How many patients do we need?

1) Determine what is to be the 'Principal outcome measure'.

Often trickier than statisticians expect! There may be several measures of equal clinical importance. Which measure is the new treatment most likely to influence? It is better to calculate numbers for each measure than to try to combine measures into a single scale.

2) Judge a realistic 'true difference' between the treatments.

Possibly the most annoying question a statistician will ask. The response is usually "If I knew the difference I wouldn't need to do a trial!" What we mean is "How much difference would there really have to be to make it worthwhile concluding that there is a difference?" For example, if the current treatment has an 80% success rate, is it worth knowing that a new treatment has an 81% success rate? This will depend on all sorts of things: gravity of outcome; side effects; relative convenience to patients and staff; and finances. How big a difference would persuade you to change routine practice?

3) Select (arbitrary) statistical power and significance levels.

Even if there really is a difference between treatments, you may be unlucky and miss it. The more patients you study, the higher the chance you won't miss it. This chance is the **statistical power** of the study. Significance level is conventionally set at 5% ($p<0.05$). Similarly, power is conventionally at least 80%, sometimes 90%, and only rarely 95%. There is no logical reason for these conventions.

4) Use appropriate equation for chosen outcome measure.

Examples using the two most common equations are overleaf. You can twist almost all studies to fit one of these. Given the guesswork in coming up with realistic treatment differences, concern with exact equations for all possible research designs seems futile. However, a good book including many more equations and nomograms is available: *Machin, Campbel, Fayers & Pinol. Sample size tables for clinical studies. Blackwells.*

5) Wonder why required number of patients is so high!

If the numbers aren't much higher than anticipated then check your calculations! Is your estimate of 'true difference' wildly optimistic? Most research is hopelessly under-powered. Only with huge advances in care can small single centre studies

establish the benefits. If you can't recruit the required numbers you might consider this as pilot work to determine whether there is sufficient promise to justify major funding. Otherwise, if you really want the answer, you may need to find collaborators locally or nationally.

CALCULATIONS
Trials with n patients in each of two groups

1) Comparing proportion of 'successes'

$$\mathbf{n = \frac{(success\ on\ A \times failure\ on\ A) + (success\ on\ B \times failure\ on\ B)}{(success\ on\ A - success\ on\ B)^2} \times magic\ number}$$

2) Comparing means of normally distributed outcome

$$\mathbf{n = \frac{2 \times (standard\ deviation)^2}{(difference\ in\ means)^2} \times magic\ number}$$

Magic Number
This value is determined by the required power and significance level. For common values it is tabulated below:

		Statistical Power						
		50%	**60%**	**70%**	**80%**	**90%**	**95%**	**99%**
Significance level	**5%**	3.8	4.9	6.2	7.8	10.5	13.0	18.4
(p-value)	**1%**	6.6	8.0	9.6	11.7	14.9	17.8	24.0

EXAMPLE 1. Consider a trial looking for improvement from 85% success rate on the current treatment (A) to 90% on a new treatment (B), and requiring an 80% chance of detecting a difference at the 5% significance level ($p<0.05$).

From first formula and table,

$$\mathbf{n = \frac{(85 \times 15) + (90 \times 10)}{(85\text{-}90)^2} \times 7.8 \approx 680\ patients\ per\ treatment}$$

NOTE that this is per treatment. What if only 700 patients in total could be recruited - how would this affect the power of the study? Working the formula backwards, we have

$$\mathbf{350 = \frac{2175}{25} \times magic\ number}$$

The magic number must be 4.0, and so, from the table, the power would be about 52%.

EXAMPLE 2. Consider a standard treatment on which patients improve by an average of 10 units although some do much better (improve by 30 units) and some deteriorate by as much as 10 units. A new treatment carries some side-effects. It is determined that, to be worthwhile, it would have to improve patients by 15 units on average.

The second formula requires knowing the standard deviation. If this is not known, one rule of thumb is that the standard deviation will be about one quarter of the range of usual scores on standard therapy. From +30 to –10 is a range of 40, so here we can guess SD=10.

$$\mathbf{n = \frac{2(10)^2}{(5)^2} \times 7.8 \approx 63\ patients\ per\ treatment}$$