Monograph:
Hernández Alava, M. and Wailoo, A. (2010) Multilevel modelling of cost data: an application to thrombolysis and primary angioplasty in the UK NHS. Discussion Paper. (Unpublished)

HEDS Discussion Paper 10/06

HEDS Discussion Paper 10/06


ScHARR
School of Health and
Related Research

Multilevel modelling of cost data: an application to thrombolysis and primary angioplasty in the UK NHS.

Mónica Hernández Alava and Allan J Wailoo
Health Economics & Decision Science, University of Sheffield

January 2010

ABSTRACT: Cost data are frequently collected from several locations and tend to be non negative and skewed. Generalised linear multilevel models provide a means of dealing with each of these issues. This paper compares several statistical models within this class using data drawn from an observational study of 3,000 patients treated for heart attack in 15 UK NHS hospitals. A number of alternative link functions and covariates were considered. We demonstrate that whilst it is important to take account of clustering in the data, the precise manner in which this is done is equally important. Models which allow for correlation between the random effects components and heteroskedasticity across all hospitals performed best in terms of model fit and made substantial differences to cost estimates.

KEYWORDS: Multilevel model; Cost data; Generalised linear latent and mixed model; heart attack; angioplasty.

JEL CLASSIFICATION: I10, C13, C16.

ADDRESS FOR CORRESPONDENCE: Allan Wailoo. Health Economics and Decision Science. School of Health and Related Research. University of Sheffield. Regent Court. 30 Regent Street Sheffield S1 4DA. Email: a.j.wailoo@sheffield.ac.uk. Tel: 00 44 114 2220729

* We are grateful to Mark DeBelder for access to the NIAP data and to all the other participating hospitals. All responsibility for the analysis and interpretation of the data presented here lies with the authors.

1 Introduction

It is a requirement of cost effectiveness analysis that estimates are based on mean costs and benefits since this is most relevant to informing decision makers of the true expected costs and benefits of adopting a particular intervention across a population. Typically however, cost data are non negative and highly skewed. It may therefore be inappropriate to employ standard methods of analysis which rely on the assumption of conditional normality in seeking to estimate mean costs conditional on treatment or other characteristics. There are two broad approaches that can be used to attempt to overcome the challenge of skewness. Transformations, such as taking the logarithm of cost, which produce a more normal distribution can be undertaken, with standard methods then conducted on the transformed data. This approach raises additional difficulties however since it is not a straightforward task to back transform the estimates to the original cost scale without introducing bias (Duan (1983)). More recently, approaches using Generalized Linear Models (GLMs) have been conducted in the health economics literature (see for example Blough and Ramsey (2000) Glick (2007)). The GLM approach is more attractive since costs can be analysed on their original scale directly. This is achieved by allowing for a variety of distributions, not necessarily normal, to be specified for costs. These two approaches have been compared by Manning and Mullahy (2001).

An additional complication is that many studies are conducted in multiple settings. This is particularly true of randomized controlled trials which are often multicentre giving rise to natural clusters within the data. Such clustering ought not be ignored. Observations from the same cluster will not be independent and failure to incorporate this may lead to misleading results. Whilst standard GLM models assume that all the observations in the sample are independent, multilevel modelling techniques (Goldstein (2005)), also referred to as hierarchical or mixed effects models, can be used to account for clustering in the data. Multilevel models are constructed by allowing one or more of the model parameters to vary across centres

and there is a small but growing literature on their use in health economics. Most of these applications have used a normal distribution to model the dependent variable (Or et al. (2004), Willan et al. (2005), Manca et al. (2005), Manca et al. (2007)) or its logarithm (Carey (2000)) but there are also some applications where GLM and multilevel models (GLMMs) have been combined to simultaneously account for both non normal distributions and clustering in the data (Burgess et al. (2000), Grieve et al. (2005), Gauthier et al. (2009), Thompson et al. (2006), Grieve et al (2009), Willan and Kowgier (2008)). Liu et al. (2009) and Cooper et al. (2007) demonstrate how implementation of GLMMs as part of a two part (also known as hurdle) modelling framework may be valuable in situations where there are a significant proportion of zero cost observations. The first part of the model is used to predict whether a patient incurs a positive cost of not, with a GLMM then fitted to the second part for patients that do.

Few of these GLMM applications allow for heterogeneity in more than one model parameter. In most cases, where the primary interest is in the effect of different treatments on costs, a single random effect is used either for the intercept or the active treatment covariate but not both. An exception is Thompson et al. (2006) who examined the costs of stroke patients in 13 international centres and the impact of incontinence, a marker of stroke severity, on those estimates. Heterogeneity is reflected in both parameters although the fact that these are treated as independent may itself be seen as a limitation. In many situations, it may be overly simplistic to have such limitations on the random structure of the model. For example, the units which give rise to the multilevel structure of the data such as countries or hospitals may well be expected to differ in terms of the costs of both control and active treatment. There are also many situations in which our interest extends to multiple treatments and their associated costs. Furthermore it is inevitable that these random parameters are themselves correlated and it would be appropriate for any cost model to reflect this in order to avoid inappropriate estimates.

This paper illustrates the application of such models. We aim to develop the GLMM framework for the analysis of cost data further by allowing for heterogeneity in multiple parameters whilst also incorporating a full correlation matrix in the random component of the model. We compare these models with standard GLMs and a GLMM without such correlations. These issues are demonstrated using cost data for UK NHS hospital heart attack patients between April 2005 and March 2006 treated either by thrombolytic (clot-busting) drugs or primary percutaneous coronary intervention (primary angioplasty or PPCI). We therefore have three random treatment parameters unlike many previous applications based on two arm RCTs. Analyses are performed using maximum likelihood estimation methods using STATA software, in contrast to almost all other applications we are aware of that have used Monte Carlo Markov Chain (MCMC) simulation methods.

Section 2 of the paper provides brief details of the study which provided the patient level data used in the analyses. Section 3 presents the GLM and GLMM models. Results for all models are provided in Section 4 whilst Section 5 discusses those results and concludes.

2 Data

Current standard treatment for acute myocardial infarction (heart attack) centres on rapid, clot-busting (thrombolytic) drug treatment. However, evidence from a number of trials dating from the early 1990's suggests that primary angioplasty may be more beneficial in terms of mortality, reinfarction and stroke (Asseburg et al. (2007)). It has been estimated that primary angioplasty represents a cost effective treatment in the UK NHS setting provided the additional time delay to treatment does not exceed approximately one hour compared to thrombolysis (Bravo Vergel et al. (2007)). However, these estimates were based on several assumptions about likely resource use. Given the organisational challenges of providing a 24

hour angioplasty service, the UK NHS set up a pilot scheme, the National Infarct Angioplasty Project (NIAP), in order to investigate the feasibility of a national roll-out of a comprehensive angioplasty service. One of the aims of the pilot was to gather resource use data to allow real-life estimates of cost to be made which in turn could be used to make updated estimates of cost-effectiveness (Goodacre (2008), Wailoo et al. (2009)).

Ten hospitals providing primary angioplasty became NIAP pilot sites and collected detailed data on the initial treatment episode and follow up information to one year for all ST - segment elevation myocardial infarction (STEMI) patients admitted between 1st April 2005 and 31st March 2006. These hospitals varied in terms of the times of day and the days of the week primary angioplasty was provided (not all offered 24/7 provision during the study period) and whether they had arrangements with local hospitals unable to provide angioplasty for patients to be transferred for primary angioplasty. These hospitals provided information on 2083 patients. As part of NIAP, one hospital that did not provide primary angioplasty also collected data for the study period. We supplemented this control site information by obtaining equivalent data for the same time period from four other hospitals who did not provide primary angioplasty for their STEMI patients. A total of 919 control patients were included.

Data were collected on treatment received, drugs administered, consumables including the number, type and make of stent, tests, length of stay, staffing of the catheter laboratories and ambulance journeys. These were costed at 2006/7 £ sterling values.

These data come from an observational study, not a randomised controlled trial. Whilst the study offers distinct advantages over RCTs because of the real life setting, differences between the patients at different hospitals and receiving different treatments is not considered in the design and is therefore a genuine concern that may be less relevant to trials. Therefore, data on a wide range of patient characteristics

were included as potential covariates to the cost analysis. Table 1 provides details of the main characteristics of the sample by NIAP and control hospitals. Statistically significant differences can be seen in several dimensions. The patients treated in NIAP hospitals were slightly younger (64 vs 66yrs), more likely to be male (71% vs 67%), more ethnically diverse with 81% caucasian compared to 98% in the control hospitals, had a greater prevalence of previous heart disease such as angina (22% vs 16%), previous percutaneous coronary intervention (PCI)(9% vs 3%) or previous coronary artery bypass graft (CABG)(4% vs 1%). There was also evidence that the NIAP treated patients had a greater prevalence of other comorbidities.

Table 1: Characteristics of patients

		NIAP sites (n=2083)		Control sites (n=919)	
		mean	std err	mean	std err
Mean age in years		63.5***	0.31	66.1***	0.44
MI severity	Mean peak troponin	32.3***	3.8	19.9***	0.71
		%	n	%	n
Ethnic group	Caucasian	80.5***	1407	98.0***	624
	Black	2.9	50	0.3	2
	Asian	11	192	1.6	10
	Oriental	0.5	9	0.2	1
Previous CHD	AMI	15.9	326	15	136
	Angina	21.8***	441	15.9***	76
	Previous PCI	8.6***	176	2.6***	23
	Previous CABG	3.8***	78	1.3***	10
Comorbidities	Hypertension	45.9***	931	39.3***	327
	Hypercholesterolaemia	40.3	789	43.3	343
	Peripheral vascular disease	4.3***	87	2.0***	12
	Cerebrovascular disease	6.4	129	5.1	44
	Asthma or COPD	11.8	238	11.4	97
	Chronic renal failure	2.5	50	2.1	19
	Diabetes	16.6***	339	11.17***	101
Left Ventricular Ejection Function	Good	55.7	547	58.3	151
	Moderate	33.3	327	34.36	89
	Poor	11.1	109	7.34	19
Male		71.1**	1479	66.5**	611

Notes: *** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level.

AMI - Acute Myocardial Infarction, CHD - Coronary Heart Disease, COPD - Chronic Obstructive Pulmonary Disease

Figure 1 displays histograms to demonstrate the distribution of the data for the individual hospitals and overall. It can be seen that there are substantial deviations from normality in the cost distributions both at the aggregate and individual hospital levels. In addition, the degree of skewness and kurtosis differs substantially by hospital. For example, the distribution for hospital 3 is leptokurtic with data in the right tail extending to around £15,000. The distribution for hospital 5 is more platykurtic but still contains data in the tail extending to the same range as hospital

Hospital 1
Hospital 2
Hospital 3
Hospital 4
Hospital 5
Hospital 6
Hospital 7
Hospital 8
Hospital 9
Hospital 10
Hospital 11
Hospital 12
Hospital 13
Hospital 14
Hospital 15
Total
Density
0
5000
10000
15000

3. The right skew is less pronounced for hospitals 6 and 13, with few observations exceeding £10,000. The mean cost of the treatment episode ranges from £2,408 to £5,329 across the hospitals.

It is therefore clear that the use of a normal distribution to model this cost data is unlikely to be appropriate but, in addition, the data may be more fully represented by approaches which permit the shape of the chosen distribution to vary across hospitals. No individual patient had a zero cost.Two part models are therefore not required in this situation.

3 Model specification

The focus of the paper is to estimate the impact of "no treatment", "thrombolysis" and "PPCI" on the cost of the initial treatment episode. We control for a number of covariates at both the individual and the hospital level to try to isolate as much as possible the effect on costs of these three variables. The GLMs presented here use a gamma distribution with an identity link. Other specifications, including different families and links, were also estimated but the gamma distribution with additive effects of the covariates on the mean was found to fit the data better according to both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

All multilevel models have been estimated using the program GLLAMM written in STATA by Rabe-Hesketh et al. (2004). This program uses maximum likelihood estimation. The marginal log-likelihood is obtained by using adaptive Gauss-Hermite quadrature to integrate out numerically the latent variables.

The estimation of four models is reported in the next section. These four models are described below in order of complexity.

Model 1: *Generalised Linear Model.* This is the simplest model where costs are modelled in a GLM framework. It ignores the multilevel structure of the data and

is only estimated for comparison purposes. Let c_{ij} be the initial cost of treatment of patient i ($i = 1$ to N) in hospital j ($j = 1$ to 15). Using a gamma distribution with mean μ_i and shape ρ and an identity link, the model for c_{ij} can then be written as:

$$c_{ij} \tilde{} \Gamma(\mu_i, \rho), \quad \mu_i = \sum_{s=1}^{3} \beta_s x_{ijs} + \beta_4 d_{ij} x_{ij1} + \beta_5 d_{ij} x_{ij2} + \beta_6 t_{ij} x_{ij3} + \sum_k \alpha_k z_{ijk}$$

where x_{ij1}, x_{ij2} and x_{ij3} are dummy variables for "no treatment", "thrombolysis" and "PPCI" respectively, z_{ijk} represent different covariates ($k = 1$ to 8) to control for differences across individuals, d_{ij} is a dummy variable equal to one if patient i went to a NIAP centre and t_{ij} is a dummy variable equal to one if the PPCI was performed out of hours. In this set-up β_4 and β_5 represent a differential effect on the initial cost of treatment between NIAP and control hospitals and β_6 is the differential effect of out of hours PPCI.

The model assumes that the effects of "no treatment" (β_1), "thrombolysis" (β_2) and "PPCI" (β_3) on costs are the same across all the hospitals in our sample. This assumption is relaxed in the following four models.

Model 2: *Generalised Linear Multilevel Model with independent random effects and level 1 clustering (heteroskedasticity).* This model generalises Model 1 by taking into account the multilevel structure of the data. Patients are grouped within hospitals which differ in many organisational and geographical factors which are likely to have a substantial bearing on overall cost. Costs within hospitals tend to vary less than costs for patients chosen at random from different hospitals. Ignoring this structure in the data may lead to making inappropriate inferences. The initial cost of the treatment for patient i in hospital j is modelled using a gamma distribution as before but now the coefficients of "no treatment", "thrombolysis" and "PPCI" are no longer the same across hospitals. It is assumed that these three coefficients vary randomly across hospitals, each one of them following its own normal distribution with its own mean and variance. This model also allows the "shape" of the gamma

distribution to differ across NIAP (ρ_1) and control hospitals (ρ_2). This extended model can be written as:

$$c_{ij} \tilde{} \Gamma\left(\mu_{ij}, \rho_r\right), \quad \mu_{ij} = \sum_{s=1}^{3} \beta_{sj} x_{ijs} + \sum_k \alpha_k z_{ijk}, \quad r = 1, 2.$$

$$\begin{pmatrix} \beta_{1j} \\ \beta_{2j} \\ \beta_{3j} \end{pmatrix} \tilde{} N \left(\begin{pmatrix} \beta_1 + \beta_4 d_{ij} \\ \beta_2 + \beta_5 d_{ij} \\ \beta_3 + \beta_6 t_{ij} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \right)$$

It is important to note that this model recognises explicitly the fact that d_{ij} and t_{ij} relate to effects at the hospital level and not at the individual level.

The parameters to be estimated in this model can be divided into those corresponding to the fixed component of the model (β_1 to β_6 and α_1 to α_8) and those corresponding to the random component (σ_1^2, σ_2^2 and σ_3^2). This can be easily seen by writing the random coefficients as:

$$\begin{aligned} \beta_{1j} &= \beta_1 + \beta_4 d_{ij} + u_{1j} \qquad & u_{1j} \tilde{} N\left(0, \sigma_1^2\right) \\ \beta_{2j} &= \beta_2 + \beta_5 d_{ij} + u_{2j} \qquad & u_{2j} \tilde{} N\left(0, \sigma_2^2\right) \\ \beta_{3j} &= \beta_3 + \beta_6 t_{ij} + u_{3j} \qquad & u_{3j} \tilde{} N\left(0, \sigma_3^2\right) \end{aligned}$$

and substituting these equations into the mean of the gamma distribution

$$\mu_{ij} = \sum_{s=1}^{3} \beta_{sj} x_{ijs} + \beta_4 d_{ij} x_{ij1} + \beta_5 d_{ij} x_{ij2} + \beta_6 t_{ij} x_{ij3} + \sum_k \alpha_k z_{ijk} + \sum_{s=1}^{3} u_{sj} x_{ijs}$$

Model 3: *Generalised Linear Multilevel Model with correlated random effects and level 1 clustering.* This model relaxes the assumption of model 2 that the random effects are uncorrelated. This assumption might be an appropriate assumption to make in some cases but may be considered particularly strong in this example. Within the same hospital, the costs of treating patients by angioplasty or thrombol-

ysis and the costs of no treatment would each be likely to be correlated. Therefore, in model 3 the three random effects are assumed to follow a joint normal distribution with means identical to model 2 but allowing for non-zero covariances.

$$c_{ij} \tilde{} \Gamma\left(\mu_{ij}, \rho_r\right), \quad \mu_{ij} = \sum_{s=1}^{3} \beta_{sj} x_{ijs} + \sum_k \alpha_k z_{ijk}, \quad r = 1, 2$$

$$\begin{pmatrix} \beta_{1j} \\ \beta_{2j} \\ \beta_{3j} \end{pmatrix} \tilde{} N \left(\begin{pmatrix} \beta_1 + \beta_4 d_{ij} \\ \beta_2 + \beta_5 d_{ij} \\ \beta_3 + \beta_6 t_{ij} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \right)$$

Model 4: *Generalised Linear Multilevel model with correlated random effects and level 1 heteroskedasticity across all hospitals.* This final model allows for different shape parameters of the gamma distribution for every hospital in the sample. The previous GLMM models allow only two gamma shape parameters, one which is identical for all NIAP hospitals and one which is identical for all control hospitals. Consideration of the cost distributions (see Figure 1) illustrates that this may not be an appropriate simplification. Model 4 can be written as:

$$c_{ij} \tilde{} \Gamma\left(\mu_{ij}, \rho_r\right), \quad \mu_{ij} = \sum_{s=1}^{3} \beta_{sj} x_{ijs} + \sum_k \alpha_k z_{ijk}, \quad r = 1, 2, \ldots, 15$$

$$\begin{pmatrix} \beta_{1j} \\ \beta_{2j} \\ \beta_{3j} \end{pmatrix} \tilde{} N \left(\begin{pmatrix} \beta_1 + \beta_4 d_{ij} \\ \beta_2 + \beta_5 d_{ij} \\ \beta_3 + \beta_6 t_{ij} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \right)$$

3.1 Hypothesis testing and model comparisons.

Parameter estimates and standard errors for the fixed parts of the multilevel models (β_1 to β_6 and α_1 to α_8) can be used to make inferences about significance using standard approaches such as t-tests. However, in general, these usual tests are not

adequate for the random parts of the models since the true variance value might be on the boundary of the parameter space, making inferences unreliable. In these cases, it is better to base inferences on the likelihood ratio statistic. To test a null hypothesis H_0 against an alternative hypothesis H_A, the following likelihood ratio statistic is used:

$$LR = -2 \ln \left(\frac{L_0}{L_A} \right)$$

where L_0 and L_A are the likelihoods for the null and alternative hypothesis respectively. Under the usual regularity conditions, this statistic has a chi squared distribution with degrees of freedom equal to the difference in the number of estimated parameters between the null and the alternative hypotheses. However, in cases where the null hypothesis involves parameters that are on the boundary of the parameter space the asymptotic distribution of the likelihood ratio statistic needs to be derived. In this situation, the likelihood ratio test is too conservative in the sense that the true p-value of the test will be smaller, as has been shown by Self and Liang (1987) , Skrondal (2004) and Verbeke and Molenberghs (2003). Therefore, caution should be exercised where the p-value indicates that there is no evidence to reject the null whereas one can be confident in the converse case.

4 Results

Initially, we included a broad range of covariates in a GLM, covering treatment type and location, other treatments given within the episode, patient demographics and comorbidities. Variables were only deleted from the model in cases where they were grossly insignificant with t-values well below one although treatment and location variables were always maintained. A quadratic effect of age on total cost was found to perform better either than using age on its own or the logarithm of age. Data on previous Acute Myocardial Infarction (AMI), angina, PCI and CABG had a large number of missing values. Two different specifications were considered: one with

all four dummy variables and another where they were combined in one dummy variable with a value of one if patients had any of previous AMI, angina, PCI and CABG. The latter model had a better fit and was therefore retained (Model 1). All other models are developed from Model 1 with results shown in Table 2.

Model 1 is the standard GLM with no account taken of the multilevel nature of the data. It can be seen from Table 2 that the cost of thrombolysis (£2,639) is only slightly higher than the cost of no treatment (£2,540) in control hospitals. Primary angioplasty is substantially more costly at £4,310. There are additional costs of £665 and £263 for thrombolysis which occurs in a NIAP hospital and for primary angioplasty that occurs out of standard working hours, respectively. Treatment cost increases with patient age. We also find substantial and significant positive effects on costs where patients have pre-existing vascular disease or other comorbidities. Where other coronary procedures are undertaken within the admission, costs are higher with CABG adding £4,950 to the total cost.

Some of the restrictions embedded in Model 1 are relaxed cumulatively in Models 2, 3, and the preferred model, Model 4[1]. The model fit as assessed by both AIC and BIC improves as these restrictions are relaxed. Similarly, using the conservative p-values, the likelihood ratio test rejects models 1, 2 and 3 in favour of model 4 with the following test statistics respectively (conservative p-values in square brackets), 1866.90[0.00], 173.40[0.00] and 162.04[0.00].

It is noticeable that as the random effects are incorporated into the models, the standard errors relating to the treatment-type covariates increase substantially. This is as expected since clustering in the data, which was masked in Model 1, reduces the amount of information that can be gained from each observation. It is also worth noting that the variances of the random effects are very large reflecting the variability across hospitals and that the correlations between the random effects are significant. Using the results of Model 4, the interval in which 95% of hospitals are

[1]Initially we used 24 integration points to calculate the integrals and increased it to 28 to check the stability of the estimated parameters.

expected to fall is £695 to £4,819 for the no treatment intercept, £1,512 to £4,303 for the thrombolysis intercept and £3,697 to £5,145 for the primary angioplasty intercept.

The relaxation of the restriction embedded in Model 1, that there is a common scale to the gamma distribution across all hospitals, is undertaken in two ways. Models 2 and 3 allow separate scale parameters for NIAP versus Control hospitals, whilst in Model 4 the scale parameter is allowed to differ across all hospitals. The improvement in the fit gained by Model 4 indicates that this greater degree of separation is preferable and that the distinction between NIAP and control sites is not the cause of across hospital differences in the scale.

Compared to Model 1, the expected episode cost for all treatments is much higher but also the differences between treatments vary. For example, the difference in cost between thrombolysis in a NIAP centre compared to primary angioplasty falls from over £1,000 in Model 1 to £760 in Model 4. We also find that the cost of primary angioplasty performed out of standard working hours is approximately £100 higher in Model 4. The impact of comorbidities also varies substantially between models 1 and 4, with substantially lower costs assigned to the presence of vascular disease or other comorbidities in the latter model.

Table 2: Parameter estimates for Models 1 to 4

	Model 1	Model 2	Model 3	Model 4
No treatment (control)	2540.16 (178.35)	2668.11 (441.80)	2788.46 (456.98)	2757.31 (309.75)
Thrombolysis (control)	2638.52 (96.09)	2934.54 (294.47)	2947.77 (304.25)	2907.65 (297.83)
PPCI (NIAP)	4310.00 (133.29)	4455.94 (171.42)	4479.11 (168.41)	4421.02 (155.99)
No treatment (NIAP)	−63.27 (189.10)	−60.12 (523.22)	−264.85 (532.71)	
Thrombolysis (NIAP)	664.94 (162.36)	501.93 (387.11)	600.45 (396.76)	754.31 (365.95)
Out of hours PPCI	263.45 (156.16)	262.01 (136.95)	213.31 (135.32)	362.80 (118.50)
(age-64)/10	390.79 (37.74)	352.38 (34.30)	350.29 (34.13)	332.33 (31.85)
((age-64)/10)^2	63.56 (20.70)	62.97 (15.07)	61.06 (15.14)	65.96 (14.57)
Previous coronary heart disease	130.73 (117.64)	189.71 (100.71)	186.63 (100.79)	217.53 (97.04)
Vascular/cerebrovascular disease	1026.15 (213.38)	972.49 (181.80)	958.20 (181.09)	842.84 (166.88)
Other comorbidities	467.84 (89.30)	251.16 (80.87)	248.85 (81.02)	198.22 (75.30)
Dead at discharge	−752.43 (170.14)	−716.04 (170.32)	−694.73 (168.62)	−704.52 (176.45)
Non primary PCI in this admission	201.50 (161.50)	267.55 (155.08)	233.57 (152.31)	287.88 (146.52)
CABG at this admission	4949.73 (1096.61)	4673.96 (905.94)	4648.99 (907.15)	4767.03 (877.15)
Scale parameter	0.34 (NA)			(1)
Scale parameter NIAP		0.23 (0.01)	0.23 (0.01)	
Scale parameter Control		0.36 (0.01)	0.36 (0.01)	
Random effects				
σ_1		833.07 (203.88)	833.25 (202.01)	1052.04 (277.55)
σ_2		607.75 (155.49)	630.40 (165.53)	711.84 (202.42)
σ_3		393.59 (118.26)	384.99 (106.17)	369.46 (109.19)
ρ_{12}			0.49 (0.33)	0.67 (0.24)
ρ_{13}			0.87 (0.18)	0.57 (0.30)
ρ_{23}			0.00 (0.46)	−0.23 (0.40)
AIC		43825.15	43819.79	43681.76
BIC		43838.61	43835.37	43705.83

Standard errors are provided in brackets. For the random part of the models they should be interpreted with caution

(1) Separate scale parameters for all 15 hospitals, range = 0.09 to 0.39

5 Discussion

In this paper, we compare different models which attempt to deal with the challenges posed by data that is both non-normally distributed and clustered. We demonstrate how a GLM can be developed to incorporate random effects and to allow for differences in the scale according to cluster using cost data from an observational study of UK NHS heart attack patients.

Health care is dominated by studies that generate data that is hierarchical in nature. To date, most applications of multilevel models in health economics have focussed on clinical trials (Willan et al. (2005), Grieve et al. (2005), Manca et al. (2005)). This is natural since, for reasons of practicality and to demonstrate generalisability, trials are frequently conducted at multiple centres, sometimes in multiple countries. Thompson et al. (2006) recognise that data generated by observational studies often also include clustering.

We demonstrate that accounting for clustering in the data through multilevel models is important but furthermore, consideration of alternative specifications of the random effects components of the statistical model is required as well as the specification of the fixed effects. In our example, we find that the preferred model is one which allows for separate random effects for treatment type by hospital, correlations between those random effects, and separate scale parameters for each hospital. Model selection is based on the AIC/BIC and Likelihood Ratio tests. Substantial differences in cost estimates are apparent between the various models.

Failure to appropriately account for both skewness and clustering in the data may result in biased estimates of treatment cost and its variance. In our example, there are substantial differences in the cost estimates between the GLM model that does not account for clustering at all and the multilevel models that do. But in addition, there are substantial differences between the various multilevel models. In particular, the differences in the estimated costs of thrombolysis versus angioplasty, and the cost of angioplasty performed in or out of hours, could be sufficient

to translate into substantial differences in cost effectiveness estimates and subsequently decision making. Indeed, our analysis of treatment costs for heart attack are primarily intended to inform a cost effectiveness model comparing primary angioplasty with thrombolysis where much of the clinical effectiveness data is rightly drawn from trials (Wailoo et al. (2009)). However, observational datasets of this type may often be more appropriate for estimating real world resource use compared to the artificial environments of clinical trials. The use of statistical models that include a variety of covariates to control for differences between patient groups is more likely to be a requirement for analysis of observational datasets given the non randomised nature of the patient samples. We considered a large number of regressors that reflect patient and hospital characteristics as well as the treatments themselves.

Most studies to date that have used multilevel models in analysis of cost or cost effectiveness data have maintained the assumption of normality (Carey (2000), Burgess et al. (2000), Willan et al. (2005)). Our study demonstrates, in common with both Thompson et al. (2006) and Manca et al. (2005), that both skewness and clustering can be incorporated simultaneously. In addition, most previous studies have not considered correlations between the random effects. Manca et al. (2005) and Manca et al. (2007) provide examples that did include such terms based on cost effectiveness data in multicentre trials. Manca et al. (2005) only uses a single treatment covariate whereas our observational dataset required the consideration of a large number of potential covariates, 14 of which are retained in the preferred model. Our study complements and extends this work by demonstrating the improvement in model fit that can be obtained between alternative specifications both of the random effects and the scale parameter.

Finally, one minor difference is in the implementation of the statistical models. Almost all published applications of multilevel models have been perfomed in a Bayesian framework, using Monte Carlo Markov Chain (MCMC) simulation

methods (Manca et al. (2007), Thompson et al. (2006), Cooper et al. (2007)) which is sometimes seen as a more straightforward means of implementing non standard models. The use of quadrature techniques for estimation have been shown to offer a convenient alternative in standard statistical packages such as the STATA module (GLAMM) in this example and in SAS (Liu et al. (2009)).

References

Asseburg, C., Bravo Vergel, Y., Palmer, S., Fenwick, E., de Belder, M., Abrams, K., et al. (2007) Assessing the effectiveness of primary angioplasty compared with thrombolysis and its relationship to time delay: a Bayesian evidence synthesis. *Heart,* 93,1244-1250

Blough, D. K. and Ramsey, S. D. (2000) Using Generalized Linear Models to Assess Medical Care Costs. *Health Services & Outcomes Research Methodology,* 1, 185-202

Bravo Vergel, Y., Palmer, S., Asseburg, C., Fenwick, E., de Belder, M., Abrams, K., Sculpher, M. (2007) Is primary angioplasty cost effective in the UK? Results of a comprehensive decision analysis. *Heart* 93, 1238-1243

Burgess, J. E., Cristiansen, C. L., Michalak, S. E. and Morris, C. N. (2000) Medical profiling: improving standards and risk adjustments using hierarchical models. *Journal of Health Economics,* 19, 291–309.

Carey, K. (2000) A multi-level modelling approach to analysis of patient costs under managed care. *Health Economics,* 9, 435–446.

Cooper, N. J., Lambert, P. C., Abrams, K. R. and Sutton, A. J. (2007) Predicting costs over time using bayesian Markov Chain Monte Carlo methods: an application to early inflammatory polyarthritis. *Health Economics*, 16, 37-56.

Duan, N. (1983) Smearing Estimate: A Nonparametric Retransformation Method, *Journal of the American Statistical Association*, 78, 605-610.

Gauthier, A., Manca, A. and Anton, S. (2009) Bayesian modelling of healthcare resource use in multinational randomized clinical trials. *Pharmacoeconomics*, 27, 1017-1029

Glick, H.A., Doshi, J.A., Sonnad, S.S. Polsky, D. (2007) Economic Evaluation in Clinical Trials, Oxford: Oxford University Press.

Goodacre, S., Sampson, F., Carter, A., Wailoo, A., O'Cathain, A., Wood, S., Capewell, S., Campbell, S. (2008) Evaluation of the National Infarct Angioplasty Project: Report for the National Co-ordinating Centre for NHS Service Delivery and Organisation R&D (NCCSDO), October 2008. Available at http://www.sdo.nihr.ac.uk/files/project/120-final-report.pdf

Goldstein, H. (2003) Multilevel Statistical Models. Third Edition. Arnold. London.

Grieve, R., Nixon, R., Thompson, S. G. and Normand, C. (2005) Using multilevel models for assessing the variability of multinational resource use and cost data. *Health Economics*, 14, 185–196.

Grieve, R., Nixon, R. and Thompson, S.G. (2009) Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized trials. *Medical Decision Making*, online first, doi:10.1177/O272989XO9341752

Liu, L., Strawderman, R. L., Cowen, M. E. and Shih, Y. T. (2009) A flexible two-part random effects model for correlated medical costs. Journal of Health Economics, online first, doi:10.1016/j.jhealeco.2009.11.010

Manca, A., Lambert, P. C., Sculpher, M. and Rice, R. (2007) Cost-Effectiveness Analysis Using Data from Multinational Trials: The Use of Bivariate Hierarchical Modeling, *Medical Decision Making,* 27, 471-490.

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

Manning, W. G. and Mullahy, J. (2001) Estimating log models: to transform or not to transform? *Journal of Health Economics*, 20, 461-494.

Or, Z., Wang, J. and Jamison, D. (2005) International differences in the impact of doctors on health: a multilevel analysis of OECD countries. *Journal of Health Economics,* 24, 531–560.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004). GLLAMM Manual. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.

Rabe-Hesketh, S. and Skrondal, A. (2002) Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2, 1-21.

Self, S. G. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *Journal of the American Statistical Association,* 82,605-610.

Skrondal, A., Rabe-Hesketh, S. (2004) Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. Boca Raton, FL; Chapman and Hall/CRC.

Thompson, S. G., Nixon, R. and Grieve, R. (2006). Addressing the issues that arise in analysing multicentre cost data, with application to a multinational study, J*ournal of Health Economics*, 25,1015-1028.

Verbeke, G. and Molenberghs, G. (2003) The use of score tests for inference on variance components. *Biometrics*, 59, 254-262.

Wailoo, A.J., Hernández, M., Goodacre, S., Sampson, et al. (2009) Primary angioplasty versus thrombolysis for acute ST-elevation myocardial infarction in the UK NHS: an economic analysis of the National Infarct Angioplasty Project Pilots. *Heart*, online first, doi:10.1136/hrt.2009.167130

Willan, A. R. and Kowgier, M.E. (2008) Cost-effectiveness analysis of a multinational RCT with a binary measure of effectiveness and an interacting covariate. *Health Economics*, 17, 777-791.

Willan, A. R., Pinto, E. M., O'Brien, B. J. , Kaul, P., Goeree, R., Lynd, L. and Armstrong, P. W. (2005) Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. *Health Economics,* 14, 327–338.