A040 - Covariate adjustment in cost-effectiveness results: an exploration of the performance of proposed methods using data from a randomised controlled trial.

Background and aims

The availability of data collected alongside randomised clinical trials has led to a steady increase in the number of economic evaluations conducted on the basis of patient-level data. At the same time, and as a result of this, a number of uncertainties have emerged, many of which relate to the most appropriate way of analysing the available data. Prominent amongst them are issues around the methods of adjusting cost-effectiveness results for important covariates. Three main approaches to covariate adjustment have emerged in the literature: Ordinary least squares (OLS) regression, seemingly unrelated regression (SUR) and Bayesian regression. This paper aims to discuss these methods and explore how they perform in analysing patient-level cost-effectiveness data.

Methods

OLS regression of net monetary benefits, SUR, and Bayesian regression analysis (with normal and gamma distributed costs) were applied to patient-level data obtained from the TASMINH2 trial. Each of the methods was employed to adjust for two sets of covariates, with the first set including covariates used to stratify the randomisation process in the trial, and the second set consisting both stratification factors and covariates which were correlated with costs and outcomes. The robustness of the estimates was assessed through an examination of the obtained Akaike Information Criterion and standard error of the incremental net monetary benefits for each model.

Results and conclusions

Results suggest that, in situations where the nature of the distribution of costs is not taken into account, SUR perform better than the remaining models. In contrast, when accounting for the skewed distribution of costs, the Bayesian model with gamma distributed costs reports the most robust estimates. With regard to covariate adjustment, using the first set of covariates (i..e. stratification factors only) improved the precision in the parameters and the model fit as compared to no adjustments. Further adjustments using the second set of covariates (i.e. stratification factors and strong predictors of costs and benefits) did not appear to improve the precision of the obtained estimates. This may be due to the relatively low or moderate correlations between the adjusted covariates and the trial endpoints. Further work to explore the generalisability of these findings would be highly useful and we would welcome comments from HESG members for suggestions on taking this research forward.

Theodoros Mantopoulos[1]; Paul Mitchell[2]; Nicky Welton[1]; Richard McManus[3]; Lazaros Andronis[2]

[1] School of Social and Community Medicine, University of Bristol

[2] Health Economics Unit, University of Birmingham

[3] Department of Primary Health Care Sciences, University of Oxford

1. Introduction

Information on health technologies' effectiveness and cost-effectiveness is nowadays used routinely to inform decisions on which treatments to adopt and provide to the population. Such information is provided by economic evaluation studies, such as cost-effectiveness analyses (CEA) which assess the health and economic impact of introducing and using a technology in clinical practice. Key input in such studies is primary evidence on the efficacy and resource use associated with the technology, which is typically captured by applied clinical research

Clinical research of a particular design—randomised controlled trials (RCTs)—have been described as *"the crown jewel"*(Fuchs, 1990 p.673) of clinical research, and they are considered as the best source of primary evidence for assessing and comparing health technologies(NICE, 2008). Given this, it is of no surprise that CEAs are increasingly conducted alongside RCTS. Such studies allow for participants to be randomly allocated in any of the alternative treatments. Owing to this, prognostic factors—such as baseline characteristics—can be balanced appropriately across treatments (Moher *et al.*, 2010). While it is typical in studies looking into the effectiveness of technologies to account for the confounding impact of baseline characteristics on the final outcomes, similar adjustments are more rear in CEA, largely due to difficulty of considering costs and effects jointly (Nixon and Thompson, 2005). In non-randomised studies, it is essential to adjust for baseline covariates, to make appropriate allowance for case-mix differences between patients in the intervention and the control group (Black, 1996), as costs and effects may be influenced by case-mix characteristics. In randomised trials adjustment for baseline covariates has two advantages: firstly, it offers space for discrepancies in patient characteristics between randomised groups and, secondly, it allows for a possibility of achieving precision in the ensuing results (Nixon and Thompson, 2005).

Flexibility in methods for performing overall CEAs is important, in order to allow for any correlation between costs and effects (O'Hagan and Stevens, 2001a) and also because cost data may be skewed (Briggs and Gray, 1998). Such analyses involve two-dimensional cost–effectiveness plane presentations (Black, 1990), depicting the cost difference in one dimension and the outcome difference in the other, while considering their variances and covariance. Another way is to calculate net monetary benefits (NMBs) by combining cost and outcome measures for each individual (Stinnett and Mullahy, 1998; Tambour *et al.*, 1998), or present the inferences as cost–effectiveness acceptability curves (CEACs) (van Hout *et al.*, 1994; Lothgren and Zethraeus, 2000). The former describes a range of inferences depending on the decision maker's 'willingness-to-pay' (WTP) for a unit of health gain (NICE, 2008). Given the data and information available, the CEAC provides the probability of the new intervention being cost-effective for any WTP value (O'Hagan *et al.*, 2000).

The above methods are typical procedures of CEA. However, in recent years different methodological extensions have been proposed in order to handle covariates. While each of these methods present distinct strengths, they are also hindered by limitations. With this in mind, this paper sets out to examine the performance of three different covariate-adjustment methods proposed for CEA: a) ordinary least squares; b) seemingly unrelated regressions and c) Bayesian regression. The exploration was undertaken on the basis of data from a large RCT, the TASMINH2 study aiming to look into the effect of self-managements in hypertensive patients, and made use of different sets of baseline covariates.

The structure of the remaining paper is organised as follows: Section 2 provides relevant literature reviews on selection of covariates and methods for covariate adjustment in economic evaluations alongside RCTs. Section 3 discusses the RCT and the data used in the analysis. In Section 4, the methodology is outlined, while in Section 5 an empirical research is undertaken and the results are presented. Section 6 provides a discussion of the results together with the limitations of the analysis. Finally, in Section 7, conclusions are drawn from the main findings and recommendations for future research are made.

2. Background

2.1 Rationale for covariate adjustment and selection of covariates

The outcomes and costs of a treatment, in practical clinical research, depend both on the treatment itself and on specific characteristics of the patient. In statistics, these characteristics are referred to as covariates. A specific use of covariates is to demonstrate and illustrate differences in costs and efficacy between patients as though the only difference between them was the treatment given (Vazquez-Polo *et al.*, 2005). In RCTs, baseline differences between the treatment groups are avoided through randomisation (Senn, 1989). Incorrect randomisation, however, can lead to imbalances in baseline covariates (Senn, 1989) leading to biased cost-effectiveness estimates (Hoch *et al.*, 2002; Manca *et al.*, 2005a; Willan and Briggs, 2006). So as to avoid imprecision, regression adjustment has been proposed (Hoch *et al.*, 2002; Manca *et al.*, 2005b; Nixon and Thompson, 2005; Willan and Briggs, 2006). This way "corrected" estimates of the effect of exposure are estimated, and theoretically, these estimates are no longer biased by any confounders (McNamee, 2005). Adjustment can reduce variation and give more precise cost-effectiveness estimates, even when covariates are balanced (Willan and Briggs, 2006).

Despite the potential benefits of controlling for baseline covariates in CEA, problems can arise in their selection (Popock *et al.*, 2002). According to Popock *et al.* (2002), baseline characteristics that (i) predict outcome, (ii) are imbalanced between groups and/or (iii) were used to stratify the randomisation, must be

considered. The use of significance testing to compare the distribution of baseline covariates in RCTs has come under some criticism because its purpose is to test whether the observed differences are due to chance, a fact that is already known as treatments are allocated by randomisation (Senn, 1994; Begg, 1990; Altman, 1985; Rothmann, 1977; Altman and Dore, 1990). Some baseline characteristics could be accurate predictive indicators of outcome and can reduce between-patient variability, although they show little difference across treatment allocations statistically (Fayers and King, 2008).

Another consideration is whether the RCT uses stratified randomisation. Stratification refers to the process of creating separate block randomisation lists for important prognostic factors in order to ensure that treatment groups are similar in important baseline characteristics (Akobeng, 2005). For RCTs that use stratified randomisation, it has been suggested that any analysis should take into account stratification factors (Cook and DeMets, 2008).In the context of trial-based CEA, Manca et al. (2005a) highlight the importance of adjusting for baseline utility in the estimation of mean quality-adjusted life-years (QALYs), as imbalance often exists in baseline utility between the treatment groups; also there is a strong correlation between baseline utility and QALYs over the follow-up period. If this imbalance is not controlled, it can result in inappropriate estimates for the parameters of interest (Manca *et al.*, 2005a).

2.2 Methods for CEA and covariate adjustment in RCTs

Description of various methods for covariate adjustment of cost-effectiveness data can be found in literature. Hoch *et al.* (2002) were the first to propose the net benefit framework to conduct cost-effectiveness analysis alongside RCTs. Its linear nature has more attractive statistical properties than the incremental cost-effectiveness ratio (ICER) as it enables the employment of regression methods. The Net Monetary Benefit (NMB) of an intervention can be calculated by subtracting the additional cost from the additional effect valued in monetary terms (Tambour *et al.*, 1998). The decision rule in the net benefit framework is that the new treatment should be adopted over the existing one if the NMB is greater than zero (Hoch *et al.*, 2002).

Willan *et al.* (2004) extended the net benefit regression model proposed by Hoch *et al.* (2002). In their analysis, they considered costs and effects jointly, assuming a bivariate normal distribution, by proposing the use of a system of seemingly unrelated regressions (SUR). SUR is a set of regression equations in which the error terms are assumed to be correlated across the regression equations (Institute for Digital Research and Education, 2013).

Unlike Hoch *et al.* (2002) and Willan *et al.* (2004), Vazquez-Polo *et al.* (2005) adopted a Bayesian framework to provide covariate-adjusted cost-effectiveness estimates. Methods for CEA of clinical trials

using a Bayesian perspective had been previously considered in the CEA literature (Briggs, 1999; O'Hagan and Stevens, 2001a; Chaloner and Rhame, 2001; O'Hagan and Stevens, 2001b) with Vazquez-Polo *et al.* (2005) extending these methods to incorporate covariates in order to reduce uncertainty about the cost-effectiveness parameters of healthcare technologies. In particular, a regression model directly on effects and costs was proposed by including patient characteristics as covariates, and by assuming that effects influence costs.

3. Example: the randomised trial and data

The tele-monitoring and self-management in the control of hypertension (TASMINH2) trial is a randomised control trial carried out in 24 general practices (GPs) in the West Midlands, United Kingdom, that examined if self-management by people with hypertension led to lower levels of blood pressure compared with usual care (McManus *et al.*, 2010). Enrolment criteria were patients aged between 35 and 85 years old with blood pressure more than 140/90 mm Hg despite antihypertensive treatment (McManus *et al.*, 2010). Randomisation was stratified by GP with minimisation for sex, baseline systolic blood pressure, and presence or absence of diabetes or chronic kidney disease (McManus *et al.*, 2010). The main clinical endpoint was change in mean systolic blood pressure between baseline and the two follow-up periods (6 months and 12 months) (McManus *et al.*, 2010). Five hundred twenty seven (n=527) patients were randomised to self-management (n=263) or usual care (n=264), of whom 463 (88%; self-management, n=227; control, n=236) were included in the analysis.

The direct costs to the health service for each individual over the 12-month clinical trial period were estimated as the sum of medication costs, outpatient costs, inpatient costs, referral costs, general practice visit costs, hospital costs, training costs, and equipment costs. Although the principal effectiveness outcome in the RCT was reduction in mean systolic blood pressure, the EQ-5D-3L (The EuroQol Group, 1990; Brooks, 1996) were collected as a health-related quality of life measure from patients. EQ-5D scores are used to calculate quality-adjusted life-years (QALYs) and use them as the effectiveness outcome in the CEA of the TASMINH2 RCT. Patient-level QALYs were calculated using the 'area under the curve' (AUC) approach (Drummond et al. 2005).

4. Methods

Three different methods for covariate adjustment are considered for the CEA of the TASMINH2 RCT: Least squares regression of net benefits, SUR, and a Bayesian regression model. The following notation is used: let c_i and e_i be the costs and effects for the i^{th} individual.

4.1 OLS regression using NMBs

Net benefits can be calculated in order to convert costs and effects to a single variable and then be used in typical regression analyses (Drummond *et al.*, 2005). In a typical linear regression model, we can have net monetary benefit as the response variable and treatment arm together with the covariates of interest as the explanatory variables (Hoch *et al.*, 2002). That is:

$$NMB_i = \alpha + \delta t_i + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \qquad (1)$$

Where, α is an intercept term, t a treatment dummy taking the value zero for the standard treatment and the value one for the new treatment, x are the p covariates of interest, and ε is a stochastic error term (Hoch *et al.*, 2002). The regression coefficient δ represents the incremental net benefit (INB) attributable to the new treatment controlling for covariates, for that WTP level (Hoch *et al.*, 2002). The INB is the difference in the mean NMB of the new treatment and the mean NMB of the standard treatment (Hoch *et al.*, 2002).

4.2 Seemingly unrelated regressions

SUR is a system of different regression equations with error terms that are assumed to be correlated across the equations (Zellner, 1962). Different sets of covariates can be included in each equation; however, in Model (2), the same individual-level covariates are used.

$$c_i = a^c + \delta^c t_i + \beta_1^c x_{i1} + \cdots + \beta_p^c x_{ip} + \varepsilon_i^c$$

$$e_i = a^e + \delta^e t_i + \beta_1^e x_{i1} + \cdots + \beta_p^e x_{ip} + \varepsilon_i^e \begin{pmatrix} \varepsilon_i^c \\ \varepsilon_i^e \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & \rho\sigma_c\sigma_e \\ & \sigma_e^2 \end{pmatrix}\right) \qquad (2)$$

Where, α are the intercept terms in each model, t a treatment dummy taking the value zero for the standard treatment and the value one for the new treatment, x are the p covariates of interest, and ε are the stochastic error terms in each model. The regression coefficient δ^c represents the incremental cost attributable to the new treatment controlling for covariates, and the regression coefficient δ^e represents the incremental effectiveness attributable to new treatment, again, controlling for covariates (Willan *et al.*, 2004). The error terms (ε) are assumed to follow a bivariate normal distribution, with mean zero and variances σ_c^2 and σ_e^2, while ρ represents the correlation between costs and effects, conditional on covariates (Willan *et al.*, 2004).

4.3 Bayesian regression model

Vasquez-Polo *et al.* (2005) described a Bayesian model for covariate adjustment and CEA of patient level data, defined as follows:

$$e_i = \beta_0 + \beta_T y_{1,i} + \beta_2 y_{2,i} + \cdots + \beta_k y_{k,i} + u_i$$

$$c_i = \delta_0 + \delta_T z_{1,i} + \delta_2 z_{2,i} + \cdots + \delta_l z_{l,i} + \delta_{l+1} e_i + v_i \quad (3)$$

where $y_i = (1, y_{1,i}, \ldots, y_{k,i})$ and $z_i = (1, z_{1,l}, \ldots, z_{l,i})$ are, respectively, (k+1) and (l+1) vectors of patient-level covariates, $y_{1,i}$ and $z_{1,i}$ are dummy variables to describe in which treatment the individual has been assigned to (zero for the standard treatment, one for the new treatment). Correlation between costs and effects is allowed by including the term $\delta_{l+1} e_i$ in the above equation. Coefficient β_T is the expected incremental effectiveness attributable to the new treatment, controlling for covariates. To calculate the incremental cost attributable to the new treatment controlling for covariates, we must sum the value of the term δ_T with that of the product of the coefficient of effectiveness and the incremental effectiveness ($\delta_{t+1} * \beta_T$) (Vasquez-Polo *et al.*, 2005). Vectors $\beta = (\beta_0, \ldots, \beta_k)$, $\delta = (\delta_0, \ldots, \delta_t, \delta_{t+1})$ are vectors of unknown coefficients, and u_i, v_i, are error terms, assumed to be independent and normally distributed with mean 0 and precision $\tau_e = 1/\sigma_e^2$ and $\tau_c = 1/\sigma_c^2$, respectively (Vasquez-Polo et al,. 2005).

In this example, equation (3) with two underlying distributions will be considered. The first assumes a normal distribution for the costs (Bayesian normal distributed model) and the second assumes a gamma distribution for the costs (Bayesian gamma distributed model). The choice of the gamma distribution is based on previous findings (Nixon and Thompson, 2004), which suggest that gamma distributions generally fit cost data better than log-normal distributions. Both models will assume a normal distribution for effectiveness.

Assuming equation (4) under a Bayesian framework, the simultaneous prior distribution on parameters β, δ, τ_e, and τ_c must be specified. The normal/gamma form for the base prior is proposed and independence between the coefficients (β, δ) and precision terms (τ_e, τ_c) is assumed (Vasquez-Polo *et al.*, 2005). The prior structure will be:

$$\pi_e(\beta, \tau_e) = \pi_{e,1}(\beta)\pi_{e,2}(\tau_e) \quad (4)$$

$$\pi_c(\delta, \tau_c) = \pi_{c,1}(\delta)\pi_{c,2}(\tau_c) \quad (5)$$

Where $\pi_{e,1}(\beta) \sim N(\beta_0, V_1^{-1})$ and $\pi_{c,1}(\delta) \sim N(\delta_0, V_2^{-1})$

and $\pi_{e,2}(\tau_e) \sim G(\alpha_e, b_e)$ and $\pi_{c,2}(\tau_c) \sim G(\alpha_c, b_c)$

where N and G denote the normal and gamma distribution, respectively, and the parameters β_0, V_1^{-1}, δ_0, V_2^{-1}, (mean and covariance matrix for each distribution) and α_e, b_e, α_c, b_c (scale and shape parameters of the prior gamma density function with which the mean and variance of the prior

distribution can be controlled via expressions α/b and α/b^2, respectively) determining the prior are specified by the expert in the analysis (Vasquez-Polo *et al.*, 2005).

4.4 Model comparison

The different statistical methods described previously are examined in pre-specified analyses: (i) without covariate adjustment; (ii) with adjustment for baseline covariates which were used in the RCT to stratify the randomisation; (iii) with adjustment for both stratification factors and main covariate effects (baseline covariates that are imbalanced and predictive of the endpoints of interest). The comparison of the different statistical models with different covariate-inclusion strategies is done by reporting the Akaike Information Criterion (AIC) (Akaike, 1974) and the Standard Error (SE) of the estimated INB (at WTP=£20000).

OLS regression and SUR models were estimated in STATA software, version 12.1 (StataCorp, College Station, Texas), while the Bayesian models were implemented using MCMC methods in WinBUGS software (Spiegelhalter *et al.*, 1999). For the Bayesian models, two parallel chains were used with different starting values. Posterior distributions for the parameters of interest were derived from 50000 iterations of the Markov chain, after an initial burn-in of 8000 iterations.

5.Results

Table 1 reports the balance of baseline covariates between the two treatment groups, measured as per cent standardised mean differences (SMDs)[1], which is invariant to sample size (Austin, 2009). There is not a pre-specified level of imbalance that should be a concern, but a SMD of more than 10% is considered to be important (Rosenbaum and Rubin, 1985; Austin, 2009). Apart from the SMD, the correlation between each covariate with costs and QALYs for the two treatment groups is reported. As it can be seen, baseline EQ-ED scores were imbalanced (SMD=30.51%), while BMI and coronary kidney disease were slightly imbalanced (SMD=10.83% and SMD=13.86%, respectively). With regard to correlations with endpoints, baseline EQ-5D scores were strongly correlated with QALYs in both treatment groups ($r_{control}$ = 0.77, $r_{intervention}$ = 0.88), while a smaller level of correlation was found between ethnicity and costs in the control group (r = 0.34), BMI and QALYs in the intervention group (r = -0.18), alcohol intake and QALYs in the intervention group (r = 0.29), and coronary kidney disease and costs in the control group (r = 0.22).

It was mentioned in the methods section that the stratification factors used in the TASMINH2 RCT were general practice, baseline systolic blood pressure, sex, and diabetes or coronary kidney disease status.

[1] The formula for calculating the SMD for a continuous covariate (x) is: $SMD_x = \frac{\mu_{x1}-\mu_{x2}}{\sqrt{(var_{x1}+var_{x2})/2}}$, where μ_{x1}, μ_{x2} and var_{x1}, var_{x2} are the means and variances for each group (Gomes *et al.*, 2012).

However, outcomes of patients in different general practices are similar (intra-class correlation coefficient for costs=0.0002; intra-class correlation coefficient for effects=0.006). Moreover, the primary outcome in the RCT was reduction in mean systolic blood pressure, which provides the rationale for adjusting for baseline systolic blood pressure. Since the effectiveness outcome in the CEA of the RCT is QALYs, it is considered more meaningful to adjust for baseline EQ-5D instead of baseline systolic blood pressure. Therefore, the covariates selected to adjust in the analysis, based on stratification factors, are *baseline EQ-5D, sex, and diabetes or coronary kidney disease status*. Furthermore, the findings in Table 1 motivate the need to adjust for the imbalanced and correlated with endpoints covariates. Thus, the covariates selected to adjust in the analysis, based on adjustment for main covariate effects and stratification factors, are baseline *EQ-5D, sex, BMI, diabetes or coronary kidney disease, ethnicity, and alcohol intake*.

Table 1. Baseline characteristics of TASMINH2 RCT. Balance of baseline characteristics, correlation with endpoints

Covariates	Intervention group (n=227)	Control group (n=236)	Standardised difference	Correlation with endpoints	
Baseline EQ-5D scores	0.81 (0.21)	0.88 (0.19)	30.51	$r_{cost\,1} = -0.08$	$r_{cost\,0} = -0.04$
				$r_{QALYs\,1} = 0.87$	$r_{QALYs\,0} = 0.77$
Male, n(%)	105 (46.30%)	108 (45.80%)	1.11	$r_{cost\,1} = -0.15$	$r_{cost\,0} = -0.11$
				$r_{QALYs\,1} = -0.07$	$r_{QALYs\,0} = -0.08$
Age (years), mean (SD)	66.10 (8.73)	65.70 (8.86)	4.34	$r_{cost\,1} = -0.01$	$r_{cost\,0} = 0.07$
				$r_{QALYs\,1} = -0.14$	$r_{QALYs\,0} = -0.01$
systolic blood pressure, mean (SD)	151.96 (11.92)	151.66 (11.88)	2.49	$r_{cost\,1} = 0.05$	$r_{cost\,0} = 0.03$
				$r_{QALYs\,1} = 0.04$	$r_{QALYs\,0} = -0.01$
Ethnicity: White, n(%)	218 (96.00%)	228 (96.60%)	3.14	$r_{cost\,1} = -0.04$	$r_{cost\,0} = 0.34$
				$r_{QALYs\,1} = -0.04$	$r_{QALYs\,0} = -0.09$
Height, mean (SD)	1.66 (0.10)	1.66 (0.09)	0.11	$r_{cost\,1} = 0.09$	$r_{cost\,0} = -0.02$
				$r_{QALYs}\,1 = -0.01$	$r_{QALYs\,0} = 0.05$
Body mass index, mean (SD)	29.46 (5.68)	30.06 (5.47)	10.83	$r_{cost}\,1 = 0.02$	$r_{cost}\,0 = -0.05$
				$r_{QALYs\,1} = -0.18$	$r_{QALYs\,0} = -0.10$
Marital status: Married (n%)	171 (77.00%)	172 (73.00%)	3.71	$r_{cost\,1} = 0.06$	$r_{cost\,0} = -0.05$
				$r_{QALYs\,1} = -0.08$	$r_{QALYs\,0} = -0.12$
Occupation: Managerial, n(%)	96 (42.30%)	106 (44.90%)	5.81	$r_{cost\,1} = -0.06$	$r_{cost\,0} = 0.06$
				$r_{QALYs\,1} = 0.06$	$r_{QALYs\,0} = 0.02$
IMD 2007 score, mean (SD)*	16.85 (13.38)	17.31 (13.88)	3.39	$r_{cost\,1} = -0.02$	$r_{cost\,0} = 0.01$
				$r_{QALYs}\,1 = -0.14$	$r_{QALYs\,0} = -0.11$
Current smoker, n(%)	16 (7.10%)	14 (6.00%)	4.53	$r_{cost\,1} = -0.06$	$r_{cost\,0} = 0.04$
				$r_{QALYs\,1} = 0.03$	$r_{QALYs\,0} = -0.10$
Alcohol intake in last year, n(%)	184 (81.10%)	188 (79.70%)	3.51	$r_{cost\,1} = 0.02$	$r_{cost\,0} = 0.03$
				$r_{QALYs\,1} = 0.29$	$r_{QALYs\,0} = 0.10$
Past medical history:					
Coronary kidney disease, n(%)	16 (7.10%)	26 (11.00%)	13.86	$r_{cost\,1} = 0.00$	$r_{cost\,0} = 0.22$
				$r_{QALYs\,1} = 0.04$	$r_{QALYs\,0} = -0.01$
Diabetes, n(%)	18 (7.90%)	16 (6.80%)	4.39	$r_{cost\,1} = 0.01$	$r_{cost\,0} = 0.03$
				$r_{QALYs\,1} = -0.01$	$r_{QALYs\,0} = -0.06$

* Index of multiple deprivation 2007 score

5.1 OLS regression using NMBs

Assessment of the overall cost-effectiveness of the TASMINH2 trial using NMB regression methods is provided in Table 2. In the model without the inclusion of covariates (Model 1.a) is evident that self-management is not cost-effective at any WTP value.

Table 2: Estimates for Net Benefit regression model

N=463 observations	**Without covariate adjustment (Model 1.a)**	**Adjusted for covariates based on stratification factors (Model 1.b)**	**Adjusted for main covariate effects and stratification factors (Model 1.c)**
	Mean [se]	Mean [se]	Mean [se]
	(95% CIs)	(95% CIs)	(95% CIs)
INB at λ=20000	-934.56 [360.09]	14.08 [217.15]	-12.06 [217.05]
	(-1642.20, -226.92)	(-412.67, 440.82)	(-438.61, 414.50)
ICER	current dominates	17443.1	23055.6
AIC	8966.5	8489.2	8488.8

INB: incremental net benefit; ICER: incremental cost-effectiveness ratio; AIC: Akaike Information Criterion; CIs: Confidence Intervals

The inclusion of covariates based on stratification factors (Model 1.b) has an impact on the estimated INBs. The WTP value above which the INBs become positive is 17443.1, which is the value of the ICER. AIC values show that model fit has improved considerably compared to the previous model. Moreover, this model provides more certain estimates as the SEs of the INB coefficients are smaller, and consequently the Confidence Intervals (CIs) are narrower. Finally, the net benefit estimates adjusted for main covariate effects and stratification factors (Model 1.c) are illustrated. In this case, similar estimates to Model 1.b are obtained; however, the ICER is 23055.6 a finding that indicates that above this value the INB is positive and self-management of hypertension cost-effective. Despite the addition of three baseline covariates, the fit of the model has not improved considerably compared the Model 1.b. More specifically, AIC values are slightly decreased (difference of 0.4) and the uncertainty in the INB estimates has slightly decreased as well (difference of 0.1).

5.2 Seemingly unrelated regressions

To examine if SUR models can produce more efficient estimates compared to OLS regression-based models in the CEA of the TASMINH2 RCT, results are presented in Table 3. Without covariate adjustment (Model 2.a), the resulting INB is -933.57 (CIs: -1625.10, -242.04). With the inclusion of stratification-based covariates (Model 2.b), an ICER of 17428.3 and an INB of 13.97 (CIs: -400.30, 428.24) are calculated. Finally, Model 2.c results in an ICER and INB of 23030 and -11.97 (CIs: -424.65, 400.71), respectively. In terms of the performance of each model, the inclusion of baseline covariates provides a better fit to the data as the AIC reports lower values and the standard errors less uncertainty in the estimated parameters in both cases. Model 2.c performs slightly better compared to Model 2.b with a decrease in AIC of 2.7 units and a reduction in the SE of the INB of 0.8.

Table 3: Estimates for Seemingly Unrelated Regression model

N=463 observations	**Without covariate adjustment (Model 2.a)** Mean [se] (95% CIs)	**Adjusted for covariates based on stratification factors (Model 2.b)** Mean [se] (95% CIs)	**Adjusted for main covariate effects and stratification factors (Model 2.c)** Mean [se] (95% CIs)
Incremental cost	107.57 [72.36]	96.03 [72.06]	90.97 [71.59]
	(-34.25, 249.39)	(-45.20, 237.25)	(-49.35, 231.28)
Incremental QALY	-0.04135 [0.01730]	0.00551 [0.00994]	0.00395 [0.00990]
	(-0.07520, -0.00750)	(-0.01390, 0.02490)	(-0.01546, 0.02335)
INB at λ=20000	-933.57 [352.82]	13.97 [211.36]	-11.97 [210.55]
	(-1625.10, -242.04)	(-400.30, 428.24)	(-424.65, 400.71)
ICER	current dominates	17428.3	23030.4
AIC	7237.6	6711.8	6709.0

INB: incremental net benefit; ICER: incremental cost-effectiveness ratio; AIC: Akaike Information Criterion; CIs: Confidence Intervals

Compared to the results of the net benefit regression analysis, all SUR models give similar results with their respected previous ones. Despite that, SUR models have considerably lower AIC values, indicating a better fit, and generate more certain results. More specifically, the AIC value for Model 2.c is 6709 and the SE of the INB is 210.55, while for Model 1.c the corresponding values are 8488.8 and 217.05. Taken these findings into consideration, it seems that SUR models are more appropriate methodological tools for CEA of individual patient data compared to OLS net benefit regression models.

5.3 Bayesian regression models

Results for the Bayesian model with the normal distribution for costs and effects are presented in Table 4 as posterior means, standard errors of the means, and 95% credible intervals (CrIs). In the model without inclusion of baseline covariates (Model 3.a), the INB, at WTP=20000, is equal to -939.20 (CrIs: -1625.00, -251.40). The inclusion of stratification-based covariates in the Bayesian model (Model 3.b) improves the model fit and the uncertainty in the estimated results as evidenced in the smaller AIC values and the smaller SEs (and narrower CrIs). Under this model, the INB is 3.96 (CrIs: -410.90, 417.10), and the ICER is 19354.8. Therefore, for a WTP value of more than 19354.8, self-management becomes cost-effective as the probability of having a positive INB is more than 50%. Finally, the inclusion of covariates based on main covariate effects and statistical factors in the Bayesian model (Model 3.c), does not provide better and more certain estimates, as the AIC of the model and the SEs of the parameters are larger than these of Model 3.b. This finding is somewhat unexpected as in the previous two models the inclusion of both sets of covariates provided, even slightly, better and more certain results. The estimated ICER value is 20674.9.

Table 4: Estimates for Bayesian model with normal distributed costs

N=463 observations	Without covariate adjustment (Model 3.a) Mean [se] (95% CrIs)	Adjusted for covariates based on stratification factors (Model 3.b) Mean [se] (95% CrIs)	Adjusted for main covariate effects and stratification factors (Model 3.c) Mean [se] (95% CrIs)
Incremental cost	113.40 [57.59]	120.00 [56.99]	94.05 [57.89]
	(0.21, 226.20)	(7.97, 231.60)	(-19.27, 207.10)
Incremental QALY	-0.04129 [0.01730]	0.00620 [0.01020]	0.00455 [0.01018]
	(-0.07530, -0.00728)	(-0.01381, 0.02617)	(-0.01548, 0.02450)
INB at λ=20000	-939.20 [349.20]	3.96 [211.40]	3.07 [211.90]
	(-1625.00, -251.40)	(-410.90, 417.10)	(-421.20, 412.40)
ICER	current dominates	19354.8	20674.9
AIC	7256.0	6758.0	6766.0

INB: incremental net benefit; ICER: incremental cost-effectiveness ratio; AIC: Akaike Information Criterion; CrIs: Credible Intervals

In comparison with the performance of the SUR models, Bayesian models with normal distributions for costs and effects underperform. That is, all three models report higher AIC values compared to their respective SUR values (difference of 18.4, 46.2, and 57 units in Models a, b, and c, respectively) and Model 3.c reports slightly higher SEs in the INB estimates compared to Model 2.c (difference of 1.35). Therefore, SUR models can obtain better and (in some cases) more certain cost-effectiveness estimates in comparison with the normal distributed Bayesian model.

Table 5 presents the findings of the Bayesian model with gamma distribution for costs and normal for effects. With regard to the model without covariates (Model 4.a), again, similar estimates with the previous models are obtained, as the INB is estimated at -940.40 (CrIs: -1624.00, -259.90). Similar to the pattern of the Bayesian model with normal distributions, the inclusion of covariates based on stratification factors (Model 4.b) provides better fit to the data (AIC=5766) and less uncertainty in the estimated INB (mean: -5.48; CrIs: -413.20, 398.70), while the inclusion of both sets of covariates (Model 4.c) does not have the same effect as, compared to Model 4.b, higher AIC values (5780) and a slightly higher SE (207.40) for the INB (mean: -30.56; CrIs: -435.1, 374.8) are obtained.

Table 5: Estimates for Bayesian model with gamma distributed costs

N=463 observations	Without covariate adjustment (Model 4.a) Mean [se] (95% CrIs)	Adjusted for covariates based on stratification factors (Model 4.b) Mean [se] (95% CrIs)	Adjusted for main covariate effects and stratification factors (Model 4.c) Mean [se] (95% CrIs)
Incremental cost	111.60 [32.36]	129.30 [30.39]	120.80 [31.14]
	(50.74, 176.20)	(71.28, 189.70)	(60.55, 182.70)
Incremental QALY	-0.04144 [0.01730]	0.00619 [0.01019]	0.00451 [0.01022]
	(-0.07540, -0.00759)	(-0.01385, 0.02609)	(-0.01549, 0.02457)
INB at λ=20000	-940.40 [348.00]	-5.48 [206.40]	-30.56 [207.40]
	(-1624.00, -259.90)	(-413.20, 398.70)	(-435.10, 374.80)
ICER	current dominates	20885.2	26773.1
AIC	6301.0	5766.0	5780.0

INB: incremental net benefit; ICER: incremental cost-effectiveness ratio; AIC: Akaike Information Criterion; CrIs: Credible Intervals

The gamma distribution assigned to costs clearly fits the data better and produces more certain estimates. More specifically, the AIC values for all three Bayesian gamma distributed models are substantially reduced

compared to both the SUR and the normal distributed Bayesian models, and the SEs of INBs are reduced as well, driven from the reduction in the SEs of the cost estimates. Therefore, the Bayesian model with the gamma distribution for costs achieve the best performance overall, by taking into account the extremely skewed distribution of costs.

6. Discussion

The increasing availability of patient-level data over the past years has led to a growing interest in undertaking economic evaluations of health care technologies alongside RCTs. This fact has led to the development of statistical methodologies in order to produce more robust and certain estimates, and therefore support the decision making process. Regression-based methods can improve the precision of the resulting estimates by adjusting for baseline characteristics. In this study, three proposed methods (OLS regression of NMBs, SUR, and Bayesian linear regression) were applied to the CEA of the TASMINH2 RCT. Moreover, in each model different sets of covariates were included in order to identify an optimal methodology for the CEA of patient level data.

The findings suggest that both cost-effectiveness estimates and the uncertainty in these estimates can differ according to the selected methodology. OLS regression of NMBs proposed by Hoch *et al.* (2002) achieved the worst performance overall, due to the poor data fit and higher uncertainty in the estimated results. Without taking into consideration the skewed distribution of costs, SUR models, proposed by Willan *et al.* (2004), performed better than the remaining models by recognising the correlation between costs and effects in the parameter estimation. Therefore, in circumstances that the distribution of costs is approximately normal, SUR models offer advantages and should be preferred for the CEA of RCTs. However, it is known that the distribution of costs usually exhibits a high degree of skewness (Altman, 1985). In such a case, the Bayesian model with gamma distributed costs should be adopted, as the model fit is considerably improved and the SEs of the incremental net benefit estimates are reduced at the same time. This finding is in agreement with Nixon and Thompson's (2005) findings, while it contradicts other findings according to which methods that assume normality are reasonably robust to skewed cost data (Willan *et al.*, 2004; Nixon *et al.*, 2010).

With regard to covariate adjustment, the findings were mixed. The inclusion of patient-level covariates that were used to stratify randomisation improved the precision and the certainty in the results of the CEA in all models; this verifies the findings of Vazquez-Polo *et al.* (2005) and their work with simulated data. The findings are also in line with previous literature (Cook and DeMets, 2008; Pocock *et al.*, 2002), which suggests that in randomised trials that a stratified randomisation was followed, the stratification factors should be considered in the analysis.

However, contradictory results were obtained when the models were adjusted for main covariate effects and stratification factors. In particular, net benefit regression and SUR models reported a slightly improved model fit and marginally reduced SEs compared to adjustment only for stratification factors, while for the Bayesian models that was not the case. In the CEA of the TASMINH2 RCT, controlling for main covariate effects was found to be inadequate. However, this may be the case generally in CEA of individual-level data (Gomes *et al.*, 2012). Here, the relatively low/moderate correlation between the adjusted covariates (except from baseline EQ-5D scores) and the endpoints may be a reason for the unconvincing results. This finding confirms Pocock *et al.* (2002) and Fayers and King (2008), who suggested that a weak correlation between a covariate and an outcome makes adjustment for that covariate irrelevant. Another factor contributing to inadequate findings can be the different prognostic relationship between treatment groups, as covariates were correlated with costs or outcomes only in one treatment group.

A number of limitations do exist in this study. A starting point of concern can be the distribution of effectiveness data. QALYs exhibited a certain degree of skewness, and thus, the normal distribution provided a poor fit. The Bayesian model enables to select a distribution for the data at hand. A commonly used distribution for QALYs, in case the time horizon of the intervention is no more than one year, is the beta distribution, ranging from zero to one (Sculpher, 2004). Nevertheless, a negative QALY value in the dataset prohibits using the beta distribution and improving the model fit. Another limitation is with regard to the fact that CEA results can be sensitive to choice of parametric model (Briggs *et al.*, 2005). Finally, the Bayesian models examined assume that costs do not influence the measure of effectiveness. Yet, there may be situations where this assumption might not hold (Vasquez-Polo *et al.*, 2005).

7. Conclusions

This study examined the relative performance of three previously proposed methods for conducting CEA alongside RCTs and incorporating covariates in the analysis. These were a net benefit regression model proposed by Hoch *et al.* (2002), a SUR model proposed by Willan *et al.* (2004), and a Bayesian model proposed by Vasquez-Polo *et al.* (2005) were compared using data from the TASMINH2 RCT. The results indicated that SUR models perform better in circumstances where the distribution of costs is approximately normal, while in cases of skewed cost data a Bayesian model with gamma distributed costs can provide more robust and precise estimates. With regard to covariate adjustment, all unadjusted models reported poor data fit and high uncertainty in the INB estimates. Adjusting for both main covariate effects and stratification factors provided unconvincing results, implying that adjustment only for baseline characteristics that were used to stratify randomisation in the RCT can provide appropriate inferences. This finding may be due to the moderate correlations between the adjusted covariates and the endpoints.

Clearly, a central consideration to any research is the generalisability of results. Thus, further research is recommended to look into these methods using data from other RCTs and from simulations to ensure the results obtained here are valid and generalisable in different settings. In conclusion, this application showed that methods based on Bayesian approaches offer an attractive alternative for CEA of RCTs and covariate selection. On this basis, it is thought that the use of such methods in economic evaluations of healthcare technologies warrants more attention.

References

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19 (6): 716–723.

Akobeng, A., K. (2005) Understanding randomised controlled trials. Archives of disease in childhood, 90: 840-844.

Altman, D., G. (1985) Comparability of randomised groups. Statistician, 34(1): 125–136.

Altman, D., T., Dore, C., J. (1990) Randomisation and baseline comparisons in clinical trials. Lancet, 335: 149-153.

Austin, P., C. (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine, 28: 3083-3107.

Begg, C., B. (1990) Significance tests of covariate imbalance in clinical trials. Control Clin Trials, 11: 223-225.

Black, W., C. (1990) The CE plane: a graphic representation of cost–effectiveness. Medical Decision Making, 10: 212–214.

Black, N. (1996) Why we need observational studies to evaluate the effectiveness of health care. British Medical Journal, 312: 1215–1218.

Briggs, A., H. (1999) A Bayesian approach to stochastic cost-effectiveness analysis. Health Economics, 8: 257–261.

Briggs, A., Gray, A. (1998) The distribution of health care costs and their statistical analysis for economic evaluation. Journal of Health Services Research and Policy, 3: 233–245.

Briggs, A., Nixon, R., Dixon, S., Thompson, S. (2005) Parametric modelling of cost data: some simulation evidence. Health Economics, 14: 421 - 428.

Brooks, R. (1996) EuroQol: the current state of play. Health Policy 37(1):53-72

Chaloner, K., Rhame, F., S. (2001) Quantifying and documenting prior beliefs in clinical trials. Statistics in Medicine, 20: 581–600.

Cook, T., D., DeMets, D., L. (2008) Introduction to statistical methods for clinical trials. Boca Raton, FL: Chapman & Hall/CRC.

Drummond, M., F., Sculpher, M., J., Torrance, G., W., O'Brien, B., J., Stoddart, G., L. (2005) Methods for the economic evaluation of health care programmes, third ed. Oxford University Press, Oxford.

Fayers, P., King, M. (2008) The baseline characteristics did not differ significantly. Quality in Life Research, 17: 1047-1048.

Gomes, M., Grieve, R., Nixon, R., Edmond S., W., Carpenter, J., Thompson, S., G. (2012) Methods for covariate adjustment in cost-effectiveness analysis that use cluster randomised trials. Health Economics, 21: 1101–1118.

Hoch, J., Briggs, A., 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, 11: 415– 430.

Institute for Digital Research and Education (2013) What is seemingly unrelated regression and how can I perform it in Stata? [Online]. Available from: http://www.ats.ucla.edu/stat/stata/faq/sureg.htm [Accessed 2 August 2013].

Lothgren, M., Zethraeus, N. (2000) Definition, interpretation and calculation of cost–effectiveness acceptability curves. Health Economics, 9: 623–630.

Manca, A., Hawkins, N., Sculpher, M. (2005a) Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Economics, 14: 487–496.

Manca, A., Rice, N., Sculpher, M., J., Briggs, A., H. (2005b) Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. Health Economics, 14: 471–485.

Matthews, J., N., S., Altman, D., Campbell, M., J. (1990) Analysis of serial measurements in medical research. British Medical Journal, 300: 230-235.

McManus, R., J., Mant, J., Bray, E., P., Holder, R., Jones, M., I., Greenfield, S., Kaambwa, B., Banting, M., Bryan, S., Little, P., Williams, B., Hobbs, F., D., R. (2010) Telemonitoring and self-management in the control of hypertension (TASMINH2): a randomised controlled trial. The Lancet, 376(9736): 163-172.

Moher, D., Hopewell, S., Schulz, K., F., Montori, V., Gøtzsche, P., C., Devereaux, P., J., Elbourne, D., Egger, M., Altman, D., G. (2010) ConSoRT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ, 340: 869-897.

National Institute for Health and Clinical Excellence (NICE). (2008) Guide to the methods of Technology Appraisal 2008 [Online]. Available from: http://www.nice.org.uk/ [Accessed 23 July 2013].

Nixon, R., M., Thompson, S., G. (2004) Parametric modelling of cost data in medical studies. Statistics in Medicine, 23: 1311–1331.

Nixon, R., M., Thompson, S., G. (2005) Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. Health Economics, 14: 1217–1229.

Nixon, R., M., Wonderling, D., Grieve, R., D. (2010) Non-parametric methods for cost-effectiveness analysis: the central limit theorem and the bootstrap compared. Health Economics, 19: 316-333.

O'Hagan, A., Stevens, J., W., Montmartin, J. (2000) Inference for the C/E acceptability curve and C/E ratio. PharmacoEconomics, 17: 339-349.

O'Hagan, A., Stevens, J., W. (2001a) A framework for cost-effectiveness analysis from clinical trial data. Health Economics, 10: 303–315.

O'Hagan, A., Stevens, J., W. (2001b) Bayesian assessment of sample size for clinical trials of cost-effectiveness. Medical Decision Making, 21: 219–230.

Pocock, S., J., Assmann, S., E., Enos, L., E., Kasten, L., E. (2002) Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statistics in Medicine, 21: 2917-2930.

Rosenbaum, P., R., Rubin, D., B. (1985) Constructing a control-group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician, 39: 33–38.

Rothman, K., J. (1977) Epidemiologic methods in clinical trials. Cancer, 39:1771-1775.

Sculpher, M., J. (2004) The use of probabilistic sensitivity analysis for decision making: the example of drug-eluting stents [Online]. Available from: http://www.york.ac.uk/che/pdf/teehtacosteff04.pdf [Accessed 15 August 2013].

Senn, S., J. (1989) Covariate imbalance and random allocation in clinical trials. Statistics in Medicine, 8: 467-475.

Senn, S., J. (1994) Testing for baseline balance in clinical trials. Statistics in Medicine, 13: 1715-1726.

Spiegelhalter, D., J., Thomas, A., Best, N. (1999) WinBUGS, version 1.2. MRC Biostatistics Unit: Cambridge, UK.

StataCorp: Stata Statistical SoFtware: Release 12 College Station, TX; 2011.

Stinnett, A., A., Mullahy, J. (1998) Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making, 18(suppl.): S68–S80.

Tambour, M., Zethraeus, N., Johannesson, M. (1998) A note on confidence intervals in cost-effectiveness analysis. International Journal of Technology Assessment in Health Care 1998, 14(3): 467–471.

The EuroQol Group (1990) EuroQol-a new facility for the measurement of health-related quality of life. Health Policy 16(3):199-208

van Hout, B., A., Al, M., J., Gordon, G., S., Rutten, F. (1994) Costs, effects and C/E ratios alongside a clinical trial. Health Economics, 3: 309–319.

Vazquez-Polo, F.J., Negrin, M., Gonzalez, B. (2005) Using covariates to reduce uncertainty in the economic evaluation of clinical trial data. Health Economics, 14: 545–557.

Willan, A., Briggs, A. (2006) Statistical Analysis of Cost-Effectiveness Data. Wiley: Chichester, UK.

Willan, A., R., Briggs, A., H., Hoch, J., S. (2004) Regression methods for covariate adjustment and subgroup analysis for non-censored cost–effectiveness data. Health Economics, 13: 461–475.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. Journal of the American Statistical Association, 57: 348–368.