



Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis

Jeffrey S. Hoch^{a,*}, Andrew H. Briggs^b and Andrew R. Willan^c

^a*Department of Epidemiology and Biostatistics, University of Western Ontario, Canada*

^b*Health Economics Research Centre, University of Oxford, UK*

^c*Department of Clinical Epidemiology and Biostatistics, McMaster University, Canada*

Summary

Economic evaluation is often seen as a branch of health economics divorced from mainstream econometric techniques. Instead, it is perceived as relying on statistical methods for clinical trials. Furthermore, the statistic of interest in cost-effectiveness analysis, the incremental cost-effectiveness ratio is not amenable to regression-based methods, hence the traditional reliance on comparing aggregate measures across the arms of a clinical trial. In this paper, we explore the potential for health economists undertaking cost-effectiveness analysis to exploit the plethora of established econometric techniques through the use of the net-benefit framework – a recently suggested reformulation of the cost-effectiveness problem that avoids the reliance on cost-effectiveness ratios and their associated statistical problems. This allows the formulation of the cost-effectiveness problem within a standard regression type framework. We provide an example with empirical data to illustrate how a regression type framework can enhance the net-benefit method. We go on to suggest that practical advantages of the net-benefit regression approach include being able to use established econometric techniques, adjust for imperfect randomisation, and identify important subgroups in order to estimate the marginal cost-effectiveness of an intervention. Copyright © 2002 John Wiley & Sons, Ltd.

Keywords cost-effectiveness analysis using regression; net-benefit framework; cost-effectiveness acceptability curve; economic evaluation; econometrics

Introduction

The development of applied health economics has progressed along two broad paths. The traditional path sees applied health economics undertaken in economics departments, employing applied econometrics methods. The second way in which health economics has developed has been in the economic evaluation of health care technologies. In these

cases, health economists have undertaken such evaluations as members of multidisciplinary teams composed of clinicians, statisticians, epidemiologists and trialists. They assist in facilitating the team's goals of producing information about the cost-effectiveness of interventions. It is perhaps of little surprise, therefore, that the development of economic evaluation alongside clinical trials owes more to medical statistics than to econometrics,

*Correspondence to: Department of Epidemiology and Biostatistics, The University of Western Ontario, London, Ontario, Canada N6A 5C1. Tel.: 519-661-2162; fax: 519-661-3766; e-mail: jeffhoch@biostats.uwo.ca

the former being characterised by methods surrounding experimental design [1], while the latter typically involves information obtained from population surveys [2]. Furthermore, cost-effectiveness analysis has traditionally been concerned with the estimation of cost-effectiveness ratios – which are not amenable to regression analysis, the mainstay of methods employed by applied econometricians.

In this paper, we explore the potential for health economists undertaking cost-effectiveness analysis to exploit the plethora of established econometric techniques through the use of the net-benefit framework – a recently suggested reformulation of the cost-effectiveness problem that avoids reliance on cost-effectiveness ratios and their associated statistical problems. This approach also allows formulation of the cost-effectiveness problem within a standard regression type framework. We go on to suggest that this formulation will encourage researchers to address, in a statistically robust way, the underlying economics of cost-effectiveness analysis in the sense of exploring the importance of covariates on the marginal cost-effectiveness of an intervention (i.e., interaction effects between the intervention and important subgroups). All too often within the context of a clinical trial, the potential for cost-effectiveness to vary at the margin is obscured by aggregation across the arms of the trial.

For many years, the recommendation for analysts conducting economic evaluation of health care interventions has been to calculate *incremental* cost-effectiveness ratios (ICERs) in order to summarise the value for money of interventions [3–5]. More recently, as economic evaluations have begun to be conducted prospectively alongside clinical trials, statistical problems associated with ratio statistics have become apparent in the interpretation of sampling uncertainty in the ICER [6–9]. Ambiguity arises with ICERs, since the value of the ratio itself is not sufficient to give unequivocal treatment recommendations. For example, a negative ICER is consistent with either a more expensive, less effective treatment or a less expensive, more effective treatment. A positive ICER indicates either a more expensive, more effective treatment or a less expensive, less effective treatment. While it is trivial to check which situation has produced the ICER estimate, great care must be taken when constructing a confidence interval so that ratios with similar signs but different interpretations are not grouped together.

The analyst who chooses to bootstrap the ICER's confidence interval – a very popular method for handling uncertainty in stochastic cost-effectiveness analysis – may choose to reorder the bootstrap replicates so that ordering reflects the decision-making implications of the replicates. Unfortunately, the negative ICERs do not obey the law of transitivity; unambiguous preference ordering is not possible in the Southeast and Northwest quadrants of the cost-effectiveness plane (the plane is illustrated in Figure 1). Hence, the construction of confidence intervals for cost-effectiveness ratios when uncertainty covers more than one quadrant of the cost-effectiveness (CE) plane can be problematic.

Recently, a new framework for cost-effectiveness analysis has been suggested; the net-benefit framework [10,11] reformulates the cost-effectiveness problem to generate a net-benefit statistic. Its linear form has more attractive statistical properties than the ICER and offers a simpler alternative for handling uncertainty in stochastic cost-effectiveness analysis. In this paper, we argue that the linear nature of the net-benefit statistic also allows the analyst to enhance economic evaluation by employing regression methods. Practical advantages include being able to identify important subgroups, adjust for imperfect randomisation and make use of established econometric techniques. The remainder of this paper is organised into four parts. The next section provides a statistical motivation for the use of regression methods in the net-benefit framework. An empirical example of the procedure is then provided in

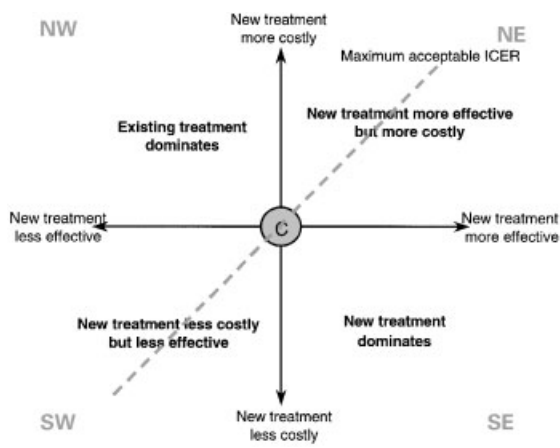


Figure 1. The cost-effectiveness plane

Section 3. The remaining sections provide discussion and suggest implications of the main results of this paper.

Statistical considerations in cost-effectiveness analysis

This section begins by specifying the usual approach of estimating an ICER in cost-effectiveness analysis, together with an outline of the challenges faced when analysing ICERs in a stochastic framework. The net-benefit framework is then introduced as a solution to these problems and a regression approach for net-benefits is presented in order to demonstrate how cost-effectiveness can be estimated from a single regression equation.

The incremental cost-effectiveness ratio

In the most common case, an economic analysis involves an evaluation of an intervention treatment (T_1) compared to a standard care treatment (T_0). Denoting the expected values of cost and effect for T_k (for $k=0$ or 1) as μ_{Ck} and μ_{Ek} respectively, the incremental cost-effectiveness ratio (ICER) comparing T_1 to T_0 is defined as

$$\text{ICER} = \frac{\mu_{C1} - \mu_{C0}}{\mu_{E1} - \mu_{E0}} = \frac{\mu_{\Delta C}}{\mu_{\Delta E}} \quad (1)$$

with the implication that the intervention offers good value for money if the ICER is below some maximum willingness to pay for health gain. That is, a decision should be made to implement the more costly, but more effective treatment intervention if

$$\frac{\mu_{\Delta C}}{\mu_{\Delta E}} < \lambda \quad (2)$$

where λ is the maximum acceptable willingness to pay per unit of health gain (or ceiling ratio).

Of course, it is never possible to know the true incremental costs and true incremental effects of an intervention, since it is impossible to simultaneously observe the costs and effects of two different treatments in the same population of patients [12]. Using sample data for economic data collected in a clinical trial setting and the analogy principle [13,14], it is possible to estimate the true,

but unobservable ICER parameter by

$$\widehat{\text{ICER}} = \frac{\bar{C}_1 - \bar{C}_0}{\bar{E}_1 - \bar{E}_0} = \frac{\Delta \bar{C}}{\Delta \bar{E}}$$

using the sample mean costs \bar{C}_k and sample mean effects \bar{E}_k for the treatment arms (i.e., $k=0, 1$).

The ICER statistic does not give sufficient information for decision-making without knowledge of the quadrant of the CE plane (or equivalently, the sign of the numerator or denominator of the ratio). This is of particular concern when attempting to calculate confidence limits for cost-effectiveness ratios where uncertainty covers more than one quadrant of the CE plane. For example, bootstrapping methods – widely applied to the problem of ICER confidence interval estimation – produce multiple ‘replications’ of the cost and effect differences. If these methods are used to calculate ICER estimates without knowledge of where on the CE plane the replicates fall, negative and positive ICERs from different quadrants may be improperly pooled together. The resulting confidence limit estimates will be misleading to the extent of the conflation of the ICER replicates from different quadrants but with similar signs.

Challenges still remain after the bootstrap replicates have been sorted into their appropriate quadrants. Some analysts have suggested that the magnitude of a negative ICER conveys no useful statistical information [10], and that confidence intervals for ICERs are meaningful only when uncertainty is restricted to one of the positive quadrants of the CE plane. Therefore when ICER replicates cover more than one quadrant of the plane, the cost-effectiveness acceptability curve approach to summarising uncertainty should be employed [9,15].

The net-benefit framework

Recently, two papers have highlighted what we call the ‘net-benefit’ approach to handling uncertainty in cost-effectiveness analysis [10,11]. The net-benefit framework starts from the premise that the ICER is only of partial assistance to decision-makers. Decision-makers must judge whether the additional effect is worth the additional cost (i.e. whether the ICER signals a ‘good deal’). This is formalised in the decision rule of Equation (2).

The equation can be rearranged to give a measure of 'net-benefit' and an associated decision rule that the new therapy should be implemented only if the net-benefits are positive.

Two formulations of net-benefit have been suggested based on Equation (1). A programme's net-monetary-benefit (NMB) is calculated by subtracting the additional cost from the additional effect valued in dollars [11]. In contrast, a programme's net-health-benefit (NHB) is calculated by subtracting the additional cost valued in effect units from the additional effect [10]. The decision rule in the net-benefit framework is that the new therapy should be implemented over the existing treatment if:

$$\text{NMB} = \lambda \cdot \mu_{\Delta E} - \mu_{\Delta C} > 0 \quad (3)$$

or, equivalently, if:

$$\text{NHB} = \mu_{\Delta E} - \mu_{\Delta C}/\lambda > 0 \quad (4)$$

where λ is used to denote the maximum willingness to pay (wtp) for an additional effect.

The same sample analogues are employed to estimate the mean effect and cost differences in order to give the estimated net-benefit statistics

$$\widehat{\text{NMB}} = \lambda \cdot \Delta \bar{E} - \Delta \bar{C}$$

$$\widehat{\text{NHB}} = \Delta \bar{E} - \Delta \bar{C}/\lambda.$$

However, in contrast to the ICER, where the variance is not defined, the variance of net-benefits estimated from sample mean cost and effects in the trial arms is simply a linear combination of two asymptotically normal variables. Therefore it can be defined as:

$$\begin{aligned} \text{var}(\widehat{\text{NMB}}) \\ = \lambda^2 \text{var}(\Delta \bar{E}) + \text{var}(\Delta \bar{C}) - 2\lambda \text{cov}(\Delta \bar{E}, \Delta \bar{C}) \end{aligned}$$

in terms of the monetary net-benefit measure, or:

$$\begin{aligned} \text{var}(\widehat{\text{NHB}}) \\ = \text{var}(\Delta \bar{E}) + \frac{1}{\lambda^2} \text{var}(\Delta \bar{C}) - \frac{2}{\lambda} \text{cov}(\Delta \bar{E}, \Delta \bar{C}) \end{aligned}$$

for the net health benefit measure. The linear nature of the NMB and NHB statistics makes them preferable to work with relative to the ICER. This is evident when constructing a $(1 - \alpha)\%$ confidence interval. For net-benefits, the interval can be determined in the standard fashion as $\widehat{\text{NB}} \pm z_{\alpha/2} \sqrt{\sigma_{\text{NB}}^2}$ where $\widehat{\text{NB}}$ is the estimated net-benefit measure with variance σ_{NB}^2 , and $z_{\alpha/2}$ is the

critical value from the standard normal distribution.

The net-benefit framework's many advantages come with a potential drawback; the net-benefit statistic is a function of λ , a value unknown to the analyst in most cases. Stinnett and Mullahy [10] consider this attribute a strength as it forces explicit consideration of the value of λ . They emphasise the importance of sensitivity analysis to examine different values of λ . Furthermore, cost-effectiveness acceptability curves estimated using the net-benefit framework and varying values of λ will exactly coincide with those calculated using an appropriate analysis on the CE plane, since the underlying cost-effectiveness decision rule is the same in each case [16]. Nevertheless, the net-benefit framework provides a much more straightforward method of calculating such acceptability curves.

A net-benefit regression approach

In this section we exploit the linear nature of the net-benefit statistic to show how net-benefits can be used to estimate cost-effectiveness within a regression framework. Without loss of generality, we use net monetary benefits on the cost scale to illustrate the approach (the results could equivalently be presented in terms of net health benefits).

Introductory textbooks emphasise the importance of taking an incremental approach [4,5] rather than comparing average cost-effectiveness ratios. A more recent contribution to the literature highlighted the fundamental problem of taking patient-level average ratios: the mean of ratios is not equal to the ratio of the means [17]. The consequence is that

$$\frac{\bar{C}_1}{\bar{E}_1} - \frac{\bar{C}_0}{\bar{E}_0} \neq \frac{\bar{C}_1 - \bar{C}_0}{\bar{E}_1 - \bar{E}_0}$$

demonstrating that the incremental ratio cannot be constructed from the difference between the average cost-effectiveness ratios in each arm of the trial.

In contrast, the difference in the mean net-benefit of the experimental treatment and the mean net-benefit of standard care treatment will give the overall incremental net-benefit statistic of Equation (3) [10]. This is straightforward to see algebraically through simple manipulation of the

net-benefit expressions

$$\begin{aligned}\bar{\text{NMB}}_1 - \bar{\text{NMB}}_0 &= (\lambda \cdot \bar{E}_1 - \bar{C}_1) - (\lambda \cdot \bar{E}_0 - \bar{C}_0) \\ &= \lambda(\bar{E}_1 - \bar{E}_0) - (\bar{C}_1 - \bar{C}_0) \\ &= \lambda \cdot \Delta \bar{E} - \Delta \bar{C} \\ &= \Delta \bar{\text{NMB}},\end{aligned}$$

for the experimental (subscript 1) and control (subscript 0) interventions. Therefore the usefulness of average net-benefit is not directly in terms of the average figures themselves, but in the simple linear relationship between average and incremental net-benefit.

The linearity of the net-benefit framework can be employed to estimate cost-effectiveness within a regression framework by defining a net-benefit value for each subject. For example,

$$\text{NMB}_i = \lambda \cdot E_i - C_i$$

where E_i and C_i are the observed effect and cost for subject i . A simple linear model for subject i 's net-monetary-benefit (NMB_i) can be formed in the following way

$$\text{NMB}_i = \alpha + \delta t_i + \varepsilon_i \quad (\text{Model 1})$$

where α is an intercept term, t a treatment dummy taking the value zero for the standard treatment and the value one for the treatment under consideration, and ε is a stochastic error term. The regression coefficient δ on the treatment dummy provides the estimate of the incremental net-benefit, $\bar{\text{NMB}}_1 - \bar{\text{NMB}}_0$, from a standard net-benefit analysis. Similarly, the standard error of the coefficient is the same as that calculated from the standard approach.

The power of this framework is that it is straightforward to add additional explanatory variables in order to examine their impact on cost-effectiveness directly. For example, we can model the patient-level net-benefit with an alternative model

$$\text{NMB}_i = \alpha + \sum_{j=1}^p \beta_j x_{ij} + \delta t_i + \varepsilon_i \quad (\text{Model 2})$$

where there are p covariates x . That is, in this model the coefficient δ on the treatment dummy gives the incremental net-benefit, and therefore the cost-effectiveness, of implementing the new treatment controlling for confounding variables. Of course in the context of an experimental design like a randomised controlled trial (RCT), the randomisation process is expected to ensure an

equal balance of both observed and unobserved confounding factors across the treatment arms.

Correcting for unbalanced allocation in observed covariates that has arisen by chance in clinical evaluation is only one advantage of adopting a regression based approach to cost-effectiveness analysis. All too often in RCT based cost-effectiveness analyses, the results are simply aggregated across the arms of the trial to provide the overall ICER without any consideration of how the ICER varies between subgroups (at the margin). Since economics is concerned fundamentally with the margin, the impact of covariates such as age, sex and disease severity on the cost-effectiveness of treatment interventions is of fundamental interest. The net-benefit regression approach outlined in this section gives an explicit method for examining marginal issues through the use of interaction terms.

Consider the model

$$\text{NMB}_i = \alpha + \sum_{j=1}^p \beta_j x_{ij} + \delta t_i + t_i \sum_{j=1}^p \gamma_j x_{ij} + \varepsilon_i \quad (\text{Model 3})$$

where the final summation is the interaction between the treatment dummy and the covariates. The magnitude and significance of the coefficients γ_j on the interaction between the covariates of the model and the treatment dummy indicate how cost-effectiveness of treatment is expected to vary at the margin; large and statistically significant γ_j 's point towards important patient subgroups. The key advantage of this framework is the ability to use standard regression techniques to examine the marginal impact of covariates on *incremental* cost-effectiveness instead of the usual approach of aggregating cost and effect differences across arms of the trial. We now illustrate the general approach with an applied example.

Example: empirical data from a randomised trial

Background

The Program in Assertive Community Treatment (PACT) is one of the most studied models of care for persons with severe and persistent mental illnesses (SPMI) [18–21]. Lehman *et al.* [22] found that an assertive community treatment (ACT)

program, relative to usual community services, reduced psychiatric inpatient days, emergency room visits, days homeless, and days in jail for homeless persons with SPMI in Baltimore, Maryland (USA). The study's rationale was that by providing potentially more expensive but coordinated, community-based care through the ACT programme, homeless persons with severe mental illnesses would spend more days in stable community housing with savings realized by shifting the patterns of care from higher cost crisis-oriented inpatient and emergency services to lower cost, ongoing ambulatory services. The results suggest that in the city of Baltimore, ACT was effective in achieving important outcomes warranting an examination of the cost-effect trade-off. Lehman *et al.* [23] conducted an economic evaluation of the ACT programme as it was implemented. Their analysis employed ICERs and provides an empirical example of the simplifying and unifying nature of the net-benefit framework.

Methods and data

Direct treatment costs across the one year intervention period were examined from the perspective of the state mental health authority. Housing status was chosen as the main effectiveness measure because of its established validity as a primary outcome for homeless persons with SPMI [24]. A day of stable housing was defined as living in a non-institutionalised setting not intended to serve the homeless (e.g., independent housing, living with family, etc.). Subjects randomised to the comparison usual care condition had access to services usually available to homeless persons in the city of Baltimore. Lehman *et al.* [23] offer more detail about the study's methodology.

One hundred forty-eight persons who were homeless with SPMI were randomised to either the experimental ACT program or to usual community services. Subjects were recruited dur-

ing a 19-month period in 1991 and 1992 from inner-city psychiatric hospitals, primary health care agencies, shelters, missions and soup kitchens. Baseline data collection included an assessment of overall mental health functioning using the Global Assessment of Functioning (GAF) Scale [25]. For this paper, we obtained complete data on 73 participants randomly assigned to the ACT program and 72 randomly assigned to usual care (comparison) services.

Results using a standard cost-effectiveness analysis

Baseline group comparisons examined differences between the two intervention groups on demographics, diagnoses and histories of homelessness at baseline [23]. Table 1 presents an abbreviated set of results. The two groups were comparable with ACT subjects being slightly older and higher functioning. In contrast, there was a greater than expected percentage of African Americans randomised to the comparison condition ($p < 0.01$).

Table 2 provides a brief statistical summary of the cost and effect data and provides a conventional cost-effectiveness analysis of the data by looking at the incremental costs and effects between the two groups. ACT subjects had lower costs and more days of stable housing, suggesting this was the dominant treatment. Due to the significant difference in the subjects between treatment arms with respect to ethnic origin, a stratified analysis is also reported in Table 2.

Of course, it is important to take into account the sampling variability of the data when interpreting the results of this analysis. Therefore the results from Table 2 are presented on the cost-effectiveness plane in Figure 2(a) with uncertainty represented by elliptical contours covering 5%, 50%, and 95% of the integrated density under the joint normal assumption. Figure 2(b) presents an acceptability curve [15], to summarise the

Table 1. Subject characteristics by treatment group

Characteristic	ACT subjects ^a	Comparison subjects
Mean age (SD)	38.96 years (9.43)	36.00 years (8.30)
Mean GAF ^b score (SD)	37.90 (9.08)	35.32 (9.06)
African American**	62%	83%

^aACT indicates assertive community treatment.

^bGAF is the global assessment of functioning score.

** $p < 0.01$.

Table 2. Sample statistics from the economic evaluation data

Group variable	Mean	SD	SE
Overall analysis:			
<i>Comparison arm (N = 72)</i>			
Cost ^a	67 400	76 500	9020
Effect ^b	159	105	12.4
	Correlation = -0.43		
<i>ACT arm (N = 73)</i>			
Cost	51 900	61 100	7160
Effect	212	104	12.2
	Correlation = -0.39		
<i>Increments^c</i>			
Cost difference	-15 500	—	11 500
Effect difference	52.7	—	17.4
	Correlation = -0.41		
Stratified analysis:			
<i>Black subjects</i>			
Cost difference	-5070	—	13 200
Effect difference	35.6	—	21.8
<i>White subjects</i>			
Cost difference	-62 700	—	33 000
Effect difference	98.1	—	39.0
Incremental net-benefits:		Net monetary benefit (SE)	
<i>Value of ceiling ratio</i>		<i>Overall</i>	<i>Black subjects</i>
$\lambda = 0^d$		15 500 (11 500)	5070 (13 200)
$\lambda = 100$		20 800 (12 300)	8630 (14 500)
$\lambda = 500$		41 900 (17 000)	22 900 (21 300)
$\lambda = 1000$		68 200 (24 500)	40 700 (31 100)
$\lambda = \infty^e$		$\infty(\infty)$	$\infty(\infty)$

^aAll costs in US dollars, all results to three significant figures.

^bAll effects in 'days of stable housing'.

^cDifferences are calculated as ACT value minus comparison value.

^dEquivalent to minus the cost difference.

^eAs $\lambda \rightarrow \infty$, NMB $\rightarrow \infty$ and se(NMB) $\rightarrow \infty$.

uncertainty in the cost-effectiveness of the intervention. The stratified analyses are presented on the cost-effectiveness plane and as acceptability curves in Figure 3.

Results using a net-benefit regression approach

An equivalent net-benefit regression approach was employed by estimating

$$\text{NMB}_i = \alpha + \delta \text{ACT}_i + \varepsilon_i,$$

Model 1 from the previous section with an ACT treatment dummy variable. The results from this model are presented in Table 3. Net monetary benefits were calculated employing λ values of \$0, \$100, \$500, and \$1000.

To estimate the behaviour of the t -statistics and p -values as $\lambda \rightarrow \infty$, a regression was run with 'days stable housing', the effectiveness measure, as the dependent variable. The coefficients from this 'effect' regression (and not the NMB regression) are reported in the final column of Table 3, since as $\lambda \rightarrow \infty$, the NMB coefficient estimates and their

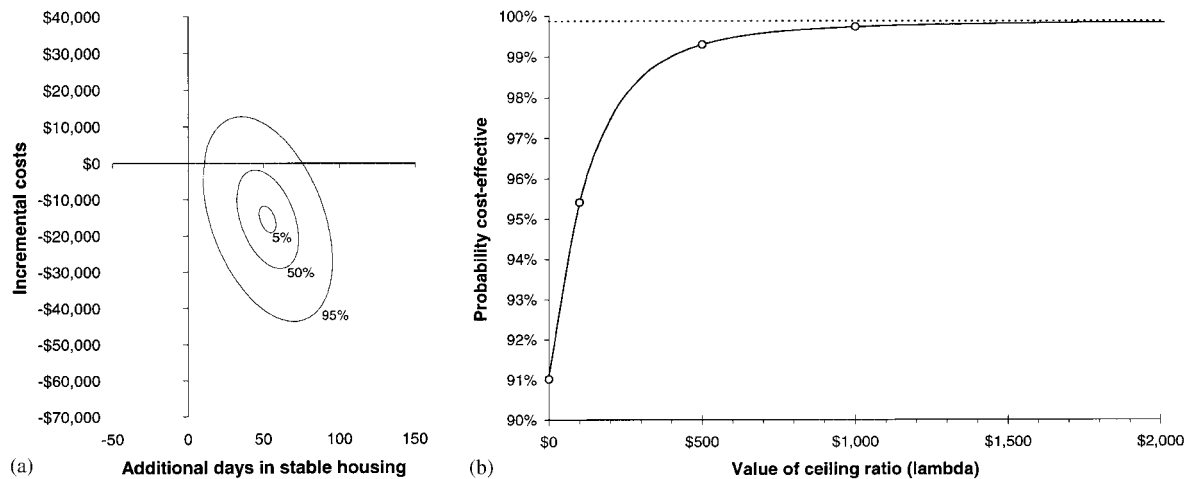


Figure 2. The cost-effectiveness plane (a) and acceptability curve (b) for the standard cost-effectiveness analysis of the ACT programme versus standard care. Uncertainty on the CE plane is represented by elliptical contours covering 5%, 50% and 95% of the joint density of cost and effect differences

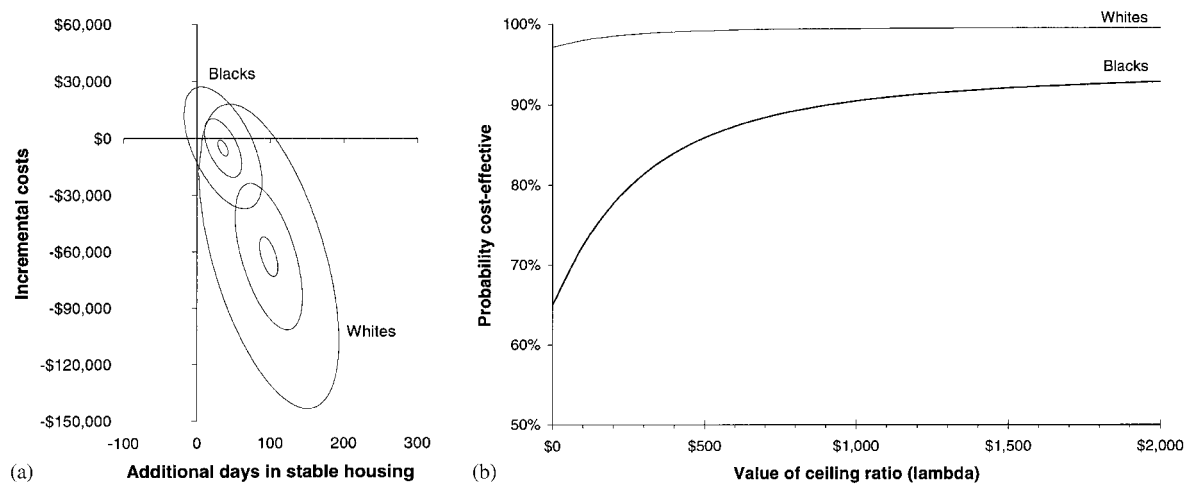


Figure 3. The cost-effectiveness plane (a) and acceptability curve (b) for the standard stratified cost-effectiveness analysis of the ACT programme versus standard care

standard errors tend to ∞ . However, as $\lambda \rightarrow \infty$, the t -ratios and p -values from the NMB regression tend to those for the 'effect' regression. Therefore, the interest in the final column of Table 3 is on the ratio of the coefficient estimate to the standard error rather than on the coefficients themselves since they are on a different scale (days of stable housing).

Note that the coefficient on the treatment dummy corresponds to the incremental net-benefit

and that the regression results are exactly equivalent to the standard approach to cost-effectiveness analysis presented in Table 2. These regression results can also be used to obtain a cost-effectiveness acceptability curve by plotting $1 - p/2$ against λ where p is the p -value from the coefficient on the ACT treatment dummy variable. (The divisor of two is employed because the acceptability curve is equivalent to a one-sided test.) These values are plotted as points in Figure 2(b),

Table 3. Simple net-benefit regression estimates (Model 1)^a

<i>N</i> = 145 explanatory variables	NMB with $\lambda = \$0^b$ [se] (<i>p</i> -value)	NMB with $\lambda = \$100$ [se] (<i>p</i> -value)	NMB with $\lambda = \$500$ [se] (<i>p</i> -value)	NMB with $\lambda = \$1000$ [se] (<i>p</i> -value)	Effect ^c [se] (<i>p</i> -value)
Constant term	-67 400 [8160] (<0.001)	-51 500 [8740] (<0.001)	12 200 [12 000] (0.313)	91 800 [17 400] (<0.001)	159 [12.4] (<0.001)
<i>Treatment dummy</i>					
ACT	15 500 [11 500] (0.179)	20 800 [12,300] (0.093)	41 900 [17 000] (0.015)	68 200 [24 500] (0.006)	52.7 [17.4] (0.003)
<i>R</i> -squared (adjusted)	0.006	0.018	0.041	0.045	0.054
<i>F</i> (1, 143)	1.82	2.85	6.04	7.75	9.13
Prob > <i>F</i>	0.179	0.094	0.015	0.006	0.003

^aAll monetary measures in US dollars, all results to three significant figures.

^bWhen $\lambda = 0$, NMB = -Cost.

^cThe coefficients from the Effect regression (and not the NMB regression with $\lambda \rightarrow \infty$) are reported since as $\lambda \rightarrow \infty$, the *p*-values for the NMB coefficient estimates are equivalent to those obtained when 'days stable housing' is the dependent variable.

and it is clear that they correspond to points on the acceptability curve calculated in the standard fashion [15]. The dashed line in Figure 2(b) shows where the acceptability curve is tending to as $\lambda \rightarrow \infty$.

Having demonstrated that a simple net-benefit regression approach is equivalent to a standard analysis, we consider some richer model specifications. It is clear from Table 1 that despite the randomisation of homeless persons with SPMI to each arm of the study, many more white people were allocated to the treatment arm. To statistically address this imbalance, we estimated the cost-effectiveness of the ACT programme adjusting for potentially confounding variables such as race, age and GAF score. This regression equation

$$NMB_i = \alpha + \beta_1 \text{black}_i + \beta_2 \text{age}_i + \beta_3 \text{gaf}_i + \delta \text{ACT}_i + \varepsilon_i$$

corresponds to Model 2 of the previous section. The estimated coefficients from this model are presented in Table 4. It is clear from the reported results of the *F*-test that the covariate adjustment has more impact on the measure of effectiveness than the measure of cost. Consequently, the adjustment has a greater impact on net-benefit regressions based on higher λ values, which place greater weight on the effect variable.

Again, in this model, it is the coefficient on the treatment dummy that gives the results of interest, and these coefficients can be used to plot the adjusted results on the CE plane and to generate a cost-effectiveness acceptability curve. These are

presented in Figure 4, where the unadjusted results are presented in light grey to aid comparison with Figure 2.

In Model 2, although coefficients are generated for the covariates, these are not of direct interest since they describe the impact on average net-benefits. Although average net-benefits are useful as a basis from which to obtain incremental net-benefits, they are not helpful for decision-making on their own. This is because they suffer from the same informational limitations associated with average cost-effectiveness ratios. From an economic point of view, the interest in the covariates is on how they affect the estimate of the intervention's incremental net-benefit (i.e. the marginal impact on incremental cost-effectiveness). To examine this, we employed a model that interacts the treatment dummy with the covariates

$$NMB_i = \alpha + \beta_1 \text{black}_i + \beta_2 \text{age}_i + \beta_3 \text{gaf}_i + \delta \text{ACT}_i + \gamma_1 \text{ACT}_i \text{black}_i + \gamma_2 \text{ACT}_i \text{age}_i + \gamma_3 \text{ACT}_i \text{gaf}_i + \varepsilon_i$$

which corresponds to Model 3 of the previous section. The results from this regression are in Table 5. These results show there is a significant interaction between race and treatment with black subjects achieving lower net-benefits from treatment in comparison to their white counterparts. The results also show that age and GAF score are potentially important covariates on the marginal cost-effectiveness of treatment, although not all of the interaction terms across the regressions are consistently significant at the

Table 4. Covariate adjusted net-benefit regression estimates (Model 2)^a

<i>N</i> = 145 explanatory variables	NMB with $\lambda = \$0^b$ [se] (<i>p</i> -value)	NMB with $\lambda = \$100$ [se] (<i>p</i> -value)	NMB with $\lambda = \$500$ [se] (<i>p</i> -value)	NMB with $\lambda = \$1000$ [se] (<i>p</i> -value)	Effect ^c [se] (<i>p</i> -value)
Constant term	-80 700 [13 900] (<0.001)	-63 800 [14 900] (<0.001)	3 890 [20 600] (0.850)	88 500 [29 500] (0.003)	169 [20.8] (<0.001)
<i>Covariates</i>					
Black (dummy)	16 400 [13 500] (0.224)	15 700 [14 400] (0.276)	13 100 [19 900] (0.531)	9 690 [28 500] (0.734)	-6.73 [20.1] (0.739)
Age	-86.0 [660] (0.896)	72 [707] (0.919)	704 [975] (0.471)	1 490 [1 400] (0.287)	1.58 [0.987] (0.112)
GAF	372 [644] (0.565)	527 [690] (0.446)	1150 [953] (0.230)	1 920 [1 370] (0.161)	1.55 [0.964] (0.110)
<i>Treatment dummy</i>					
ACT	18 400 [12 100] (0.132)	22 600 [13 000] (0.084)	39 600 [17 900] (0.029)	60 900 [25 700] (0.019)	42.51 [18.2] (0.021)
<i>R</i> -squared (adjusted)	<0.001	0.005	0.031	0.074	0.071
<i>F</i> (4, 140)	0.96	1.2	2.15	2.81	3.75
Prob $> F$	0.431	0.316	0.078	0.028	0.006

^aAll monetary measures in US dollars, all results to three significant figures.

^bWhen $\lambda = 0$, NMB = -Cost.

^cThe coefficients from the Effect regression (and not the NMB regression with $\lambda \rightarrow \infty$) are reported since as $\lambda \rightarrow \infty$, the *p*-values for the NMB coefficient estimates are equivalent to those obtained when 'days stable housing' is the dependent variable.

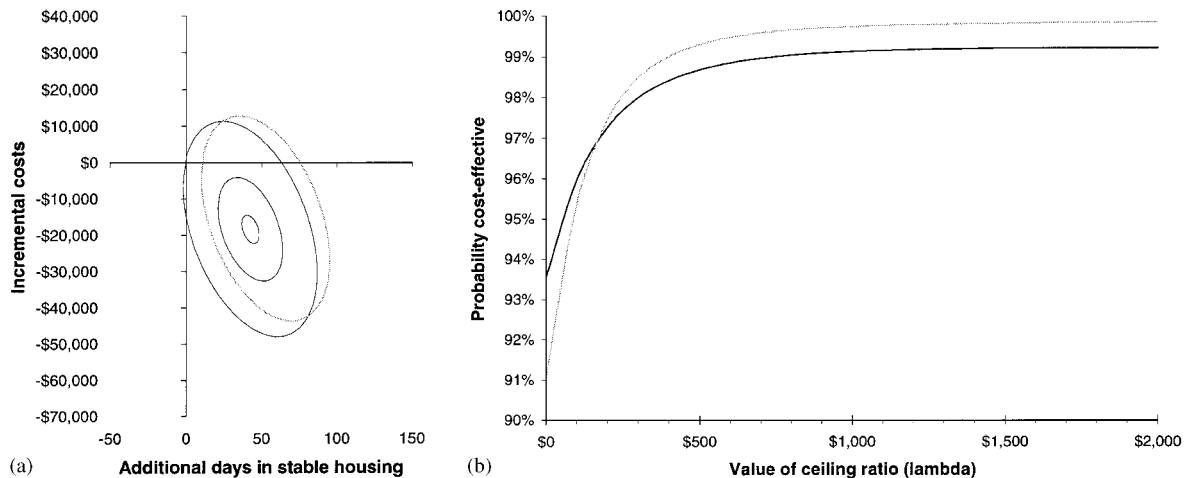


Figure 4. The cost-effectiveness plane (a) and acceptability curve (b) for the covariate adjusted cost-effectiveness analysis of the ACT programme versus standard care. The light grey ellipse and curve in each panel shows the position of the unadjusted estimates from Figure 2 (only the 95% ellipse is shown for comparison)

5% level. To see how these results can be used consider Figure 5, which shows uncertainty on the cost-effectiveness plane and cost

acceptability curves for the white versus black subjects at average age and GAF score. The implications of these findings are discussed further in the next section.

Table 5. Covariate adjusted net-benefit regression estimates with treatment interaction (Model 3)^a

<i>N</i> = 145 explanatory variables	NMB with $\lambda = \$0^b$ [se] ^d (<i>p</i> -value) ^d	NMB with $\lambda = \$100$ [se] ^d (<i>p</i> -value) ^d	NMB with $\lambda = \$500$ [se] (<i>p</i> -value)	NMB with $\lambda = \$1000$ [se] (<i>p</i> -value)	Effect ^c [se] (<i>p</i> -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)
<i>Covariates</i>					
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 ^c	1300 [1220] (0.283)	1580 [1300] (0.227)	2620 [1470] (0.077)	3930 [2120] (0.067)	2.61 [1.53] (0.091)
GAF ^c	1260 [940] (0.181)	1560 [990] (0.118)	2760 [1350] (0.043)	4260 [1940] (0.030)	2.99 [1.40] (0.034)
<i>Treatment dummy</i>					
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)
<i>Treatment-covariate interactions</i>					
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 ^c	-2860 [1560] (0.031)	-3100 [1690] (0.068)	-4050 [1930] (0.038)	-5230 [2780] (0.063)	-2.37 [2.01] (0.240)
ACT · GAF ^c	-2280 [1140] (0.049)	-2600 [1230] (0.036)	-3900 [1870] (0.038)	-5540 [2690] (0.042)	-3.26 [1.94] (0.095)
<i>R</i> -squared (adjusted)	0.069	0.082	0.150	0.112	0.097
<i>F</i> (7, 137)	2.53	2.83	3.46	3.60	3.20
Prob > <i>F</i>	0.018	0.009	0.002	0.001	0.004

^aAll monetary measures in US dollars, all results to three significant figures.

^bWhen $\lambda = 0$, NMB = -Cost.

^cThe coefficients from the Effect regression (and not the NMB regression with $\lambda > ?$) are reported since as $\lambda > ?$, the *p*-values for the NMB coefficient estimates are equivalent to those obtained when 'days stable housing' is the dependent variable.

^dHuber-White robust standard errors and *p*-values corrected for heteroskedasticity.

^eThe continuous variables Age and GAF have been centred around their respective means.

Regression diagnostics

We tested for heteroskedasticity [26] in the five varieties of the fully interacted model (Model 3). Results for the test of nonconstant variance were not uniform. For $\lambda = \$0$ and $\$100$, we rejected the null hypothesis of homoskedasticity at $p < 0.001$. For $\lambda = \$500$, $\$1000$ and $\lambda \rightarrow \infty$ (i.e., 'days stable housing', the effectiveness measure, as the dependent variable), the respective *p*-values were 0.06, 0.45 and 0.58, and we were unable to reject the null hypothesis at the 5% level.

Residuals were calculated and analysed. For $\lambda \leq \$1000$, we were able to reject joint skewness and kurtosis tests for normality; however, at $\lambda = \$1000$, the residuals' kurtosis was not significantly different from that of a normally distributed

variable. In the regression where $\lambda \rightarrow \infty$, the joint null hypothesis that the residuals were normally distributed with zero skewness ($p = 0.37$) and kurtosis equal to three ($p = 0.03$) could not be rejected at the 5% level ($p = 0.06$). We also examined the residuals for patterns by plotting them against predicted values of the dependent variable. Visual inspection of the residuals confirmed concerns about heteroskedasticity; Figure 6 shows a cone-like pattern evident when $\lambda = 0$.

We examined potential influential observations by calculating "leverage points" and *DFBETAs* [27]. In general, leverage values greater than twice the average may require closer inspection; for this analysis the cutoff value was $2 \times 8/145 \approx 0.11$. There were 12 subjects with leverage scores ≥ 0.11 . The highest score was 0.21 for a subject

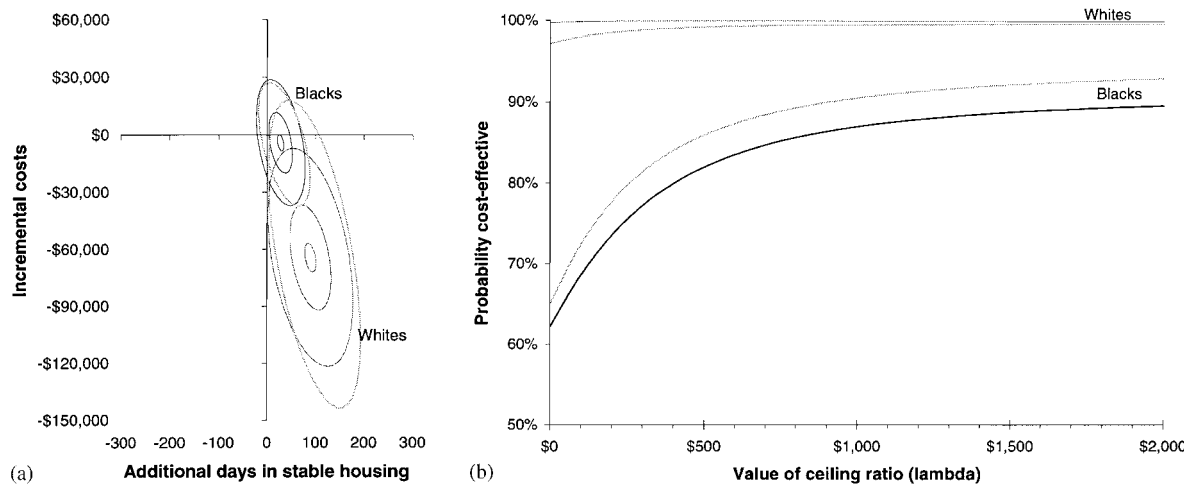


Figure 5. The cost-effectiveness plane (a) and acceptability curve (b) for regression-based cost-effectiveness analysis of the ACT programme versus standard care by race. The light grey ellipses and curves in each panel shows the position of the unadjusted estimates from the standard stratified analysis in Figure 3

who did not receive the intervention treatment but had an exceptionally good outcome accompanied by exceptionally low costs. Using DFBETAs calculated for both the cost and outcome regressions, we examined subjects' influence on the estimate of direct effect of ACT (i.e., δ). For the cost regression ($\lambda = 0$), there were four subjects (all control and all black) whose influence was consistent with shifting the ACT coefficient ≥ 0.20 deviations downwards (the largest absolute shift was by 0.84); conversely, there were seven subjects (six control and all black) whose inclusion was estimated to shift the ACT coefficient ≥ 0.20 deviations upwards (the largest shift was by 0.54). When DFBETAs were calculated for the case where $\lambda \rightarrow \infty$, four subjects (three control and all black) had values consistent with shifting the ACT coefficient ≥ 0.20 deviations downwards (the largest absolute shift was by 0.46); conversely, there were three subjects (all control and all black) whose inclusion was estimated to shift the ACT coefficient ≥ 0.20 deviations upwards (the largest shift is by 0.40).

Given the results of the preceding regression diagnostics, we reanalysed Model 3 using *rreg*, the robust regression command in STATA [28]. The *rreg* command computes iteratively reweighted least squares where the weights are based on absolute residuals; iterations were stopped when the maximum change in weights dropped below

0.01. While the results are not presented in detail here, we briefly sketch the main differences between the results obtained by ordinary least squares and by robust regression (additional details are available from the authors). For $\lambda = \$0$ and $\$100$, the coefficient for the ACT indicator was no longer statistically significant. For $\lambda = \$500$ and $\$1000$, the coefficients for both the black indicator variable and its interaction term with ACT were no longer statistically significant. On the other hand, the coefficient for age became statistically significant at the 5% level. For the outcome regression (i.e., where $\lambda \rightarrow \infty$), there were no substantive changes. In general, when $\lambda \leq \$1000$ the robustly estimated coefficients were smaller in absolute terms for the overall effect of ACT (i.e., both the direct effect and the interaction terms were robustly estimated as closer to zero). When $\lambda \rightarrow \infty$, the overall effect of ACT was larger in absolute terms.

Discussion

The problems associated with statistical estimation of ICERs are now well documented in the health economic literature. The net-benefit framework, originally suggested as a method for handling uncertainty in economic evaluation, has been

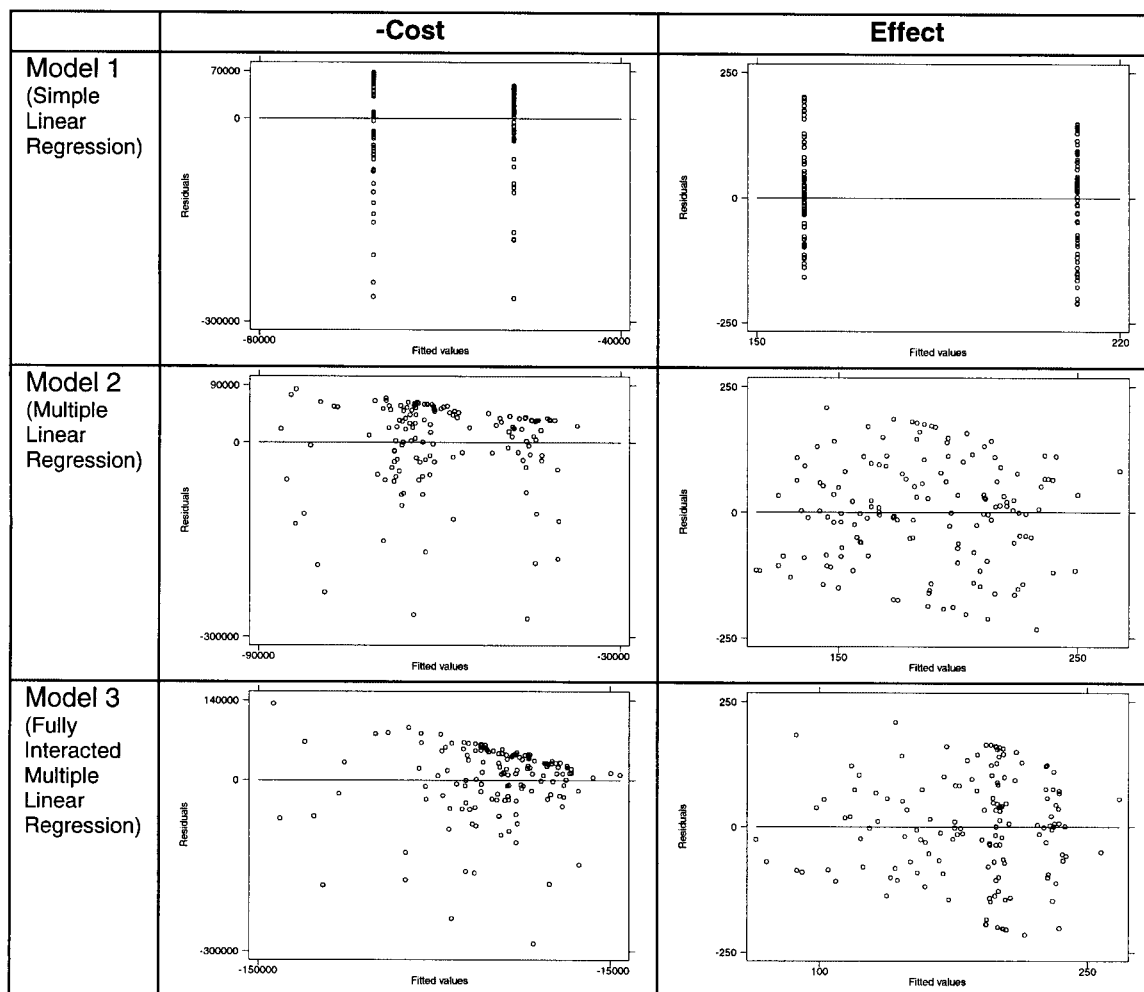


Figure 6. Residual plots for the regressions on -cost (equivalent to NMB when $\lambda = 0$) and effect (equivalent to NMB when $\lambda \rightarrow \infty$) for each of the three models

extended here to show how the net-benefit statistic can allow cost-effectiveness to be estimated directly in a regression framework. Within the context of simple regression analysis with a single (dummy) variable for the treatment intervention, the results are entirely equivalent to the standard approach to cost-effectiveness analysis. The advantages of the framework outlined in this paper come from the ability to move beyond simple regression modelling to explore the impact of covariates on marginal cost-effectiveness through interaction terms.

As recognised in the original economic evaluation of these data [23], the observed difference in

cost and effect between black and white subjects is noteworthy. The authors suggested one interpretation of the data is that the pattern of usual care for homeless persons with SPMI varies according to race in Baltimore. The researchers found that in spite of these differences, ACT tended to reduce excess use of crisis services and to increase the use of under-utilised ambulatory services for both groups. The authors concluded that the overall lower efficiency of ACT for black subjects in producing stable housing suggested that more attention should be given in the programme to differences between races on patterns of homelessness and service utilisation. These important

policy findings were discovered by considering race as a covariate in the economic evaluation.

The importance of the net-benefit regression approach to cost-effectiveness analysis described here is that it is possible to estimate the marginal effect of race on the cost-effectiveness of the programme while controlling for other covariates. The graphical presentation in Figure 5 shows the effect of race to be even more important than the original simple stratified analysis would suggest, since controlling for the effect of additional covariates further separates the estimated joint densities on the cost-effectiveness plane.

In our empirical example, both the negative ICER estimates (i.e. the dominance of the ACT programme compared to standard care) and the small sample sizes once the analysis was stratified underscored the potential of the net-benefit framework. There are other situations in which researchers may benefit from the convenience of the net-benefit regression approach, since it includes all of the strengths inherent in the standard net-benefit framework. One example is when the cost-effectiveness data are characterised by a large amount of statistical uncertainty, and bootstrapped replicates of the ICER cover two non-adjacent quadrants of the CE plane. Adopting the net-benefit regression approach allows the analyst to skip the *ad hoc* solution of reordering the bootstrap replicates that is sometimes employed when constructing a confidence interval.

However, the net-benefit regression approach also shares the limitations associated with the net-benefit framework. It is an unavoidable fact that decision-making for cost-effectiveness analysis will depend on λ , the maximum willingness to pay per unit of effect. Analysts may also be cautious of normative statements based on λ values in light of the fact that the net-benefit framework has the same prescriptive limitations as an ordinary cost-effectiveness analysis. The net-benefit approach simplifies some of the statistical aspects of cost-effectiveness analysis, but it is still cost-effectiveness analysis and subject to its inherent limitations [10].

An additional challenge inherent in using the net-benefit regression approach involves handling violations of the assumptions of classical normal linear regression. Regression diagnostics can help to identify outliers. In this paper, we used influence statistics, such as DFBETAs, to determine which subjects were driving the results and in what direction and also employed robust regression.

The data in this paper featured small sample size (common in many research areas) and non-normality (non-negative outcome data and skewed cost data). Because the net-benefit variable is a weighted combination of the cost and outcome data, highly skewed cost data may be less of a problem as λ increases and outcomes are valued more (as was evident in our empirical example). An important implication is that standard statistical tests relying on normality may not be valid for small values of λ .

The net-benefit regression approach presented in this paper is largely one of convenience. Regression methods have not been used widely to estimate cost-effectiveness analysis in economic appraisals conducted alongside clinical trials. However, the theory for bivariate regression already exists and the knowledgeable analyst will be able to calculate entirely equivalent analyses without forming a net-benefit regression directly [29,30]; however, if decisions are to be made based on the analysis, the λ value must be incorporated. Nevertheless, the ability to run net-benefit regressions directly on a statistical package is a practical advantage. Although OLS yields BLUE estimates of the net-benefit coefficients, it may be possible to improve upon the efficiency of the OLS estimates by adopting more sophisticated regression methods like the minimum distance estimator framework expressed in Chamberlain [31] or the generalised method of moments framework described in Hansen [32]. Future econometric work might investigate this promising direction.

Summary

This paper has discussed how the net-benefit framework can simplify the statistical work involved in economic evaluation (e.g., avoiding problems associated with ratio statistics) and also offer insights (e.g., exploring the importance of covariates on the marginal cost-effectiveness of an intervention). While it is true that the simplifying linearity comes at the cost of conditioning the analysis on values of λ , this may not be such a crucial limitation especially if the decision-maker can be assumed to know λ . We have suggested augmenting the standard net-benefit framework by utilising a regression approach, so that the potential strengths of both methods may be incorporated into a unified framework. The

marriage of econometrics and economic evaluation brings together something old (the regression framework), something new (the net-benefit framework), something borrowed (the decision-maker's λ) to produce something BLUE.

Acknowledgements

The authors are grateful to Anthony Lehman, Professor of Psychiatry and Director of the Center for Mental Health Services Research at the University of Maryland, for the use of the data. Valuable comments on an earlier draft of this paper were received from Carolyn Dewa, Vivian Ho, Willard Manning, Allan Pollock, Arthur Sweetman, Stephanie Van Bebber, Tiemen Woutersen, participants at the International Health Economics Association's Third International Conference, and two anonymous referees. The views expressed and any omissions are, however, those of the authors alone. Dr. Hoch is a recipient of the Ontario Ministry of Health and Long Term Care Career Scientist Award and also received funding from the National Sciences and Engineering Research Council of Canada. Dr. Briggs was funded by a Joint UK MRC/Southeast Region Training fellowship.

References

1. Armitage P, Berry G. *Statistical Methods in Medical Research*, (3rd edn). Blackwell Scientific Publications: Oxford, 1994.
2. Jones AM. Health econometrics. In *Handbook of Health Economics*, Newhouse JP, Culyer AJ (eds). Elsevier Science Ltd.: Amsterdam, 2000.
3. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977; **296**: 716–721.
4. Weinstein MC, Fineberg HV. *Clinical Decision Analysis*. W.B. Saunders: Philadelphia, 1980.
5. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press: Oxford, 1987.
6. O'Brien BJ, Drummond MF, Labelle RJ *et al.* In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care* 1994; **32**: 150–163.
7. Chaudhary MA, Stearns, SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat Med* 1996; **15**: 1447–1458.
8. Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Econ* 1997; **6**: 243–252.
9. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ* 1998; **7**: 723–740.
10. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decision Making* 1998; **18** (Special Issue on Pharmacoeconomics): S68–S80.
11. Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *Int J Technol Assessment Health Care* 1998; **14**(3): 467–471.
12. Mullahy J, Manning WG. Statistical issues in cost-effectiveness analyses. In *Valuing Health Care*, Sloan FA (ed.) Cambridge University Press: Cambridge, 1995; 149–184.
13. Goldberger AS. *Econometric Theory*. Wiley: New York, 1964.
14. Manski CF. *Analog Estimation Methods in Econometrics*. Chapman & Hall: New York, 1988.
15. van Hout BA, Malwenn JA, Gordon GS *et al.* Costs, effects, and C/E ratios alongside a clinical trial. *Health Econ* 1994; **3**: 309–319.
16. Löthgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Econ* 2000; **9**: 623–630.
17. Stinnett AA., Paltiel AD. Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means. *Med Decision Making* 1997; **17**: 483–489.
18. Stein LI, Test MA. Alternative to mental hospital treatment: I. Conceptual model, treatment program, and clinical evaluation. *Arch Gen Psychiatry* 1980; **37**: 392–397.
19. Olfson M. Assertive community treatment: an evaluation of the experimental evidence. *Hospital Community Psychiatry* 1990; **41**: 634–641.
20. Burns BJ, Santos AB. Assertive community treatment: an update of randomized trials. *Psychiatr Serv* 1995; **46**: 669–675.
21. Scott JE, Dixon LB. Assertive community treatment and case management for schizophrenia. *Schizophrenia Bull* 1995; **21**: 657–668.
22. Lehman AF, Dixon LB, Kernan E *et al.* A randomized trial of assertive community treatment for homeless persons with severe mental illness. *Arch Gen Psychiatry* 1997; **54**: 1038–1043.
23. Lehman AF, Dixon LB, Hoch JS *et al.* Cost-effectiveness of assertive community treatment for homeless persons with severe mental illness. *Br J Psychiatry* 1999; **174**: 346–352.
24. Newman SJ. The severely mentally ill homeless: housing needs and housing policy. The Johns Hopkins University Institute for Policy Studies. Occasional paper no. 12, 1992.

25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders DSM-III-R* (3rd revised edn). American Psychiatric Association: Washington, DC, 1987.
26. Cook RD, Weisberg S. *Residuals and Influence in Regression*. Chapman & Hall: New York, 1982.
27. Belsley D, Kuh E, Welsch RE. *Regression Diagnostics*. Wiley: New York, 1980.
28. StataCorp. *Stata Statistical Software: Release 6.0*. Stata Corporation: College Station, TX, 1999.
29. 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**: 181–187.
30. Grieve R, Porsdal V, Hutton J, Wolfe C. A comparison of the cost-effectiveness of stroke care provided in London and Copenhagen. *Int J Technol Assessment Health Care* 2000; **16**: 684–695.
31. Chamberlain G. Multivariate regression models for panel data. *J Econometrics* 1982; **18**: 5–46.
32. Hansen L. Large sample properties of generalized method of moments estimators. *Econometrica* 1982; **50**: 1029–1054.