TCPInnovationsLtd
SCIENCE INTO TECHNOLOGY

Publications

Published December 31st, 2011

David Grainger
TCP Innovations

David Grainger is an academic in the Department of Medicine, Cambridge University, researching mechanisms underpinning chronic inflammatory diseases. He is also a leading consultant to the pharmaceutical industry through TCP Innovations Ltd, and is the Chief Scientific Officer of Funxional Therapeutics Ltd, a Cambridge-based biotechnology company he founded in 2005, which develops novel anti-inflammatory drugs. He delivers his often iconoclastic opinions on recent trends in life sciences industries through the Drug Baron blog.

The [Statistical] Power and the Glory

The worst possible outcome from a clinical trial is a negative result with a drug that actually works in the chosen indication. These 'false negatives' almost invariably destroy real value, and for all but the largest pharmaceutical companies spell the death not only of the programme but also the company itself.

In theory, drawing the short straw and getting a negative outcome when the drug actually works should be a stroke of real bad luck. In a properly powered trial, the odds for such an outcome should be less than one in five. But under-powering trials, at least in Phase IIa, is so prevalent that the number of 'false negatives' may be as high as one in every two early stage trials.

While the impact on the individual company is usually devastating, the impact on the pharmaceutical industry as a whole is no less severe. DrugBaron contends that more attention to the detail of trial power could double the productivity of the sector as a whole, catapulting drug development from a poorly performing sector to the very top of the pile.

The most egregious examples of under-powered trials result from overlooking the need for power calculations altogether. For the most part, its not because the clinical development team lacks an understanding of the importance of correctly powering their trial, but from a misunderstanding of the degree of comfort that accompanies a previous positive result.

A common approach to early-stage trial design is to copy a previous positive study involving a drug that went on to achieve regulatory approval. If you use the same end-point in the same patient population and your drug is at least as good as the drug previously approved (which, clearly, it needs to be in order to be commercially valuable) then surely their positive trial must have been suitably powered. No. Absolutely not.

If your CMO justifies his chosen trial design by pointing to a previous successful trial, follow DrugBaron's advice and sack him on the spot.

Their trial might only have been powered to detect an effect of that magnitude one time in two, or one time in three, not the usual minimum power we aim for of four positives out of five. In other words, they may very well have been lucky. They may have taken a gamble on a trial that comes up positive one time in two and hit the jackpot. But you need not be so lucky. Indeed, if it's a popular trial design that's been used quite a few times, then even a couple of positive results in the literature is no where near enough to demonstrate adequate power.

The simple fact is that estimating the power of a design requires a look at the distribution of possible outcomes when the same trial is run time after time. A single event gives almost no indication what that distribution looks like. We almost never run the same trial twice (even if we re-use the same design it is almost always with a different drug – and since we don't know whether the test agent really works or not, we cannot tell which negative outcomes were 'true negatives' and which were 'false negatives' due to lack of statistical power).

So lets assume you were smart enough to know that copying a successful trial was no way to guarantee sufficient power. You need to do a proper power calculation, right? Right. And the formula required, at least for a simple two group comparison, is very simple indeed. But that's where the simplicity ends, unfortunately. The issue is not the using the right formula, but plugging in the right numbers.

The challenge is the estimate of the variance in the end-point among untreated (or placebo-treated subjects). All the other numbers required are straightforward to find or estimate and plug straight in. But where to get an estimate of the variability among placebo-treated subjects? The literature, of course. For most of us, outside of the global pharmaceutical companies at least, this is the only possible source of reference values.

A small biotech usually has just the one shot on goal, a single trial with a binary outcome. So it is even more important to ensure that trial has sufficient statistical power.

Yet it isn't as easy, or as reliable, as it seems. For a start, few studies are really true replications: the technical means of measuring the end-point may differ, the population under study may be subtly different: any one of a myriad of small differences (even down to the technical competence of the people performing the trial) can, and will, affect the variability between subjects.

Worst of all, there is a massive positive publication bias, which means that the data most readily to hand is usually from trials from positive outcomes. These trials were the "lucky" ones, the ones where the variability among the control subjects was small enough to allow the impact of the treatment to reach statistical significance. In short, positive trials have tighter control groups than you get "on the average". If you assume your control group will be as tight as those in the positive studies, you will under-power your trial. And when your control group comes in with the "average" dispersion, you will get a negative result – even if your drug really works.

In practice, all the factors are tipped in the same direction, making the actual variability you will see in your trial larger than the variability seen in a "typical" published trial. And if you don't allow for that, and most people don't, you will perform yet another underpowered early stage trial. (Just as an aside, its worth remembering that these effects are much bigger for early stage trials simply because the sample sizes are typically smaller – of the order of tens of subjects per group – than later stage trials with typically hundreds per group. As the sample size gets bigger, then each time the trial is run the variability in the control groups will be much more similar each time. In statistical speak, the sample standard deviation approaches the population standard deviation).

The solution is simple enough: bigger trials are better.

And since our current tools for estimating how big the group size should be have, in most people's hands, a tendency to significantly under-estimate trial size, its better to err on the side of caution. After all, you know what damage a 'false negative' will do to your accumulated shareholder value.

If DrugBaron's advice looks a little technical, and seems to be aimed at Chief Medical Officers and others involved in clinical trial design and implementation, rather than at investors and chief executives, then stop and think again. Investors are often guilty of seizing any excuse to reduce costs and time (since time translates into rising costs in any case). If your CMO is asking for a bigger trial

than conventional wisdom would hold to be necessary, perhaps its time to remember that the cost of cutting the size of the trial may be the loss of everything invested in the company, if the resulting trial is under-powered. Instead, perhaps you should think yourself fortunate to have such an insightful clinical development team, and back their insight with the extra cash to deliver a proper test of your precious asset.

Positive trials have tighter control groups than you get "on the average". If you assume your control group will be as tight as those in positive studies, you will under-power your trial.

By contrast, if your CMO justifies his chosen trial design by pointing to a previous successful trial, follow DrugBaron's advice and sack him on the spot. And if he ran a power calculation, ask him where he got his estimate of the variance in the control group from. If he points to a published positive study, its time to start looking for his replacement.

It is also important to remember that statistical power of your trial design matters more in a small biotech than in a large pharmaceutical company. Typically, having decided to advance a particular compound into the clinic, the larger companies commit to several "human pharmacology" or "translational medicine" studies to learn about their compound in patients. If one or two of these studies happen to be under-powered, its unlikely to kill the programme, because the decision to move forward will be on a "weight of evidence" basis across multiple studies. By contrast, a small biotech usually has just the one shot on goal, a single trial with a binary outcome.

So if your life hangs by a single thread, make sure it's a good strong one.

© The copyright to this article is held by TCP Innovations Ltd (2011). You are expressly licensed to distribute this article in any form, including electronic transmission, provided that the format of the article including the acknowledgement of authorship and this copyright statement are not altered or removed.