while pharmaceutical companies want to endorse their enormous investments in drug research with significant p-values. Unfortunately, there is a tendency to forget that p-values are only tools for interpreting the results and not the results themselves. Some examples will illustrate this point.

Before conducting a comparative study between two interventions, a prior calculation of the sample size necessary to detect real differences between the two arms is usually required. The rationale for doing so is that if the number of patients is too low, the results will be nonsignificant, even if real differences exist, while if it is too high, the study will waste time and resources. However, the resulting sample size should be taken with caution.

For example, for a study designed to compare a new drug with placebo or a standard treatment, we want to know the number of patients required in each of the two arms. To make this calculation it is necessary to know the α and β errors, the expected rate of response in the control group (based on data from the literature or previous experiments), and the difference between the two treatments that we consider clinically relevant. With these data, many statistical software programs will yield a sample size which we assume is definitive. However, the values that we have entered into the computer are far from exact; they are based on our best guesses and may vary considerably, depending on the sources consulted. Therefore, the sample size arrived at should be regarded as giving a gross idea of the sort of numbers required for the study. The figure illustrates the different sample sizes necessary for a study depending on the rate of response that we have chosen for the control (or experimental) group. As it can be seen, small differences in our appreciation may result in marked differences in the number of patients required. In other words, the floor we are stepping on is not steady, but rather quicksand, and, consequently, we cannot be too confident on it.

Megatrials involving a large number of patients are common in certain clinical situations, such as the evaluation of the efficacy of antiretroviral drugs in human immunodeficiency virus (HIV) infection. The need to find new treatments for this serious disease and pressures from the pharmaceutical industry have promoted many trials involving thousands of patients each, which are expected to yield statistically significant results. By including these large numbers of patients, significant results will be obtained but, in many cases, they may not be too relevant to clinical practice. For example, a trial with 2000 patients in each arm that compares the efficacy of a new treatment versus placebo will yield statistically significant results if the response to placebo is 1% and that of treatment only 2% (p=0.009, χ² test Yates corrected). Despite the significance of the p-value, such a small difference is clinically irrelevant; even higher differences may be offset by other factors such as toxicity, drug interactions, cost, and impact on the quality of life.

Two groups undergoing different interventions can never have the same outcome. We would be able to ‘demonstrate’ either the efficacy or the deleterious effect of almost every drug by simply increasing the sample size or the number of statistical tests by subgrouping. Should we consider recommending as standard care for AIDS patients, say, daily treatment with cefotaxime, because a large megatrial involving thousands of patients has found a marginal but statistically significant increase in the survival time of only a few days? Of course not. The benefit for the patients would be minimal, and it would not counterbalance the many inconveniences of such a treatment.

On the contrary, a trial of treatment versus placebo for a relatively uncommon disease that includes only 10 patients in each arm, will not yield significant results if the response to placebo is 10% and that of treatment as high as 60% (p=0.06, χ² test, Yates corrected). Even a 80% difference between the two treatments will be statistically nonsignificant if only five patients are included in each arm. Although it is clear that we cannot regard the results of such a small trial as conclusive, they are more clinically significant that those of the above-mentioned megatrial. If another group of investigators reported similar results with a similarly...
small sample, we could be reasonably confident about the efficacy of the experimental treatment, regardless of its nonsignificant p-value. The absence of a statistically significant effect does not mean that this effect does not exist. Accordingly, many 'negative' studies should not be considered as such only because of nonsignificant results. What actually matters to patients and clinicians is the real difference between the two treatments and not the p-values, because statistical significance is not synonymous with clinical significance.

Another reporting strategy, commonly seen in abstracts and in HIV-related trials, consists of the description of the relative instead of the absolute benefit. Thus, if the rate of progression to an endpoint during a period of time is 6% in the control group, and that of the experimental group is 3%, the true difference of only 3% may be reported as a 50% improvement. This style of reporting does not differentiate between this quantitatively small difference and that obtained between two treatments that had a response rate of 45% and 90%, respectively, although, the meaning of the 50% improvement in these two experiments is very different. How information is framed greatly affects its interpretation, and both patients and physicians will have a tendency to choose the experimental treatment if the results are presented in relative as opposed to absolute terms.

Finally, another problem arises with the interpretation of meta-analyses. This epidemiologic procedure is a powerful tool for finding statistically significant differences by combining several similar trials in order to obtain a large sample size. One of the 'weaknesses' of meta-analysis resides precisely in its potency. If several large and well-designed trials have not found a conclusive benefit of one treatment over another, it is probably because the benefit is marginal or not very significant clinically; this clinical significance will not be substantially increased by a meta-analysis demonstrating statistically significant differences. I would not feel very comfortable if I had a disease whose treatment depended on the result of a meta-analysis.

So, how should we use statistics in clinical medicine? In my opinion, to help us analyse the data critically while we rely on clinical judgement. We work in a field full of uncertainties and consequently feel rather comfortable when we have an exact number or a p-value to help us to make decisions. However, numbers in biological sciences are not as exact as they seem, because many uncontrolled factors have a major influence on them. In fact, p-values are a measure of our degree of uncertainty when we make declarations, and they should not dictate, by themselves, our actions. Without doubt, statistics are necessary for both research and clinical practice, but they are only a tool, not an objective. Although, of course, it should not be taken literally, the message of the sentence "If your experiment needs statistics to demonstrate something, you ought to have done a better experiment" should not be forgotten.

JULIO COLLAZOS
Section of Infectious Diseases and Service of Internal Medicine,
Hospital de Galdakao, Vizcaya, Spain

Keywords: statistics; clinical trials; meta-analysis

Summary points

- p-values are useful for interpreting the results, but they do not constitute the results themselves; statistics are a tool, not an objective
- the absence of a statistically significant effect does not mean that this effect does not exist; absence of evidence is not evidence of absence
- meta-analysis is a powerful tool that is able to find statistically significant results by combining several trials. However, it is possible that these results may not be clinically relevant
- the way in which results are presented greatly affects their interpretation; results should always be expressed in absolute terms
- what matters is the actual difference between two interventions and not the p-value; statistical significance is not synonymous with clinical significance

References

Statistics in medicine: some considerations from a clinician's point of view.

J. Collazos

doi: 10.1136/pgmj.74.876.577

These include:

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/