Sample size calculation in economic evaluation

What is CREST?

The Centre for Health Economics Research and Evaluation (CHERE) at UTS has been contracted by Cancer Australia to establish a dedicated **Cancer Research Economics Support Team (CREST)** to provide high quality, expert advice and support to Multi-site Collaborative Cancer Clinical Trials Groups.

Factsheets

CREST will produce a series of factsheets as resources for cancer collaborative group researchers wishing to include economic evaluation in their clinical trials.

Authors: Stephanie Knox
Prepared: August 2012

SUMMARY

Clinical trials usually calculate required sample size on the basis of clinical outcomes. Economic evaluation includes both costs and treatment effects, and so a clinically based sample size may not provide adequate power for economic evaluation.

This factsheet provides information about the implications of underpowered economic evaluations, and how they may be addressed.

- Economic evaluation is not typically concerned with hypothesis testing, but is more about estimation and so can still provide useful information even when under-powered
- Economic evaluations typically require larger sample sizes for adequate power than a typical clinical study
- While the power calculation should always be done and reported in an economic evaluation, an underpowered economic evaluation can still provide valuable information about the costs and benefits of new treatments.

For more information about CREST, or for other factsheets in this series, please see our website:

www.chere.uts.edu.au/crest

Calculating sample size and power for economic evaluation in clinical trials

Why calculate sample size and power for economic evaluation?

In many cases, when incorporating economic evaluation (EE) into clinical trials the sample size has already been calculated based on the expected effect size of the clinical outcome of interest. The EE includes both costs as well as treatment effects, and may be more interested in secondary patient outcomes such as Quality Adjusted Life Years (QALYs). Therefore a trial that is powered for a specific clinical outcome will not necessarily have the same power for EE.

A power calculation can be useful to establish if the trial design has a reasonable chance of demonstrating cost-effectiveness if the treatment is in fact cost-effective (Petrou & Gray 2011; Briggs & Gray 1998; Glick et al 2011a), and to inform the sensitivity analysis.

If there is insufficient power to reach conclusions about cost-effectiveness it may be ethically and/or economically infeasible to embark on EE in a trial. In this case, it may be necessary to modify the sample size to improve the power for EE. However, even when a trial is under-powered for EE, the power calculation provides information about the degree of uncertainty around the estimates of cost-effectiveness.

This is because when thinking about power and EE it is important to distinguish between estimation and hypothesis testing. Usually EE is not concerned with rejecting the null hypothesis in the classical way, but is primarily concerned with estimating the level of cost-effectiveness. The amount of uncertainty around this estimate is then explored with sensitivity analysis.

Factors that affect sample size for economic evaluation

Principles and assumptions for sample size calculations for cost-effectiveness are similar to calculations for clinical effect size (Glick et al 2011a). Economic evaluation is interested in the ratio of treatment outcome to costs. So for trials that incorporate EE we need the power to estimate the joint distribution of patient outcomes (either clinical or QALY) and costs (Petrou & Gray 2011). Including costs in power calculations usually but not always, leads to larger required sample sizes than those required for estimating clinical effects alone. Some of the factors in EE that affect sample size include:

- EE often requires a longer duration of follow up, which can lead to drop outs, and therefore require a larger initial sample size. However longer follow up may also lead to sustained benefits in terms of clinical or QALY outcomes with decreasing costs, thus reducing the required sample size.

Sample size calculation in economic evaluation

- Because EE is usually concerned with estimation of effect size rather than hypothesis testing, sensitivity analysis is used instead of confidence intervals to assess the possible range of estimated values. Sensitivity analysis and uncertainty around estimates usually leads to a larger required sample size.
- Under some conditions, the joint distribution of costs and effects may be significant even if the difference in clinical outcomes and the cost differences are not individually significant (Petrou & Gray 2011).

Maximum Willingness to Pay

Medicare calculations of power and sample size are calculated against some value of Maximum Willingness to Pay (WTP) for a unit of treatment effect. The sample size needs to give enough power for the upper confidence limit of the Incremental Cost Effectiveness ratio (ICER) of a cost-effective treatment to fall below the value of Maximum WTP.

The treatment effect is usually expressed in QALYs, and a decision should be made a priori about the Maximum WTP for 1 QALY gained. The Maximum WTP should be considered carefully as in some ways it is hypothetical. It may be defined from a societal perspective, as an opportunity cost, or as an arbitrary threshold, and this choice is often dictated by the budget in question. Other clinical endpoints such as survival may also be used (Willan 2011), and similarly the Maximum WTP for one unit of survival needs to be set.

Incremental Net Benefit

The Incremental Cost Effectiveness ratio (ICER) is defined as:

$$ICER = \frac{Cost_{New} - Cost_{Comparator}}{Effectiveness_{New} - Effectiveness_{Comparator}}$$

There are concerns about calculating power and sample size using the ICER since an ICER cannot be interpreted unless the sign of the change in costs and the change in effects are known (Willan 2011). Instead the Incremental Net Benefit (INB) is a measure that is easier to interpret and work with, since a positive INB means that the treatment is cost-effective. We still need to set a maximum WTP to calculate INB:

$$INB = \left((Effect_{New} - Effect_{Comparator}) * WTP\right) - (Cost_{New} - Cost_{Comparator})$$

Calculating Sample Size

We need enough power for the 95% CI to exclude maximum WTP per QALY if the treatment is in fact cost-effective. This formula is given as (from Glick et al 2011a):

$$n = \frac{2*(Z_{\alpha} + Z_{\beta})^2\left[sd_c^2 + (W*sd_q)^2 - (2W\rho*sd_c sd_q)\right]}{(WQ-C)^2}$$

Where:

Z_{α} is the Z-statistic for the level of Type I error (usually set at 95%)

Z_{β} is the Z-statistic for the level of Type II error (usually set at 80%)

sd_q, sd_c are the std deviations for each group for treatment effect and cost respectively

W is the Maximum Willingness to Pay

Q is the expected mean difference in treatment effectiveness

C is the expected mean difference in treatment cost

ρ is the expected correlation of the difference in cost *(C)* and effect *(Q).* This is a measure of the covariance of changes in effectiveness and changes in cost. Negative covariance, where cost decreases with increasing effectiveness result in a larger sample size. Positive covariance where cost increases with increasing effectiveness result in smaller sample sizes

Note that the denominator is the square of the INB. If you substitute only clinical effects into the formula, it returns to a calculation of sample size for clinical effects.

Calculating Power

If the sample size has already been calculated on clinical effect sizes then you may want to estimate the power of the trial to detect cost-effectiveness. The trial will need reasonable power if it is to be suitable to incorporate economic evaluation.

$$Z_\beta = \sqrt{\frac{n*(WQ-C)^2}{2*(sd_c^2+(W*sd_q)^2-(2W\rho*sd_c sd_q)}} - Z_\alpha$$

The power of the study is then the area under the standard normal distribution that is to the left of Z_β.

Variations in Maximum Willingness to Pay

Glick et al (2011a, 2011b) demonstrated that power and sample size do not necessarily increase and decrease monotonically with increases and decreases in Maximum Willingness to Pay. Therefore it is important to fix the Maximum Willingness to Pay at a realistic level before calculating sample size and power. In multinational trials where Maximum Willingness to Pay and resources may differ between countries you may need to calculate sample size or power at the actual value of Maximum Willingness to Pay for each country since the effect on power cannot be extrapolated.

Using one-tailed tests

Willan (2011) argues that in RCTs where the comparator is usual care we are interested in whether the treatment improves outcomes compared to the comparator. So the null hypothesis is that the difference in treatment effects is ≤ 0 (the treatment does not improve outcomes) and the alternative hypothesis is that the treatment improves outcomes compared to the comparator. Therefore one-tailed tests are appropriate since the test is for improvement of outcomes over standard.

The advantage of using a one-tailed test is that it allows for a smaller sample size for a specified level of α. Most clinical trials will continue to use 2-tailed tests for sample size calculations to detect clinically important treatment effects. However a one-tailed test may be appropriate in power calculations to establish the feasibility of EE based on the established sample size.

Sample size calculation in economic evaluation

Cluster randomised trials

For cluster randomised trials the calculation of power needs to account for the design effect of the intra-cluster correlation (ICC). There is a body of literature on calculating sample size and power for clinical outcomes of cluster randomised trials, but little has been written on cluster randomised trials in EE (Gomes, Ng et al. 2012). As the design will have an effect on sample size and power in EE it should be accounted for when calculating sample size and power. For further reading on this topic see the references at the end of this document.

Values for populating power calculations in economic evaluations

In order to complete the power or sample size calculation, values to populate the equations will be required. Existing published values for clinical effect size and costs can be used, particularly where these come from similar studies. If the comparator is usual care then the comparator costs may be available in published evaluations or may be derived from existing administrative datasets or patient logs. For new treatments the costs of treatment may not be available and assumptions will need to be made about possible range of costs. Power calculations across a range of expected costs will then provide a sensitivity analysis on the expected power for EE. Finally, assumptions will also need to be made about the direction and strength of the correlation of costs and clinical effects.

Bayesian approaches to calculating power in economic evaluations

The classical approach to calculating sample size discussed above has generally been criticised for relying on arbitrarily set error rates and uncertainties over clinically important effect sizes. Bayesian approaches that exploit value of information have been recently proposed for calculating sample size in economic evaluation. Willan (2011) argues that incorporating decision theory in sample size calculations has better construct validity for economic evaluation. And in practical terms using prior information can result in smaller estimated sample sizes. See the references at the end of this document for further reading on this topic.

The implications of under-powered economic evaluation

It may be the case that a sample size calculation demonstrates that the planned sample size will not provide sufficient power for an economic evaluation in a clinical trial. Ideally the planned sample size would be revised, to allow adequate power for the EE, however this is not always possible. There may be logistical issues such as patient recruitment or financial constraints which limit the available sample.

Consideration needs to be given to the overall purpose of the economic evaluation. In some cases, it may be appropriate to include an underpowered EE in a trial. Trials of new treatments may have high levels of uncertainty around clinical benefit and costs, and so the collection of additional information

regarding these, despite being underpowered for analysis of cost effectiveness may be extremely useful. In addition, the power calculation itself can provide useful information to the sensitivity analysis about the amount of uncertainty around estimates of cost-effectiveness.

Conclusion

Estimating the power of a clinical trial for EE can be useful to assess the feasibility of undertaking EE in a particular trial. Sample size and power calculations in EE are an extension of power calculations for clinical effect sizes. However in EE we are calculating the power to estimate the joint distribution of costs and treatment effects. We need more information for estimating power in EE, namely expected costs of treatments, the expected covariance of treatment effects and costs, and the maximum willingness to pay for the treatment effect. While the power calculation should always be done and reported in an economic evaluation, an underpowered economic evaluation can still provide valuable information about the costs and benefits of new treatments.

For more information

For more information on any part of this factsheet, please contact:

crest@uts.edu.au

References

Briggs, A.H., & Gray, A.M. (1998). Power and sample size calculations for stochastic cost-effectiveness analysis. *Med Decis Making*, 18 (2 Suppl), S81-92.

Gardiner, J.C., Sirbu, C.M., & Rahbar, M.H. (2004). Update on statistical power and sample size assessments for cost-effectiveness studies. *Expert Rev Pharmacoecon Outcomes Res*, 4 (1), 89-98.

Glick, H.A. (2011). Sample size and power for cost-effectiveness analysis (part 1). *Pharmacoeconomics*, 29 (3), 189-198.

Glick, H.A. (2011). Sample size and power for cost-effectiveness analysis (part 2): the effect of maximum willingness to pay. *Pharmacoeconomics*, 29 (4), 287-296.

Petrou, S., & Gray, A. (2011). Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ*, 342, d1548.

Willan, A.R. (2001). Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. *Control Clin Trials*, 22 (3), 228-237.

Willan, A.R. (2011). Sample size determination for cost-effectiveness trials. *Pharmacoeconomics*, 29 (11), 933-949.

Gomes M, Ng ES et al (2012). Developing Appropriate Methods for Cost-Effectiveness Analysis of Cluster Randomised Trials. *Med Decis Making,* **32**(2):350-361.