

# **Addressing the issues that arise in analysing multicentre cost data, with application to a multinational study**

Simon G. Thompson<sup>1</sup>, Richard M. Nixon<sup>1</sup>, Richard Grieve<sup>2</sup>

<sup>1</sup> MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK

<sup>2</sup> Health Services Research Unit, London School of Hygiene and Tropical  
Medicine, London, UK

## **Abstract**

Differences in the mean, spread and skewness of cost data collected from different countries present problems for analysis and interpretation. This paper develops generalised linear multilevel models to estimate the effect of patient and national characteristics on costs. Using gamma distributions and multiplicative effects for patient characteristics fitted the data better than models that assumed normal distributions or estimated additive effects. A multilevel gamma model is employed to allow for heterogeneity in the effects of patient case-mix across centres. Analysis of multinational cost data must recognise differences in mean, spread and skewness across centres, as well as the data's hierarchical structure.

**Running title:** Analysing multicentre cost data

**Keywords:** Cost data, multicentre studies, multinational studies, health economic evaluation, generalised linear models, hierarchical models

**Corresponding author:** Professor S. G. Thompson, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK. Tel +44 1223 330366, Fax +44 1223 330388, Email [simon.thompson@mrc-bsu.cam.ac.uk](mailto:simon.thompson@mrc-bsu.cam.ac.uk)

17<sup>th</sup> March 2005, submitted to *Journal of Health Economics*

## 1. Introduction

Cost data, whether collected in randomised trials or observational studies, are usually highly skewed (Briggs & Gray, 1998). While methods of analysis that assume a normal distribution may therefore be inappropriate (O'Hagan & Stevens, 2003), it is acknowledged that, to meet the needs of decision-makers, the focus should be on estimating mean costs (Thompson & Barber, 2000). The appeal of using generalised linear models for regression analysis of cost data is that they allow non-normal distributions to be specified, permit flexible modelling of covariates, and enable inferences to be made directly about the mean cost (Manning & Mullahy, 2001; Barber & Thompson, 2004).

Although an increasing number of economic evaluations are conducted on a multinational basis (Willke et al., 1998; Cook et al., 2003), there has been little attention given to which statistical methods are appropriate for dealing with the hierarchical nature of the cost data that arises. Variations in clinical practice, technical efficiency, factor prices and allocative efficiency across health care settings lead to differences in both the mean and spread of costs (Drummond et al., 1992; O'Brien, 1997; Raikou et al., 2000; Danzon & Chou, 2000; Bernt, 2000; Localio et al., 2001). In addition, differences in patient case-mix may lead to variation in costs (Phelps & Mooney, 1993). However, this clustering of patients within settings has generally been ignored in health economics (Newhouse, 1994; Jones, 2000), and in the analysis of cost and cost-effectiveness data in particular (Willke et al., 1998). For example, in economic evaluations, cost data are routinely pooled across national and international health care settings to provide overall estimates of the incremental cost-effectiveness of different technologies (Glick et al., 1998). This leads to difficulties in the interpretation of the results, and can limit the use of these economic evaluations in decision-making (Spath et al., 1999).

Multilevel models that acknowledge the hierarchical structure of data have been widely used in education and health services research (Goldstein, 1995; Leyland & Goldstein, 2001), and there are a few examples of their use in health economics (Burgess et al., 2000; Carey, 2000; Grieve et al., 2005; Willan et al., 2005; Manca et al., 2005; Or et al., 2005). A potentially important problem with these conventional multilevel models is that they assume residuals are normally distributed, and that the data have a similar spread across the settings. While generalised linear models have been shown as appropriate for tackling these problems when analysing non-hierarchical data (Manning & Mullahy, 2001; Barber & Thompson, 2004) or

longitudinal data (Seshamani & Gray, 2004), they have not previously been applied to the analysis of hierarchical cross-sectional data.

This issue is addressed here by using generalised linear multilevel models, and illustrated by examining the effect of patient and national-level characteristics on costs collected from health care centres in different countries. These methods are also applicable to estimating the effect of an intervention on costs in a randomised multinational trial.

The paper is structured as follows. Section 2 describes the study and dataset used. Section 3 presents simple models for estimating overall mean costs and the effect of a single patient-level variable on costs assuming homogeneity across centres. These models are extended in Section 4 to allow for heterogeneity using random effects and multiplicative covariate effects, and in Section 5 to include multiple patient-level variables. Section 6 includes a national-level covariate and considers the remaining heterogeneity across centres. Section 7 discusses the implications of the models presented.

## **2. Study and dataset used**

As an example, we use data from an observational study of the costs of providing stroke care across 13 centres in 10 countries in Western and Eastern Europe (Grieve et al., 2001a; Grieve et al., 2001b). The centres prospectively recruited cases who suffered a first-ever stroke during a one-year period in 1996-97. Patient-level characteristics were collected on standard forms; these included pre-stroke living conditions, stroke severity measures (incontinence during the first week after stroke, paralysis at hospital admission) and stroke subtype (ischaemic infarction, intra-cerebral haemorrhage, or unspecified stroke). One national-level variable is also considered, namely the proportion of gross domestic product spent on public health care.

Each patient was followed up for three months. The unit costs for resource use items were collected separately at each centre during site visits (Grieve et al., 2001a). These were initially reported in local currencies (1998 prices) and then converted to US

dollars using the Gross Domestic Product purchasing power parity (PPP) index to adjust for differences in the opportunity cost of resources across the economies concerned (Kavanos & Mossialos, 1999). As a sensitivity analysis, costs were also converted using a technology-specific PPP that adjusts more fully for international differences in factor prices for the technology concerned (Wordsworth & Ludbrook, 2004). Total three-month costs were calculated for each patient. Patients who died during the three months were excluded from the dataset, leaving 1298 patients for this analysis.

### *2.1 Description of the cost data*

The distributions of patients' costs in each of the 13 centres are shown in Figure 1, and summarised in Table 1. The mean costs varied widely across centres, from around \$600 to \$10,000, with the lowest means in Eastern European centres. The spread of the distributions also varied greatly, with standard deviations (SDs) ranging from around \$400 to \$8000. Generally, the centres with smaller standard deviations were those with the smaller means. The distributions of the cost data were positively skewed in almost all the centres.

Because cost data are skewed, distributions other than the normal are usually more appropriate. The gamma and log-normal distributions are often applied to cost data for this reason, and may yield different conclusions than when assuming a normal distribution (O'Hagan & Stevens, 2003; Nixon & Thompson, 2004). A comparison of these distributions is shown superimposed on the cost data for each centre in Figure 1. It is apparent that the gamma and log-normal distributions, while not perfect, fit the data in most of the centres better than the normal distributions. In the analyses that follow, we focus on the normal distribution because of its simplicity, and the gamma distribution because it better represents the data; we comment further on the use of the log-normal distribution in the discussion.

## **3. Overview of statistical methods and simple models**

The statistical methods developed below are based on generalised linear models. Being non-standard extensions, however, they were implemented using Markov

Chain Monte Carlo methods in WinBUGS (Gilks et al., 1996). Vague uniform priors were used for all means, for all regression coefficients, for the within-centre log standard deviations in the normal models, for the shape parameters in the gamma models, and for the standard deviations of any random effects. Posterior means and 95% credible intervals are presented. The Bayesian deviance (minus twice log likelihood, averaged over the uncertainty in the parameters) was used as a measure of model fit (Gilks et al., 1996); a lower deviance of 4 or more per extra parameter is generally regarded as a material improvement. For clarity, we present a sequence of analyses together with their results.

### 3.1 Simple models for overall costs

In order to describe the models, we set up the following notation. Let  $y_{ij}$  be the cost for patient  $i$  in centre  $j$  ( $j=1 \dots 13$ ). To allow extension to more complex models later, we write that  $y_{ij}$  has a mean  $\theta_{ij}$  that can depend on the individual ( $i$ ) as well as the centre ( $j$ ). The simplest model that allows for differences across centres is:

$$y_{ij} \sim \text{Normal}(\theta_{ij}, \sigma^2), \theta_{ij} = \mu_j \quad (1)$$

Here we are assuming that each centre has a different mean cost  $\mu_j$  and that the costs have the same SD  $\sigma$  in each centre. This latter assumption can be avoided by allowing a different SD  $\sigma_j$  in each centre:

$$y_{ij} \sim \text{Normal}(\theta_{ij}, \sigma_j^2), \theta_{ij} = \mu_j \quad (2)$$

Model (2) is the representation given by the superimposed normal distributions in Figure 1.

To allow for the positive skewness, we use a gamma distribution instead of a normal distribution. A gamma distribution can be parameterised in terms of its mean  $\theta$  and its ‘shape’  $\rho$ ; the SD of such a gamma distribution is  $\theta/\sqrt{\rho}$ . So the model can be written:

$$y_{ij} \sim \text{Gamma}(\theta_{ij}, \rho_j), \theta_{ij} = \mu_j \quad (3)$$

Model (3) is the representation given by the superimposed gamma distributions in Figure 1.

Formal comparison of models (1) to (3) in terms of deviance is shown in Table 2. Model (1) assuming the same SDs across centres fits very poorly, and we do not pursue it further, while the gamma model (3) fits better than the normal model (2).

### 3.2 Simple models for a single patient-level variable assuming homogeneity

We now extend these models to investigate the effect of a patient-specific variable, whether individuals were incontinent, on costs. Since incontinence is an indicator of more severe stroke and functional impairment, it is not surprising that the costs were on average higher for incontinent patients than continent patients in each centre (Table 1). However, the mean difference in costs between these groups varied greatly across centres, from around \$300 to \$9000, with the smaller differences generally corresponding to centres with lower overall mean costs.

First we extend model (2) to allow a mean difference in cost between patients who are incontinent ( $x_{ij}=1$ ) and those that are not ( $x_{ij}=0$ ):

$$y_{ij} \sim \text{Normal}(\theta_{ij}, \sigma_j^2), \theta_{ij} = \mu_j + \beta x_{ij} \quad (4)$$

$\beta$  represents the average mean cost difference between incontinent and continent patients. A similar extension of model (3) is:

$$y_{ij} \sim \text{Gamma}(\theta_{ij}, \rho_j), \theta_{ij} = \mu_j + \beta x_{ij} \quad (5)$$

In these models the effect of incontinence ( $\beta$ ) is assumed to be the same in each of the 13 centres. It can be shown that they are equivalent to fitting separate models for each centre, and then combining the 13 estimates for the effect of incontinence using a ‘fixed effect’ meta-analysis that does not allow for possible heterogeneity across centres (Thompson, 1994).

The results from models (4) and (5) are shown in Table 2. They fit better than models (2) and (3) respectively, because there is a clear effect of incontinence on costs. As before, the gamma model (5) fits the data better than the normal model (4). However the estimate of the effect of incontinence, for example in model (4), is to raise costs by about \$600, which does not appear to be an appropriate average of the results in Table 1. The reason is evident from Figure 2 which shows that, on an absolute scale, the estimated effect of incontinence is more precisely estimated in those centres with

lower average costs (such as centres 8, 12 and 13). These get the vast majority of the weight in the meta-analysis across centres. The gamma model (5) produces almost as extreme a weighted average. The effects of incontinence vary enormously from centre to centre, as is clear from the non-overlapping credible intervals in Figure 2, and the assumption of no heterogeneity across centres therefore misrepresents the data.

#### 4. Addressing heterogeneity in the effect of a single patient-level variable

##### 4.1 Random effects model

We consider two ways of addressing the above problem of heterogeneity. The first is to allow directly for heterogeneity in the effect of incontinence across centres. For example, extending model (5):

$$\begin{aligned} y_{ij} &\sim \text{Gamma}(\theta_{ij}, \rho_j), \theta_{ij} = \mu_j + \beta_j x_{ij} \\ \beta_j &\sim \text{Normal}(\beta, \sigma_\beta^2) \end{aligned} \quad (6)$$

Here the  $\beta_j$  are the absolute mean differences in costs for each centre, between the incontinent and continent patients, as given in Table 1. We have allowed these to vary across centres, by assuming they follow a normal distribution with mean  $\beta$  (an ‘average’ of the mean differences, that is the focus for estimation) and SD  $\sigma_\beta$ . This is a multilevel, or hierarchical, model (Goldstein, 1995; Grieve et al., 2005). The analysis is equivalent to fitting separate models for each centre, and then combining the 13 estimates for the effect of incontinence using a ‘random effects’ meta-analysis that now allows for the heterogeneity across centres (Thompson, 1994).

The results from this model (Table 2) are completely different from model (5). First it fits considerably better, as assessed by the deviance. Second, the effect of incontinence is estimated as around \$3300, a much more reasonable ‘average’ of the differences shown in Table 1 and Figure 2. The width of the credible interval for this overall estimate of the effect of incontinence (Figure 2) increases to take account of the extreme heterogeneity across the 13 centres.

##### 4.2 Use of multiplicative rather than additive effects

Another way of trying to accommodate the heterogeneity across centres is to change the scale on which the effect of incontinence is measured. Rather than supposing that the effect of incontinence is to *add* a certain cost in each centre, it could be assumed that the effect is to *multiply* the costs by a certain factor in each centre. The ratios of the mean costs between the incontinent and continent groups, shown in Table 1, are somewhat less variable (ranging from around 1.2 to 2.7-fold) than the absolute differences. These ratios are not strongly related to the overall mean cost in each centre.

Such an analysis can be achieved by using a log link (Manning & Mullahy, 2001; Barber & Thompson, 2004), so that we model the log of the means  $\theta_{ij}$ :

$$y_{ij} \sim \text{Gamma}(\theta_{ij}, \rho_j), \log(\theta_{ij}) = \mu_j + \beta x_{ij} \quad (7)$$

Now  $\exp(\beta)$  is the ratio of costs between the incontinent and continent groups, and is assumed to be the same in each centre. Again, by using a multilevel model, we can allow for potential differences in the effect of incontinence (expressed as a ratio) across centres, as in a random effects meta-analysis:

$$\begin{aligned} y_{ij} &\sim \text{Gamma}(\theta_{ij}, \rho_j), \log(\theta_{ij}) = \mu_j + \beta_j x_{ij} \\ \beta_j &\sim \text{Normal}(\beta, \sigma_\beta^2) \end{aligned} \quad (8)$$

Now  $\exp(\beta_j)$  corresponds to the ratio of costs between the incontinent and continent groups in each centre, and  $\exp(\beta)$  is an ‘average’ ratio.

The results of models (7) and (8) are given in Table 2, and the meta-analysis of ratios depicted in Figure 3. Model (7) has a substantially lower deviance than model (5), and therefore a common ratio fits the data better than a common absolute difference. The estimated ratio of 1.80 appears to be a reasonable summary of the centre-specific ratios (Figure 3). The centres with low costs now do not take such a large proportion of the weight in the meta-analysis. Nevertheless, there is still some heterogeneity across centres: model (8) has a lower deviance than model (7). The credible interval for the overall estimated ratio of 1.76 is now somewhat wider in model (8) than model (7), reflecting this degree of heterogeneity across centres.

Out of models (4) to (8) considered for estimating the effect of incontinence, model (8) seems the most appropriate. It is based on assuming gamma distributions for the

cost distributions within centres, which is better than assuming normal distributions. Ratios are more consistent than absolute difference across centres, although it is still appropriate to allow for the differences across centres even on the ratio scale. Models (6) and (8) fit equally well in the formal terms of deviance, but the evidence for (and extent of) heterogeneity is much more extreme for the absolute differences than for the ratios. So the ‘overall’ estimate of the additive effect of incontinence is difficult to interpret in the face of the extreme heterogeneity in Figure 2. Thus we conclude that the most appropriate representation of the overall effect of incontinence is to multiply the costs by 1.76 (95% credible interval 1.50 to 2.05).

This model can also be used to provide more appropriate estimates of the centre-specific effects of incontinence on costs, than the crude figures in Table 1. By borrowing information across centres about the effect of incontinence, while acknowledging that there is still some heterogeneity, we can derive ‘shrunk’ estimates (Goldstein, 1995; Willan et al., 2005) depicted as dotted lines in Figure 3. These are pulled in towards to the overall mean, compared to the crude estimates, and are more precise (having shorter credible intervals).

## 5. Models for multiple patient-level variables

The above analyses can easily be extended to consider multiple patient-specific variables. The purpose could be, for example, to remove the confounding of other correlated variables on the effect of incontinence, as well as potentially to improve precision. As an example, we include dummy variables for paralysis at admission ( $z_{1ij}$ ), for ischaemic vs. haemorrhagic vs. unknown stroke subtype ( $z_{2ij}$  and  $z_{3ij}$ ), and for living alone vs. living with others vs. living in a nursing home ( $z_{4ij}$  and  $z_{5ij}$ ). The focus is still on estimating the effect of incontinence on costs, so we extend model (8) as:

$$\begin{aligned} y_{ij} &\sim \text{Gamma}(\theta_{ij}, \rho_j), \log(\theta_{ij}) = \mu_j + \beta_j x_{ij} + \sum_k \gamma_k z_{kij} \\ \beta_j &\sim \text{Normal}(\mu_\beta, \sigma_\beta^2) \end{aligned} \quad (9)$$

The coefficients  $\gamma_1 \dots \gamma_5$  are here assumed to be homogeneous across centres, but the model could be extended to allow these to vary across centres using random effects.

The results from this model are shown in the last line of Table 2; the model fits better because the new variables are important predictors of costs, but the effect of incontinence is similar to that in model (8). Thus there was little confounding of the effect of incontinence by the other variables in this example.

### 6. Models that include higher-level variables

There is often interest in whether certain centre or national-level characteristics are related to the differences in average costs across centres (Grieve et al., 2005). Our example considers how a national-level variable, in this case the proportion of Gross Domestic Product spent on public health care (%GDP), can be included in the model. It is likely that centres from those countries spending a higher proportion of GDP on health care have higher average patient costs; indeed this is revealed in Figure 4. When examining the association of such higher-level variables with cost per patient it is necessary to acknowledge the hierarchical structure of patients within centres; otherwise the association will be over-precisely estimated (Grieve et al., 2005; Manca et al., 2005). However, while investigating differences across centres, we also want to simultaneously remove the effects of differences in patient case-mix across centres.

In all models (1) to (9) above, the adjusted centre means ( $\mu_j$ ) take distinct and unrelated values for each centre. An alternative, for example modifying model (9), is to allow the  $\mu_j$  to vary randomly across centres:

$$\begin{aligned} y_{ij} &\sim \text{Gamma}(\theta_{ij}, \rho_j), \log(\theta_{ij}) = \mu_j + \beta_j x_{ij} + \sum_k \gamma_k z_{kij} \\ \beta_j &\sim \text{Normal}(\beta, \sigma_\beta^2), \mu_j \sim \text{Normal}(\mu, \sigma_\mu^2) \end{aligned} \quad (10)$$

Here the adjusted centre log means are assumed to follow a normal distribution, with SD  $\sigma_\mu$ .

This model can be extended to see whether %GDP (denoted  $w_j$  for each centre  $j$ ) explains any of the variability across centres:

$$\begin{aligned} y_{ij} &\sim \text{Gamma}(\theta_{ij}, \rho_j), \log(\theta_{ij}) = \mu_j + \alpha w_j + \beta_j x_{ij} + \sum_k \gamma_k z_{kij} \\ \beta_j &\sim \text{Normal}(\beta, \sigma_\beta^2), \mu_j \sim \text{Normal}(\mu, \sigma_\mu^2) \end{aligned} \quad (11)$$

The quantity  $\exp(\alpha)$  represents the ratio of average centre costs per unit increase in %GDP. The amount by which  $\sigma_\mu^2$  has decreased in model (11) compared to model

(10) is the extent to which %GDP has explained the systematic (case-mix adjusted) differences in mean costs across centres.

The estimate of  $\alpha$  is 0.59 (95% CI 0.02 to 1.18); the quantity  $\exp(\alpha)$  therefore estimates a 1.8-fold increase in mean costs per 1% increase in GDP spent on national health care, compatible with the crude association shown in Figure 4. The estimate of  $\sigma_{\mu}^2$  decreases from 0.91<sup>2</sup> in model (10) to 0.78<sup>2</sup> in model (11), so that %GDP is estimated to explain 27% of the variability in mean costs across centres.

## ***7. Discussion***

The statistical analysis of cost data from multicentre studies is problematic especially when the centres are located in different countries. As we have shown, if inappropriate methods are used, the results will be misleading. This applies to estimating the association of either patient or higher-level variables with cost; in both cases the hierarchical nature of the data needs to be acknowledged, as well as the differences in mean, spread and skewness across centres. Variability in costs can occur across centres within one country (Coyle & Drummond, 2001), and the methods developed here may also be useful for single-country multicentre studies.

Apart from their role in multicentre economic evaluation, the methods developed may also be applied to studies that compare provider performance, adjusting for differences in patients' characteristics. Recent studies in this area have recognised the hierarchical nature of data on health service utilisation and health service costs, and have used multilevel models to try to identify factors associated with differences in performance (Carey, 2000; Burgess et al., 2000). However, these studies assumed that the data were normally distributed and, as our study suggests, making this assumption may not be appropriate. As studies comparing provider performance have increasing access to data on patient-level costs and covariates, methods are required that can assess the effects of patient-level covariates whilst recognising the structure and distribution of the data. The methods presented would therefore seem applicable to this broader area of research.

In our example, expressing the effect of patient-level characteristics on costs as a relative effect (using a generalised linear model with a gamma distribution and a log link) seemed appropriate. However this will not always be the case, and the choice of distribution and link function may depend on the particular category of cost data under consideration. For example, drug costs set by pharmaceutical companies might be additive rather than multiplicative, and models with some covariates acting additively and some acting multiplicatively might have to be considered (Basu & Rathouz, 2005). Similarly gamma distributions might not always be appropriate. In our example, using gamma distributions for costs, while clearly better than normal distributions, did not represent the cost data well in every centre. In certain centres (for example numbers 7 and 9 in Figure 1), there were ‘lumps’ at the right end of the distribution corresponding to patients who stayed in hospital for the full three months of follow-up. While a two-part model (Mullahy, 1998) might have produced a better fitting distribution, its implementation for hierarchical data is complex (Seshamani & Gray, 2004). We did not pursue using a log-normal distribution for two reasons: first because it is not part of the generalised linear model family of distributions, which naturally allow covariate adjustment on the mean cost directly, and secondly because we have found that it tends to give too high a weight to extreme costs way beyond the high costs actually observed in the data (Thompson & Nixon, 2005).

In our analysis, the PPP currency conversion used was based on the GDP PPP index; it reflected general differences in opportunity costs across the countries concerned. This conversion factor has been recommended for converting prices to a common currency in preference to medical care specific PPPs that are biased toward the pharmaceutical sector (Kavanos & Mossialos, 1999). In this instance, where labour costs dominate, the GDP PPP conversion failed to adjust for price differences between Western and Eastern European centres. As an alternative, we converted costs from local currencies to US dollars using a technology-specific PPP, designed to minimise differences in those factor prices specific to the technology concerned (Wordsworth & Ludbrook, 2004). However this did not change our results. There were still large differences in mean and spread of costs across centres, and the heterogeneity in both the absolute and relative effects of incontinence on costs were almost as large as in our original analysis. It is preferable to base costings on local unit costs as we did in our study (Grieve et al., 2001a), rather than apply the same unit

costs to all centres (Mark et al., 1995; Schulman et al., 1996; Johannesson et al., 1997; Glick et al., 1998). The latter may obscure the need for analytical techniques that reflect the resultant cost differences.

Multilevel statistical models have been used to analyse multinational cost data in a randomised trial (Willan et al., 2005), to model incremental net benefit across centres (Manca et al., 2005), and to compare resource use and costs across health care providers (Burgess et al., 2000; Carey, 2000). However, all these applications used the conventional assumption of normality, which is usually inappropriate for cost data applications. This can not only affect the estimates of the population mean cost (Nixon & Thompson, 2004), but also the estimates of case-mix coefficients. The conventional assumption that the variance of costs is the same in all centres will also often be inappropriate. Addressing the skewness in hierarchical cross-sectional data is novel; only one other recent paper presents a similar GLM approach for the bivariate analysis of cost-effectiveness in the context of multicentre randomised trials (Nixon & Thompson, 2005).

As others have also commented (Willan et al., 2005; Manca et al., 2005), one advantage of such multilevel models is their ability to use information from all centres to produce shrunken centre-specific estimates. An essential assumption underlying the validity of such estimates is that the quantities being estimated in different centres are ‘exchangeable’ – that knowledge of the centre alone does not, *a priori*, give any information about the magnitude of the quantity. In our example of estimating the effects of incontinence on costs, this assumption seems far more plausible when the effects are considered on a relative than an absolute scale; on an absolute scale one would generally expect the effects of patient-level variables to be greater in high-cost countries. Whether assuming normality to represent any heterogeneity across centres is fully appropriate is difficult to assess in our example since there were only 13 centres (Hardy & Thompson, 1998). Although some have mentioned concerns about the consistency of parameter estimates in multilevel models (Blundell & Windmeijer, 1997), there is no problem in our application because the numbers of patients per centre is large.

We conclude that the use of generalised linear models in health economics can be extended to the analysis of hierarchical cross-sectional data. These methods recognise the clustering in the data, but also differences in the skew and spread of the data across centres. The models are therefore more appropriate for drawing inferences in studies with patient-level data and hierarchical study designs, such as multicentre or multinational studies and cluster randomised trials, which are increasingly being employed in practice.

**Note:** The WinBUGS computer code for implementing models (8) and (11) are available on [www.mrc-bsu.cam.ac.uk/richard/personal](http://www.mrc-bsu.cam.ac.uk/richard/personal)

## References

- Barber JA, Thompson SG, 2004. Multiple regression of cost data: use of generalised linear models. *Journal of Health Services Research and Policy* 9, 197--204.
- Basu A, Rathouz PJ, 2005. Estimating marginal and incremental effects on health outcomes using flexible link and variance function parameters. *Biostatistics* 6, 93--109.
- Bernt E, 2000. International comparisons of pharmaceutical prices: what do we know, and what does it mean? *Journal of Health Economics* 19, 283--287.
- Blundell R, Windmeijer F, 1997. Cluster effects and simultaneity in multilevel models. *Health Economics* 6, 439--443.
- 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.
- Burgess JE, Christiansen CL, Michalak SE, Morris CN, 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.
- Cook JR, Drummond M, Glick H, Heyse JF, 2003. Assessing the appropriateness of combining economic data from multinational clinical trials. *Statistics in Medicine* 22, 1955--1976.
- Coyle D, Drummond MF, 2001. Analyzing differences in the costs of treatment across centres within economic evaluations. *International Journal of Technology Assessment in Health Care* 17, 155--163.
- Danzon P, Chao LW, 2000. Cross-national price differences for pharmaceuticals: how large, and why? *Journal of Health Economics* 19, 159--195.
- Drummond MF, Bloom BS, Carrin G et al, 1992. Issues in the cross-national assessment of health technology. *International Journal of Technology Assessment in Health Care* 8, 671--82.
- Gilks WR, Richardson S, Spiegelhalter DJ, 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall, London.

- Glick HA, Willke R, Polsky D et al, 1998. Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage. *International Journal of Technology Assessment in Health Care* 14, 145--160.
- Goldstein H, 1995. *Multilevel statistical models*. Edward Arnold, London.
- Grieve R, Dundas R, Beech R, Wolfe CDA, 2001a. The development and use of a method to compare the costs of acute stroke across Europe. *Age and Ageing* 30, 67--72.
- Grieve R, Hutton J, Bhalla A et al, 2001b. A comparison of the costs and survival of hospital-admitted stroke patients across Europe. *Stroke* 32, 1684--1691.
- Grieve R, Nixon RM, Thompson SG, Normand C, 2005. Using multilevel models for assessing the variability of multinational resource use and cost data. *Health Economics* 14, 185--196.
- Hardy RJ, Thompson SG, 1998. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* 17, 841--856.
- Johannesson M, Jonsson B, Kjekshus J et al, 1997. Cost effectiveness of simvastatin treatment to lower cholesterol levels in patients with coronary heart disease. Scandinavian Simvastatin Survival Study Group. *New England Journal of Medicine* 303, 332--336.
- Jones AM, 2000. Health econometrics. In: Culyer AJ, Newhouse JP (eds). *Handbook of Health Economics, Vol 1*. Elsevier, Amsterdam.
- Kavanos P, Mossialos E, 1999. International comparisons of health care expenditures: what we know and what we do not know. *Journal of Health Services Research and Policy* 4, 122--126.
- Leyland AH, Goldstein H (eds), 2001. *Multi-level Modelling of Health Statistics*. Wiley, Chichester.
- Localio AR, Berlin JA, Ten Have TR, Kimmel SE, 2001. Adjustments for center in multicenter studies: an overview. *Annals of Internal Medicine* 135, 112--123.
- Manca A, Rice N, Sculpher MJ, Briggs AH, 2005. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Economics*, published online 14 Jun 2004.
- Manning WG, Mullahy J, 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* 20, 461--494.
- Mark DB, Flatky MA, Califf RM et al, 1995. Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for

- acute myocardial infarction. *New England Journal of Medicine* 332, 1418--1424.
- Mullahy J, 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 17, 247--81.
- Newhouse J, 1994. Frontier estimation: How useful a tool for health economics? *Journal of Health Economics* 13, 317--322.
- Nixon RM, Thompson SG, 2004. Parametric modelling of cost data in medical studies. *Statistics in Medicine* 23, 1311--1331.
- Nixon RM, Thompson SG, 2005. Incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Economics*, in press.
- O'Brien BJ, 1997. A tale of two (or more) cities: geographic transferability of pharmaco-economic data. *The American Journal of Managed Care* 3, S33--S39.
- O'Hagan A, Stevens JW, 2003. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Economics* 12, 33--49.
- Or Z, Wang J, Jamison D, 2005. International differences in the impact of doctors on health: a multilevel analysis of OECD countries. *Journal of Health Economics*, in press.
- Phelps CE, Mooney C, 1993. Variations in medical practice use, causes and consequences. In: Arnoud RJ, Rich RF, White WD (eds). *Competitive approaches to health care reform*. The Urban Institute Press, Washington DC, pp 139--178.
- Raikou M, Briggs A, Gray A, McGuire A, 2000. Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. *Health Economics* 9, 191--198.
- Schulman K, Buxton M, Glick H et al, 1996. Results of the economic evaluation of the first study. A Multinational Prospective economic Evaluation. *International Journal of Technology Assessment in Health Care* 12, 698--713.
- Seshamani M, Gray AM, 2004. A longitudinal study of the effects of age and time to death on hospital costs. *Journal of Health Economics* 23, 217--235.
- Spath H-M, Carrere M-O, Fervers B, Philip T, 1999. Analysis of the eligibility of published economic evaluations for transfer to a given health care system:

- Methodological approach and application to the French health care system.  
Health Policy 49, 161--177.
- Thompson SG, 1994. Why sources of heterogeneity in meta-analysis should be investigated. British Medical Journal 309, 1351--1355.
- Thompson SG, Barber JA, 2000. How should cost data in pragmatic randomised trials be analysed? British Medical Journal 320, 1197--1200.
- Thompson SG, Nixon RM, 2005. How sensitive are cost-effectiveness analyses to choice of parametric distributions? Medical Decision Making, in press.
- Willan AR, Pinto EM, O'Brien BJ et al, 2005. Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrunken estimation: the Canadian ASSENT-3 economic analysis. Health Economics 14, 327--338.
- Willke RJ, Glick HA, Polsky D, Schulman K, 1998. Estimating country-specific cost-effectiveness from multinational clinical trials. Health Economics 7, 481--493.
- Wordsworth S, Ludbrook A, 2004. Comparing costing results in across-country economic evaluations: the use of technology-specific purchasing power parity. Health Economics 14, 93--99.

## Legends for Figures

**Figure 1:** Histograms of three-month cost data (PPP-adjusted \$1000) for stroke patients in 13 centres, with superimposed normal (continuous line), gamma (dashed line) and log-normal (dotted line) distributions.

**Figure 2:** Absolute differences in mean costs (PPP-adjusted \$) between incontinent and continent stroke patients: estimates with 95% credible intervals in each centre, and combined across centres.

**Figure 3:** Relative differences in mean costs between incontinent and continent stroke patients: estimates with 95% credible intervals in each centre, and combined across centres. Dotted lines represent posterior ‘shrunk’ estimates in each centre.

**Figure 4:** Relationship between crude mean costs (PPP-adjusted \$, log scale) for stroke patients and proportion of GDP spent on public health care (%GDP) across 13 centres; symbol size proportional to the precision of each log mean.

Figure

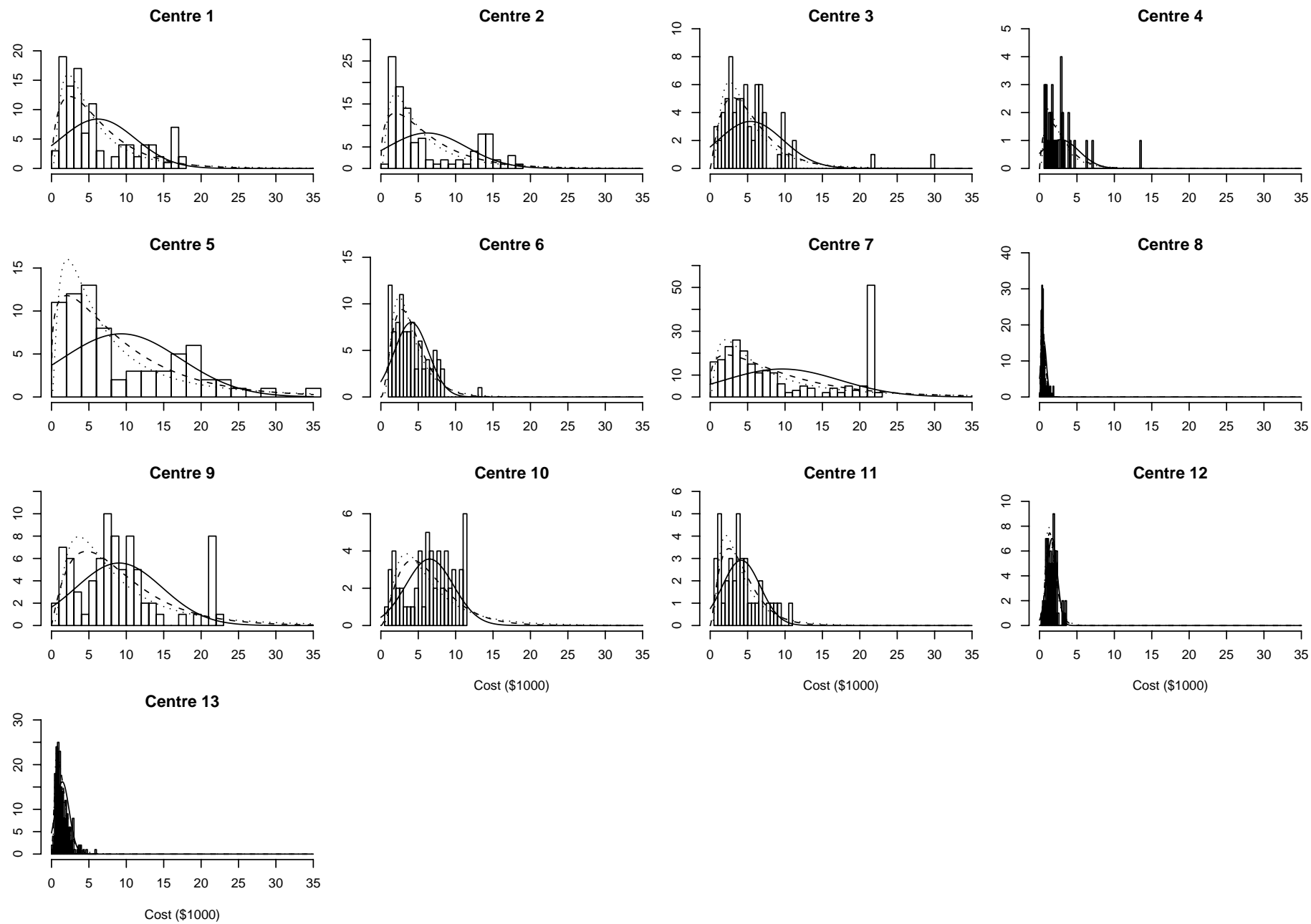


Figure 1:

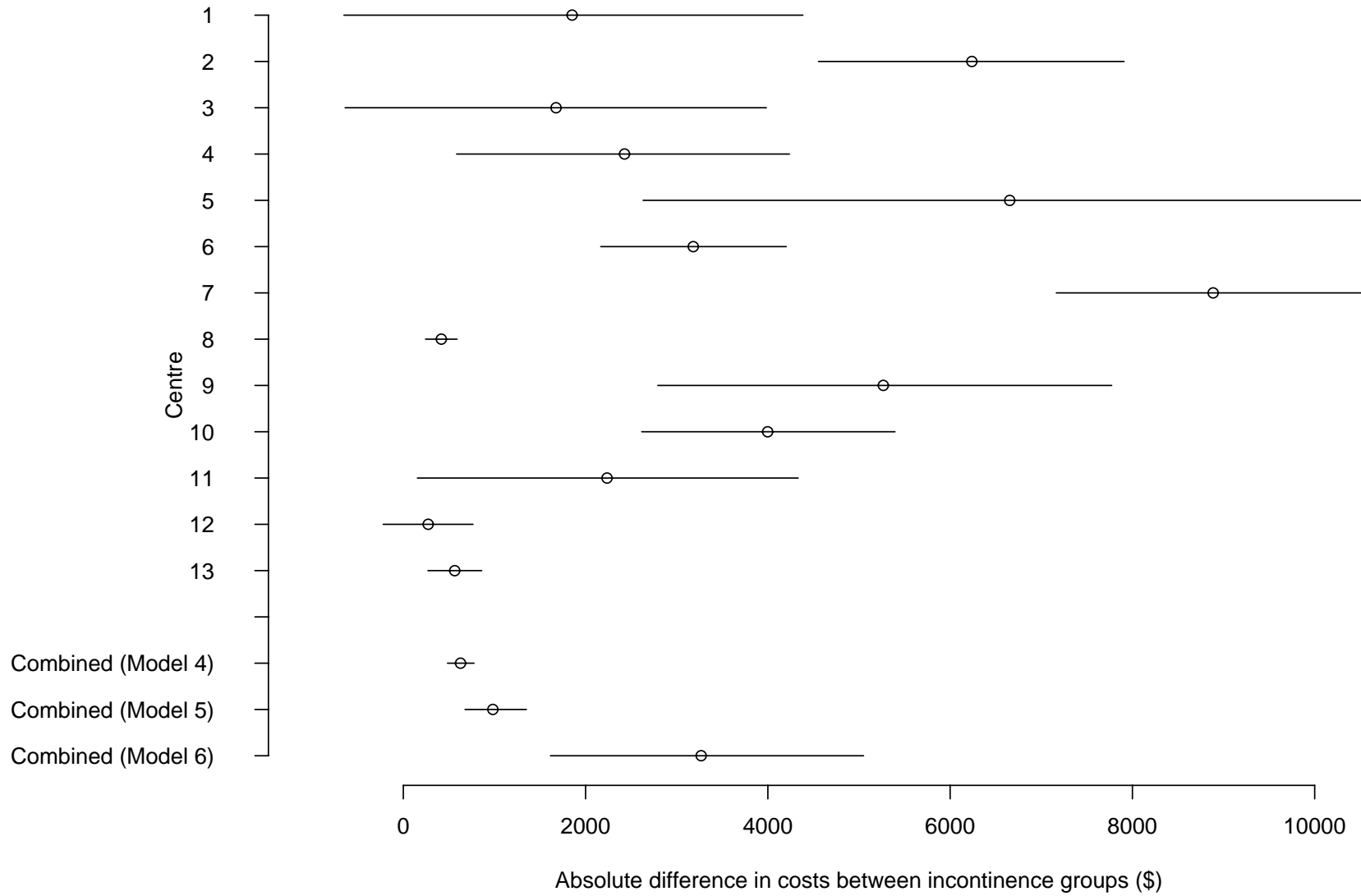


Figure 2:

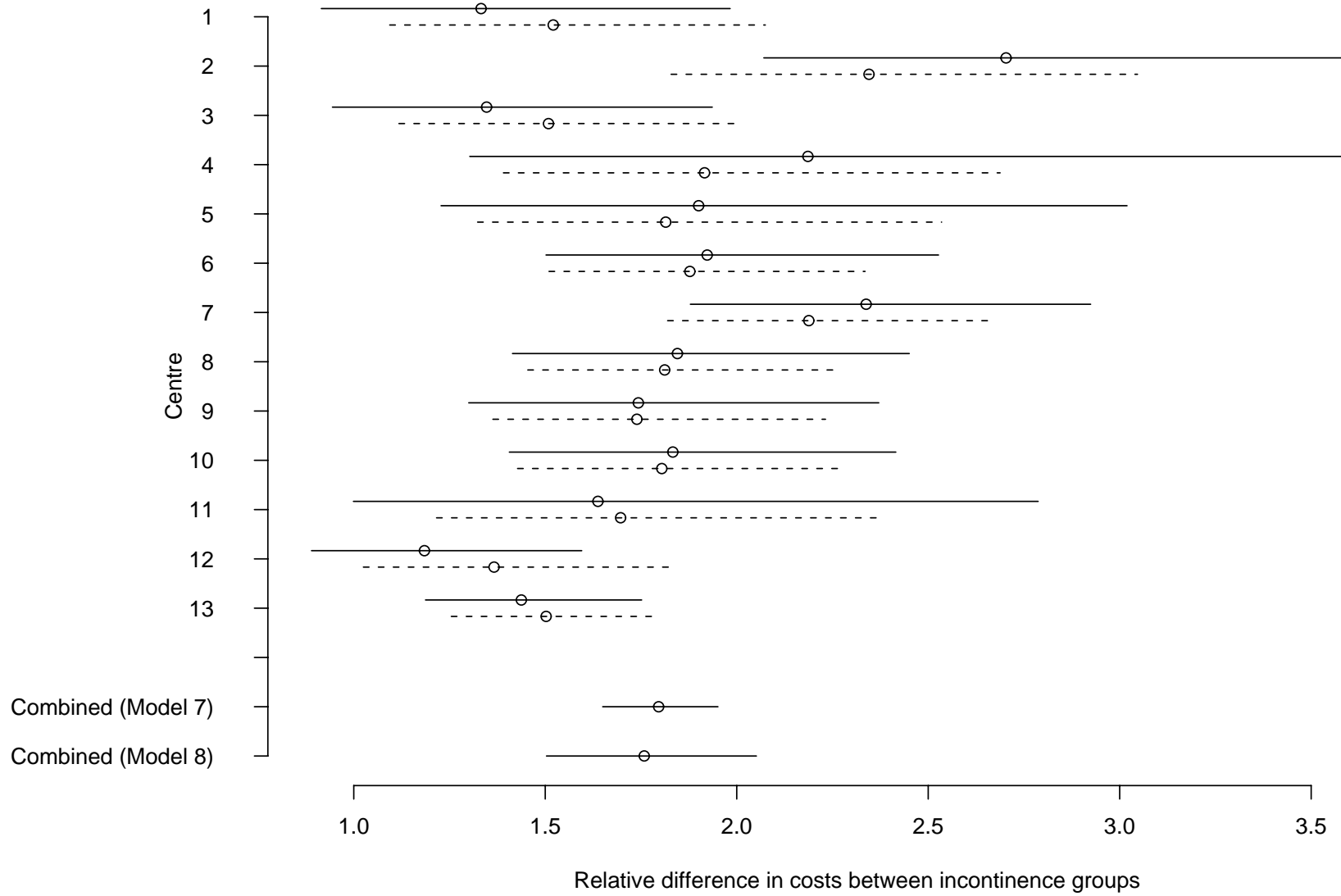


Figure 3:

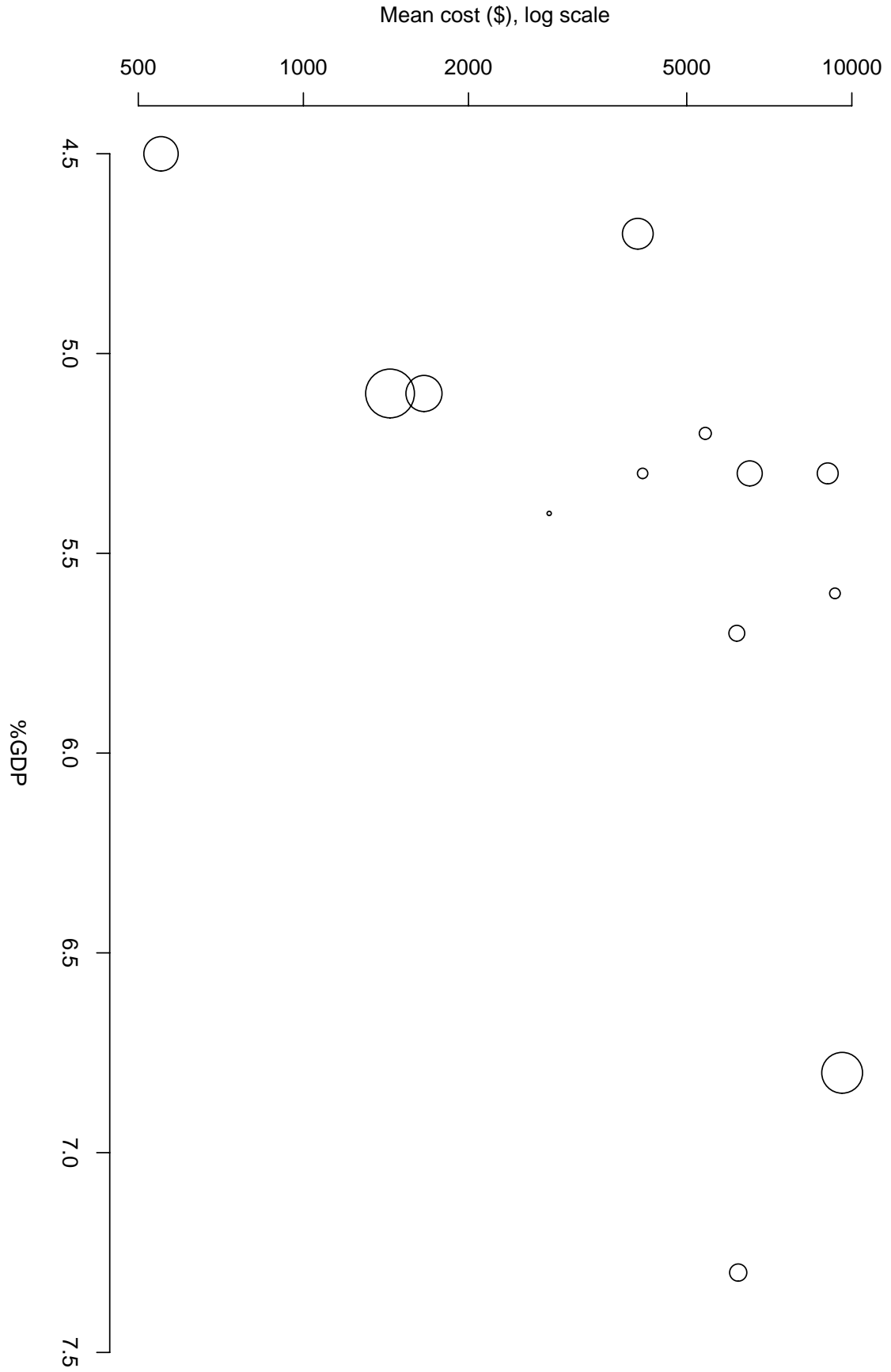


Figure 4:

**Table 1:** Costs (PPP-adjusted US dollars) over 90-days in 1298 patients with incident stroke

Centre	Number of patients	Mean cost	SD	Mean cost in patients without incontinence	Mean cost in patients with incontinence	Difference in means	Ratio of means
1. France	105	6204	5015	5867	7728	1861	1.32
2. Italy	109	6170	5293	3660	9878	6219	2.70
3. Portugal	73	5405	4353	4996	6653	1657	1.33
4. Spain	32	2805	2514	2123	4550	2427	2.14
5. Britain	73	9311	7983	7575	14245	6670	1.88
6. Poland	92	4072	2304	3451	6629	3179	1.92
7. Denmark	246	9600	7706	6673	15562	8889	2.33
8. Latvia	145	550	370	498	914	416	1.83
9. Finland 1	81	9036	5815	7216	12483	5267	1.73
10. Finland 2	57	6515	3215	4831	8832	4001	1.83
11. Finland 3	37	4155	2565	3732	5964	2231	1.60
12. Lithuania 1	62	1659	711	1615	1889	274	1.17
13. Lithuania 2	186	1439	919	1299	1863	563	1.43

**Table 2:** Fit of different models and their estimates of the effect of incontinence on costs (PPP-adjusted US dollars)

Model *	Deviance	Effect of incontinence (95% credible interval)
(1) Normal, with same standard deviations	7800	-
(2) Normal	6260	-
(3) Gamma	5725	-
(4) Normal, common absolute effect of incontinence	6185	Difference = 630 (480 to 780)
(5) Gamma, common absolute effect of incontinence	5645	Difference = 980 (680 to 1350)
(6) Gamma, random absolute effect of incontinence	5517	Difference = 3270 (1620 to 5050)
(7) Gamma, common relative effect of incontinence	5537	Ratio = 1.80 (1.65 to 1.95)
(8) Gamma, random relative effect of incontinence	5516	Ratio = 1.76 (1.50 to 2.05)
(9) Gamma, random relative effect of incontinence, adjusted for other patient case-mix variables	5444	Ratio = 1.63 (1.37 to 1.92)

\* See text for full details