Sample size calculation for economic evaluation
Gerhart Knerer

Introduction
Evidence suggests that evaluations powered to detect differences in clinical outcomes may be under-powered to detect large differences in cost (Gray et al., 1997). The evaluation of new interventions, however, should be adequately powered to address economic questions, even if over-powered with respect to clinical outcomes. Indeed, commentators have suggested that under-powered studies are unethical (Briggs, 2000).

In recent years, there has been a proliferation of papers concerned with methods for sample size calculation in economic evaluations (see, for example, Briggs and Gray, 1998; Laska et al., 1999; Willan and O’Brien, 1999). The newer methods call for calculations to be based on both costs and effects and, generally, require more information than methods for calculating sample sizes for effect or cost differences alone. This paper explores one such method in order to highlight some of the issues involved.

Sample size methodology
Willan (2001) has described a method based on the net benefit concept (Stinnett and Mullahy, 1998, provide a detailed explanation of net benefits). The formula to derive sample sizes for estimating incremental net benefit is reproduced below. Beginning with an explanation of notation:

- $Z$ in the equation $(Z_{1-\alpha} + Z_{1-\beta})^2$ refers to the area under the normal curve distribution for one or two tailed significance tests ($\alpha$ and power $(1-\beta)$). For a 2-sided test where $\alpha=0.05$ and power $(1-\beta)=0.90$, the corresponding value for $(Z_{1-\alpha} + Z_{1-\beta})^2$ is equal to 10.507 (Machin et al., 1997).
- $\text{Var}(\epsilon)$, $\text{Var}(\zeta)$, and $\text{Cov}(\epsilon, \zeta)$ denote, respectively, the variance of effect, variance of cost and covariance between effects and costs observed for patients receiving new treatment (T) and standard treatment (S). Variance measures the spread or variability in the data. Covariance measures the degree of linear dependence between two variables.
- Lambda ($\lambda$) represents society’s willingness-to-pay for a unit gain in effectiveness.
- Delta ($\delta$) is defined as the smallest clinically important difference.

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{\delta^2} \left( X_T + X_S \right)$$

where:

$$X = \text{Var}(\epsilon) + \frac{\text{Var}(\zeta)}{\lambda^2} - \frac{2\text{Cov}(\epsilon, \zeta)}{\lambda}$$

Example
A hypothetical example is outlined for an evaluation seeking to compare two treatments for mild and moderate depression in general practice — generic anti-depressants plus supportive GP care or placebo plus supportive GP care. Supportive care is defined as the establishment of a positive relationship with the patient and may include referral to practice-based counsellors or psychology services. The hypothesis is that treatment with generic anti-depressants plus supportive care is more cost-effective than supportive care alone for people with
mild and moderate depression, defined as a score of 10 to 29 on the Beck Depression Inventory (BDI) (Beck et al., 1988).

**Variance-Covariance** In order to populate the above formula, estimates of the variance of costs and effects must be identified from previous trials or unpublished sources (for example, NICE submissions). This assumes that variance estimates derived from one trial are applicable to another, given similarity of treatments. An in-depth search revealed only one suitable study (Bower et al., 2000). In this randomised controlled trial, cost and effect differences were compared across three treatments — usual GP care, cognitive behavioural therapy (CBT) and non-directive counselling (NDC). For the purpose of the current example, we assume that NDC or CBT plus antidepressants is equivalent to our experimental treatment whilst NDC or CBT without antidepressants is equivalent to our control treatment. The BDI and societal costs were collected over a 12-month follow-up. Trial estimates of variance and covariance are summarised in Table 1.

**Delta (δ)** Careful thought must be given to the value of delta — the smallest clinically important difference. Discussions with clinicians suggest that a difference of 3–5 points on the BDI can be viewed as clinically important. Therefore, the differences in BDI scores used in the current study are 3 and 5 points.

It should be noted that there is a preference for a more ‘universal’ delta, rather than one specific to the treatments being compared. In other words, there is a preference for generic, rather than disease specific, measures of patient preferences (Willan, 2001). Although the EQ-5D measure of health-related quality of life was included in the original study (Bower et al., 2000), it is doubtful in the current context whether this would provide superior estimates to those derived using BDI scores. As Bennett points out, general health status and mobility instruments are of ‘limited usefulness for describing depression’ (Bennett et al., 2000).

**Lambda (λ)** Some uncertainty surrounds the appropriate value to assign to society’s willingness-to-pay for a unit gain in effectiveness — lambda. Rather than just apply the commonly cited value of £30,000 (Timmins, 2001), the approach taken here is to explore the uncertainty surrounding this figure by varying the value of lambda.

### Results

Populating the above equation, approximately 235 patients per arm are required to give 90% power to detect an incremental net benefit equivalent to a 3-point difference in BDI scores at α=0.05 and δ=£30,000. To detect an incremental net benefit equivalent to a 5-point difference, 85 patients per arm are needed. These figures are stable for values of lambda in the range £500–£30,000 (Figure 1).
Conclusions

More strenuous attempts should be made to ensure that adequate power becomes the rule rather than the exception in economic evaluations. Although sample size calculations for cost-effectiveness require more information than calculations for costs or effectiveness alone, this example shows that such calculations are feasible, given adequate prior information on costs and effects.

References


Economic evaluation of treatments for eating disorders

CEMH researchers are involved in three evaluations of treatments for eating disorders, each of which employs a randomised controlled design to assess costs, effectiveness and patient acceptability.

One multi-centre trial compares specialist inpatient treatment, specialist outpatient treatment and general management in Child and Adolescent Mental Health Services for adolescent anorexia nervosa. The work is undertaken in collaboration with the Departments of Child and Adolescent Psychiatry at the Universities of Liverpool and Manchester.

Contact: Sarah Byford

There are also two multi-centre studies funded by the PPP Medical Healthcare Trust with lead collaborators at the Institute of Psychiatry. The first of these compares the effectiveness and cost-effectiveness of cognitive guided self-care versus family therapy for adolescents with bulimia nervosa. The second will explore brief multiple family day treatment, inpatient care and outpatient family therapy for anorexia nervosa.

Contact: Jennifer Beecham