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

WHAT DOES 'STATISTICALLY SIGNIFICANT' MEAN?

'To be honest, it's a tricky idea. It can tell us if the difference between a drug and a placebo or between the life expectancies of two groups of people, for example, could be just down to chance . . . It means that a difference as large as the one observed is unlikely to have occurred by chance alone.

Statisticians use standard levels of "unlikely". Commonly they use significant at the 5% level (sometimes written as p=0.05). In this case a difference is said to be 'significant' because it has a less than 1 in 20 probability of occurring if all that is going on is chance.'

Spiegelhalter D, quoted in: *Making Sense of Statistics*. 2010. www.senseaboutscience.org

be explained by the play of chance. But that doesn't mean that it is necessarily of practical importance. If a healthy man's chance of having a heart attack is already very low, taking a drug to make it even lower may be unjustified, particularly since aspirin has side-effects, some of which – bleeding, for example – are occasionally lethal.[1] On the basis of the evidence from the systematic review we can estimate that, if 1,000 men took an aspirin a day for ten years, five of them would avoid a heart attack during that time, but three of them would have a major haemorrhage.

OBTAINING LARGE ENOUGH NUMBERS IN FAIR TESTS OF TREATMENTS

Sometimes in tests of treatments it is possible to obtain large enough numbers from research done in one or two centres. However, to assess the impact of treatments on rare outcomes like death, it is usually necessary to invite patients in many centres, and often in many countries, to participate in research to obtain