



Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations

Richard M. Nixon and Simon G. Thompson*

MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge, UK

Summary

Background: Overall assessments of cost-effectiveness are now commonplace in informing medical policy decision making. It is often important, however, also to investigate how cost-effectiveness varies between patient subgroups. Yet such analyses are rarely undertaken, because appropriate methods have not been sufficiently developed.

Methods: We propose a coherent set of Bayesian methods to extend cost-effectiveness analyses to adjust for baseline covariates, to investigate differences between subgroups, and to allow for differences between centres in a multicentre study using a hierarchical model. These methods consider costs and effects jointly, and allow for the typically skewed distribution of cost data. The results are presented as inferences on the cost-effectiveness plane, and as cost-effectiveness acceptability curves.

Results: In applying these methods to a randomised trial of case management of psychotic patients, we show that overall cost-effectiveness can be affected by ignoring the skewness of cost data, but that it may be difficult to gain substantial precision by adjusting for baseline covariates. While analyses of overall cost-effectiveness can mask important subgroup differences, crude differences between centres may provide an unrealistic indication of the true differences between them.

Conclusions: The methods developed allow a flexible choice for the distributions used for cost data, and have a wide range of applicability – to both randomised trials and observational studies. Experience needs to be gained in applying these methods in practice, and using their results in decision making. Copyright © 2005 John Wiley & Sons, Ltd.

Keywords cost-effectiveness analysis; covariate adjustment; subgroups; multicentre studies; hierarchical models

Introduction

Cost-effectiveness analysis using data from individual patients is now an established process for informing health policy decision making. For example, the results from randomised trials often include an overall cost-effectiveness evaluation [1]. While it is common for analyses of effectiveness outcomes, such as survival, to be extended to

consider the effects of baseline covariates, this is uncommon for cost-effectiveness. This is principally because of the technical difficulties of considering costs and effects together in one analysis. In this paper we propose a general set of methods to allow such analyses to be undertaken.

Adjustment for baseline covariates is crucial in non-randomised studies, in order to make proper

*Correspondence to: MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK.
E-mail: simon.thompson@mrc-bsu.cam.ac.uk

allowance for case-mix differences between those patients receiving one intervention and those receiving another [2]. Case-mix characteristics may influence both costs and effects, and unadjusted results are likely to be biased. In randomised trials the rationale for adjusting for baseline variables is both to allow for any chance imbalance in patient characteristics between randomised groups, as well as the possibility of gaining precision in the resulting inferences [3].

It is often likely that the cost-effectiveness of an intervention varies importantly across different groups of patients. For example, patients at high risk may gain the most benefit without the cost of the intervention being much increased [4]. To investigate this, cost-effectiveness needs to be compared between separate subgroups of patients defined by particular characteristics. Examples include the comparison of cost-effectiveness between men and women, between patients of different ages, or between patients in different centres in a multicentre study.

Statistical methods for performing overall cost-effectiveness analyses need to be flexible enough to allow for potential correlation between costs and effects [5], and for the fact that cost data are usually very skewed [6]. The inferences from such analyses involve presentation on the cost-effectiveness plane (CE-plane) [7], as incremental net benefits (INB) [8,9], or as cost-effectiveness acceptability curves (CEAC) [10,11]. The latter explicitly describes a range of inferences depending on the purchaser's 'willingness-to-pay' per unit of health gain [1]. The CEAC provides the probability, given the data available and any prior information, that the new intervention is cost effective for any willingness-to-pay value.

The above are now the standard tools of cost-effectiveness analysis. Their extension to handle covariates has received little previous attention. Work by Hoch and colleagues [12] was based on analysing the net benefit for fixed willingness-to-pay values, rather than costs and effects jointly as a two-dimensional outcome. Willan and colleagues [13] directly considered costs and effects jointly, but assumed that they were normally distributed. Manca and colleagues [14] have considered a hierarchical model in the context of multicentre data, but only for the net benefit. We return to the limitations of these methods in the discussion. The Bayesian methods we present here consider costs and effects jointly, and allow for the skewness typical of cost data; these are important extensions

to previously available methods. We develop the methods in the context of two-arm randomised trials of a new intervention against standard treatment, where effects and costs are assessed for each individual over a fixed period of time. However, the methods have wider applicability, for example to non-randomised studies.

Example

The data we use to exemplify the methods come from a randomised trial of case management in psychotic patients [15]. The trial compared intensive intervention (in which managers had a case load of 10–15 patients) with the standard (case loads of 30–35 patients). The rationale for the trial was that, by providing more intensive support, the patients' use of health and other services might be reduced, and clinical outcomes might be improved. The societal cost for each individual over 24 months was estimated according to recorded resource use (including social, hospital and community components) and fixed unit costs [16]. While clinical outcomes were assessed, the principal effectiveness outcome was the number of days in hospital for psychiatric problems over 24 months [15]. Although this is not a utility measure, it was used in the original cost-effectiveness analysis [16] and is suitable for exemplifying the methods being proposed. As part of the baseline data, information was collected on the number of days each patient spent in hospital in the 24 months prior to randomisation. Patients were recruited from four centres in the UK, and data on costs and effects were available for 667 of the 708 patients randomised (335 in the intervention group and 332 in the control group).

The distributions of the effectiveness outcome (days in hospital) are shown in Figure 1, together with superimposed fitted normal distributions (which are symmetric) and gamma distributions (which are skewed). The gamma distributions clearly fit the data much better than the normal distributions. The cost data are shown in Figure 2, and again gamma fit much better than normal distributions. A summary of the sample mean effects and costs is shown in Table 1. The overall incremental effectiveness is close to zero ($\Delta_E = +0.47$ days), while the overall incremental cost is positive ($\Delta_C = \text{£}1849$) but not statistically significant. Hence, overall, intensive case management appears to be slightly more expensive than

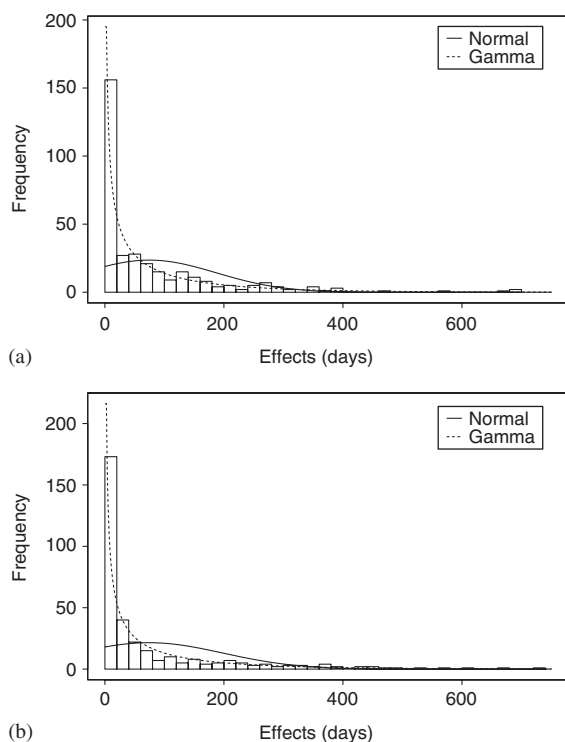


Figure 1. *Effectiveness data distributions*: Effectiveness data (days in hospital over 24 months) in the case management trial, with fitted densities from normal and gamma distributions: (a) control group, (b) intervention group

standard case management, without changing duration of subsequent hospitalisation. The point estimate of the incremental cost–effectiveness ratio (ICER = Δ_C/Δ_E) is $1849/0.47 = \text{£}3946/\text{day}$ in hospital. The data set is unusual both in the skewed nature of the effectiveness outcome, and in the substantial correlations between the cost and effectiveness outcomes (0.71 and 0.81 in the control and intervention groups, respectively) which arise because the duration of hospitalisation is a strong cost driver [16].

Overall cost–effectiveness

We first consider methods for overall cost–effectiveness [5], and extend this later for handling covariates and subgroups. We denote by E_{ij} the effectiveness outcome, and by C_{ij} the cost, for the j th person in arm i of the trial ($i = 1$ control, $i = 2$ intervention). These follow some distributions

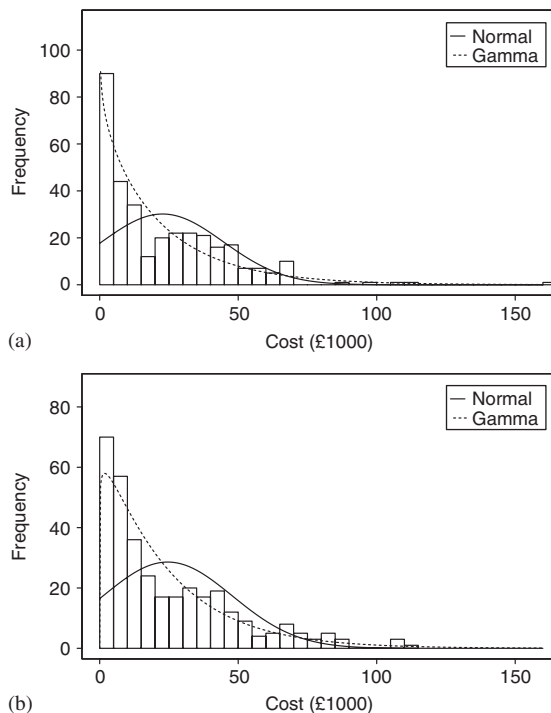


Figure 2. *Cost data distributions*: Cost data in the case management trial, with fitted densities from normal and gamma distributions: (a) control group, (b) intervention group

with unknown mean (represented by ϕ) and standard deviation (represented by σ) in each arm of the trial. We use subscript E for effects, and C for costs:

$$E_{ij} \sim \text{Dist}(\phi_{Eij}, \sigma_{Ei})$$

$$C_{ij} \sim \text{Dist}(\phi_{Cij}, \sigma_{Ci}) \tag{1}$$

If we use normal distributions then, for example, the effects in the control group are normally distributed with mean ϕ_{E1j} and standard deviation σ_{E1} . There are a total of four distributions, one for each of costs and effects in each of the control and intervention arms, as depicted in Figures 1 and 2. In a simple model without covariates, we allow for the potential correlation between costs and effects by writing:

$$\phi_{Eij} = \mu_{Ei} + \beta_i(C_{ij} - \phi_{Cij})$$

$$\phi_{Cij} = \mu_{Ci} \tag{2}$$

Thus, in arm i , the mean cost is the same (μ_{Ci}) for all the patients in that arm, while the predicted

Table 1. Summary data in the case management trial: Mean effects and costs by randomised group, incremental effects and costs, and incremental cost-effectiveness ratio (ICER); overall, by intelligence subgroup, and by centre

| | Effectiveness: days in hospital over 2 years | | | | | | Societal costs over 2 years (£) | | | | | | ICER (£ per day in hospital) |
|----------------------|--|-----------------|----------------|-----------------|--------------------------|--------------|---------------------------------|-----------|-----------|-----------|-------------|--|------------------------------|
| | Intensive | | Standard | | Incremental ^a | | Intensive | | Standard | | Incremental | | |
| | Mean (SD) | Mean (SD) | Mean (SE) | Mean (SD) | Mean (SE) | Mean (SD) | Mean (SD) | Mean (SE) | Mean (SD) | Mean (SE) | Mean (SE) | | |
| Overall (n = 667) | 74.19 (124.52) | 74.66 (112.45) | 0.47 (9.19) | 24 553 (23 408) | 22 704 (22 000) | 1849 (1755) | | | | | | | 3946 |
| Intelligence groups | | | | | | | | | | | | | |
| Normal (n = 453) | 85.93 (131.72) | 65.91 (100.21) | -20.02 (11.01) | 25 529 (24 539) | 21 030 (19 955) | 4498 (2109) | | | | | | | -225 |
| Borderline (n = 100) | 48.23 (97.96) | 105.77 (159.52) | 57.54 (26.16) | 23 808 (20 583) | 28 983 (30 719) | -5175 (5394) | | | | | | | -90 |
| Centres | | | | | | | | | | | | | |
| Centre 1 (n = 185) | 73.45 (138.44) | 64.27 (85.09) | -9.18 (16.88) | 24 161 (25 036) | 19 435 (22 743) | 4727 (3507) | | | | | | | -515 |
| Centre 2 (n = 149) | 79.92 (129.85) | 60.58 (114.19) | -19.35 (20.01) | 25 455 (24 773) | 25 194 (23 492) | 261 (3916) | | | | | | | -13 |
| Centre 3 (n = 189) | 66.67 (112.49) | 91.95 (125.24) | 25.28 (17.29) | 23 347 (19 777) | 22 935 (19 882) | 412 (2940) | | | | | | | 16 |
| Centre 4 (n = 144) | 78.43 (116.40) | 78.86 (121.95) | 0.43 (19.90) | 25 590 (24 335) | 24 102 (22 191) | 1488 (3803) | | | | | | | 3472 |

^a Incremental effectiveness is standard minus intensive, as a shorter stay in hospital is better.

effect (ϕ_{Eij}) depends, through the parameter β_i , on how much the cost C_{ij} is above the mean cost μ_{Ci} . The subtraction of ϕ_{Cij} in the first line of (2) ensures that μ_{Ei} remains interpretable as the overall mean effect in arm i of the trial.

Any bivariate normal distributions for costs and effects can be represented by Equations (1) and (2), the correlations being provided for by the regression coefficients β_i which are allowed to be different in the two arms of the trial. The model, for a two-arm trial, has 10 parameters that have to be estimated from the data. These are the mean and standard deviation of effects and costs, and the regression coefficient, in each arm.

The advantage of writing the model in this general form is two-fold. The first is that it directly extends to distributions other than the normal. This is particularly important for cost data, because they are typically extremely skewed, and ignoring this can lead to misleading results [17]. In the example below, we used skewed gamma distributions to represent the cost data of Figure 2. In fact, in our example, we also use gamma distributions for the skewed effectiveness data in Figure 1. Other distributions, such as the log-normal, could also be used for positively skewed distributions, and we return to this point in the discussion. The model even allows more flexibility, since it does not even require that the same distributions are used in both arms of the trial. The second advantage is that it naturally extends to handling covariates, as explained later.

Deriving cost-effectiveness summaries

Fitting these models by conventional maximum likelihood is straightforward for the (bivariate) normal case [18], but is difficult when using gamma distributions. Hence, we apply Bayesian methods, details of which are given in the Appendix. Inferences are provided for the incremental cost ($\Delta_C = \mu_{C2} - \mu_{C1}$), and the incremental effect ($\Delta_E = \mu_{E2} - \mu_{E1}$), between the two arms of the trial. (For the effectiveness outcome in the case management trial, we reverse the sign of Δ_E as in Table 1, since shorter hospital stays are better.) The distribution of (Δ_C, Δ_E) on the CE-plane is represented by a contour plot, achieved by smoothing their joint distribution [19]. The incremental net benefit, $INB(K) = K \times \Delta_E - \Delta_C$, is a function of the willingness-to-pay value K . The CEAC is a plot of the probability, given the data,

that the intervention is cost effective against K , that is of the probability that $INB(K) > 0$. Model fit was assessed by calculating deviances (minus twice log-likelihood); lower values indicate a better fit to the data.

Example

Assessments of the overall cost-effectiveness in the case management trial are shown on the CE-plane in Figure 3(a). These depict the distributions of incremental costs against incremental effects, first when normal distributions are used to model the original costs and effects data (solid line), and secondly when gamma distributions are used (dashed line). The deviances (Figure 3, legend) indicate, not surprisingly given the distributions in Figures 1 and 2, that the model assuming gamma

distributions fits the data vastly better than the model assuming normal distributions. The inferences are not identical in the two cases. For the gamma distributions, the mean cost difference is estimated as about £1500, rather than near £2000 as in the normal case. This is because the correlations between costs and effects influence the results differently when assuming different underlying distributions. Also the mean cost difference is less precise (wider contours in the vertical direction) when using gamma distributions.

The consequent effect of the different distributional assumptions on the CEACs is shown in Figure 3(b). Whereas the CEAC assuming normal distributions starts low and rises to around 0.5 (solid line), the CEAC assuming gamma distributions starts somewhat higher (dashed line). Formally therefore, in this case, there would be no difference in the overall decision according to these different choices of distribution (since the probability that intensive case management is cost effective remains below 0.5 at any willingness-to-pay). However, one would clearly prefer to present the results based on assuming gamma distributions, since this model fits the data so much better.

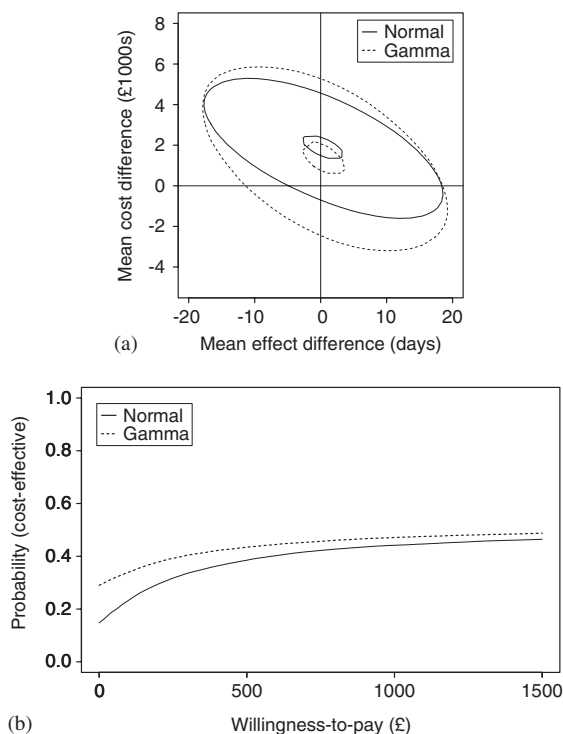


Figure 3. Overall cost-effectiveness: (a) Contour plots of cost-effectiveness distributions (5 and 80% contours shown): intensive versus standard case management. Costs and effects are either both fitted by normal distributions (solid line, deviance 22950), or both by gamma distributions (dashed line, deviance 20950). (b) Corresponding cost-effectiveness acceptability curves (CEAC)

Covariate adjustment

Important baseline variables in an economic evaluation might include age, sex, ethnic group, clinical severity and functional status. In cost-effectiveness analyses of randomised trials, adjusting for baseline variables predictive of either effects or costs (or both) should improve precision [20]. We first consider adjusting for a single continuous baseline variable; we take age as an example. We denote by X_{ij} the age of patient i in arm j of the trial, and centre the variable by subtracting the overall mean age ($X_{..}$) across both arms of the trial. We write the centred variable as $x_{ij} = X_{ij} - X_{..}$, and adjust both costs and effects for age by extending Equation (2):

$$\begin{aligned} \phi_{Eij} &= \mu_{Ei} + \beta_i(C_{ij} - \phi_{Cij}) + \gamma_E x_{ij} \\ \phi_{Cij} &= \mu_{Ci} + \gamma_C x_{ij} \end{aligned} \tag{3}$$

Here γ_E is the regression coefficient of effects on age. By assuming that this is the same in both the intervention and control arms, an adjustment for age is obtained. Similarly, γ_C is the regression

coefficient of costs on age. Precision is gained because of the reduction in the standard deviations σ_{Ei} and σ_{Ci} in Equation (1) consequent upon including age as a predictor. x_{ij} is centred in order that the interpretation of μ_{Ei} and μ_{Ci} (and hence of Δ_C and Δ_E) remains the same as before, and inferences for cost-effectiveness follow as previously described.

Adjustment for a set of continuous variables can be similarly carried out. If the S variables are denoted $X_{1ij} \dots X_{Sij}$, with centred versions $x_{1ij} \dots x_{Sij}$, then we replace $\gamma_E x_{ij}$ in (3) by $\gamma_{E1} x_{1ij} + \dots + \gamma_{ES} x_{Sij}$ and similarly for $\gamma_C x_{ij}$. For a dichotomous baseline variable such as sex, we define two dummy variables $X_{1ij} = 1$ if male and $X_{2ij} = 1$ if female (and zero otherwise). Then model (3) is adapted as

$$\begin{aligned} \phi_{Eij} &= \mu_{Ei} + \beta_i(C_{ij} - \phi_{Cij}) + \gamma_{E1} X_{1ij} + \gamma_{E2} X_{2ij} \\ \phi_{Cij} &= \mu_{Ci} + \gamma_{C1} X_{1ij} + \gamma_{C2} X_{2ij} \end{aligned} \quad (4)$$

Constraints are needed on the 2 parameters γ_{E1} and γ_{E2} (and γ_{C1} and γ_{C2}) to ensure that the interpretation of μ_{Ei} and μ_{Ci} (and hence of Δ_C and Δ_E) remains the same as before, as described in the Appendix. For a multi-category variable, for example ethnic group with five categories, we define five dummy variables $X_{1ij} \dots X_{5ij} = 0$ or 1, where $X_{sij} = 1$ only if the i th patient in arm j is in the s th category. An arbitrary mix of continuous, binary and categorical variables can be similarly handled. It is not necessary to adjust costs and effects for the same set of variables in (3) and (4).

Example

We consider adjusting the cost-effectiveness analysis in the case management trial for a single baseline variable, the number of days each patient had been in hospital in the 24 months prior to randomisation. The crude correlations between this variable with costs were 0.40 and 0.42, and with effects were 0.31 and 0.30, in the control and intervention arms of the trial, respectively. The contour plots for cost-effectiveness are shown on the CE-plane in Figure 4(a), unadjusted from model (2) and adjusted from model (3), based on assuming gamma distributions for both costs and effects. The adjusted incremental effect is now around 5 days, rather than 0 days, due to a modest imbalance between randomised groups in the baseline hospitalisation variable. There is a small

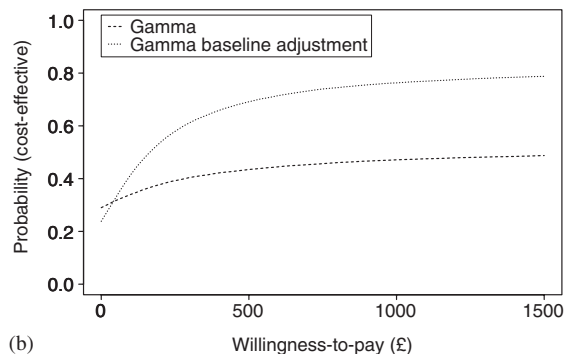
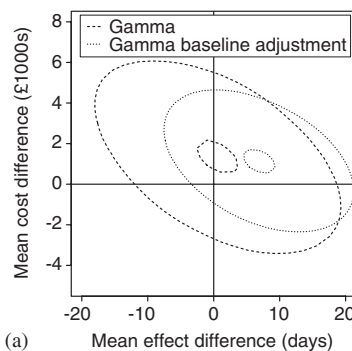


Figure 4. Covariate adjustment: (a) Contour plots of cost-effectiveness distributions (5 and 80% contours shown): intensive versus standard case management. Costs and effects are both fitted by gamma distributions, without adjustment (dashed line, deviance 20 950) or with adjustment for prior days in hospital at baseline (dotted line, deviance 20 830). (b) Corresponding cost-effectiveness acceptability curves (CEAC)

reduction in the spread of these contour plots, in both the vertical and horizontal directions, indicating some increase in precision for both costs and effects. The improvement in fit is also evidenced by the decrease in deviance (Figure 4, legend). The consequent change in the CEAC in Figure 4(b), comparing the unadjusted (dashed) and adjusted (dotted) lines, comes about both through adjustment for baseline imbalance and increased precision. The increase in precision corresponds to a steeper CEAC, giving more definitive inferences. Moreover, at a willingness-to-pay of greater than £200, a decision maker would favour intensive case management since the adjusted analysis indicates a probability greater than 0.5 that it is cost effective.

Finding good predictors of subsequent costs is notoriously difficult, in both health economic evaluation [21] and in insurance risk assessment

[22]. Prior resource use is sometimes the only really important predictor of subsequent resource use. This was the situation in the case management trial: an extensive analysis of the baseline predictors of costs revealed that the previous number of days in hospital was the only predictor strong enough to materially reduce the residual variance of the cost data [23]. In many economic evaluations, it may prove difficult in practice to find variables sufficiently highly correlated with either costs or effects to substantially improve the precision of overall estimates of cost-effectiveness.

Differences between subgroups

Cost-effectiveness may well differ between groups of patients, for example between men and women, or between patients of different ages [24]. Given limited healthcare budgets, it may often be crucial to identify those patient groups with the greatest cost-effectiveness so that resources can be targeted appropriately [4,25]. We first consider whether cost-effectiveness differs according to a continuous baseline variable, such as age. We use the centred continuous variable x_{ij} as above, together with a new variable I_i for group i : $I_1 = 0$ for the control group and $I_2 = 1$ for the intervention group. We modify model (3) as

$$\begin{aligned} \phi_{Eij} &= \mu_{Ei} + \beta_i(C_{ij} - \phi_{Cij}) + \gamma_E x_{ij} + I_i \delta_E x_{ij} \\ \phi_{Cij} &= \mu_{Ci} + \gamma_C x_{ij} + I_i \delta_C x_{ij} \end{aligned} \tag{5}$$

We have replaced the regression coefficients γ_E by two regression coefficients, γ_E in the control group and $\gamma_E + \delta_E$ in the intervention group, and similarly for γ_C . If the regression coefficients are the same in the two arms (corresponding to $\delta_E = 0$ and $\delta_C = 0$), then we have reproduced the previous model (3). In model (5) however, the mean effect difference Δ_E is greater by an amount δ_E per year of age. Similarly, the mean cost difference Δ_C is greater by δ_C per year of age. So model (5) provides different estimates of cost-effectiveness according to age, rather than an overall estimate for all ages combined as in model (3). The improvement in fit, comparing model (5) to model (3) is a test of ‘interaction’, which summarises the evidence from the data that the cost-effectiveness is really different according to age.

For a categorical variable, such as sex, we define X_{1ij} and X_{2ij} as in (4) above. We modify model (4) as

$$\begin{aligned} \phi_{Eij} &= \mu_{Ei} + \beta_i(C_{ij} - \phi_{Cij}) + \gamma_{E1} X_{1ij} + \gamma_{E2} X_{2ij} \\ &\quad + I_i(\delta_{E1} X_{1ij} + \delta_{E2} X_{2ij}) \\ \phi_{Cij} &= \mu_{Ci} + \gamma_{C1} X_{1ij} + \gamma_{C2} X_{2ij} \\ &\quad + I_i(\delta_{C1} X_{1ij} + \delta_{C2} X_{2ij}) \end{aligned} \tag{6}$$

The same constraints are used as before (see Appendix). Model (6) provides different estimates of cost-effectiveness in men and women, rather than an overall estimate for men and women combined as in model (4). Interactions according to multi-category variables can be handled in a similar way as before.

Models (5) and (6) are not the same as doing separate analyses in different subgroups, for a number of reasons. First, common standard deviations σ_{Ei} and σ_{Ci} in Equation (1) are assumed across subgroups. Similarly, common interrelationships between costs and effects (indicated by the parameters β_i) are assumed. The model allows for the direct inclusion of continuous variables, which cannot be addressed by subgrouping. Interactions for costs and effects do not have to be considered for the same set of variables. Lastly, interactions with multiple variables can be simultaneously investigated.

Example

It was observed in the case management trial that intensive treatment might be more effective in those psychotic patients with borderline intelligence [26], and less costly [27], as compared to those with normal intelligence. We here investigate this subgroup finding with respect to cost-effectiveness, our analysis being restricted to the 553 patients who undertook the baseline assessment of intelligence used in the trial. In the smaller group of patients with borderline intelligence, intensive case management improved both costs and effects markedly, while if anything the reverse was true for the normal intelligence group (Table 1).

Separate CE-plane contours are shown for the two subgroups in Figure 5(a), derived from model (6), again using underlying gamma distributions for both costs and effects. These are in line with the

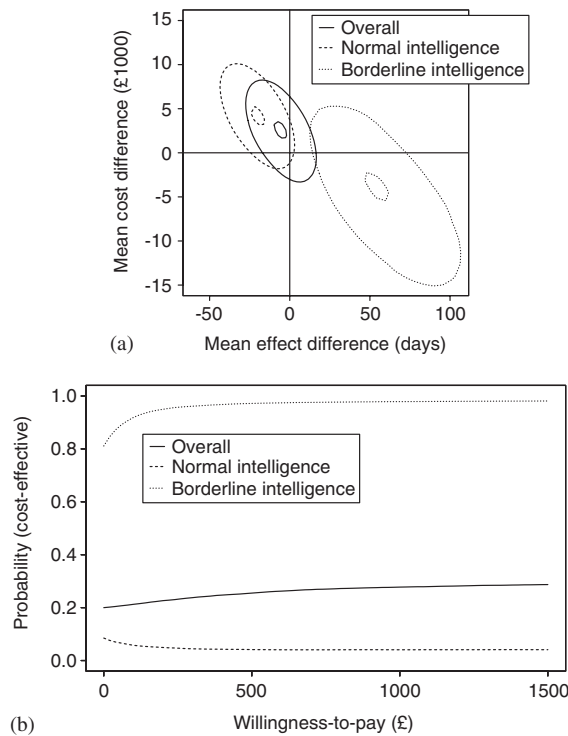


Figure 5. Subgroup analysis: (a) Contour plots of cost-effectiveness distributions, overall and by intelligence subgroup (5 and 80% contours shown): intensive versus standard case management. Costs and effects are both fitted by gamma distributions. (b) Corresponding cost-effectiveness acceptability curves (CEAC)

crude results in Table 1. The wide spread of contours in the borderline intelligence subgroup is caused by the small number of patients, but despite this the results suggest that intensive case management may be differentially cost effective in the two subgroups. The difference between the deviance of model (6) and that of model (4), based on the restricted set of patients for whom the intelligence assessment was available, is 7.8. Comparing this to a chisquared statistic on 2 degrees of freedom (corresponding to the 2 extra parameters estimated) yields $P = 0.02$. This provides moderately strong evidence that the cost-effectiveness of intensive case management differs between those with borderline and normal intelligence. The corresponding CEACs are shown in Figure 5(b): for all willingness-to-pay values, a decision maker would favour intensive case management for the borderline intelligence sub-

group (CEAC above 0.5), but not for the normal intelligence subgroup (CEAC below 0.5).

Differences between centres

It is likely that cost-effectiveness, while showing some similarity, may not be totally homogeneous across centres in a multicentre study [28]. The reasons may include the different levels of service, clinical care and infrastructure available. Lack of homogeneity may be particularly apparent when centres come from different countries [29]. While centre differences could simply be considered as an example of interaction with a categorical covariate, it is in general more reasonable to suppose at least a degree of similarity in cost-effectiveness between centres [30]. For example, the most extreme centre-specific cost-effectiveness estimates, amongst many centres, are likely to be more extreme than is truly the case, by chance alone. This provides a rationale for the use of random centre effects [30], whereby the cost-effectiveness estimates for each centre are shrunk towards the overall average for the whole study; their precision is also improved since information is 'borrowed' from other centres [31]. The amount of shrinkage depends on the estimated precision within each centre and true heterogeneity between centres. Such 'hierarchical model' analyses are widely advocated in multicentre studies of effectiveness outcomes [32], and are starting to appear in analyses of multicentre cost data [33] and INB [14]. However, they have not yet been directly applied to costs and effects considered as a joint bivariate outcome.

We implement such an analysis as an extension to model (6). As before, for a set of S centres, we define S dummy variables, with $X_{sij} = 1$ only if the i th patient in arm j is in the s th centre. For each centre, $s = 1 \dots S$, there are corresponding regression coefficients γ_{Es} , γ_{Cs} , δ_{Es} and δ_{Cs} in model (6); δ_{Es} and δ_{Cs} represent the set of differences between centres in incremental effects, and in incremental costs, respectively. The final stage is to model a degree of similarity between centres by assuming that $(\delta_{Es}, \delta_{Cs})$ are 'random effects' drawn from a bivariate normal distribution. This is the same idea as in 'random effects' meta-analysis, where effects in different studies are assumed to be similar but not identical, and are modelled as being drawn from some distribution [31]. Further details of the

model are given in the Appendix, together with information on the priors used in the Bayesian analysis.

Example

Patients in the case management trial came from four centres in the UK [15]. Different centres showed apparently different incremental effects and (to a lesser extent) costs (Table 1), although tests of interaction for effects and costs were both not significant [15, 16]. The point estimates of the ICERs varied considerably. Using model (6) in which differences between centres are simply represented by a variable with four categories gives the contours on the CE-plane shown in Figure 6(a); the overall results across all centres is also shown. The contours for each separate centre are more widely spread out than for the overall

results because of the smaller numbers of patients. The middles of the contours are in line with the centre-specific estimates in Table 1. One centre (centre 3) shows a high probability of cost-effectiveness, while the other centres do not. This is also clear from the corresponding CEACs in Figure 6(b).

The analogous results using a hierarchical model for centre differences are shown in Figure 7. Comparing Figure 7(a) with Figure 6(a), which are drawn on the same scale, reveals that the centre-specific contours are now estimated to be more similar, and that they are more precise. This results from the ‘borrowing of information’ about cost-effectiveness across centres [31]. Similarly, in Figure 7(b), the centre-specific CEACs are closer to the overall CEAC. Even so, in centre 3, intensive case management would be judged cost effective for a willingness-to-pay above about £200

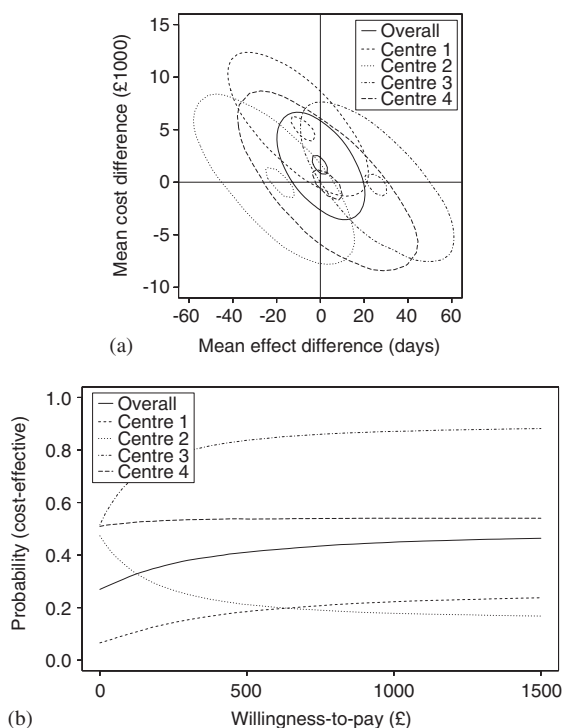


Figure 6. Crude centre differences: (a) Contour plots of cost-effectiveness distributions, overall and by centre (5 and 80% contours shown): intensive versus standard case management. Costs and effects are both fitted by gamma distributions. (b) Corresponding cost-effectiveness acceptability curves (CEAC)

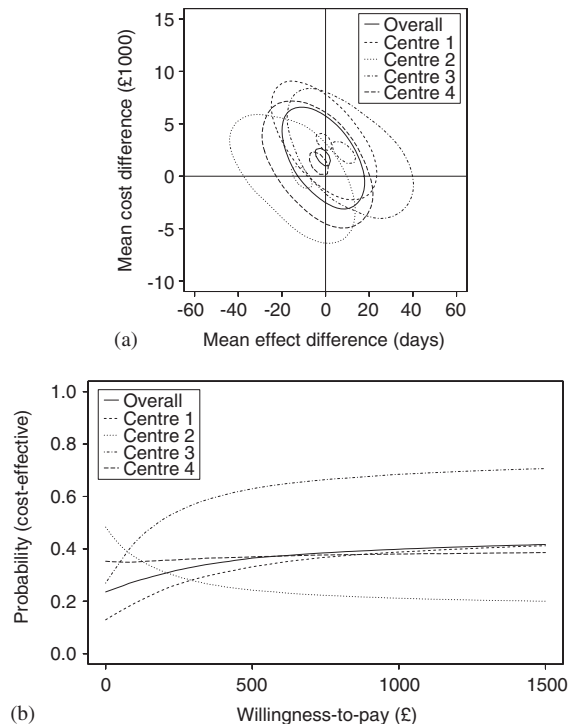


Figure 7. Estimated centre differences using a hierarchical model: (a) Contour plots of cost-effectiveness distributions, overall and by centre (5 and 80% contours shown): intensive versus standard case management. Costs and effects are both fitted by gamma distributions. (b) Corresponding cost-effectiveness acceptability curves (CEAC)

(CEAC above 0.5), while for the other centres this is not the case.

Discussion

This paper has proposed methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. These have rarely been undertaken in the past, no doubt due to the technical difficulty of addressing both effects and costs simultaneously and of allowing for the typically skewed distribution of cost data. Similar analyses for effectiveness data alone, such as survival or other clinical outcomes, are well established [34]. There is a limited amount of work undertaking such analyses for cost data alone, for example based on the use of generalised linear models [35,36]. In our example, we used a gamma distribution for costs and an identity link function, whereby the covariates acted additively on the predicted mean, in line with recent analyses of the same data [37]. The novel aspect of our paper lies in the simultaneous handling of costs and effects, to provide cost-effectiveness inferences in the form of distributions on the cost-effectiveness plane or as CEACs. This is done within a flexible parametric Bayesian framework, in which a variety of distributions can be used for the underlying skewed cost data [5,17]. The adoption of a Bayesian analysis has the additional advantage that informative priors for parameters, if that were considered appropriate in a particular application, could be easily introduced.

Development of methods for handling covariates in cost-effectiveness evaluations has been considered in two previous publications. Work by Hoch and colleagues [12] was based on deriving a one-dimensional summary using the net benefit statistic. For a specific values of K , the willingness-to-pay, the net benefit was calculated for each individual in a study. Then intervention effects, covariate adjustment, and interaction effects were estimated by ordinary least squares regression. This method suffers some limitations. First separate analyses are needed across a range of values of K , so that no overall assessment of the importance of covariates or subgroups is provided. Secondly ordinary least squares regression assumes that net benefits are normally distributed. This is unreasonable when net benefits are highly skewed (because of either skewed costs or effects).

Willan and colleagues [13] directly considered the two-dimensional aspects of costs and effects. However, their analyses were limited to normal distributions, and assumed constant variances between intervention and control arms, for both costs and effects. They also assumed that the correlation of costs and effects were the same in the intervention and control arms. Although they investigated the potential impact of skewness by undertaking simulations, it is known that covariate effects can be sensitive to skewness [37]. These limitations make such models unlikely to be fully appropriate for real data. One important aspect of our models is that the distribution used for the cost data (and the effects data) can be chosen carefully for the data at hand. We have shown elsewhere that such choice can materially affect the conclusions about incremental costs [17]. The sensitivity of results to choice of distribution will carry through to the more complex cost-effectiveness analyses incorporating covariates that we consider here. While others have considered log-normal distributions for cost data [38,39], we did not do so here in this paper, partly in the interests of brevity, partly because the gamma distribution fitted our cost data better [17], and partly because, in common with others [36], we have found that the log-normal distribution often has too heavy a right-hand tail for reliable estimation of population mean costs.

In undertaking subgroup analyses in practice, two important issues must be acknowledged. The first is that subgroups should be directly compared. This is the virtue of a test of interaction such as we have employed, rather than simply testing the results in separate subgroups [40]. The second is that, if many subgroups are investigated or if they are not specified in advance of looking at the data, there is a danger of at least some false-positive results [41]. Conventionally, such 'data-dredging' can be controlled by pre-specifying subgroups in which different health policy decisions could reasonably be made, by limiting the number of subgroups investigated, and by using a more stringent level of statistical significance before subgroup differences are accepted. Although statistical significance has been argued to be somewhat irrelevant to economic decision making [42], it is unlikely that subgroup differences in cost-effectiveness would be accepted without strong direct evidence. Also, the presence (or absence) of interactions for either costs or effects by themselves does not imply the presence

(or absence) of an interaction for cost–effectiveness [43].

The use of hierarchical models for estimating the differences between centres in cost–effectiveness have been considered previously only by Manca and colleagues [14]. They, however, worked on the one-dimensional net benefit scale, performing analyses for a range of specific values of K , the willingness-to-pay. This suffers similar limitations as discussed above regarding the work of Hoch and colleagues [12], namely separate analyses across a range of K are needed, no overall assessment of between-centre heterogeneity is made, and normality of net benefits is assumed. Our consideration of how to model between-centre differences in cost–effectiveness analyses is apparently novel. These extend similar analyses, advocated for effectiveness outcomes alone, to cost–effectiveness. We should expect that centres in a multicentre study show results that are to a degree similar, and therefore the idea of shrinkage of observed centre-specific results towards the overall mean for all centres is intuitively reasonable. It also avoids the problems of multiple comparisons mentioned above when there are many centres. We have provided one way of undertaking this analysis, but should note a caveat in our example. There were only four centres, and it was difficult to estimate the parameters of the bivariate normal distribution assumed for the centre effects (see Appendix). The method would work more reliably if there were more centres.

The methods we have proposed still have some limitations. We have not addressed the handling of missing data, nor their application to censored cost and survival data [44]; these are important extensions that need to be developed. While our methods are technically more difficult to implement than those recently proposed by Willan and colleagues [13], they are more flexible and so have the capacity to be more appropriate for real data. Experience with using these models in practice is now needed, partly to investigate whether their additional flexibility leads to greater reliability in the cost–effectiveness conclusions derived.

Acknowledgements

We thank Sarah Byford and the UK700 Group for the use of the case management trial data in this paper, and

Andrew Willan, Eleanor Pinto and two referees for their helpful comments.

Appendix

(a) Implementation of Bayesian methods

The methods were implemented using Markov chain Monte Carlo (MCMC) methods in the software BUGS [45]. Bayesian methods require prior distributions for the parameters of the distributions. Here we use priors intended to be non-informative, so that the resulting inferences only depend on the data. Specifically, we use uniform priors for all the μ 's, β 's, γ 's and δ 's, as well as for the $\log(\sigma)$'s in the normal distributions. We parameterised the gamma distribution in terms of its mean μ and shape ρ , and put a uniform prior on ρ . Posterior distributions of quantities of interest for inferences about cost–effectiveness were derived from 20 000 iterations of the Markov chain, after an initial 5000 iterations were discarded to ensure convergence [45]. Example BUGS code is available from www.mrc-bsu.cam.ac.uk/personal/richard/downloads.html.

(b) Models for categorical variables and differences between centres

The constraints needed in model (4) to ensure the correct interpretation of μ_{Ei} and μ_{Ci} (and hence of Δ_C and Δ_E) are $\gamma_{E1}X_{1..} + \gamma_{E2}X_{2..} = 0$ and $\gamma_{C1}X_{1..} + \gamma_{C2}X_{2..} = 0$, where $X_{s..}$ is the mean of X_{sij} over all the study participants. Similarly, for a categorical variable with S levels represented by dummy variables X_{sij} ($s = 1 \dots S$), the necessary constraints are $\sum_s \gamma_{Es} X_{s..} = 0$ and $\sum_s \gamma_{Cs} X_{s..} = 0$. For model (6), we also require $\delta_{E1}X_{1..} + \delta_{E2}X_{2..} = 0$ and $\delta_{C1}X_{1..} + \delta_{C2}X_{2..} = 0$. Similarly, for a categorical variable we require $\sum_s \delta_{Es} X_{s..} = 0$ and $\sum_s \delta_{Cs} X_{s..} = 0$. These constraints, that the coefficients sum to zero over all the study participants, are a commonly available option in statistical software.

For S centres, the differences in incremental effects (δ_{Es}) and in incremental costs (δ_{Cs}) between centres are the 'centre effects'. So, in modelling them as random effects, we assume the set of (δ_{Es} , δ_{Cs}) are drawn from a bivariate normal

distribution with means zero, variances τ_E^2 and τ_C^2 with correlation ρ_{EC} . To maintain the correct interpretation of the δ_E 's, they are finally rescaled to $\delta_{Et} = \delta_{Et} - \sum_s \delta_{Es} X_{s..}$ for each $t = 1 \dots S$, and similarly for the δ_C 's. This ensures that $\sum_s \delta_{Es}^* X_{s..} = 0$ and $\sum_s \delta_{Cs}^* X_{s..} = 0$ as required. μ_{Ei} is also rescaled to $\mu_{Ei}^* = \mu_{Ei} + I_i \sum_s \delta_{Es} X_{s..}$, and similarly for μ_{Ci} . In our example there were only four centres with which to estimate the 3 parameters of the bivariate normal distribution. So to implement the analysis, we had to make the assumption that ρ_{EC} was zero, while using uniform priors for τ_E and τ_C .

References

1. National Institute for Clinical Excellence. *Guide to the Methods of Technology Appraisal*. National Institute for Clinical Excellence: London, 2004.
2. Black N. Why we need observational studies to evaluate the effectiveness of health care. *Br Med J* 1996; **312**: 1215–1218.
3. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med* 1989; **8**: 467–475.
4. Rothwell PM, Warlow CP. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. *Lancet* 1999; **353**: 2105–2110.
5. O'Hagan A, Stevens JW. A framework for cost-effectiveness analysis from clinical trial data. *Health Econ* 2001; **10**: 303–315.
6. Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. *J Health Serv Res Policy* 1998; **3**: 233–245.
7. Black WC. The CE plane: a graphic representation of cost-effectiveness. *Med Decis Making* 1990; **10**: 212–214.
8. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998; **18**(suppl.): S68–S80.
9. Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *Int J Technol Assess Health Care* 1998; **14**: 467–471.
10. van Hout BA, Al MJ, Gordon GS, Rutten F. Costs, effects and C/E ratios alongside a clinical trial. *Health Econ* 1994; **3**: 309–319.
11. Lothgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Econ* 2000; **9**: 623–630.
12. 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**: 415–430.
13. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004; **13**: 461–475.
14. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Econ* 2005; **14**: 471–485.
15. Burns T, Creed F, Fahy T, Thompson S, Tyrer P, White I. Intensive versus standard case management for severe psychotic illness: a randomised trial. UK 700 Group. *Lancet* 1999; **353**: 2185–2189.
16. UK700 Group. cost-effectiveness of intensive v. standard case management for severe psychotic illness: UK700 case management trial. *Br J Psychiatry* 2000; **176**: 537–543.
17. Nixon R, Thompson SG. Parametric modelling of cost data in medical studies. *Stat Med* 2004; **23**: 1311–1331.
18. Willan AR, Lin DY. Incremental net benefit in randomised clinical trials. *Stat Med* 2001; **20**: 1563–1574.
19. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *J R Stat Soc Ser B* 1991; **53**: 683–690.
20. Altman DG. *Practical Statistics for Medical Research*. Chapman and Hall: London, 1991.
21. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of health care interventions. *Health Technol Assess* 1999; **3**: 1–134.
22. Duan N. Smearing estimate: a non-parametric retransformation method. *J Am Stat Assoc* 1983; **78**: 605–610.
23. Byford S, Barber JA, Fiander M, Marshall S, Green J. Factors that influence the cost of caring for patients with severe psychotic illness: report from the UK700 trial. *Br J Psychiatry* 2001; **178**: 441–447.
24. Russell LB, Sisk JE. Modeling age differences in cost-effectiveness analysis: a review of the literature. *Int J Technol Assess Health Care* 2000; **16**: 1158–1167.
25. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Econ* 2003; **12**: 421–427.
26. Tyrer P, Hassiotis A, Ukoumunne O, Piachaud J, Harvey K. Intensive case management for psychotic patients with borderline intelligence. *Lancet* 1999; **354**: 999–1000.
27. Hassiotis A, Ukoumunne OC, Byford S, Tyrer P, Harvey K, Piachaud J, Gilvarry K, Fraser J. Intellectual functioning and outcome of patients with severe psychotic illness randomised to intensive case management. Report from the UK700 trial. *Br J Psychiatry* 2001; **178**: 166–171.

28. Willke RJ, Glick HA, Polsky D, Schulman K. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Econ* 1998; **7**: 481–493.
29. Cook JR, Drummond M, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. *Stat Med* 2003; **22**: 1955–1976.
30. Skene AM, Wakefield JC. Hierarchical models for multicentre binary response studies. *Stat Med* 1990; **9**: 919–929.
31. Carlin JB. Meta-analysis for 2×2 tables: a Bayesian approach. *Stat Med* 1992; **11**: 141–158.
32. Brown HK, Kempton RA. The application of REML in clinical trials. *Stat Med* 1994; **13**: 1601–1617.
33. Grieve R, Nixon RM, Thompson SG, Normand C. Using multilevel models for assessing the variability of multinational resource use and cost data. *Health Econ* 2005; **14**: 185–196.
34. Pocock SJ. *Clinical Trials: a Practical Approach*. Wiley: Chichester, 1983.
35. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ* 2001; **20**: 461–494.
36. Deb P, Burgess JF. *A quasi-experimental comparison of econometric models for health care expenditures*. Hunter College Department of Economics Working Papers 2003; 212.
37. Barber JA, Thompson SG. Multiple regression of cost data: use of generalised linear models. *J Health Serv Res Policy* 2004; **9**: 197–204.
38. O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Econ* 2003; **12**: 33–49.
39. Zhou XH. Inferences about population means of health care costs. *Stat Methods Med Res* 2002; **11**: 327–339.
40. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; **355**: 1064–1069.
41. Oxman AD, Guyatt GH. A consumers guide to subgroup analyses. *Ann Intern Med* 1992; **116**: 78–84.
42. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999; **18**: 341–364.
43. Johnston K, Gray A, Moher M, Yudkin P, Wright L, Mant D. Reporting the cost-effectiveness of interventions with nonsignificant effect differences: example from a trial of secondary prevention of coronary heart disease. *Int J Technol Assess Health Care* 2003; **19**: 476–489.
44. Willan AR, Chen EB, Cook RJ, Lin DY. Incremental net benefit in randomized clinical trials with quality-adjusted survival. *Stat Med* 2003; **22**: 353–362.
45. Spiegelhalter DJ, Thomas A, Best N. *WinBUGS version 1.2*. MRC Biostatistics Unit: Cambridge, UK, 1999.