Dancing with the devil:

the importance of sub-group analysis in economic evaluation

Andrew H. Briggs
Health Economics Research Centre
Department of Public Health
University of Oxford

&

Andrew R. Willan, *McMaster University*
Jeffrey S. Hoch, *University of Western Ontario*
Henry A. Glick, *University of Pennsylvania*

Abstract:

The standard pitfalls associated with *ad hoc* sub group analyses in clinical evaluation are well known. Since variation in clinical effects is argued to be largely 'quantitative' rather than 'qualitative' (i.e. associated with a change in the size of an effect not the direction of that effect), the suggestion is that sub-group effects are not important for clinical decision-making. By contrast, in economics we are concerned with diminishing marginal returns and therefore 'quantitative' differences between patient groups in clinical effect can lead to 'qualitative' differences in cost-effectiveness for those groups. This suggests that sub-group analysis is of critical importance for economic evaluation and the policy decision that such evaluations seek to inform. However, we must remain mindful of the potential pitfalls in terms of inappropriate inference, especially when studies were not powered to detect sub-group differences. Building on previous work and examples, we explore the use of empirical Bayes methods for estimating sub-group effects while 'shrinking' estimates to account for the greater expected variance associated with dividing the data.

Paper presented to the Health Economists' Study Group Conference, University of East Anglia, January 2002.

1. Introduction

Informal discussions of the value for money of health care interventions often proceed without explicit statements relating to the eligible patient population to receive the treatment in question. However, the characteristics of the patient population for whom the intervention is advocated will often be an important determinate of the overall cost-effectiveness of the intervention. This has long been acknowledged by health economists undertaking secondary evaluations of health care interventions. For example, Williams' early illustrative analysis emphasised the difference in cost-effectiveness of coronary artery bypass grafting depending on the underlying severity of patients' heart disease (Williams 1985). Yet in primary analyses, typically characterised by an economic analysis performed alongside a clinical trial, the sub-group analyses that are crucial to identifying how incremental cost-effectiveness varies at the margin are often discouraged.

The aim of this paper is to argue a role for sub-group analysis in economic evaluation while remaining mindful of the reasons that sub-group analyses are generally discouraged in clinical evaluation – in particular, the dangers of inappropriate inference. We begin by reviewing the traditional orthodoxy in clinical evaluation and argue that sub-group analysis is more important in economic evaluation. We go on to review a recently advocated approach to sub-group analysis in cost-effectiveness studies based on net-benefit regression and show how the approach can be adapted further to estimate cost-effectiveness in sub-groups while 'shrinking' to adjust for the greater variance associated with dividing the data. Two examples are introduced: the first demonstrates the approach using sub-groups defined by patient characteristics; the second is a multi-national study showing that the approach can be applied to sub-groups defined by country.

2. The traditional orthodoxy in clinical trials

It is probably fair to say that the general recommendation relating to sub-group analysis in clinical evaluation is *don't do it*! This is never quite stated explicitly: instead we find that guidance on the subject suggests that sub-groups are pre-specified at the design stage, that the choices of sub-groups to examine are justified (i.e. that

there is some a priori reason to suppose that there may be differential treatment effects) and that sufficient patients are recruited to give power to detect differences at conventional levels of power and significance (for a useful overview of interpreting sub-group analyses in clinical evaluation see Oxman & Guyatt 1992). Of course, in practice, these criteria turn out to be quite stringent with the practical result that few studies can meet these criteria.

When considering potential differences between sub-groups of patients it is common to distinguish between *qualitative* differences between groups where the treatment effect acts in difference directions (i.e. treatment is beneficial for some groups, but harmful for others) from *quantitative* differences where treatment effects act in the same direction (i.e. the effect only differs in magnitude not in sign). One argument for not conducting sub-group analysis routinely in clinical evaluation is that true qualitative differences are rare – and examples are often given of apparent qualitative differences being shown to be spurious when additional data are collected. The implication being that if treatment differences are only quantitative, then conclusions or policy recommendations would not vary between the sub-groups under analysis.

While this may be true in a clinical evaluation, the quantitative differences (from the clinical evaluation perspective) may induce qualitative differences in an economic evaluation. Consider Figure 1 that shows the results of three subgroup analyses on the cost-effectiveness plane. For the sake of simplicity, we suppose that cost is estimated without error and is the same for each sub-group. The sub-groups differ only in terms of the effect difference and the horizontal 'I' bars for each group represent the uncertainty in the estimation of the effect differences. In any clinical evaluation, sub-group A would be considered to have a treatment effect that represented a qualitative difference from sub-groups B and C, which have only quantitative differences from each other. However, in an economic evaluation, it is sub-groups A & B that are considered quantitatively different (neither are cost effective, although treatment is less cost-effective in sub-group A than in sub-group B), which are qualitatively different from sub-group C (where treatment is cost-effective).

Since in economic evaluation the size of the treatment effect is crucial in estimating the cost-effectiveness of the intervention, it is important that the absence of qualitative treatment effects does not become a reason for not examining the potential for treatment effects to differ at the clinical margin. Note, however, that it is not sufficient to simply look at whether there are important differences in costs and effects independently. Consider Figure 2, which shows the results of two sub-group analyses on the cost-effectiveness plane. The confidence intervals for the two groups on either cost or effect difference overlap suggesting that there is no significant difference in the magnitude of either costs or effects independently. However, the joint distributions are clearly separated and fall either side of the £30,000/LYG line (our assumed threshold or 'ceiling ratio' for decision-making), indicating that treatment might be judged cost-ineffective for sub-group A, but cost-effective for sub-group B.

3. Net-benefit regression for sub-group estimation

In a previous paper we have described a regression approach to cost-effectiveness analysis (Hoch, Briggs, & Willan 2002), in which we argued that a direct regression method offered a robust approach to analysing sub-groups in cost-effectiveness studies. In this section we briefly introduce the methods and main results from that previous paper before building on the approach and suggesting an alternative method of sub-group analysis in the next section.

We are cautioned against formulating patient-level cost-effectiveness ratios since the average of individual cost-effectiveness ratios does not equal the ratio of the averages (Gold et al. 1996; Stinnett & Paltiel 1997). Furthermore, the difference in average ratios does not equal the incremental cost-effectiveness ratio (Stinnett & Paltiel 1997). By contrast, the net-benefit statistic has the desirable property that averaging over individual net-benefits is exactly equivalent to forming a net-benefit from averaged costs and effects (Stinnett & Mullahy 1998). Furthermore the difference of two average net-benefits is the incremental net-benefit between them. This is shown algebraically below.

$$\frac{1}{N_1}\sum_{i=1}^{N_1} NMB_i - \frac{1}{N_0}\sum_{j=1}^{N_0} NMB_j = \frac{1}{N_1}\sum_i \left(l \cdot E_i - C_i\right) - \frac{1}{N_0}\sum_j \left(l \cdot E_j - C_j\right)$$

$$= \left(l \cdot \frac{1}{N_1}\sum_i E_i - \frac{1}{N_1}\sum_i C_i\right) - \left(l \cdot \frac{1}{N_0}\sum_j E_j - \frac{1}{N_0}\sum_j C_j\right)$$

$$= \left(l \cdot \bar{E}_1 - \bar{C}_1\right) - \left(l \cdot \bar{E}_0 - \bar{C}_0\right)$$

$$= l\left(\bar{E}_1 - \bar{E}_0\right) - \left(\bar{C}_1 - \bar{C}_0\right)$$

$$= l \cdot \Delta\bar{E} - \Delta\bar{C}$$

$$= \Delta N\bar{M}B$$

This straightforward result allows the net-benefit framework to be used to directly estimate cost-effectiveness within a regression framework. Net-benefit regression involves formulating, for each patient in the data set, that individual patient's net-benefit. On the monetary scale this is achieved by multiplying the patient's effect measure by the decision-maker's willingness to pay per unit of effect (or ceiling ratio), l, and subtracting the observed cost for that patient. Consider the following net-benefit regression model:

$$NMB_i = a + \sum_{j=1}^{p} b_j x_{ij} + d\,t_i + t_i \sum_{j=1}^{p} g_j x_{ij} + e_i$$

where there are p covariates x and t represents a treatment dummy. The coefficient d represents the treatment effect in terms of incremental net-benefit, which is modified by the g coefficients that represent the treatment by covariate interactions. It was argued in our previous paper that testing these treatment interactions allowed a robust approach to sub-group analysis in economic evaluation (Hoch, Briggs, & Willan2002), in the sense of appropriately testing the difference between sub-groups rather than the inappropriate approach of testing the significance of individual sub-group effects which can lead to invalid conclusions (Altman 1991).

The net-benefit regression approach was illustrated using an example of an Assertive Community Treatment (ACT) programme for homeless persons with serious mental illness in Baltimore (Lehman et al. 1997). The rationale for the study was that by offering the ACT programme to homeless persons with serious mental illness would

spend more time in 'stable' housing and less time in emergency care and inpatient crisis care departments. Lehman and colleagues found ACT, relative to usual community services, reduced psychiatric inpatient days, emergency room visits, days homeless, and days in jail. The authors went on to present a standard cost-effectiveness analysis in a subsequent article (Lehman et al. 1999). Taking a net-benefit regression approach to the problem had the advantage of adjusting for some imbalance in the randomisation procedure and allowing for a comprehensive sub-group analysis (Hoch, Briggs, & Willan 2002). The net-benefit regression model outlined above was applied to the ACT data using age and GAF score (an assessment of overall mental health functioning using the Global Assessment of Functioning scale (American Psychiatric Association 1987)) as continuous covariates together with a dummy variable to indicate persons of African American descent. The results of this analysis are presented in Table 1 and show five regressions results for lambda values of $0, $100, $500, $1000 and infinity (the final regression reports net health benefit (NHB) rather than net monetary benefit (NMB), which is equivalent to running the regression on effect only). The results show that the overall treatment effect is significant across all five regressions and that there are marginally significant interaction terms for the NMBs with lower lambda values.

On the basis of these results, we might consider that there are important sub-group analyses to be considered. Indeed, in the previous paper we presented a graphical sub-group analysis on race (Hoch, Briggs, & Willan2002). While this was the least significant term of the treatment interactions, the original cost-effectiveness study had presented an analysis stratified on race (Lehman et al. 1999) and being dichotomous the race variable was a straightforward covariate for which to present a sub-group analysis. However, while this approach avoids the naïve mistakes than can occur if sub-groups are not compared directly, it is still true that the interaction terms could be achieving significance by chance and it is important to acknowledge this problem if spurious sub-group analyses are to be avoided.

4. An alternative approach to sub-groups: random effects regression

In this section we consider an alternative approach that builds on the analysis in the previous section. To recap, employing net-benefit regression methods has allowed us

to directly test the importance of sub-groups on the cost-effectiveness through the interaction of treatment effect with covariates. While indicating that there may be important differences in cost-effectiveness between sub-groups of the data, no consideration has been given to the relatively small sample size and the possibility that sub-group differences may be spurious (i.e. that they may have arisen by chance).

In geographical epidemiology, the interest is in considering how the rate of particular events (mortality or the prevalence of a disease or condition) vary across different areas (1992). While this is a legitimate concern, if we split a large area into many smaller areas we are dividing the data into mutually exclusive partitions – any estimated event rates within these areas will be based on a smaller sample size than then event rate for the whole region and therefore will be expected to differ by chance. A common solution to this problem has been to employ empirical Bayes methodology which effectively provides an estimate of the event rate in each area while adjusting for the higher than expected variance when dividing the data (Brown & Prescott 1999). In practice this can be estimated using a multi-level model where areas belong to regions (which may belong to higher level regions and so on). In the case of a one-level model, this is equivalent to fitting a random effects model where the random effect is specified across regions.

This analogy with methods in geographic epidemiology suggests an equivalent approach that could be applied to sub-group analysis in cost-effectiveness analysis. From a policy perspective, sub-groups of patients should be mutually exclusive – so we split the data into mutually exclusive sub-groups and fit a random effects model that adjusts for the fact that differences between sub-groups could have arisen purely by chance. This approach has the effect of 'shrinking' the raw estimates of cost-effectiveness in the individual sub-groups back toward the pooled estimate of cost-effectiveness for the whole group.

In terms of the ACT data, we had three covariates: age, race and GAF score. Age and GAF score were dichotomised about their median value to create eight mutually exclusive sub-groups of data. These eight groups are listed in Table 1 together with the code for the group (used to identify the group in later figures) and the number of patients in each group.

Using this split of the data, two net-benefit regression models were fitted. The first was a fixed effect model that included treatment by sub-group interactions for seven dummy variables (corresponding to each sub-group relative to a base sub-group). This model corresponds to the 'raw' estimate of cost-effectiveness in the individual groups. The second had exactly the same specification, except that a random effect was fitted across the sub-groups in the data. This random effect corresponds to an estimate of the expected variance corresponding to dividing the data and has the effect of 'shrinking' the estimates toward the overall mean.

To illustrate the difference in the estimates obtained from the two procedures, consider first just two of the sub-groups: older blacks with high GAF scores (B-O-H) versus young whites with low GAF scores (W-Y-L). The net-benefit curves as a function of the ceiling ratio are plotted in Figure 3 for these two sub-groups assuming (a) fixed effects and (b) random effects. Also plotted is the overall pooled result (shown by the dotted net-benefit curve) and the 95% confidence limits for net-benefits in the sub-groups. Comparing the two panels of Figure 3 clearly shows that the impact of the random effects model specification is to simultaneously shrink the point estimate of net-benefit toward the overall net-benefit and to shrink the variance of that estimate. The effect on the cost-effectiveness acceptability curves for the same two sub-groups is shown in Figure 4. This shows that the consequence of employing the random effects specification is a little less predictable, with the acceptability curves not necessarily becoming closer to the overall acceptability curve. This is because the shape of the curve is related not only to the point estimate of net-benefit and its variance but also to its proximity to zero net-benefits. Thus for the B-Y-L group, while the random effect specification shrinks the variance, since it also shrinks the point estimate toward zero, there is more uncertainty surrounding the net-benefit being positive.

Having illustrated the approach with two sub-groups, the full results for all eight sub-groups are presented in Figure 5, in terms of net-benefit, and in Figure 6 as acceptability curves, again for both (a) fixed effect and (b) random effect specifications. Note from Table 2 that there were only four young whites with high GAF scores and that these all happened to be in the treatment group. This means that

under a fixed effect specification it was impossible to estimate the cost-effectiveness due to a lack of comparator. However, in the random effects analysis it was possible to estimate the cost-effectiveness – although reassuringly, given the low numbers, the estimate is shrunk back to almost exactly that for the overall pooled analysis.

5. A multinational study in the presence of censoring

The analysis presented in the previous section related to sub-groups in a particular clinical trial that were defined by patient characteristics. However, the same approach could be taken where the sub-groups of interest were related to individual centres in a multi-centre study or individual countries in a multinational study (or indeed a combination of different centres in different countries). In this section we present another example where this time the sub-groups of interest are different countries in a multinational trial. In principle, the methods of the previous section could be applied directly except that in this example (as commonly arises in economic evaluation alongside clinical trials) there was censoring of patients in relation to both the costs and outcomes of interest.

The handling of censored cost data in economic analyses alongside clinical trials has been the focus of much research interest in recent years (Carides, Heyse, & Iglewicz 2000; Etzioni et al. 1999; Fenn et al. 1995; Lin et al. 1997). However, methods for simultaneously handling censoring costs and effects in the same equation are only just beginning to emerge and theses methods do not yet allow for covariate adjustment (Willan & Lin 2001). In the presence of censoring, therefore, it is not possible to simply employ the methods of the previous section.

Data were available on the costs and QALYs of a treatment intervention for severe heart failure from a randomised trial of patients in 16 countries. Five categories were formed from the four countries that recruited more than 100 patients and the amalgam of the remaining 11 countries. The Kaplan-Meier sample average estimator was employed to estimate the costs and QALYs in the presence of censoring and net-benefits were calculated for the overall pooled analysis and the five country groups – these are plotted on the cost-effectiveness plane in Figure 7(a). The variances for the censor-adjusted estimates were obtained using non-parametric bootstrapping.

Let $\hat{q}_i$ be the estimated net-benefit in the i = 1 to 5 country groups, with $\hat{s}_i^2$ representing the corresponding variance estimates for each country group. In order to estimate the analogue of the shrinkage estimates from the previous section we assume that $\hat{q}_i \sim N(q_i, s_i^2)$ and that $q_i \sim N(t, s^2)$. The overall pooled analysis gives us our estimate $\hat{t}$. An estimate of s^2 is given by the between country variance estimated as a weighted average of the squared deviations from the overall pooled estimate,

$$\hat{s}^2 = \frac{\sum_{i=1}^{5} w_i (\hat{q}_i - \hat{t})^2}{W} \cdot \left(1 + \frac{1}{5}\right)$$

where $w_i = \frac{1}{\hat{s}_i^2}$, $W = \sum_{i=1}^{5} w_i$ and (1+1/5) is a bias correction factor to account for the small number of observations from which the variance is estimated.

This gives the shrinkage estimators for the individual country net-benefits and variance as

$$\hat{q}_i^* = \frac{\hat{s}_i^2}{\hat{s}^2 + \hat{s}_i^2} \cdot t + \frac{\hat{s}^2}{\hat{s}^2 + \hat{s}_i^2} \cdot \hat{q}_i$$

$$\hat{s}_i^{2*} = \frac{\hat{s}^2}{\hat{s}^2 + \hat{s}_i^2} \cdot \hat{s}_i^2$$

These shrinkage estimators can then be used in place of the raw estimates, $\hat{q}_i$ and $\hat{s}_i^2$. The shrunken point estimates of cost and QALYs are presented on the cost-effectiveness plane in Figure 7(b).

It should be clear from the above formulation the reason why these estimators are also described as empirical Bayes estimators. The shrinkage estimators above are obtained by assuming a prior distribution $N(t, s^2)$ in conjunction with the individual country estimates of net-benefit and its variance and this prior distribution is estimated from the data.

Acceptability curves for the overall pooled results and the individual countries are presented in Figure 8 for (a) the unadjusted estimates, and (b) the shrinkage estimates.

It is clearly apparent from the shrinkage estimates that the cost-effectiveness for individual countries does not differ in important ways from the overall analysis and that much of the difference between Figure 7(a) and Figure 8(a) can be attributed to the expected variance associated with dividing the data.

6. Discussion

That the characteristics of the patient population for an intervention under evaluation will often be important determinants of the overall cost-effectiveness of that intervention has long been acknowledged by health economists undertaking secondary evaluations of health care interventions. Yet in economic analyses performed alongside a clinical trial, the sub-group analyses that are crucial to identifying how incremental cost-effectiveness varies at the margin are often discouraged. The concern is one of making inappropriate inference about differences that do not exist, but which have simply arisen by chance when splitting the data. The dangers of trawling the data for *ad hoc* sub-group effects is exemplified in the classic example from Collins and colleagues, who showed that in an RCT of thrombolytic therapy Scorpios had a four-fold increase in benefit from treatment compared to patients born under all other star-signs combined (Collins et al. 1987).

The purpose of this paper was not to encourage ill-informed sub-group analyses in economic evaluations, rather the aims were firstly to argue that sub-group analyses are more important in economic than in clinical evaluation, and secondly, to propose a methodology that is robust to dividing the data. Stark illustration of this is given in the first example of an intervention for homeless persons: where sub-groups analysed where very small, the degree of shrinkage back toward the overall mean is great. Only where effects are much greater than would be expected by chance do the shrinkage estimators show a sustained difference between the sub-groups.

Choice of sub-group to examine is likely to be of critical importance – particularly from the point of view of policy-making. The choice of sub-group in the ACT example was essentially arbitrary with continuous variables split about their median to facilitate illustration of the approach. In practice, the sub-groups should relate to identifiable groups that are appropriate to distinguish politically. For example, in the

ACT study it may be invidious to formulate policy that offers differential treatment on the grounds of ethnic origin. In which case analysing sub-groups by race becomes questionable for policy (although these analyses may be appropriate to better understand why different racial groups respond differentially to treatment).

A common recommendation for the interpretation of sub-group analyses in clinical evaluation is that the reason for sub-group effects has scientific plausibility. There is a danger when employing demographic characteristics to define sub-groups that these are really proxies for more complex effects. For example, mortality rates for heart disease have historically been higher for men than for women. However, this effect can be explained by risk factors such as diet, smoking and stress levels at work. Having controlled for these factors the observed differences between the sexes is no longer apparent – suggesting that any treatment decision that assumed a difference between the sexes in their propensity to develop heart disease would be flawed. In general, risk factors/severity scores for a disease or condition are likely to be among the characteristics most important for defining relevant sub-groups in economic evaluation. However, the focus on life expectancy as an outcome is likely to mean that demographic characteristics such as age and sex may continue to have an important role.

Sub-group analysis is a difficult business, and the methods proposed here should not be applied unthinkingly. However, policy-makers do need to consider the evidence for heterogeneity in cost-effectiveness between sub-groups and it may be that shrinkage estimation methods provides a robust basis from which to consider the extent to which incremental cost-effectiveness may vary at the margin.

References

Elliott P, Cuzick J, English D and Stern R (Editors). 1992. *Geographical & environmental epidemiology: methods for small area studies*. Oxford: OUP.

Altman, D. G. 1991. *Practical statistics for medical research*. London, Chapman & Hall.

American Psychiatric Association 1987, *Diagnostic and Statistical Manual of Mental Disorders DSM-III-R*, 3rd edn, American Psychiatric Association, Washington, DC.

Brown, H. & Prescott, R. 1999, *Applied mixed models in medicine* John Wiley & Sons ltd, Chichester.

Carides, G. W., Heyse, J. F., & Iglewicz, B. 2000, "A regression-based method for estimating mean treatment cost in the presence of right-censoring", *Biostatistics*, vol. 1, no. 3, pp. 299-313.

Collins, R., Gray, R., Godwin, J., & Peto, R. 1987, "Avoidance of large biases and large random errors in the assessment of moderate treatment effects: the need for systematic overviews", *Stat.Med.* , vol. 6, no. 3, pp. 245-254.

Etzioni, R. D., Feuer, E. J., Sullivan, S. D., Lin, D., Hu, C., & Ramsey, S. D. 1999, "On the use of survival analysis techniques to estimate medical care costs", *J.Health Econ.*, vol. 18, no. 3, pp. 365-380.

Fenn, P., McGuire, A., Phillips, V., Backhouse, M., & Jones, D. 1995, "The analysis of censored treatment cost data in economic evaluation", *Medical Care*, vol. 33, no. 8, pp. 851-863.

Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. 1996, *Cost-effectiveness in health and medicine* Oxford University Press, New York.

Hoch, J. S., Briggs, A. H., & Willan, A. 2002, "Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis", *Health Economics*, in press.

Lehman, A. F., Dixon, L. B., Hoch, J. S., DeForge, B. R., Kernan, E., & Frank, R. G. 1999, "Cost-effectiveness of Assertive Community Treatment for Homeless Persons with Severe Mental Illness.", *British Journal of Psychiatry*, vol. 174, pp. 346-352.

Lehman, A. F., Dixon, L. B., Kernan, E., DeForge, B. R., & Postrado, L. T. 1997, "A randomized trial of assertive community treatment for homeless persons with severe mental illness", *Archives of General Psychiatry*, vol. 54, pp. 1038-1043.

Lin, D. Y., Feuer, E. J., Etzioni, R., & Wax, Y. 1997, "Estimating medical costs from incomplete follow-up data.", *Biometrics*, vol. 53, no. 2, pp. 419-434.

Oxman, A. D. & Guyatt, G. H. 1992, "A consumer's guide to subgroup analyses", *Annals of Internal Medicine*, vol. 116, no. 1, pp. 78-84.

Stinnett, A. A. & Mullahy, J. 1998, "Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis", *Med.Decis.Making*, vol. 18, no. 2 Suppl, p. S68-S80.

Stinnett, A. A. & Paltiel, A. D. 1997, "Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means.", *Med.Decis.Making*, vol. 17, no. 4, pp. 483-489.

Willan, A. R. & Lin, D. Y. 2001, "Incremental net benefit in randomized clinical trials", *Stat.Med.*, vol. 20, no. 11, pp. 1563-1574.

Williams, A. 1985, "Economics of coronary artery bypass grafting", *Br.Med.J*, vol. 291, no. 6491, pp. 326-329.

Table 1
Covariate adjusted net-benefit regression estimates with treatment interaction (From Hoch et al, *in press*)[a]

N = 145 Explanatory Variables	NMB with ë = $0 [b] [se] [d] (p-value) [d]	NMB with ë = $100 [se] [d] (p-value) [d]	NMB with ë = $500 [se] (p-value)	NMB with ë = $1000 [se] (p-value)	Effect [c] [se] (p-value)
Constant Term	-109,000 [30,100] (<0.001)	-95,000 [31,900] (0.003)	-38,900 [28,700] (0.178)	31,200 [41,400] (0.452)	140 [29.9] (<0.001)
Covariates:					
black (dummy)	54,300 [32,700] (0.089)	57,500 [33,500] (0.089)	70,400 [31,500] (0.027)	86,400 [45,400] (0.059)	32.1 [32.7] (0.329)
age [e]	1,300 [1,220] (0.283)	1,580 [1,300] (0.227)	2,620 [1,470] (0.077)	3,930 [2,120] (0.067)	2.61 [1.53] (0.091)
gaf [e]	1,260 [940] (0.181)	1,560 [990] (0.118)	2,760 [1,350] (0.043)	4,260 [1,940] (0.030)	2.99 [1.40] (0.034)
Treatment dummy:					
ACT	64,400 [32,300] (0.042)	73,400 [33,200] (0.029)	109,000 [34,400] (0.002)	154,000 [49,700] (0.002)	89.8 [35.8] (0.013)
Treatment-covariate interactions:					
ACT· black	-60,200 [34,400] (0.082)	-66,400 [36,500] (0.071)	-91,000 [39,600] (0.023)	-122,000 [57,100] (0.035)	-61.5 [41.2] (0.137)
ACT· age [e]	-2,860 [1,560] (0.031)	-3,100 [1,690] (0.068)	-4,050 [1,930] (0.038)	-5,230 [2,780] (0.063)	-2.37 [2.01] (0.240)
ACT· gaf [e]	-2,280 [1,140] (0.049)	-2,600 [1,230] (0.036)	-3,900 [1,870] (0.038)	-5,540 [2,690] (0.042)	-3.26 [1.94] (0.095)
R-squared (adjusted)	0.069	0.082	0.150	0.112	0.097
F(7, 137)	2.53	2.83	3.46	3.60	3.2
Prob > F	0.018	0.009	0.002	0.001	0.004

[a] All monetary measures in U.S. dollars, all results to three significant figures.
[b] When ë= $0, NMB = - Cost.
[c] The coefficients from the Effect regression (and not the NMB regression with ë) are reported since as ë , the *p*-values for the NMB coefficient estimates are equivalent to those obtained when 'days stable housing' is the dependent variable.
[d] Huber-White robust standard errors and *p*-values corrected for heteroskedasticity.
[e] The continuous variables age and gaf have been centred around their respective means.

Table 2
Splitting the data into mutually exclusive sub-groups by covariate information

Description	Code	Sample size
black-young-lowgaf	B-Y-L	34
black-young-highgaf	B-Y-H	26
black-old-lowgaf	B-O-L	24
black-old-highgaf	B-O-H	21
white-young-lowgaf	W-Y-L	11
white-young-highgaf	W-Y-H	4
white-old-lowgaf	W-O-L	11
white-old-highgaf	W-O-H	14

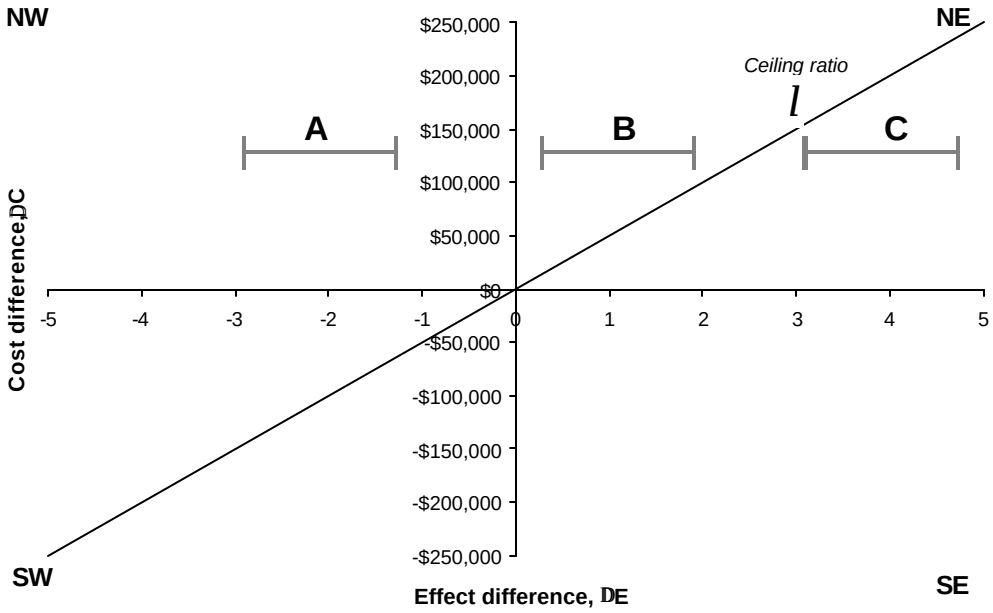

Figure 1
'Qualitative' and 'quantitative' differences in clinical and economic evaluation

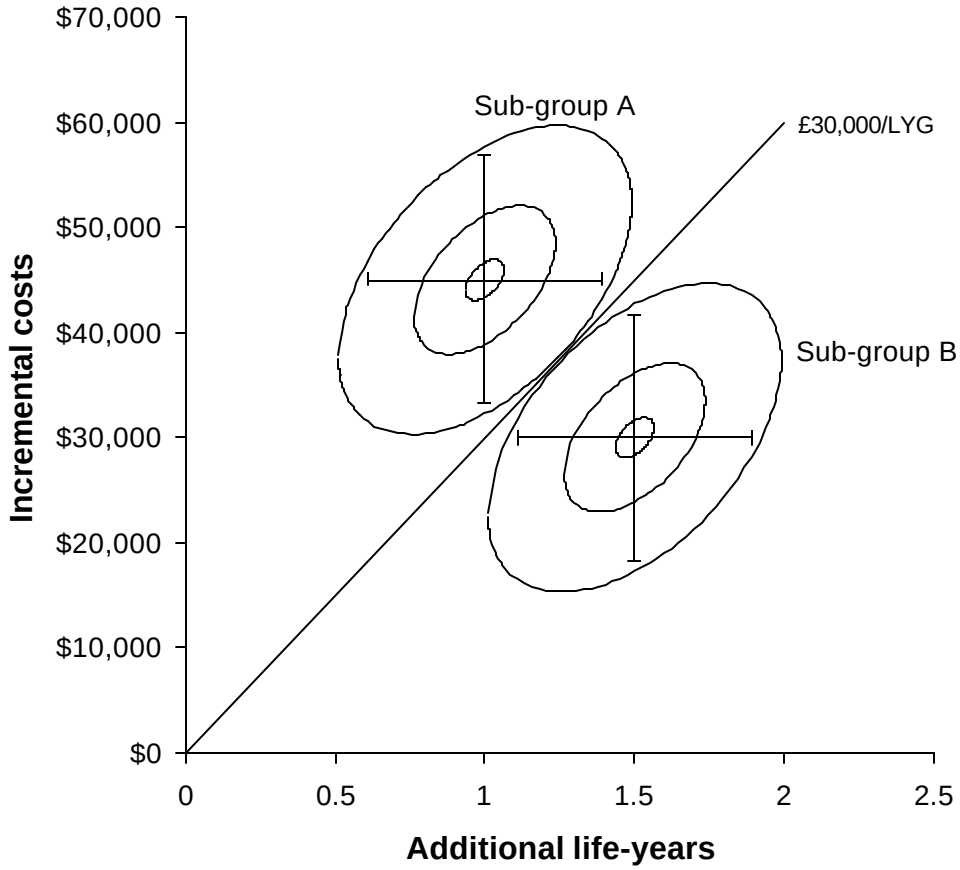

Figure 2
Sub-groups in cost-effectiveness analysis versus sub-group analysis in costs and effects independently

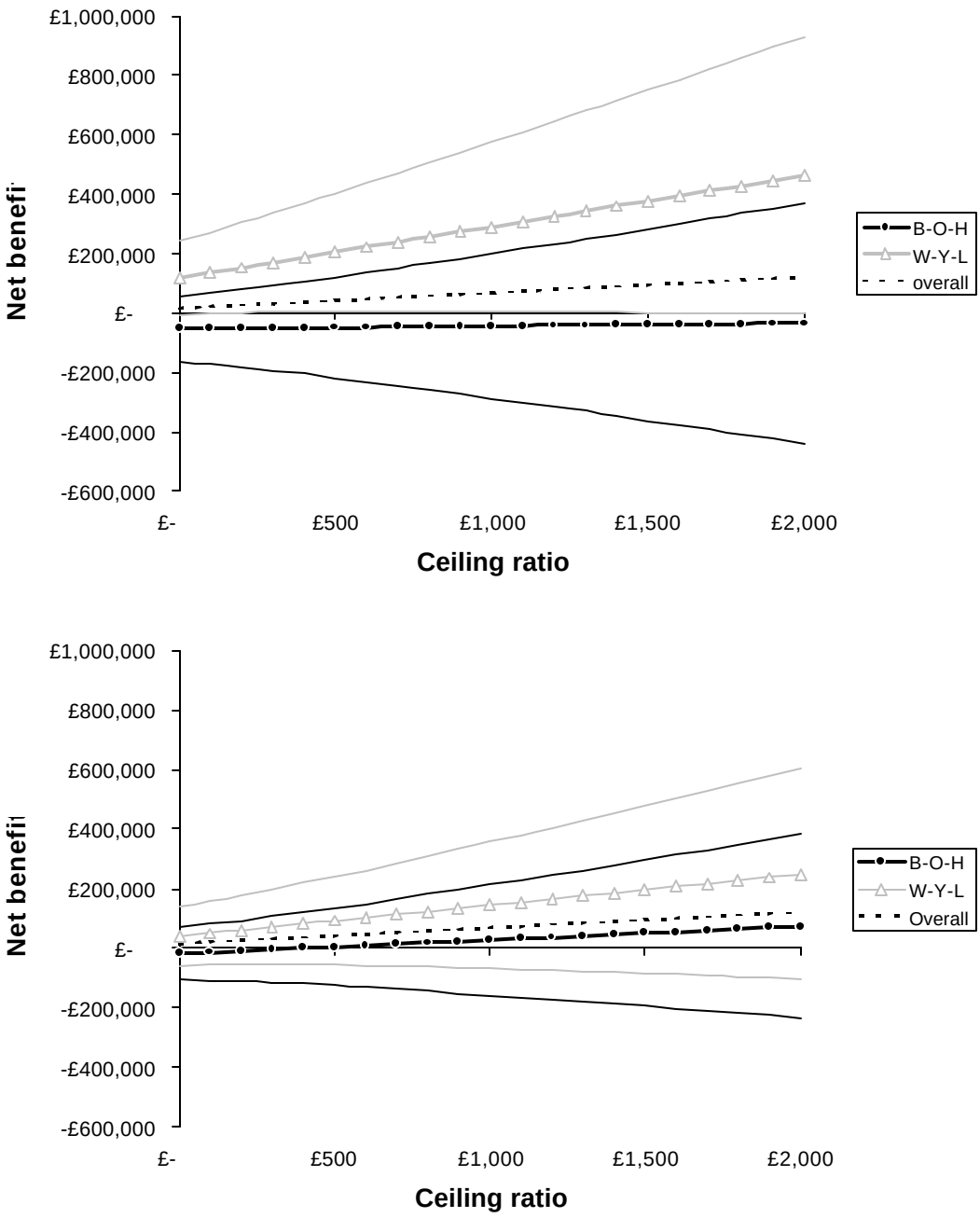

Figure 3
Sub-group analysis two groups on net-benefit (as function of ceiling ratio) assuming (a) fixed effects and (b) random effects

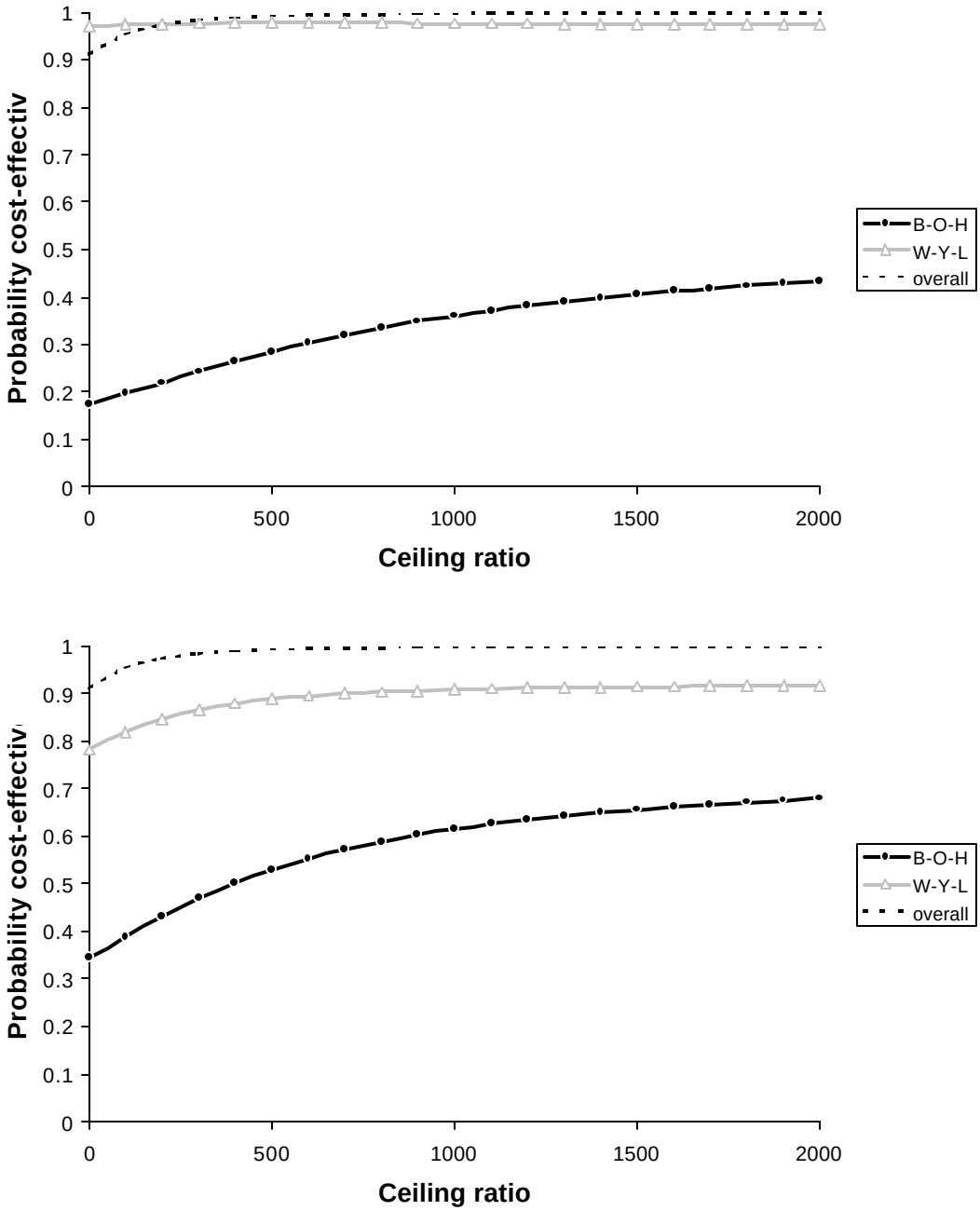

Figure 4
Cost-effectiveness acceptability curves for two sub-groups assuming (a) fixed effects and (b) random effects

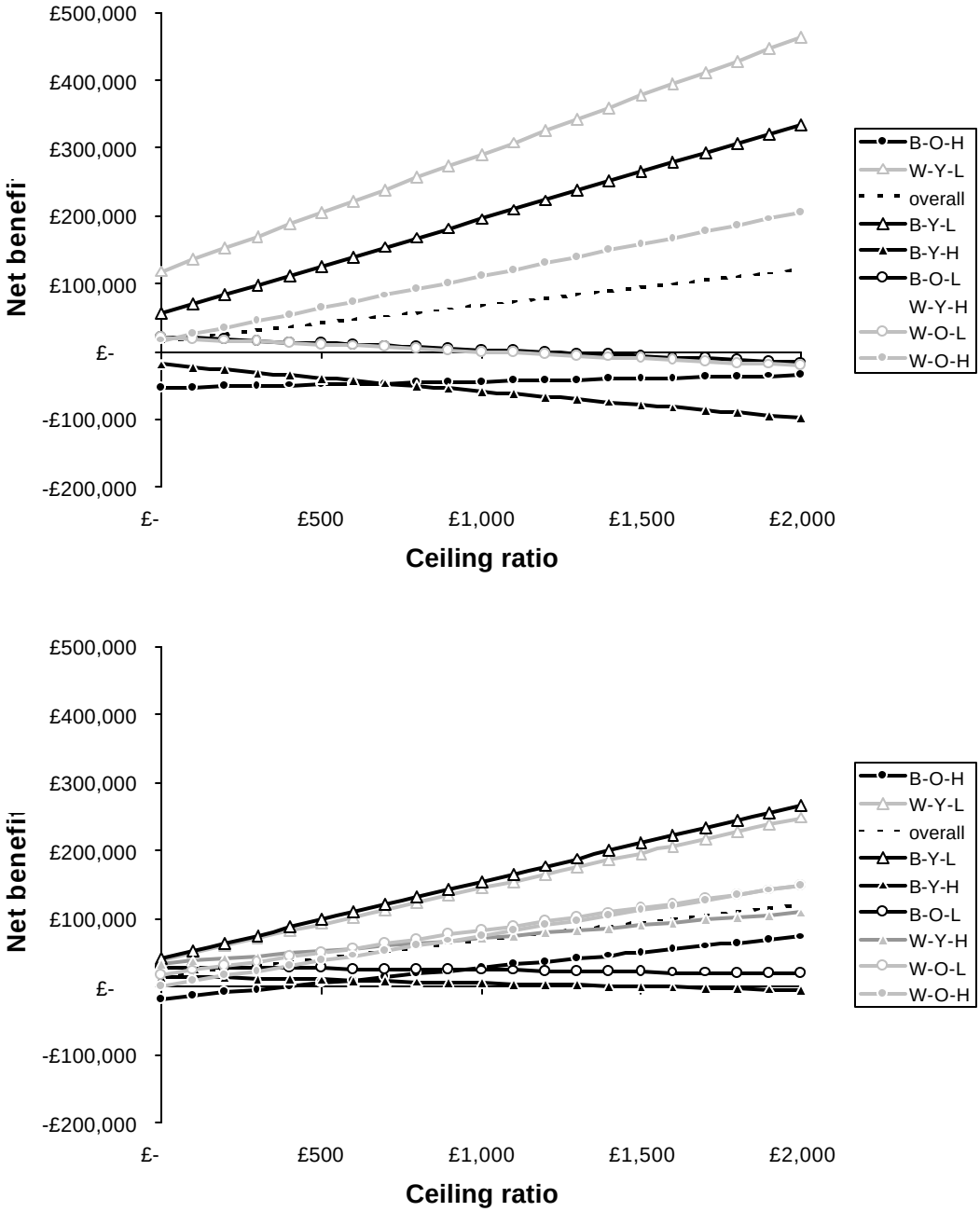

Figure 5
Sub-group analysis for all groups on net-benefit (as function of ceiling ratio) assuming (a) fixed effects and (b) random effects

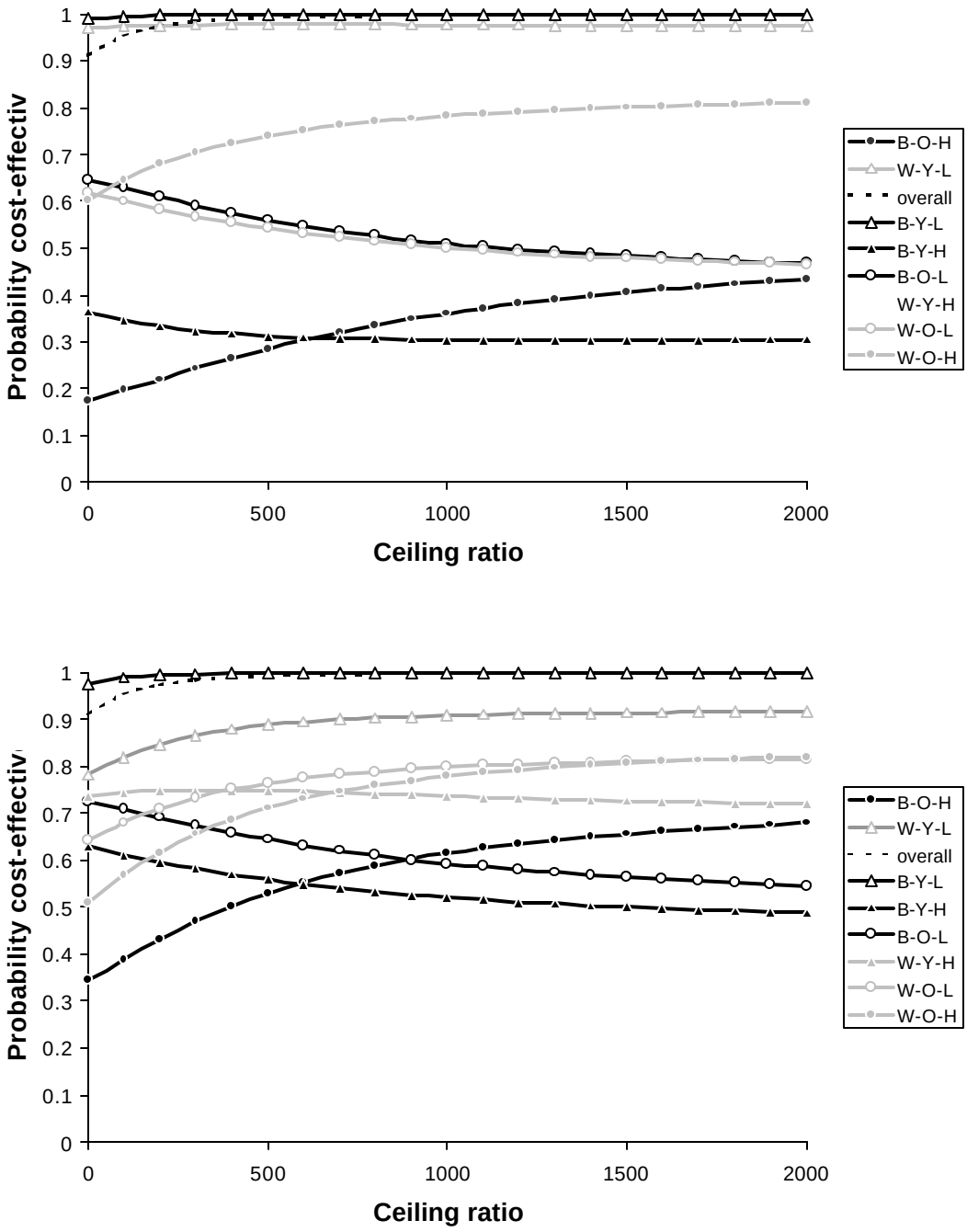

Figure 6
Cost-effectiveness acceptability curves for all sub-groups assuming (a) fixed effects and (b) random effects

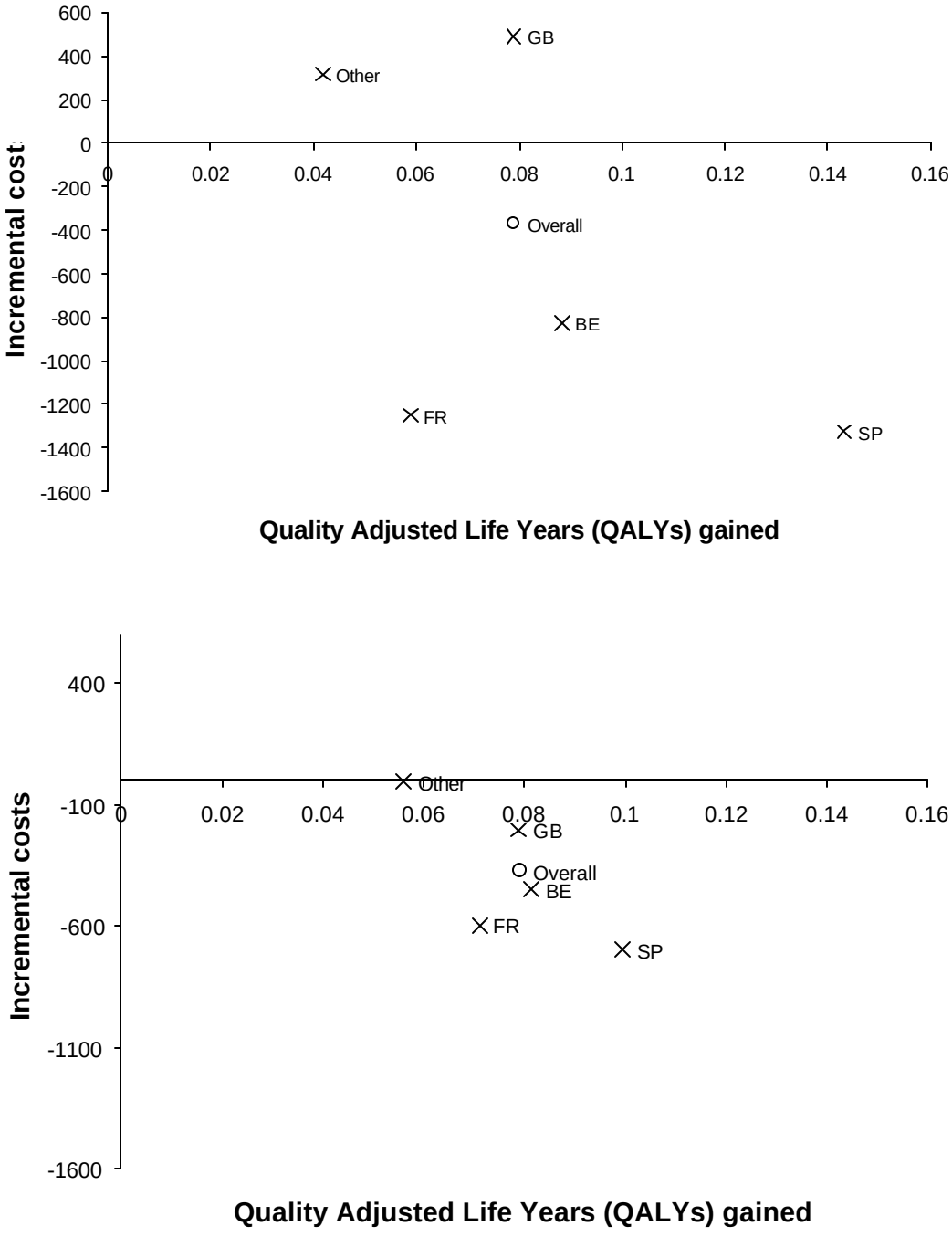

Figure 7
Overall pooled result and country specific point estimates for the multinational study of a heart failure drug for (a) unadjusted and (b) shrinkage estimators plotted on the CE plane

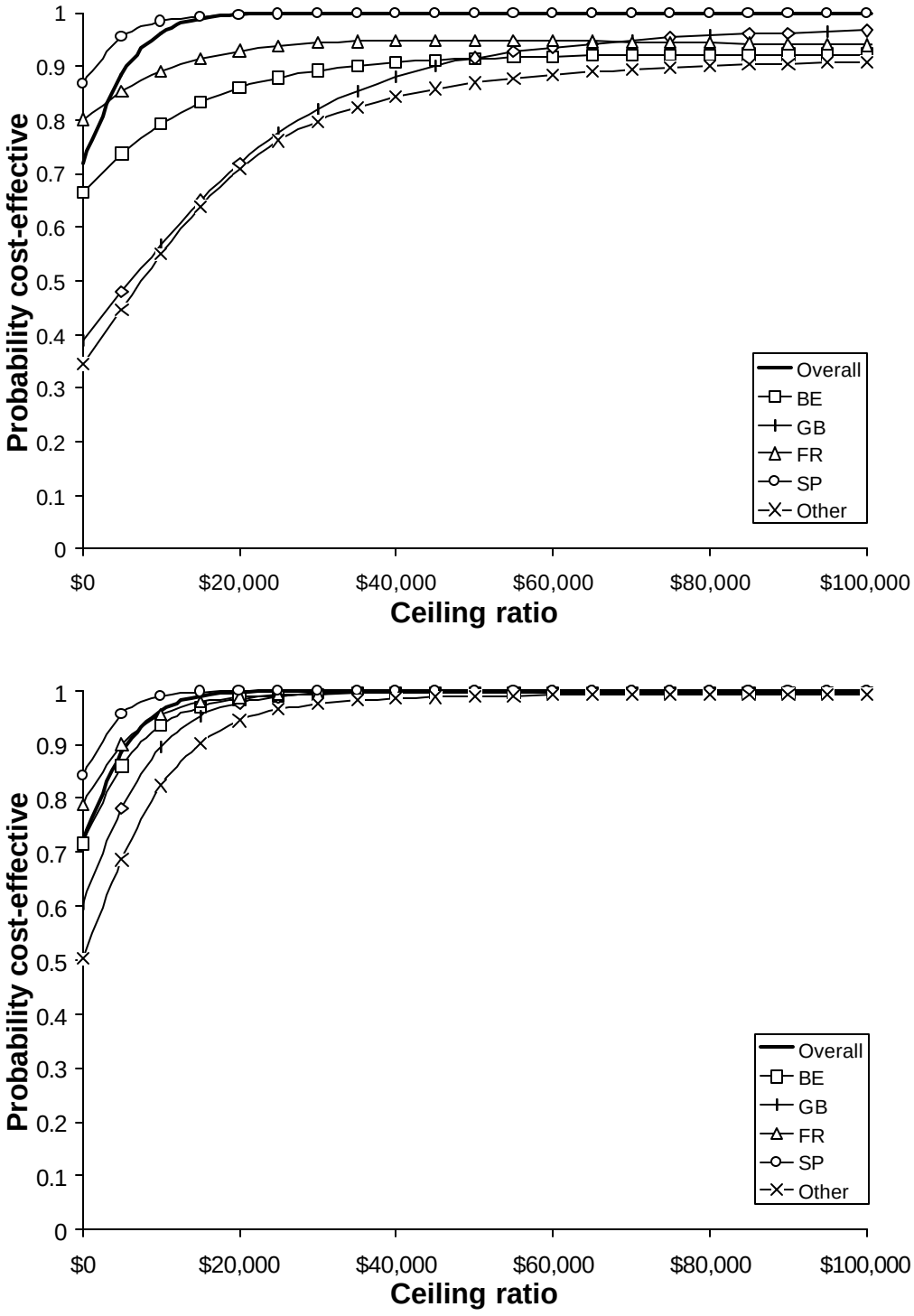

Figure 8
Cost-effectiveness acceptability curves for the overall pooled result and country specific results for the multinational study of a heart failure drug using (a) unadjusted and (b) shrinkage estimators