	Treatment A	Treatment B	Risk Ratio (A:B =)
Number who died	4	6	(4:6 =) 0.67
	10	10	

Table 1. **Does this small study provide a reliable estimate of the difference between Treatment A and Treatment B?**

one group rather than the other. If the comparison was repeated in other small groups of patients, the numbers who died in each group might be reversed (6 against 4), or come out the same (5 against 5), or in some other ratio – just by chance.

But what would you expect to see if exactly the same proportion of patients in each treatment group (40% and 60%) died after 100 patients had received each of the treatments (Table 2)? Although the measure of difference (the risk ratio) is exactly the same (0.67) as in the comparison shown in Table 1, 40 deaths compared with 60 deaths is a more impressive difference than 4 compared with 6, and less likely to reflect the play of chance. So, the way to avoid being misled by the play of chance in treatment comparisons is to base conclusions on studying sufficiently large numbers of patients who die, deteriorate or improve, or stay the same. This is sometimes referred to as 'the law of large numbers'.

	Treatment A	Treatment B	Risk Ratio (A:B =)
Number who died	40	60	(40:60 =) 0.67
Out of (total)	100	100	

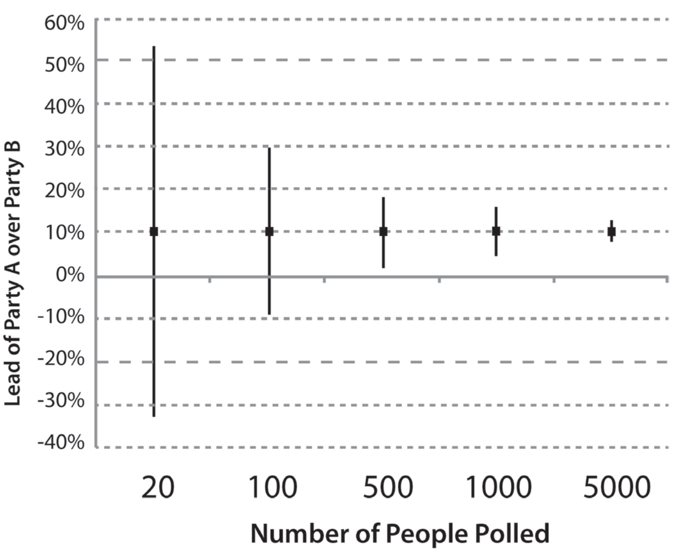
Table 2. **Does this moderate-sized study provide a reliable estimate of the difference between Treatment A and Treatment B?**

ASSESSING THE ROLE THAT CHANCE MAY HAVE PLAYED IN FAIR TESTS

The role of chance can lead us to make two types of mistakes when interpreting the results of fair treatment comparisons: we may either mistakenly conclude that there are real differences in treatment outcomes when there are not, or that there are no differences when there are. The larger the number of treatment outcomes of interest observed, the lower the likelihood that we will be misled in these ways.

Because treatment comparisons cannot include everyone who has had or will have the condition being treated, it will never be possible definitively to find the 'true differences' between treatments. Instead, studies have to produce best guesses of what the true differences are likely to be.

The reliability of estimated differences will often be indicated by 'Confidence Intervals' (CI). These give the range within which the true differences are likely to lie. Most people will already be familiar with the concept of confidence intervals, even if not by that name. For example, in the run-up to an election, an opinion poll may report that Party A is 10 percentage points ahead of Party B; but the report will then often note that the difference between the parties could be as little as 5 points or as large as 15 points. This 'confidence interval' indicates that the true difference between the parties is likely to lie somewhere between 5 and 15 percentage points. The larger the number of people polled, the less the uncertainty there will be about the results, and therefore



The 95% Confidence Interval (CI) for the difference between Party A and Party B narrows as the number of people polled increases.

the narrower will be the confidence interval associated with the estimate of the difference.

Just as one can assess the degree of uncertainty around an estimated difference in the proportions of voters supporting two political parties, so also one can assess the degree of uncertainty around an estimated difference in the proportions of patients improving or deteriorating after two treatments. And here again, the greater the number of the treatment outcomes observed – say, recovery after a heart attack – in a comparison of two treatments, the narrower will be the confidence intervals surrounding estimates of treatment differences. With confidence intervals, 'the narrower the better'.

A confidence interval is usually accompanied by an indication of how confident we can be that the true value lies within the range of estimates presented. A '95% confidence interval', for example, means that we can be 95% confident that the true value of whatever it is that is being estimated lies within the confidence interval's range. This means that there is a 5 in 100 (5%) chance that, actually, the 'true' value lies outside the range.

WHAT DOES A 'SIGNIFICANT DIFFERENCE' BETWEEN TREATMENTS MEAN?

Well, this is a trick question, because 'significant difference' can have several meanings. First, it can mean a difference that is actually important to the patient. However, when the authors of research reports state that there is a 'significant difference' they are often referring to 'statistical significance'. And 'statistically significant differences' are not necessarily 'significant' in the everyday sense of the word. A difference between treatments which is very unlikely to be due to chance – 'a statistically significant difference' – may have little or no practical importance. Take the example of a systematic review of randomized trials comparing the experiences of tens of thousands of healthy men who took an aspirin a day with the experiences of tens of thousands of other healthy men who did not take aspirin. This review found a lower rate of heart attacks among the aspirin takers and the difference was 'statistically significant' – that is, it was unlikely to