Case Study: Design of a Large Safety Study

Executive Summary

Occasionally, due to concerns relating to the occurrence of a rare but serious adverse event in patients taking a marketed medicine, large safety studies are mandated by the prevailing regulatory authority. Such studies typically aim to rule out an unacceptable level of excess risk and can be very challenging to design and size. In one such example, the regulatory authority proposed a substantial study of 34,000 patients to rule out a 1.5 fold increase in the risk of a rare event. This was successfully challenged by statistical argument, leading to a revised study size of 11,700 patients which, while still a large study, was more operationally feasible and financially viable to conduct.

Ruling out Excess Risk

In general, when dealing with events occurring at very low rates, it is important to recognize that small changes in absolute risk often translate to deceptively large changes in relative risk. For example, for an event occurring at a rate of 0.75% per year, an absolute increase of 0.25% translates to a relative risk (RR) of 1.33.

For this latter example, if it is decided to design a study to rule out a RR increase of 1.5, translating to an absolute risk difference of 0.375%, a total of 34,000 patients would be required. It is interesting to note the most extreme difference between drug and control that could be observed and yet still rule out a 1.5 fold increase in risk is just 0.12%, i.e. an excess of 12 cases per 10,000 patients treated with drug. A difference in absolute risk any higher than this and the 95% upper confidence limit for the relative risk will not fall below 1.5, and so this level of increased risk could not be ruled out.

If, on the other hand, the level of risk to rule out was set at 2.0, sample size would be reduced by 66% percent to 11,700 patients. In this case the most extreme difference between drug and control that could be observed and yet still rule out an excess risk would be 0.19%, i.e. an excess of 19 cases per 10,000 patients treated with drug.

The relative advantages of ruling out different levels of risk are highlighted below.

Study size to exclude a 1.5 to 2.5 fold increase in relative risk for a rare event occurring at a rate of 0.75% per year

Δ=RR to rule out	Power	E=req'd number events	N=total number patients	Δ_{obs}= maximum observed RR to rule out Δ	Corresponding maximal observed split of events (control vs drug)	Corresponding observed event rates (control vs drug)	Corresponding difference in observed event rates and 97.5% upper 1-sided CL
1.5	90%	256	34,087	1.174	118 vs 138	0.69% vs 0.81%	0.12%, 95% CL = 0.30%
2.0	90%	87	11,664	1.315	38 vs 49	0.65% vs 0.85%	0.19%, 95% CL = 0.50%
2.5	90%	50	6,675	1.437	21 vs 29	0.63% vs 0.87%	0.24%, 95% CL = 0.65%

α=2.5% 1-sided used throughout

With respect to trial size, reducing the RR to be ruled out from 2.0 to 1.5 results in an almost 3- fold increase in sample size from approximately 11,700 to 34,000 patients, for what is apparently little gain in information. Given a true background rate of 0.75% and a RR of 2.0, the highest observed event rate on drug that could be observed and yet still rule out an increase in risk is 85 per 10,000 patients. Yet this decreases only marginally to 81 per 10,000, when the RR to rule out is reduced to 1.5 and sample size is tripled.

Statistical evaluation of the trial proposed by the regulatory authorities therefore demonstrates that ruling out a relative risk of 2.0 can provide sufficient information to rule out small, absolute differences in risk, and that this can be achieved at trial sizes that are within the range of what might be considered operationally feasible. The use of a 1.5 relative risk results in a very challenging 3 fold increase in trial size. The feasibility and timeliness of delivering data from a trial of such magnitude is highly questionable.

Outcome

Following face-face discussion with the regulatory authority, senior clinical and statistical regulators were persuaded by the statistical argument for a reduced study size seeking to rule out a 2 fold increase in risk with drug. Several other large pharma companies who were also initially asked to perform a very large study on their product in the same pharmacologic class also benefited from this decision to opt for a relatively smaller, more operationally feasible clinical study.

More Detail on the Statistical Methodology

When ruling out excess risk, the objective of a study is to test the following non-inferiority hypothesis: H_0: RR $\geq \Delta$ vs H_1:RR $< \Delta$ where RR is relative risk and Δ is the level of excess risk to be ruled out. The most appropriate sample size methodology to address this hypothesis is the log-rank test. The log-rank, unlike the simple Binomial, correctly takes into account not only the number of events but the timing of events. The number of events, E, required to rule out a RR of size Δ with 90% power and 1-sided significance level 2.5% is $E \approx \frac{42}{(\log\Delta)^2}$ [1]. Given E, the total number of patients required is then $N = \frac{E}{\lambda_{control}}$ where $\lambda_{control}$ is the event rate per unit time on control. It is of interest to note that to rule out a true RR of size Δ demands the *observed* relative risk is not more than $\Delta_{obs} = 0.4 \times \Delta$ [2]. Further, since with low event rates, the observed relative risk $\approx$ ratio of events observed in each arm, then $\Delta_{obs} \cong \frac{E_{drug}}{E_{control}}$, where E_{drug} and $E_{control}$ are the observed number of events on drug and control, with $E_{drug} + E_{control} = E$. This means that for a trial powered to provide E events in N = 2n patients to rule out a true RR of Δ is it possible to calculate the most extreme split of E total events that can be observed and yet still rule out a true RR of Δ [2]. This split is approximately $E_{control} = \frac{E}{1+\Delta_{obs}}$ and $E_{drug} = \frac{E \times \Delta_{obs}}{1+\Delta_{obs}}$ corresponding to observed events rates of $p_{control} = E_{control}/n$ and $p_{drug} = E_{drug}/n$, giving a difference $p_{drug} - p_{control}$ with $SE = \frac{p_{drug}(1-p_{drug})}{n} + \frac{p_{control}(1-p_{control})}{n}$.

From this, the upper CL for the most extreme excess risk that could be tolerated and yet still rule out a RR of Δ can be calculated. This latter measure is useful when weighing the practical value of sample sizes to rule out differing values of Δ.

References

1. Rubinstein RV, Gail MH and Santner JT. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *J Chron Dis* 1981; **34**:469-479.

2. Carroll KJ. Back to basics: explaining sample size in outcome trials, are statisticians doing a thorough job? *Pharmaceut. Statist* 2009; **8**: 333–345.

KJC Statistics Ltd
Sovereign House
Bramhall Village
Cheshire
SK7 1AW
UK
Email: kevin.carroll@kjcstatistics.com
Phone: +44 (0) 161 478 4949
Mobile: +44 (0) 7578 324 824