# Economic evaluation of factorial randomised controlled trials: Why the method of analysis matters

**Helen Dakin and Alastair Gray**

## Abstract

Whereas most randomised controlled trials (RCTs) compare just two interventions, factorial trials simultaneously compare at least four combinations of two or more factors in the same trial: e.g. comparing placebo, Treatment A, Treatment B and A+B combination therapy. Factorial trials address several questions simultaneously, can evaluate interactions between treatments and generally have greater statistical power than three-arm trials evaluating three of the same treatments. The main statistical approaches for analysing clinical endpoints from factorial trials are regression, at-the-margins analysis (comparing outcomes for all patients receiving A with all those not receiving A) and within-the-table analysis (treating each combination of factors as a separate treatment). Regression analysis is recommended when interactions are present, although at-the-margins analysis is commonly used.

In this paper, we review these three techniques and show, with worked examples, how the analysis method affects results and conclusions. We identify the main issues associated with using these methods to conduct economic evaluations of factorial trials, such as the size of interactions with respect to QALYs and costs and the challenges raised by regression-based cost-effectiveness analysis on multi-arm trials. In particular, we argue that the nature of the decision problem in economic evaluation is different from analysis of clinical endpoints, as economic evaluation focuses on estimation and identifying which treatment option maximises net benefits at any given ceiling ratio. Using worked examples, we show that at-the-margins analysis may be inconsistent with this objective and can produce misleading conclusions for economic evaluations of factorial trials.

## 1. Features, prevalence, advantages and disadvantages of factorial trials

Most randomised controlled trials (RCTs) compare two interventions; by contrast, factorial RCTs simultaneously evaluate two or more "factors" and interactions between them. The factors under investigation may include comparisons between placebo and two or more different interventions, with "2x2" factorial trials randomising patients to receive placebo, A, B or A+B combination therapy (Table 1); for example, the AFORRD trial evaluated atorvastatin, omega-3 fatty acids, the combination of the two and neither in patients with diabetes (1). Factors may also include aspects of dose or treatment administration, with some factorial trials comparing several dosing regimens for one or more drugs. More complex factorial designs, such as dose-ranging 4x3 designs (2) or 2x2x2 designs evaluating three drugs and associated combinations (3) have also been used in clinical trials.

**Table 1:** Illustration of a 2x2 factorial trial comparing two treatments, showing the cell notation used in the paper

| | **Placebo for A** | **Treatment A** |
|---|---|---|
| **Placebo for B** | **00**<br>(Placebo for A + Placebo for B)<br>Mean outcome = $\mu_{00}$ | **A0**<br>(Treatment A + Placebo for B)<br>Mean outcome = $\mu_{A0}$ |
| **Treatment B** | **0B**<br>(Placebo for A + Treatment B)<br>Mean outcome = $\mu_{0B}$ | **AB**<br>(Treatment A + Treatment B)<br>Mean outcome = $\mu_{AB}$ |

Factorial trials enable multiple questions to be addressed in the same study (4, 5): for example, we can assess the effectiveness of A and B individually and the effect of A+B combination therapy. In particular, factorial trials allow us to investigate *interactions* between treatments (6-10), providing our sample is large enough (11, 12). Interactions indicate that the effect of A differs depending on whether or not B is given and mean that the effect of using A and B in combination is not equal to the sum of the individual effects; this may arise due to non-compliance (6) or pharmacokinetic, biological or behavioural mechanisms (6, 12, 13) and may affect our conclusions. The interaction term equals the observed outcomes in cell AB in Table 1 minus what we would expect if treatment effects were additive. For example, if A monotherapy increased utility by 0.10 and B increased utility by 0.20, there would be zero interaction if utility was 0.30 higher in cell AB than cell 00. Interactions may be divided into three categories (6, 9, 10):

- ***Super-additive interactions*** mean that the effect of the combination is *greater* than the sum of the parts; in this case the interaction term will have the same sign as the main effects (e.g. if mean utility in cell AB is 0.4 higher than cell 00, interaction=0.1).
- ***Sub-additive interactions*** indicate that the effect of the combination is *less* than the sum of the parts; the interaction term therefore has the opposite sign from the main effects of treatment (e.g. if mean utility is 0.25 higher in cell AB than cell 00, interaction=-0.05).
- ***Qualitative interactions*** indicate that that the effect of at least one of the treatments under investigation changes *sign* (not just magnitude) depending on whether or not the other therapy is given: for example, A might (in theory) increase utility by 0.10 when used alone but reduce utility by 0.05 when added to B, such that mean utility in cell AB was only 0.15 higher than cell 00 (interaction=-0.15).

If there are no interactions, factorial designs provide greater power for the same sample size than three-arm studies evaluating the same treatments and (in theory) could provide the same information as two two-arm studies at substantially lower cost (6, 11, 12, 14, 15). However, as we will discuss in Section 2, the extent to which we can realise these benefits and obtain

consistent, unbiased and statistically efficient estimates of treatment effects depends on the form of the analysis we use and whether or not the treatments interact.

Factorial experiments have been used in agriculture since the 19th century (4, 16), although the term "factorial" was probably first used by Fischer in 1935 (16, 17). This experimental design has been applied in a wide variety of fields, including medicine, with the first RCT on health technologies being published in the 1950's (18). Since then, factorial trials have continued to play a small but important role in medical research, accounting for 1-5% of all RCTs (5, 10, 12, 19, 20).

Although factorial studies are well established in medicine and more than 60 factorial RCTs with trial-based economic evaluations have now been published, we are aware of no previous studies evaluating the methods for conducting economic evaluations of factorial trials. In this paper, we will first outline the methods typically used to analyse clinical outcomes from factorial trials, before highlighting the difficulties associated with applying these methods to economic evaluations using a worked example.

## 2. Analytical methods for clinical outcomes from factorial design trials

Three methods are commonly used to analyse clinical outcomes from factorial trials: within-the-table analysis, at-the-margins analysis and regression analysis (5, 12).

To illustrate these three analytical methods, we generated two simulated sets of results for a clinical outcome measure (e.g. diastolic blood pressure) within a hypothetical 2x2 factorial trial (Table 2). Both datasets assume the same true main effects for drugs A and B, but simulation 1 assumes that the true interaction between A and B equals zero, while simulation 2 assumes a sub-additive interaction. Data were generated in Microsoft Excel 2003 by randomly sampling from a normal distribution defined by the "true" mean and standard deviation (SD) for each cell in the design, and assuming 250 patients in each cell.

***Within-the-table*** (or inside-the-table) analysis treats each cell within the factorial design as a separate treatment strategy (5). For a 2x2 factorial design, outcomes are therefore estimated and presented separately for each of the four treatment arms based only on the patients randomised to that combination of treatments (Table 2), allowing the reader to see the effect

of interactions directly. Within these examples, the outcomes for cell A0 are equal to the mean (and SD) across the 250 patients receiving A monotherapy. The ("simple") effect of treatment A is based on a straightforward comparison between outcomes for those patients randomised to A monotherapy (cell A0) and those randomised to placebo (cell 00); its mean equals the difference in means between these two groups, while its variance equals the sum of the variances for these two groups. The within-the-table estimate of treatment effect is always unbiased (regardless of interactions) (21, 22), but implicitly assumes that the effect of A in the presence of factor B provides no information on the effect of A in the absence of factor B. As result, the within-the-table approach bases treatment effects on only half of the available data and therefore produces less precise estimates than the at-the-margins approach (22) (unless there is a qualitative interaction) and provides no more statistical power than a non-factorial design.

**Table 2:** Simulated results from a 2x2 factorial trial: within-the-table approach

| | | **Placebo for A** | **Drug A** |
|---|---|---|---|
| Mean outcome (SD) for simulation 1: zero interaction between A and B | Placebo for B | 108.4 (19.7)* | 100.2 (18.4)* |
| | Drug B | 89.3 (20.1)* | 81.4 (20.9)* |
| (Simple) effect of treatment A (SE) | | -8.20 (1.70)* | |
| (Simple) effect of treatment B (SE) | | -19.14 (1.78)* | |
| Mean outcome (SD) for simulation 2: sub-additive interaction between A and B | Placebo for B | 110.0 (19.8)* | 97.8 (20.2)* |
| | Drug B | 89.4 (19.5)* | 83.8 (19.5)* |
| (Simple) effect of treatment A (SE) | | -12.24 (1.79)* | |
| (Simple) effect of treatment B (SE) | | -20.65 (1.75)* | |

* Significantly greater than zero ($p<0.05$); SD, standard deviation; SE, standard error.

The ***at-the-margins approach*** (Table 3) treats the factorial trial as though it were two overlapping two-arm RCTs (5). Results are typically presented separately for each factor (e.g. for A vs placebo of A and for B vs placebo of B), rather than for the individual cells of the factorial design. The ("main") effect of treatment A is simply based on the difference between the mean outcomes across all patients receiving drug A (cells A0 and AB in Table 1) and the mean outcomes for all patients who did not receive drug A (cells 00 and 0B). The variance around the treatment effect is equal to the variance across cells A0 and AB plus the variance across cells 00 and 0B. This approach is statistically efficient, as treatment effects are based on all patients; standard errors are therefore lower than for the within-the-table approach (Tables 2 and 3).

**Table 3:** Results of the two simulated datasets using the at-the-margins approach

| | Treatment A | | Treatment B | |
|---|---|---|---|---|
| | **Placebo** | **Active drug** | **Placebo** | **Active drug** |
| ***Simulation 1: zero interaction between A and B*** | | | | |
| Mean outcome (SD) | 98.8 (22.1)* | 90.8 (21.8)* | 104.3 (19.5)* | 85.4 (20.8)* |
| (Main) effect of treatment (SE) | -8.02 (1.39)* | | -18.96 (1.28)* | |
| ***Simulation 2: sub-additive interaction between A and B*** | | | | |
| Mean outcome (SD) | 99.7 (22.1)* | 90.8 (21.0)* | 103.9 (20.9)* | 86.6 (19.6)* |
| (Main) effect of treatment (SE) | -8.93 (1.37)* | | -17.34 (1.28)* | |

* Significantly greater than zero ($p<0.05$); SD, standard deviation; SE, standard error.

However, since the at-the-margins approach estimates the mean treatment effects for A averaged across patients receiving treatment B and those who do not receive B, it implicitly assumes that there is no interaction between treatments and that the groups are balanced. At-the-margins estimates of treatment effects are therefore biased whenever the interaction does not equal zero (6, 12, 21, 22) (bias equals half of the true interaction (22)), which may lead to incorrect conclusions. Sub-additive interactions also reduce statistical power (5, 6, 12, 23-25), while qualitative interactions can completely distort the conclusions of at-the-margins analyses (10, 12, 13, 25). This bias causes the at-the-margins estimates for simulation 2 (with a sub-additive interaction) to differ substantially from the within-the-table estimates (Table 2), whereas the results for simulation 1 are similar. Since they average treatment effects over the observed numbers of patients in each cell, at-the-margins estimates are also sensitive to the number of patients randomised to each treatment and will not reflect average outcomes in clinical practice unless this matches the distribution of patients in the clinic.

***Regression techniques*** or analysis of variance (ANOVA) provide an alternative to these two simpler techniques; regression allows us to control for covariates and make predictions and provides a flexible range of analytical tools. In principle, any regression technique can be used to analyse factorial trials, although continuous clinical outcome measures are generally analysed using ordinary least squares (OLS) or ANOVA (26), whereas dichotomous measures (e.g. mortality) are generally analysed using logistic regression, which assumes that treatment effects are multiplicative (12). However, for simplicity, we focus on OLS within this paper.

For a 2x2 factorial trial, the simplest regression model (Equation 1) includes just two explanatory variables: one dummy (A) indicating whether or not factor A is given and another for factor B. This simple model implicitly assumes that there is no interaction

between the treatments. The regression coefficients $\beta_A$ and $\beta_B$ represent the (main) treatment effects for A and B. If patient numbers are identical in each group, the point estimates for mean treatment effects from this model will be identical to those from the at-the-margins method, although standard errors will differ unless standard deviations are also identical across groups. This can be observed in Table 4, which show the results of fitting this regression model to the simulated datasets using the `regress` command in Stata version 11.0 (StataCorp, College Station, Texas). However, regression will be more appropriate than the at-the-margins approach if groups are not perfectly balanced (i.e. if patient numbers differ between study arms (12)).

$$y_i = \beta_o + \beta_A A_i + \beta_B B_i + \varepsilon_i \quad (1)$$

**Table 4:** Results of the two simulated datasets using OLS regression with/without an interaction term

| | | Mean (SE) for simulation 1 (zero interaction) | | Mean (SE) for simulation 2 (sub-additive interaction) | |
|---|---|---|---|---|---|
| | | **Without interaction term** | **With interaction term** | **Without interaction term** | **With interaction term** |
| Treatment effect for A ($\beta_A$) | | -8.02 (1.25)* | -8.20 (1.77)* | -8.93 (1.25)* | -12.24 (1.76)* |
| Treatment effect for B ($\beta_B$) | | -18.96 (1.25)* | -19.14 (1.77)* | -17.34 (1.25)* | -20.65 (1.76)* |
| Constant term ($\beta_0$) | | 108.3 (1.08)* | 108.4 (1.25)* | 108.4 (1.08)* | 110.0 (1.25)* |
| Interaction term ($\beta_{AB}$) | | N/A | 0.36 (2.50) | N/A | 6.62 (2.49)* |
| Predictions from regression model | Placebo | 108.3 (1.08)* | 108.4 (1.25)* | 108.4 (1.08)* | 110.0 (1.25)* |
| | Drug A + Placebo | 100.3 (1.08)* | 100.2 (1.25)* | 99.4 (1.08)* | 97.8 (1.25)* |
| | Placebo + Drug B | 89.4 (1.08)* | 89.3 (1.25)* | 91.0 (1.08)* | 89.4 (1.25)* |
| | Drug A + Drug B | 81.3 (1.08)* | 81.4 (1.25)* | 82.1 (1.08)* | 83.8 (1.25)* |

* Significantly greater than zero ($p<0.05$); SE, standard error.

However, unlike the at-the-margins method, the regression approach can readily be adapted to allow for interactions by including an interaction term equal to the product of the dummies for the individual factors (Equation 2). Within this model, the coefficients $\beta_A$ and $\beta_B$ represent the main effects for A and B, adjusted for the interaction between the two factors. Point estimates for $\beta_A$ and $\beta_B$ from OLS regression will be identical to the treatment effects estimated from the within-the-table approach (even if the groups are unbalanced) and standard errors will be similar, but not necessarily identical (Table 4). This regression model also directly estimates the interaction effect (albeit with less precision than main effects): i.e. the coefficient $\beta_{AB}$, which equals the difference between the observed mean outcomes for group AB and $\beta_0 + \beta_A + \beta_B$.

$$y_i = \beta_o + \beta_A A_i + \beta_B B_i + \beta_{AB} A_i B_i + \varepsilon_i \quad (2)$$

Unlike the at-the-margins approach, regression enables us to predict mean outcomes for each cell in the factorial design. The predicted values and associated standard errors shown in

Table 5 were estimated using the `predict` post-estimation command within Stata; the predicted values from the regression model with an interaction term match those estimated using the within-the-table approach (Tables 3 and 5).

However, just as results differ depending on whether the at-the-margins approach or the within-the-table approach is adopted, the results of regression analyses are affected by whether or not an interaction term is included. In particular, regression models without an interaction term will produce biased estimates of all coefficients and predictions unless the true interaction between the two factors equals zero. By contrast, the regression model with an interaction term (like the within-the-table approach) will always provide unbiased estimates (providing the usual assumptions for the regression model hold). However, including an interaction term reduces statistical efficiency if there is genuinely no interaction, increasing standard errors around both treatment effects and predicted values (Table 5).

Since the at-the-margins approach and regression without an interaction term are more efficient but prone to bias, statisticians analysing clinical endpoints of factorial trials commonly use a two-stage procedure (12, 21, 22, 27). In the first stage, the interaction effect is calculated (e.g. using regression analysis including an interaction term). Within-the-table or regression analysis with an interaction term will then comprise the primary analysis if the interaction term reaches a pre-defined level of significance (which may be higher than the conventional 0.05 level (21, 25)), while at-the-margins analysis or regression without an interaction term will be used if the interaction is not significant (21, 22, 27). However, this approach relies upon the arbitrary choice of significance level, equates absence of proof of an interaction with proof of absence and is hindered by the fact that very few trials are powered to detect interactions (12). Consequently, the two-stage approach may miss non-significant but important interactions that could bias estimates from analyses omitting interaction terms (13). Some authors have suggested taking a Bayesian approach that uses informative priors for the interaction to allow compromises between including and excluding the interaction term (25, 28, 29). However, despite the potential for bias, the at-the-margins approach or regression without an interaction term remains widely-used for analysis of clinical endpoints (5) and several key papers recommend this approach for analysis of primary trial data and for meta-analyses of factorial trials based on observations that very few clinical trials observe important or statistically significant interactions (5, 7).

## 3. Issues raised by economic evaluation of factorial studies

Although the methods for analysing clinical outcomes from factorial trials are well established (5, 12), nothing has yet been published on the appropriate form of analysis for economic evaluations of factorial trials. Preliminary results of our ongoing systematic review of economic evaluations of factorial trials suggest that current practice is split evenly between within-the-table, at-the-margins and regression approaches, although around 40% of studies do not clearly state the methods of analysis and some ignore the factorial aspect of the trial design and treat the trial as several separate two-arm studies (equivalent to conducting at-the-margins analysis). However, we argue that correct analysis of the factorial design is even more important for trial-based economic evaluation than for analysis of clinical endpoints, as economic evaluation focuses on estimation rather than hypothesis testing and as large interactions are likely to occur more commonly for costs, quality-adjusted life-years (QALYs) and net benefits than for clinical endpoints. Furthermore, although trials will often be underpowered to detect interactions in costs, QALYs or net benefits, such interactions may affect the framework we use for healthcare decision-making and change study conclusions.

### *3.1. Causes of interactions in costs, QALYs and net benefits*

For clinical outcomes (such as blood pressure or the log-odds of events), interactions may arise due to non-compliance (6) or pharmacokinetic, biological or behavioural mechanisms (6, 12, 13). Interactions may also be an artifact of the scale of analysis (6, 10, 13): if two factors have a multiplicative effect, whereby the effect of A and B in combination is equal to the product (not the sum) of the individual effects of A and B, there will be no interaction on a logarithmic scale, but a positive interaction on a natural scale (equating to a super-additive interaction if main effects are also positive [since main effects and the interaction have the same sign], or a sub-additive interaction if main effects are negative). These mechanisms are likely to be even more important for costs, QALYs and net benefits, frequently producing interactions in economic outcomes even when there is no interaction for clinical endpoints.

Large super-additive interactions for costs are likely within trials comparing several doses or dosing frequencies, since the factors (e.g. drug, dose or dosing frequency) may have a multiplicative rather than additive effect: for example, in a trial evaluating once-daily drug A

(£10), twice-daily drug A (£20), once-daily drug B (£50) and twice-daily drug B (£100), increasing dosing frequency will increase costs by £10 for A and £50 for B (interaction: £40).

Interactions will be observed for both costs and QALYs when treatments reduce the chance of clinical events that increase costs and reduce quality or length of life (30), since the effect of interventions on binary outcomes (e.g. events or deaths) are often additive on a log-odds scale, but not on a natural scale (12, 31); as we show in Section 3.3, this is likely to result in super-additive interactions for costs and sub-additive interactions for QALYs.

For QALYs, interactions may occur even if the interventions are used to treat/prevent different diseases in the same patient group – for example if both treatments affect the risk of death. Adding additional healthcare interventions may also give diminishing marginal returns for utilities (32, 33), as people may experience a larger improvement in quality of life from the first intervention than they will from any subsequent one, leading to a sub-additive interaction: the EQ-5D tariff, for example, includes the N3 interaction term and utility improves more when moving from level 3 to level 2 on any domain than moving from level 2 to level 1 (34). Super-additive interactions for QALYs are also possible if one intervention extends life expectancy and therefore increases the QALY gains from another intervention that improves quality of life.

Super-additive interactions for costs and sub-additive interactions for QALYs will combine to produce even larger sub-additive interactions on a net benefit scale. Indeed, qualitative interactions (whereby treatment A increases net benefits when used alone but reduces net benefits when added to B) may be common for net benefits, occurring whenever treatment A is cost-effective compared with no treatment but A+B is not cost-effective compared with A.

Although logarithmic or other transformations may reduce interactions in some circumstances (6, 10), this will frequently not be the case for total costs, total QALYs or net benefits, since these outcomes are the sum of several components that may be additive or multiplicative (e.g. drug cost, cost of events and cost of side-effects) and as transformations are unlikely to eliminate qualitative interactions (10). Furthermore, transforming costs and QALYs prior to analysis causes difficulties with interpreting coefficients and transforming predictions back to a natural scale (30, 35-37).

### *3.2. Effect of interactions on the framework for healthcare decision-making*

Factorial trials also raise important issues regarding when treatments should be considered mutually exclusive and how we use incremental cost-effectiveness ratios (ICERs) in healthcare decision-making. Johannesson (38) states that:

> "Two programmes are independent if the costs and effects of a programme are not affected by whether the other programme is implemented or not, e.g. a treatment for ulcer and a treatment for cancer. Two programmes are mutually exclusive if the costs and/or effects of one programme are affected by whether the other programme is implemented or not, e.g. two alternative drug therapies to lower blood pressure."

Similar definitions are used elsewhere (39-41), although some authors only describe treatments as mutually exclusive if patients who have received one of the therapies will not receive the other(s) (42). Johannesson's definition suggests that interactions are the key determinant of whether treatments should be considered mutually exclusive or independent and suggests that whenever treatments interact we should treat the cells of our factorial trial (e.g. no treatment, A, B and A+B) as mutually exclusive options, comparing the treatment combinations incrementally and selecting the single strategy that maximises net benefit for that population. Conversely, this definition implies that if there is no interaction between treatments, we should treat A and B as independent treatment options and compare their ICERs against the ceiling ratio separately, adopting those treatments with ICERs below our ceiling ratio (38, 40, 43, 44). However, it is currently unclear how much evidence (or what level of statistical significance) is required to conclude that there is an interaction and whether we can pre-specify the form of analysis for economic evaluations of factorial trials before interactions are known. More generally, our observation that any interventions given to the same patient population may produce interactions for QALYs even if they target different diseases raises the wider question of whether any treatments for the same population can be considered independent. We plan to investigate these issues further in future work.

As well as making different assumptions about interactions, the different forms of analysis for factorial trials lend themselves to different decision rules for economic evaluation. The at-the-margins approach implicitly treats the two factors in a 2x2 factorial design as independent options, calculating separate ICERs for A vs no A and for B vs no B that can be independently compared against the ceiling ratio and used to make separate decisions on the two treatments. By contrast, the within-the-table approach treats the four combinations in the

factorial design as mutually exclusive options, estimating costs, QALYs and net benefits for each of the four options and allowing us to plot the results on the cost-effectiveness plane, identify the frontier and select the option that maximises net benefits at our ceiling ratio.

### *3.3. Worked example*

We simulated data on costs and QALYs for the 1,000 participants in our hypothetical 2x2 factorial trial to illustrate the impact that the choice of analytical method can have on the conclusions of economic evaluation using a simple model (details available on request). Drugs A and B were both assumed to reduce the risk of clinical events (e.g. stroke), with zero interaction on a log-odds scale, but reduce patients' quality of life slightly due to side-effects. Each event was assumed to increase healthcare costs and reduce QALYs; the cost and QALY loss from each event was assumed to vary randomly between patients but not differ systematically between treatments.

Since treatments have a multiplicative effect on the odds of having an event, we observe a sub-additive interaction on the number of events (Table 5) with a positive interaction term and negative main effects. Since each event increases costs but reduces QALYs, we see a positive interaction for costs and a negative interaction for QALYs. After allowing for additive effects (i.e. drug costs and the disutility of treatment side-effects), main effects are positive for costs and QALYs (i.e. both A and B increase NHS costs and improve health); overall, we therefore see a large super-additive interaction for costs and a sub-additive interaction for QALYs (Table 5). Combining costs and QALYs to estimate total net benefits produces a qualitative interaction at a £30,000/QALY ceiling ratio (Rc), since A increases net benefits (i.e. is cost-effective) when used alone, but reduces net benefits when added to B. However, as is commonly observed in trial-based economic evaluations, there is substantial variability and uncertainty in costs, QALYs and net benefits that reduces the power to draw inferences about either main effects or interactions.

**Table 5:** Worked example of economic evaluation of 2x2 factorial design: group means

| | **Placebo for A + placebo for B (n=250)** | **Drug A + placebo for B (n=250)** | **Drug B + placebo for A (n=250)** | **Drug A + Drug B (n=250)** |
|---|---|---|---|---|
| Mean no. events per patient (SD) | 7.2 (2.3) | 5.9 (2.4) | 5.3 (2.1) | 4.5 (2.1) |
| Mean cost per patient (SD) | £87,804 (£33,508) | £98,324 (£32,408) | £109,109 (£28,851) | £125,015 (£29,958) |
| Mean QALYs per patient (SD) | 18.1 (6.6) | 18.9 (5.9) | 19.6 (6.0) | 19.8 (5.6) |
| Mean total net benefits per patient at Rc=£30,000 (SD) | £455,010 (£211,746) | £470,155 (£186,507) | £479,504 (£188,881) | £468,985 (£174,617) |

Simulated trial data were analysed in Stata using within-the-table and at-the-margins approaches and OLS regression with and without an interaction term (Tables 6-9). Costs and QALYs were analysed using seemingly unrelated estimation, which combines the results of two or more separate regression models (45); when the `suest` post-estimation command is used to link OLS models of costs and QALYs that have been run using the same data, standard errors are adjusted to allow for the correlation between costs and QALYs but estimates of mean coefficient values are identical to standard OLS. OLS was used for simplicity in these analyses, although costs, QALYs and net benefits were positively skewed ($p \leq 0.001$) and variances were lower for patients receiving B. The effect of using alternative regression models will be investigated in future research.

At-the-margins analysis suggested that both A and B are cost-effective treatments for this condition when they are considered to be independent options: prescribing A (with or without B) cost £25,530/QALY gained compared with not giving A, while B cost £20,189/QALY gained compared with not giving B (Table 6). Based on a £30,000/QALY ceiling ratio, we might therefore conclude that both A and B are cost-effective treatments and should be recommended. Furthermore, if treated these treatments as independent options, we may conclude that since A and B are each cost-effective and have positive incremental net benefits (INB) at a £30,000/QALY ceiling ratio, both A and B should be adopted and that use of A+B will maximise net benefits. However, this inference would be incorrect in this example as there is a qualitative interaction for net benefits. More generally, interactions of any size or direction will bias the ICERs calculated using the at-the-margins approach, which may invalidate our conclusions about the cost-effectiveness of A and B. At-the-margins estimates also show the effect of A averaged over patients receiving B and those not receiving B, giving estimates that are driven by the proportion of patients randomised to B; as such, at-the-margins estimates do not reflect the incremental costs, effects or net benefits that would be seen in current practice or indicate the cost-effectiveness of giving A alone.

Regression analyses without interaction terms replicated the mean treatment effects calculated using at-the-margins analysis (Table 7), although standard errors differed. However, unlike at-the-margins analysis, regression enables prediction of group means and their standard errors (Table 7, Figure 1A). These predictions can be used to consider the cost-effectiveness of the four treatment strategies as mutually-exclusive options, draw conclusions about dominance and identify which of the four options maximises net benefit, providing that

we are confident that the model used is appropriate. In this example, the predicted values from regression suggest that A is extendedly dominated by a combination of B and placebo (Figure 1A) and explicitly predict that A+B is the optimal treatment at a £30,000/QALY ceiling ratio, costing £25,530/QALY gained compared with B. However, these predictions rely on the same assumptions as at-the-margins analysis and will be biased if interactions are present.

**Table 6:** Worked example of economic evaluation of 2x2 factorial trial: at-the-margins approach

| | **Treatment A** | | **Treatment B** | |
|---|---|---|---|---|
| | **Placebo (n=500)** | **Active drug A (n=500)** | **Placebo (n=500)** | **Active drug B (n=500)** |
| Mean cost per patient (SD) | £98,456 (£33,005)* | £111,669 (£33,917)* | £93,064 (£33,348)* | £117,062 (£30,439)* |
| Difference in cost (SE) | £13,213 (£2,116)* | | £23,998 (£2,019)* | |
| Mean QALYs per patient (SD) | 18.9 (6.4)* | 19.4 (5.8)* | 18.5 (6.3)* | 19.7 (5.8)* |
| Difference in QALYs (SE) | 0.52 (0.38) | | 1.19 (0.38)* | |
| Mean total net benefits per patient at Rc=£30,000 (SD) | £467,257 (£200,813)* | £469,570 (£180,480)* | £462,582 (£199,470)* | £474,245 (£181,783)* |
| INB at Rc=£20,000 (SE) | £2,313 (£12,075) | | £11,662 (£12,069) | |
| Incremental cost/QALY | £25,530 | | £20,189 | |

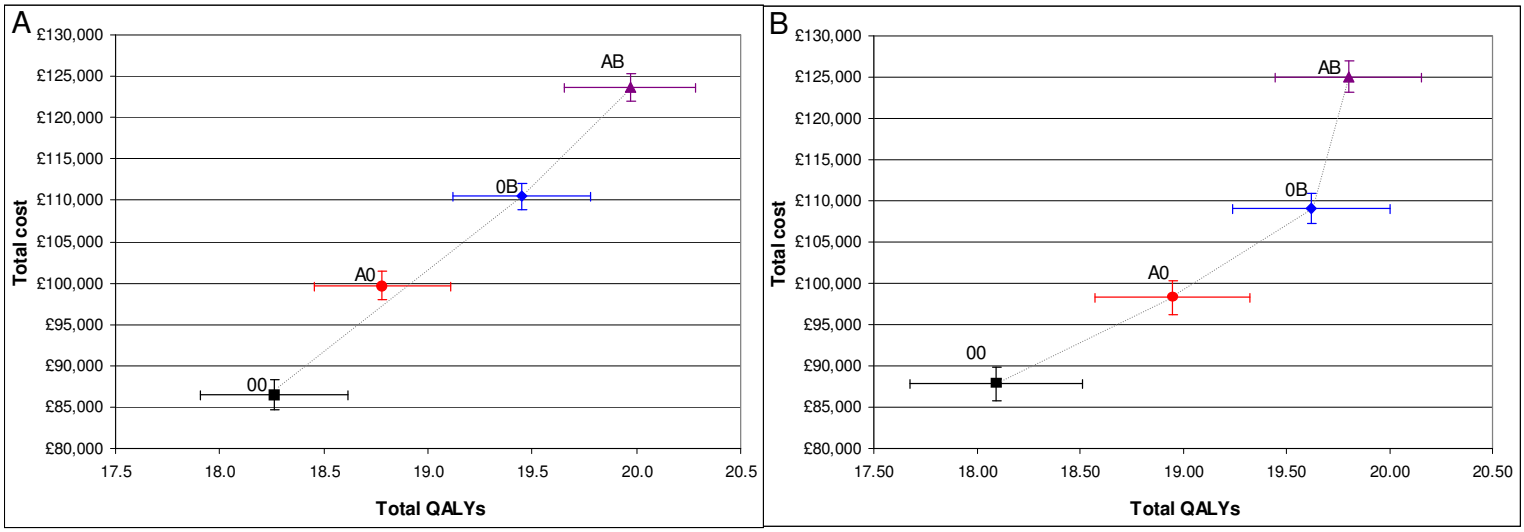
* Significantly greater than zero (p<0.05); SD, standard deviation; SE, standard error.

**Table 7:** Worked example of economic evaluation of 2x2 factorial trial: OLS regression without interaction term

| | | **Total cost/ patient†** | **Total QALYs/ patient†** | **Net benefit/ patient‡** | **Cost/QALY** | | |
|---|---|---|---|---|---|---|---|
| | | | | | **vs placebo** | **vs A** | **vs B** |
| Treatment effect for A (SE) | | £13,213 (£1,974)* | 0.52 (0.38) | £2,313 (£12,075) | - | - | - |
| Treatment effect for B (SE) | | £23,998 (£1,974)* | 1.19 (0.38)* | £11,662 (£12,075) | - | - | - |
| Constant term (SE) | | £86,457 (£1,794)* | 18.26 (0.35)* | £461,426 (£10,457)* | - | - | - |
| Predicted mean outcome (SE) | Placebo for A + placebo for B (n=250) | £86,457 (£1,794)* | 18.3 (0.35)* | £461,426 (£10,457)* | - | - | - |
| | Drug A + placebo for B (n=250) | £99,670 (£1,753)* | 18.8 (0.33)* | £463,739 (£10,457)* | £25,530 | - | - |
| | Drug B + placebo for A (n=250) | £110,456 (£1,624)* | 19.5 (0.33)* | £473,088 (£10,457)* | £20,189 | £16,070 | - |
| | Drug A + Drug B (n=250) | £123,669 (£1,664)* | 20.0 (0.31)* | £475,402 (£10,457)* | £21,809 | £20,189 | £25,530 |

* Significantly greater than zero (p<0.05); SE, standard error.
† Results are based on seemingly unrelated regression, which estimates robust standard errors.
‡ Based on a ceiling ratio of £30,000/QALY.

**Figure 1:** Cost-effectiveness plane showing predicted costs and outcomes from regression analyses on the same simulated data. A: Regression without interaction term. B: Regression with interaction term. Error bars show standard errors from regression analysis. The dotted grey line represents the cost-effectiveness frontier, based on the assumption that the treatments are mutually exclusive.




However, the predictions from regression models are valid only if the functional form is correct. In this example, we observe large interactions for costs (p=0.172) and QALYs (p=0.377) and a qualitative interaction for net benefits (p=0.288, Table 8). None of the interaction effects reach statistical significance; given that most two-arm trials are not powered to detect significant differences in mean costs or net benefits (46-48) and the variance around the interaction term is four-fold higher than that for main effects (11, 22), this finding is likely to be common among economic evaluations of factorial trials. However, these interactions nonetheless introduce substantial bias into at-the-margins estimates and those from regression without an interaction term.

Including an interaction term completely changes the conclusions for this example. When we consider each of the four treatments as separate mutually-exclusive treatment options using regression analysis with an interaction term (Table 8) or within-the-table approach (Table 9), we see that although A is cost-effective compared with placebo (costing £12,297/QALY gained) and B is cost-effective compared with A (costing £16,070/QALY gained), A+B costs £88,573/QALY gained compared with B (Figure 1B). On that basis, maximising net benefits at a £30,000/QALY ceiling ratio would require us to adopt treatment B (not A+B as at-the-margins analysis suggested). However, including the interaction term also substantially increased standard errors around treatment effects and predicted costs and benefits for each group, which will affect statistical inferences, cost-effectiveness acceptability curves and the value of collecting additional information.

**Table 8:** Worked example of economic evaluation of 2x2 factorial design: OLS regression with an interaction term

| | | Total cost/ patient† | Total QALYs/ patient† | Net benefit/ patient‡ | Cost/QALY | | |
|---|---|---|---|---|---|---|---|
| | | | | | vs placebo | vs A | vs B |
| Treatment effect for A (SE) | | £10,520 (£2,944)* | 0.86 (0.56) | £15,145 (£17,076) | - | - | - |
| Treatment effect for B (SE) | | £21,305 (£2,792)* | 1.53 (0.57)* | £24,494 (£17,076) | - | - | - |
| Interaction (SE) | | £5,386 (£3,945) | -0.68 (0.77) | £-25,664 (£24,149) | - | - | - |
| Constant term (SE) | | £87,804 (£2,116)* | 18.09 (0.42)* | £455,010 (£12,074)* | - | - | - |
| Predicted mean outcome (SE) | Placebo for A + placebo for B (n=250) | £87,804 (£2,116)* | 18.1 (0.42)* | £455,010 (£12,074)* | - | - | - |
| | Drug A + placebo for B (n=250) | £98,324 (£2,047)* | 18.9 (0.37)* | £470,155 (£12,074)* | £12,297 | - | - |
| | Drug B + placebo for A (n=250) | £109,109 (£1,822)* | 19.6 (0.38)* | £479,504 (£12,074)* | £13,956 | £16,070 | - |
| | Drug A + Drug B (n=250) | £125,015 (£1,892)* | 19.8 (0.35)* | £468,986 (£12,074)* | £21,809 | £31,375 | £88,573 |

* Significantly greater than zero (p<0.05); SE, standard error.
† Results are based on seemingly unrelated regression, which estimates robust standard errors.
‡ Based on a ceiling ratio of £30,000/QALY.

**Table 9:** Worked example of economic evaluation of 2x2 factorial design: within-the-table approach

| | Placebo for A + placebo for B (n=250) | Drug A + placebo for B (n=250) | Drug B + placebo for A (n=250) | Drug A + Drug B (n=250) |
|---|---|---|---|---|
| Mean cost per patient (SD) | £87,804 (£33,508)* | £98,324 (£32,408)* | £109,109 (£28,851)* | £125,015 (£29,958)* |
| Incremental cost per patient vs placebo (SE) | N/A | £10,520 (£2,948)* | £21,305 (£2,797)* | £37,211 (£2,843)* |
| Mean QALYs per patient (SD) | 18.1 (6.6)* | 18.9 (5.9)* | 19.6 (6.0)* | 19.8 (5.6)* |
| Incremental total QALYs per patient vs placebo (SE) | N/A | 0.86 (0.56) | 1.53 (0.57)* | 1.71 (0.55)* |
| Mean total net benefits per patient at Rc=£30,000 (SD) | £455,010 (£211,746)* | £470,155 (£186,507)* | £479,504 (£188,881)* | £468,985 (£174,617)* |
| INB vs placebo at Rc=£30,000 (SE) | N/A | £15,145 (£17,846) | £24,494 (£17,946) | £13,976 (£17,358) |
| Cost/QALY vs placebo | - | £12,297 | £13,956 | £21,809 |
| Cost/QALY vs Drug A + placebo for B | - | - | £16,070 | £31,375 |
| Cost/QALY vs Drug B + placebo for A | - | - | - | £88,574 |

* Significantly greater than zero (p<0.05); SD, standard deviation; SE, standard error.

The bias present within at-the-margins estimates and regression without an interaction term will prevent us from maximising the health gained from our budget. These approaches suggest that A+B should be given to our population; spending £1 billion on this treatment strategy will allow us to treat 8,000 patients and accrue 16,000 QALYs. By contrast, choosing the treatment that maximises net benefits based on the within-the-table approach or regression with an interaction term (B monotherapy) will allow us to treat 9,165 patients and

accrue 18,000 QALYs: 2,000 more than we would achieve using the biased at-the-margins estimates and making separate decisions on A and B.

The different analytical frameworks may also produce conflicting conclusions in different ways from this example. In particular, qualitative interactions for costs and/or QALYs could cause A or B to dominate A+B even if at-the-margins analysis finds both A and B to be cost-effective. At-the-margins analysis may also find A to cost less than £30,000/QALY when within-the-table analysis shows that the strategy maximising net benefits at this ceiling ratio is to give neither A nor B. Monotherapy with treatment A may also maximise net benefits without being cost-effective in at-the-margins analyses.

## 4. Discussion: which approach to choose?

This paper represents a preliminary examination of an under-researched topic that forms part of HD's DPhil. The initial results demonstrate that the choice of analytical method affects the conclusions of the economic evaluation and can affect the decisions that would be made regarding resource allocation.

For analyses of clinical endpoints, a two-step approach to hypothesis testing is typically conducted, whereby at-the-margins analysis (or regression without an interaction term) is used unless the interaction term reaches a pre-specified level of significance (5, 12, 21, 22, 25). If we adopted this approach in the worked example shown in Section 3.3, we would base our conclusions on regression without an interaction term since interaction terms were not statistically significant. Adopting this two-step approach would maximise statistical power (21, 22), ensure that we used all available data to estimate main effects and prevent our conclusions from being affected by an apparent interaction effect that arose entirely due to chance. However, as we have shown, omitting non-significant interaction terms would have introduced substantial bias into the analysis that would change the conclusions of our economic evaluation and prevent us from allocating healthcare resources efficiently, resulting in the loss of 2,000 QALYs per £1 billion spent.

There are several reasons why the two-stage approach is inappropriate for economic evaluation. Firstly, interactions for clinical endpoints are generally small (5) and qualitative interactions are rare (9); by contrast, interactions in costs and QALYs are likely to be

substantially larger and qualitative interactions may be common for net benefits (Section 3.1). Since the bias introduced by omitting the interaction term is proportional to the true interaction effect (22), the potential for bias in clinical analyses is likely to be negligible, whereas at-the-margins estimates of costs, effects, net benefits and ICERs are prone to substantially larger biases that may change the conclusions.

Secondly, clinical analyses focus on hypothesis testing and assessing which treatment(s) have evidence of superior efficacy, whereas economic evaluation concerns estimation. The clinical conclusions of RCTs regarding whether or not treatments are more effective than their comparators are generally driven more by statistical inference than the magnitude of the treatment effect; at-the-margins analysis maximises the chance of detecting significant differences and is unlikely to distort these clinical conclusions unless there is a qualitative interaction. By contrast, estimation is central to economic evaluation, since we must trade off the estimated health gains against the estimated cost and compare our ICER against the ceiling ratio; bias within our ICER estimate may therefore affect our conclusions about whether treatment is cost-effective. Conversely, hypothesis testing is of less (or arguably no) importance to resource allocation decisions, since we will maximise health gains from the available budget by choosing the treatment with the highest expected net benefits regardless of statistical significance (42, 49).

Thirdly, the two-step approach equates absence of proof with proof of absence and relies upon an arbitrarily chosen significance level. Statistical tests for interactions are particularly problematic as very few factorial trials are powered to detect interactions in the primary clinical endpoint (12), let alone net benefits. In this example, even though the interaction for net benefits is not statistically significant ($p=0.288$), the point estimates suggest that there is a qualitative interaction that affects the conclusions of our analysis.

The two-stage approach may therefore be appropriate for analysis of clinical endpoints, where it may be preferable to accept a small amount of bias in order to gain statistical power, since conclusions are driven by statistical significance, not point estimates. However, we suggest that it is more appropriate to use an unbiased, inefficient approach for economic evaluation than an efficient, biased one since interactions may be substantially larger, our conclusions are driven more by point estimates than p-values and analyses may not be powered to exclude qualitative interactions. Within economic evaluation, it may therefore be

appropriate to include an interaction term unless interactions are proven to be negligible, rather than excluding it unless interactions are proven to be important. However, although unbiased, within-the-table analysis or regression with an interaction term may nonetheless give the wrong answer by chance (especially for small samples) and would systematically overestimate the value of information and bias cost-effectiveness acceptability curves in cases where there is genuinely no interaction.

Regression-based cost-effectiveness analysis also raises a number of challenges, since costs and QALYs are rarely normally distributed (30, 37, 50) and frequently heteroskedastic (37, 50), regression cannot be conducted directly on the ICER due to its statistical properties (30, 50-52), analyses of net benefit depend on the unknown value of the ceiling ratio and analyses of costs and effects should allow for the correlation between these endpoints (30). Applying regression-based CEA to factorial trials raises further issues, such as how to analyse results if the relationship between the factors is multiplicative rather than additive or if the importance of the interaction term differs between different components of cost and effect or varies with ceiling ratio. Transformation (30, 37), GLM models (30), seemingly unrelated regression (30) and/or statistical modelling of specific components of cost and effect (30, 53) may help to address these issues. We plan to explore these methods further in future work.

## 5. Conclusions

In economic evaluations based on factorial RCTs, it is essential to give appropriate consideration to how the factorial design affects the statistical analysis. The choice of analytical approach is likely to be of greater importance for economic evaluation than for clinical outcomes as economic evaluation focuses on estimation not hypothesis testing and as interactions may be larger for costs, QALYs and net benefits than clinical outcomes. The clinical approach of assuming interactions to be zero unless proven otherwise may therefore be inappropriate for economic outcomes. Instead, it may be preferable to conduct within-the-table analysis or regression analyses including an interaction term unless interactions are shown to be negligible.

## 6. Acknowledgements

The authors would like to thank Sarah Wordsworth and Elizabeth Stokes for their comments on the manuscript.

## 7. References


1. Holman RR, Paul S, Farmer A, et al. Atorvastatin in Factorial with Omega-3 EE90 Risk Reduction in Diabetes (AFORRD): a randomised controlled trial. Diabetologia. 2009; 52: 50-9.
2. Goldberg MR, Offen WW. Pinacidil with and without hydrochlorothiazide. Dose-response relationships from results of a 4 x 3 factorial design study. Drugs. 1988; 36 Suppl 7: 83-92.
3. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 Collaborative Group. Lancet. 1995; 345: 669-85.
4. Fisher RA. The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain. 1926; 33: 503-13.
5. McAlister FA, Straus SE, Sackett DL, et al. Analysis and reporting of factorial trials: a systematic review. JAMA. 2003; 289: 2545-53.
6. Brittain E, Wittes J. Factorial designs in clinical trials: the effects of non-compliance and subadditivity. Stat Med. 1989; 8: 161-71.
7. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. http://wwwcochrane-handbookorg/. Version 5.0.2. The Cochrane Collaboration, September 2009.
8. Bernaerts K, Gysemans KP, Nhan Minh T, et al. Optimal experiment design for cardinal values estimation: guidelines for data collection. Int J Food Microbiol. 2005; 100: 153-65.
9. Stampfer MJ, Buring JE, Willett W, et al. The 2 x 2 factorial design: its application to a randomized trial of aspirin and carotene in U.S. physicians. Stat Med. 1985; 4: 111-6.
10. Armitage P, Berry G, Mathews JNS. Statistical methods in medical research. 4th ed. Malden, MA: Blackwell Science Ltd., 2002.
11. Byar DP. Factorial and reciprocal control designs. Stat Med. 1990; 9: 55-63; discussion 63-4.
12. Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. BMC Med Res Methodol. 2003; 3: 26.
13. Lubsen J, Pocock SJ. Factorial trials in cardiology: pros and cons. Eur Heart J. 1994; 15: 585-8.
14. Sheikh A, Smeeth L, Ashcroft R. Randomised controlled trials in primary care: scope and application. Br J Gen Pract. 2002; 52: 746-51.
15. Korttila K, Apfel CC. Factorial design provides evidence to guide practice of anaesthesia. Acta Anaesthesiol Scand. 2005; 49: 927-9.
16. Yates F, Mather K. Ronald Aylmer Fisher. Biographical Memoirs of Fellows of the Royal Society of London. 1963; 9: 91-120.
17. Fisher RA. The Design of Experiments. London: Collier Macmillan, 1935.
18. Chalmers TC, Eckhardt RD, Reynolds WE, et al. The treatment of acute infectious hepatitis. Controlled studies of the effects of diet, rest, and physical reconditioning on the acute course of the disease and on the incidence of relapses and residual abnormalities. J Clin Invest. 1955; 34: 1163-235.
19. Hopewell S, Dutton S, Yu LM, et al. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010; 340: c723.
20. Silagy CA, Jewell D. Review of 39 years of randomized controlled trials in the British Journal of General Practice. Br J Gen Pract. 1994; 44: 359-63.
21. Hung HM, Chi GY, O'Neill RT. Efficacy evaluation for monotherapies in two-by-two factorial trials. Biometrics. 1995; 51: 1483-93.
22. Hung HM. Two-stage tests for studying monotherapy and combination therapy in two-by-two factorial trials. Stat Med. 1993; 12: 645-60.
23. Slud EV. Analysis of factorial survival experiments. Biometrics. 1994; 50: 25-38.
24. Bria E, Di Maio M, Nistico C, et al. Factorial design for randomized clinical trials. Ann Oncol. 2006; 17: 1607-8.
25. Simon R, Freedman LS. Bayesian design and analysis of two x two factorial clinical trials. Biometrics. 1997; 53: 456-64.
26. Ottenbacher KJ. Interpretation of interaction in factorial analysis of variance design. Stat Med. 1991; 10: 1565-71.
27. Simon GE, Manning WG, Katzelnick DJ, et al. Cost-effectiveness of systematic depression treatment for high utilizers of general medical care. Arch Gen Psychiatry. 2001; 58: 181-7.

28. Welton NJ, Ades AE, Caldwell DM, et al. Research prioritization based on expected value of partial perfect information: a case-study on interventions to increase uptake of breast cancer screening. J R Statist Soc A. 2008; 171: 807-41.
29. Brown J, Welton NJ, Bankhead C, et al. A Bayesian approach to analysing the cost-effectiveness of two primary care interventions aimed at improving attendance for breast screening. Health Econ. 2006; 15: 435-45.
30. Willan AR, Briggs AH. Statistical analysis of cost-effectiveness data. Chichester: Wiley, 2006.
31. Eckermann S, Coory M, Willan AR. Indirect comparison: relative risk fallacies and odds solution. J Clin Epidemiol. 2009; 62: 1031-6.
32. Brazier J, Ratcliffe J, Salomon JA, et al. Chapter 8: Methods for obtaining health state values: generic preference-based measures of health and alternatives. In: Brazier J, ed., Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press, 2007.
33. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care. 2002; 40: 113-28.
34. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997; 35: 1095-108.
35. Kilian R, Matschinger H, Loeffler W, et al. A comparison of methods to handle skew distributed cost variables in the analysis of the resource consumption in schizophrenia treatment. J Ment Health Policy Econ. 2002; 5: 21-31.
36. Moran JL, Solomon PJ, Peisach AR, et al. New models for old questions: generalized linear models for cost prediction. J Eval Clin Pract. 2007; 13: 381-9.
37. Glick HA, Doshi JA, Sonnad SS, et al. Economic evaluation in clinical trials. Oxford: Oxford University Press, 2007.
38. Johannesson M, O'Conor RM. Cost-utility analysis from a societal perspective. Health Policy. 1997; 39: 241-53.
39. Hunink M, Glasziou P, Siegel J, et al. Decision making in health and medicine: integrating evidence and values. Cambridge, UK: Cambridge University Press, 2001.
40. Johannesson M. Theory and Methods of Economic Evaluation of Health Care. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.
41. Phillips C. What is cost-effectiveness? (http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Cost-effect.pdf). February 2009.
42. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. Oxford: Oxford University Press, 2005.
43. Weinstein MC. Principles of cost-effective resource allocation in health care organizations. Int J Technol Assess Health Care. 1990; 6: 93-103.
44. Johannesson M, Weinstein MC. On the decision rules of cost-effectiveness analysis. J Health Econ. 1993; 12: 459-67.
45. StataCorp. Suest - seemingly unrelated estimation. Stata Base Reference Manual: Release 11. College Station, TX: StataCorp LP, 2009.
46. Briggs A. Economic evaluation and clinical trials: size matters. Bmj. 2000; 321: 1362-3.
47. Backhouse ME. Use of randomised controlled trials for producing cost-effectiveness evidence: potential impact of design choices on sample size and study duration. Pharmacoeconomics. 2002; 20: 1061-77.
48. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: ISPOR RCT-CEA Task Force Report. Value Health. 2005; 8: 521-33.
49. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999; 18: 341-64.
50. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ. 2002; 11: 415-30.
51. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ. 1998; 7: 723-40.
52. Polsky D, Glick HA, Willke R, et al. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Econ. 1997; 6: 243-52.
53. Briggs AH, Bousquet J, Wallace MV, et al. Cost-effectiveness of asthma control: an economic appraisal of the GOAL study. Allergy. 2006; 61: 531-6.