1 Minerva. *BMJ* 1996; **312:** 1310.
2 Unwin NC. Cycle helmets—when is legislation justified? *J Med Ethics* 1996; **22:** 41–5.
3 Davis RM, Pless B. Evidence shows that cyclists should wear helmets. *BMJ* 1996; **313:** 629.
4 Davis A. Increasing the number of cyclists is more important. *BMJ* 1997; **314:** 69.
5 Waterston T. Motorists are the cause of the problem. *BMJ* 1997; **314:** 69.
6 Robinson DL. Australian laws making helmets compulsory deterred people from cycling. *BMJ* 1997; **314:** 69–70.
7 Keatinge R. Better to control the demand for fast cars. *BMJ* 1997; **314:** 70.
8 Hillman M. Health benefits of cycling greatly outweigh loss of life years from deaths. *BMJ* 1997; **314:** 70.
9 Hall GH. Do gooders' intolerance is counter productive to their aims. *BMJ* 1997; **314:** 70.
10 Robinson DL. Head injuries and bicycle helmet laws. *Accid Anal Prev* 1996; **28:** 463–75.
11 Thompson DC, Rivara FP, Thompson R. Effectiveness of bicycle safety helmets in preventing head injuries. *JAMA* 1996; **276:** 1968–73.

Introduction to statistics—1. The confidence interval

Robert W Platt

Alert readers will note that the Instructions to Authors state 'In statistical analyses 95% confidence intervals should be used where appropriate'. In applied statistics, it is customary to present the results of experiments in the form of p values. An alternative, with much to offer, is the confidence interval. In this column, I will define the confidence interval both in mathematical terms and in layman's language, and discuss appropriate and inappropriate uses of confidence intervals.

Put simply, the confidence interval (or confidence limits) can be thought of as a range of values or interval that contains the 'true value' 95% of the time. It can be argued that in fact hypothesis testing is a poor use of statistical techniques, and does not really give a researcher the information he or she needs.[1] Hypothesis tests (and p values) simply give information about the likelihood of a certain dataset under the null hypothesis. We are almost never interested in the null hypothesis, which usually is that there is no difference between two or several groups (or, in regression, that a dependent variable is not associated with a particular independent variable). What we are almost always interested in is the most plausible value of the true difference (or regression coefficient) and the degree of certainty in our estimation. Confidence intervals give both of these; hypothesis tests provide neither. So, in simple terms, a confidence interval can be thought of as an interval that includes the true value 95% of the time.

In statistical language, a 100 $(1-\alpha)$% confidence interval for a parameter q is given by (q_l, q_h) where the probability

$$P(q_l < q < q_h) = 1 - \alpha.$$

Thus a confidence interval is basically a formula to calculate the interval or range, which, when an experiment is repeated many times, the fraction of these intervals that cover the true value approaches $1-\alpha$. It is important to note that q is a fixed number—the true underlying parameter (or value)—and so it does not change and is not random. Therefore, the definitions do not refer to probability about q, but about its limits.

Using the formula enables us to say that based on our data, the interval gives the most plausible values for the true parameter. We say 'plausible' and avoid making probability statements about q because it is fixed and can't change. Remember too, that the statement depends completely on the data from your study.

What are the uses of confidence intervals and what do they allow us to do that other methods (hypothesis testing and p values) don't? First, unlike a hypothesis test, the confidence interval provides a measure of location—it gives the most plausible estimate of the true parameter, given the observed data. Second, it can be used to gauge significance—if a 95% confidence interval excludes the value 0 we conclude that 0 is not a likely value of the true parameter. Finally, and perhaps most importantly, a confidence interval gives information about the certainty of estimates and the amount of unexplained variability. A wide confidence interval implies a lot of unexplained variability and thus much uncertainty. In hypothesis testing, a small p value indicates a large effect, low variability, or both, but it is not self evident which is most important. Confidence intervals transparently give this information by showing the effect size AND the uncertainty.

Confidence intervals can also be used when designing a study to choose appropriate sample sizes. In the same way that you can calculate the sample size for a given power level for a certain effect size, you can also calculate the sample size for a given confidence interval width at a certain effect size, and hence specify the required level of precision.

In sum, there are many advantages to presenting data using confidence intervals, and some drawbacks to simply reporting p values. In addition, confidence intervals are easily calculated and available as options with most software packages. For these reasons, I recommend they be used routinely.

McGill/Montreal Children's Hospital Research Institute, 2300 Tupper, Montreal, PQ H3H 1P3, Canada

Correspondence to:
Dr Platt.

References and further reading

1 Oakes M. *Statistical inference.* Chestnut Hill, MA: Epidemiology Resources Inc, 1986.
2 Gardner MJ, Altman DG. *Statistics with confidence.* London, UK: BMJ, 1989.
3 Swinscow TDV. *Statistics at square one.* London, UK: BMJ, 1980.