A Multivariate Multilevel Analysis of Health Care Performance

Katharina Hauck, Andrew Street

Centre for Health Economics
University of York
YORK YO10 5DD
England

Katharina Hauck: kh126@york.ac.uk
Andrew Street: ads6@york.ac.uk

Draft, not for quotation

Paper presented at the 1st Franco-British Meeting in Health Economics

Centre de Conférences Internationales

Paris

14-16 January 2004

Acknowledgements

The authors are grateful to the English Department of Health for financial support.

The views expressed are those of the authors only.

Abstract

Aims The performance of health authorities in the English National Health Service (NHS) is assessed against a variety of indicators in order to provide a balanced picture of organisational performance. However, performance assessments often neglect the possibility that organisational achievement may be correlated across indicators. In this paper we analyse differences in performance across NHS organisations allowing for correlations across performance indicators due to unobservable factors. We aim to establish whether allowance for correlation changes the estimated contribution to performance made by NHS organisations.

Methods We analyse variations in performance of health authorities using multilevel models estimated as a multivariate system of equations. This approach allows us to estimate a separate equation for each performance indicator, but allowing for correlations across all equations. We compare the estimates of health authority performance with those derived from ordinary least squares (OLS) and multilevel models.

Data We use data on 13 performance indicators together with socio-demographic characteristics of 4985 English electoral wards, which are nested within 186 district health authorities. The dataset is merged from several data sources from 1990/91. The indicators measures health outcome, clinical quality, access and efficiency. In order to accommodate variations in performance that arise from variations in population characteristics, we condition on socio-economic factors and health needs.

Findings Correlations across indicators are statistically significantly. The performance rankings generated by the different estimation techniques differ for most health authorities, with the magnitudes varying by indicator. We find evidence of correlation across indicators, and this leads to different results when assessing performance with multivariate multilevel models as compared to multilevel or OLS models. Estimating variations in performance using multivariate models can change the rankings of health authorities substantially.

Introduction

Long has it been recognised that the pursuit of multiple objectives by public sector organisations makes it difficult to assess and compare their performance. Nevertheless, accurate performance assessment of such organisations is essential if they are to be held properly to account for discharging their public functions. Commonly performance assessment of such organisations is a relatively unsophisticated process, consisting of measurement against a set of objectives that may or may not be aggregated into a single index of organisational performance. This type of approach neglects the possibility that organisational achievement may be correlated across objectives. This correlation may be positive, if progress against one indicator simultaneously advances another, perhaps because good management promotes all-round performance. But the correlation may be negative if trade-offs are involved, such as when resources have to be diverted from one activity in order to meet some other objective.

In this paper we consider methods designed to provide a balanced picture of organisational performance in the presence of multiple objectives. We compare techniques to assess the performance of health authorities in the English National Health Service against a variety of performance indicators. Our data are reported at electoral ward level, with wards being clustered within health authorities, hence the data have a hierarchical (or multilevel) structure. The multivariate analysis proceeds by recognising that the outcome variables themselves are clustered and, thus, can be considered the lowest level of analysis (Yang *et al.*, 2002; Gilthorpe and Cunningham, 2000). Following this approach, we consider the set of multiple performance indicators (level 1) as clustered within electoral wards (level 2) which are themselves clustered with health authorities (level 3).

We build up our methodology in three stages. First, we ignore the hierarchical structure of the data and the possibility that correlations may exist across performance indicators. Treating the health authority as the unit of analysis, we estimate a multiple regression model separately for each performance indicator, allowing for the possibility that variations in performance might be attributable to differences in population demography, socio-economic conditions and health care needs. Levels of observed performance above (below) those predicted by the regression model are interpreted as indicative of above (below) average performance by the health authority. Second, recognising the "wards within health authorities" hierarchical structure to the data, we separately identify a health authority effect from a ward-level effect on performance for each indictor. The ward effect is interpreted as random variation and the health authority effect as indicative of health authority performance. Finally, correlations across performance indicators are considered by estimating a multivariate, multilevel model that treats the each performance indicator as part of a system of equations.

The paper is structured as follows. We describe the data in the next section and then detail the models to be estimated. Estimation results are presented in section four, together with a comparison of the results provided by the different estimation procedures in terms of the impact upon model parameters and the relative performance rankings of health authorities. Concluding comments are offered in the final section.

Data

Thirteen indicators of health authority performance are available (Table 1). Variously, these address health system objectives relating to improving health outcomes, ensuring clinical quality, fostering access and making efficient use of resources. All have been used in official assessments of NHS performance but the set of indicators analysed here is not intended as an exhaustive list of health system objectives (Hauck *et al.*, 2003). Most of the indicators are expressed as ratios, in which the actual value is standardised against national values to take account of the composition of the local population, in terms of age, gender and casemix. For all indicators, a higher value implies worse performance.

Table 1 around here

Even after standardisation, differences in performance may arise from additional variations in population characteristics. To account for this possibility, we condition the performance indicators on the needs indices developed by Carr-Hill *et al*. and used by the Department of Health since 1995 for the geographical allocation of resources for spending on hospital and community health services (Carr-Hill *et al.*, 1994). The needs indices comprise various socio-demographic variables that have been shown to be associated with the utilisation of health care. For all but one performance indicator, the needs index (NEEDAC) comprises five variables that characterise the social characteristics of the ward population, these being: the proportion of those of pensionable age living alone, the proportion of dependents in single carer households, the standardised illness ratio (ages 0-74) for residents in households only, the standardised mortality ratio (ages 0-74), and the proportion of the economically active population that is unemployed. An alternative index of needs (NEEDNAC) is used to estimate psychiatry costs, this index taking account of the proportion of persons in lone parent households, the proportion of dependents in households with no carer, the proportion of the adult population permanently sick, the proportion of residents born in the New Commonwealth, the proportion of those of pensionable age living alone, and the standardised mortality ratio (ages 0-74). Both indices are centred around zero by subtracting the grand mean from the needs index for each ward.

The dataset is merged from several sources, including the 1991 census of population, NHS administrative data, the 1990/91 Hospital Episode Statistics, a database of all hospital inpatient episodes, and the 1990/91 Health Service Indicators package, which includes summary indicators of health authority performance. The data are available for each of the 4985 English electoral wards (although there are some missing data). Descriptive data are provided in Table 2. Wards are clustered geographically within 186 health authorities covering populations of around 250,000. Our interest is in assessing the contribution to observed performance made by these health authorities.

Table 2 around here

Methods

Aggregate model

For comparison purposes only, we first estimate a conventional multiple regression model for each performance indicator in which the population-weighted ward-level data are aggregated to the relevant health authority. The health authority is thereby considered the unit of analysis and the model takes the following form:

$$y_k = \beta_0 x_0 + \beta_1 x_{1k} + u_k ,$$

where y_k represents the population-weighted mean ward value of performance indicator y across all wards in the k-th health authority, x_0 is a constant and x_{1k} represents socio-economic conditions, as measured by the needs index x_1 in the k-th health authority. The intercept parameter β_0 can be interpreted as the level of performance realised in a health authority with average socio-economic conditions. For those performance indicators defined as ratios, a positive slope parameter β_1 can be interpreted as suggesting that worse socio-economic conditions are associated with levels of performance lower than expected given the age and sex standardised population characteristics of the health authority. For the performance indicators GPPACS, DCRATE and WTLONG, which are not defined as ratios, a positive β_1 parameter implies that more challenging socio-economic conditions are associated with poorer performance as measured by these indicators.

The term u_k is the random error for the k-th health authority, assumed to have zero mean and constant variance (σ_u^2). We can interpret u_k as the parallel departure from the mean regression line of the k-th health authority. Small estimated values of u_k indicate health authorities with close to average performance, after controlling for the socio-economic situation of its wards. Large positive (negative) estimated values of u_k represent health authorities with an observed performance markedly below (above) that predicted on the basis of socio-economic conditions. If there is a large variation in u_k, this suggests marked differences in performance among health authorities. Health authorities can be ranked according to the value of u_k. Interest lies in estimating the parameters β_0 , β_1 and σ_u^2 . The model is estimated by ordinary least squares using STATA 7.

This formulation ignores the hierarchical structure of the data by taking means (weighted for population size) for y_k and x_1 over all wards within a health authority (Rice and Leyland, 1996). There are a number of drawbacks to this. First, we fail to make full use of the available information. Second, there is a danger of committing aggregation or ecological fallacy, in which a relationship found at the aggregate level may not exist among the units or individuals from which the data have been aggregated. For example, the average wait for hospital admission for residents of health authority may be no different from the national average. However, this may disguise the possibility that waiting times depend on where people live *within* the health authority, with people in one area of facing substantially longer waits and those in another area enjoying substantially better access than average. Third, failure to account for clustering results in underestimates of standard errors, which undermines significance tests.

Multilevel (ML) models

Recognising the hierarchical structure of the data, we next analyse the data at ward level. Wards are clustered within health authorities and these wards are likely to share closer similarities to one another than with wards elsewhere. For example, their shared geographical location may impinge on the levels of performance achieved. Importantly for estimation, this clustering implies that wards cannot be considered independent observations. Rather their common (perhaps unobservable) similarities imply correlation among wards within each health authority, and this correlation invalidates classical OLS estimation because the *iid* assumption is not met. To account for inter-dependence among wards, we define a two-level random intercept model as follows:

$$y_{jk} = \beta_0 x_0 + \beta_1 x_{1jk} + u_{0k} + e_{0jk} ,$$

where y_{jk} represents performance indicator y in the j-th ward within the k-th district, x_0 is a constant and x_{1jk} represents the needs index x_1 in the j-th ward in the k-th health authority. The size of wards varies considerably (range 2,041 – 33,073 people) so estimation incorporates population weights. In this context, weighting must be applied to the error components also (Hauck *et al.*, 2003). The weights applied at ward-level are calculated as $w_j = \dfrac{\frac{1}{\sqrt{n_{jk}}}}{ave\left(\frac{1}{\sqrt{n_{jk}}}\right)}$ and those at health authority level as

$w_k = \dfrac{\frac{1}{\sqrt{n_k}}}{ave\left(\frac{1}{\sqrt{n_k}}\right)}$ where n_{jk} is the population in ward j in health authority k, n_k is the population in health authority k, and *ave*(.) denotes the average across the quantities contained in parentheses.

The terms u_{0k} and e_{0jk} are error components such that u_{0k} relates to the k-th health authority (interpreted as the k-th health authority's level of relative performance) and e_{0jk} is the random error for the j-th ward within the k-th health authority. Both error components are assumed to have zero mean and constant variance ($\sigma^2_{u_o}, \sigma^2_{e_o}$). We place the same interpretation upon u_{0k} as we did on u_k under the earlier formulation. but the two estimates are notably different. Most importantly, the likelihood is that $\sigma^2_{u_o} \geq \sigma^2_u$ because, the greater the extent of clustering, the greater will be the underestimation of standard errors by OLS models (Rice and Leyland, 1996). The ML formulation may also lead to a change in the relative ranking of health authorities, depending on whether ward-level effects differ across health authorities. The influence of health authorities is likely to be conditional upon which indicator is considered. The intra-class correlation coefficient, ρ_u, is used to assess the proportion of total variance attributable to health authority effects, and is calculated as:

$$\rho_u = \sigma^2_{u_o}(\sigma^2_{u_o} + \sigma^2_{v_o})^{-1}, \qquad 0 < \rho_u < 1$$

Larger values ρ_u of are indicative of greater potential for the health authority to influence the value of the relevant performance indicator.

The computer package Mlwin version 2.1 is used for estimation(Rasbash *et al.*, 2000).

Multivariate multilevel (MVML) model

The multilevel model described above is calibrated separately for each performance indicator. This ignores the possibility that levels of achievement might be correlated across indicators. This correlation may be positive, if progress against one indicator simultaneously advances another, perhaps because good management promotes all-round performance. But the correlation may be negative if trade-offs are involved, such as when scarce resources that might be employed to achieve one objective are re-directed to pursue another. Analysis that recognises the possibility of simultaneity in the pursuit of multiple objectives will provide a more balanced picture of organisational achievement than a piecemeal analysis.

The multilevel framework can be extended to consider multiple outcomes simply by recognising that the performance indicators themselves are clustered, in this context within wards. By considering the performance indicators as the lowest tier in the data hierarchy, the possibility of within-ward and within-health authority correlation among indicators can be assessed. Thus the multivariate multilevel model is actually conceptualised as a three-level multilevel model, in which the set of I performance indicators (level 1) are clustered within wards (level 2), which are themselves clustered within health authorities (level 3). The model is given as:

$$y_{ijk} = x_0 \sum_{t=1}^{I} \beta_{0t} z_{it} + x_{1jk} \sum_{t=1}^{I} \beta_{1t} z_{it} + \sum_{t=1}^{I} u_{0tk} z_{it} + \sum_{t=1}^{I} e_{0tjk} z_{it}, \qquad t = 1,\ldots,I$$

where

$$z_{it} = \begin{cases} 1 \; for \; t = i \\ 0 \; otherwise. \end{cases}$$

Thus, y_{ijk} is the i-th performance indicator for the j-th ward clustered within the k-th health authority, and t indexes the set of performance indicators. The z's are dummy variables used to distinguish between each of the performance indicators.

The error terms u_{0tk} and e_{0tjk} are both assumed to have zero mean and constant variance ($\sigma^2_{u,i}, \sigma^2_{e,i}$). e_{0tjk} represents the random error for performance indicator i in the j-th ward, which is then summed across all performance indicators. The health authority residuals, u_{0tk}, are aggregated across the set of performance indicators to provide an estimate of overall performance.

The covariance for the *i*-th and *p*-th performance indicators within a health authority is given by:

$$\mathrm{cov}(u_{0ik}, u_{0rk}) = \sigma^2_{u,ip} .$$

These estimates of covariance can be used to calculate the degree of correlation *r* among performance indicators within health authorities:

$$r = \frac{\sigma^2_{u,ip}}{\sqrt{\sigma^2_{u,i} + \sigma^2_{u,p}}} .$$

If the correlation is positive, this implies that a health authority that has better than average performance for indicator *i* also has above average performance for indicator *p*. A negative correlation implies that above average performance for the one indicator comes at the expenses of poorer performance for the other. This correlation is interpreted as being due to unobservable influences on performance, such as the "managerial competency" of the health authority or the shared influence of environmental conditions.

If there are correlations among performance indicators, the residuals from the ML models, $u_{0k,}$ and the residuals from the MVML model, u_{0tk}, will differ for the same performance indicator *i*. We conduct a Likelihood Ratio Test to determine whether the correlations among residuals are jointly zero or not. The test statistic is given as:

$$\lambda = 2\left(LLF_{MVML} - \left(\sum_{t=1}^{I} LLF_{ML} \right) \right), \qquad t = 1,\ldots,I.$$

where $\mathrm{LLF}_{\mathrm{MVML}}$ is the log-likelihood function for the multivariate multilevel model, and $\mathrm{LLF}_{\mathrm{ML}}$ is the log-likelihood function for a multilevel model applied to a single performance indicator. Asymptotically, λ is distributed as the chi-square distribution with 156 degrees of freedom (because we have (13^2-13) zero correlation assumptions imposed on the restricted model). A significant test statistic indicates that estimation as a MVML model is preferable to separate estimation of a set of ML models, and implies the presence of correlation among performance indicators. (Note that the test cannot be used to identify the particular performance indicators among which the correlation exists).

Results

Model parameters

Estimation results pertaining to the model parameters β_0 and β_1 are shown in Table 3. For the majority of performance indicators defined as ratios, β_1 is significantly positive, implying that areas with worse than average socio-economic conditions are likely to have levels of achievement below that expected given the age-sex standardised characteristics of the population. This suggests that age-sex standardisation alone is insufficient to capture the health care requirements of the population.

Table 3 around here

Parameter estimates are in close agreement for the ML and MVML models, but differences are evident between these models and the aggregate OLS model. For example, the aggregate model estimates a negative β_1 coefficient for waiting times for routine surgery (WTSURG) and the percentage of people waiting in excess of 12 months for hospital admission (WTLONG), implying that as socio-economic characteristics of the area worsen, performance against these indicators improves; and a negative, though insignificant, β_1 coefficient for the day case rate, suggesting that as socio-economic conditions worsen, more patients are treated on a day case basis. All three associations are contrary to expectations and the ML and MVML models find that these relationships hold in the opposite direction. These differences imply that the relationships estimated by the aggregate model are contaminated by ward-level effects.

For both the ML and MVML models, the estimates of the β_1 coefficients are significant at the 5% level for all indicators (with the exception of WTRADIO under the ML specification), and they are positive for all indicators except for WTRADIO, GPACCS and ELECTEPS. The positive β_1 parameter estimate for a ratio indicator such as SMR064 indicates that areas with poorer socio-economic conditions have higher death rates among those under 64 years than would be expected solely on the basis age-sex standardisation of the population. For GPACCS, ELECTEPS, DCRATE and WTLONG worse socio-economic conditions in the ward are associated with better accessibility to general practitioners (GPs), higher levels of elective surgery, a lower proportion of routine surgery undertaken on a day case basis and a higher percentage of patients on the waiting list waiting longer than one year.

Correlations across performance indicators

The Likelihood Ratio test comparing the ML and MVML models clearly rejects the null hypothesis of jointly zero correlations among the residuals ($8_{df=156}$=8649, p=0.000). This indicates that the MVML model improves inference by allowing explicitly for correlations among the performance indicators. The correlation coefficients for the health authority effects across the various indicators are presented in table 4. Coefficents with an asterix are significant at the 5% level.

Table 4 around here

For the ‘Health Outcome’ indicators, SMR064, SMR6574 and SIR074, we find a statistically significant weak positive correlation (ρ=0.19) between the residuals for SMR064 and SMR6574 but a negative correlation (ρ=-0.74) between SMR064 and SIR074. These correlations imply that in an area with above-average mortality rates for ages 0-64, mortality rates for ages 65-74 will be above average also but rates of chronic illness will be lower than average. There is a statistically significant positive correlation (ρ=0.37) between the two ‘Clinical Quality’ indicators, DEATHS and EMOLD, implying that areas with a higher proportion of emergency admissions also report more deaths following hospital surgery. The three waiting time indicators, WTSURG, WTRADIO and WTLONG, are all positively correlated with an almost perfect correlation between WTSURG and WTLONG (ρ=0.95).

Health Authority Effects

Taking account of the hierarchical nature of the data and of the possibility of correlation among performance indicators may have an impact on the extent to which variations in performance can be attributed to health authorities. Table 5 provides details of the variance components for the three models. As can be seen, moving from an aggregated OLS model to the ML specification results in an increase in the variance, as expected. For the ML model ρ_u provides an indication of the extent to which variation in performance, after conditioning upon socio-economic factors, can be attributable to the health authority. This varies according to the performance indicator. It appears that health authorities have a substantial role in determining performance as measured by waiting times and day case rates, with around 70% of variation in performance against these indicators being attributed to health authorities. In contrast, the impact of health authorities on mortality rates is considerably less, explaining only 13% of the variation in mortality rates for the under 65s and 8% of that for those in the 65-74 age group. These estimates of health authority effects are little changed by the move to the MVML formulation.

Table 5 around here

Table 6 summarises the absolute differences between rankings generated respectively a result of taking account of hierarchical effects (by comparing the OLS and the ML models), correlation across indicators (by comparing the ML and MVML models) and in total (by comparing the OLS and MVML models). The impact of changing model specification varies according to the indicator, with average movements resulting from consideration of hierarchical effects ranging from 6.5 places (out of 186) for SMR6574 to 21.8 places for PSYCOST. With the exception of SMR6574, taking account of the hierarchical structure of the data has greater influence on rankings than does taking account of correlations across indicators, with $AVRG_{OLS\text{-}ML} > AVRG_{ML\text{-}MVML}$. The effect on rankings of these refinements to the specification may cancel out, so the overall impact may be less than that of the refinements taken separately.

Table 6 around here

Individual health authorities may be effected by these refinements more than others, and there appear to be substantial impacts for particular health authorities, as indicated by the maximum changes reported in Table 6. Compared to the aggregate OLS model, the MVML model produces a maximum change in ranking against each performance indicator that ranges from 31 to 108 places. The sensitivity of the relative performance of each health authority to specification decisions can be illustrated graphically. Figures 1 to 3 plot the difference in rankings for each of the 186 health authorities for three of the indicators, SMR064, WTRADIO and PSYCOST. The rank for the aggregate OLS model is plotted on the diagonal from the 'bottom left corner' (best performance) to the 'top right corner' (worst performance) of each figure. The ranks deriving from the ML and MVML models are indicated respectively by a diamond and triangle. The lines connecting these points depict the range in rank for each individual health authority, with longer lines indicating greater sensitivity in individual ranks to the choice of model specification.

Figure 1 shows that the estimated relative performance of the majority of health authorities in terms of standardised mortality rates for those under 65 (SMR064) is sensitive to model specification, with long lines throughout the series. General volatility is apparent for SMR6574, SIR074, DEATHS, WTSURG, WTLONG, and PSYCOST, while rankings appear fairly stable for EMOLD, GPACCS and DCRATE.

In contrast, when considering waiting times for radiotherapy (WTRADIO), volatility is much more in evidence for health authorities ranked poorly under the OLS specification, as illustrated by Figure 2. The same is true for the MATCOST and ELECTEPS performance measures. Figure 3, which plots variation in ranks for the cost of psychiatry (PSYCOST), illustrates that most of the variation in ranks stems from a failure to consider the hierarchical nature of the data. The ML and MVML ranks are in close agreement, suggesting that this performance indicator is not highly correlated with any of the others.

Conclusion

By failing to account for hierarchical data structures or correlations across performance indicators, analyses based on aggregate OLS models can lead to serious misrepresentations of relative performance. The extent of this misrepresentation varies according to the indicator, but some general observations can be made. First, it will greater for those indicators over which health authorities have limited influence, simply because OLS estimates will be more contaminated by random effects.

Second, misrepresentation will be less for those indicators that are highly correlated in a positive direction with other measures of performance. In such cases, joint analysis merely reinforces the assessment of separate (OLS or ML) models. The sensitivity of estimates of relative performance will be more substantial in the presence of negative correlations among indicators because achievement against one measure may be counter-acted by lesser achievement against another. This is likely to be an important source of error when analysing organisations with multiple conflicting objectives. We conclude, therefore, that consideration should be given to analysing organisation performance on small area or individual level data and by considering organisational objectives simultaneously.

References

Carr-Hill, R.A., Hardman, G., Martin, S., Peacock, S., Sheldon, T.A. and Smith, P.C. (1994) *A formula for distributing NHS revenues based on small area use of hospital beds*. York: Centre for Health Economics, University of York.

Gilthorpe, M.S. and Cunningham, S.J. (2000) The application of multilevel, multivariate modelling to orthodontic research data. *Community Dental Health*. **17**: 236-242.

Hauck, K., Rice, N. and Smith, P.C. (2003) The influence of health care organisations on health system performance. *Journal of Health Services Research and Policy*. **8**: 68-74.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M.*, et al* (2000) *A User's Guide to MLwiN, V2.1*. London: Institute of Education.

Rice, N. and Leyland, A. (1996) Mulitlevel models: applications to health data. *Journal of Health Services Research and Policy*. **1**: 154-164.

Yang, M., Goldstein, H., Browne, W. and Woodhouse, G. (2002) Multivariate multilevel analyses of examination results. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. **165**: 137-146.

Table 1: Performance indicators

PERFORMANCE INDICATORS	DESCRIPTION
HEALTH OUTCOME	
SMR064	Standardised mortality ratio for ages 0-64 *Ratio of observed deaths from all causes in an area to the expected equivalent given the local age/sex profile and national averages*
SMR674	Standardised mortality ratio for ages 65-74 *Ratio of observed deaths from all causes in an area to the expected equivalent given the local age/sex profile and national averages*
SIR074	Limiting long standing illness for ages 0-74 *Ratio of observed number of people reporting limiting illness in an area to the expected equivalent given the local age/sex profile and national averages*
CLINICAL QUALITY	
EMOLD	Emergency admissions of elderly people *Ratio of the rate of over 65 emergency admissions originating from an area to the expected given the age, sex and specialty of a patient and national averages*
DEATHS	Deaths following hospital surgery *Ratio of 30 day perioperative mortality after elective and non-elective surgery to the expected equivalent given the age, sex and case severity of a patient*
ACCESS	
WTSURG	Waiting time for routine surgery *Ratio of actual waiting time in days for routine surgery to the expected equivalent given the age, sex and specialty of a patient and national averages*
WTRADIO	Waiting time for radiotherapy *Ratio of actual waiting time in days for radiotherapy to the expected equivalent given the age, sex and specialty of a patient and national averages*
WTLONG	Percentage of those on waiting list waiting for 12 months or more *Proportion of elective surgery admissions waiting for more than one year standardised for patient characteristics*
GPACCS	Accessibility to general practitioners (GPs) *Indicator of relative accessibility given the supply of GPs, the distance to surgeries and the competition from local populations*
ELECTEPS	Number of elective surgery episodes *Ratio of standard surgery procedures originating from an area to the expected equivalent given the age, sex and specialty of a patient*
EFFICIENCY	
DCRATE	Day case rate *Proportion of elective episodes in routine surgery treated as day cases standardised for patient characteristics*
MATCOST	Maternity costs *Ratio of specialty specific fixed and variable costs for episodes to the expected equivalent given national averages*
PSYCOST	Psychiatry costs *Ratio of specialty specific fixed and variable costs for episodes to the expected equivalent given the age and sex of a patient and national averages*

Table 2: Descriptive statistics of performance indicators

Variable	Obs	Mean	Std. Dev.	Min	Max
SMR064	4972	0.997	0.286	0.173	2.453
SMR674	4972	0.997	0.231	0.229	2.424
SIR074	4972	0.990	0.210	0.416	2.467
EMOLD	4967	0.102	0.362	0.055	9.849
DEATHS	4967	0.997	0.422	0.000	258.650
WTSURG	4461	1.011	0.206	0.443	1.949
WTRADIO	4967	0.933	0.105	0.001	1.000
WTLONG	4967	7.376	2.988	0.792	23.913
GPACCS	4972	0.528	0.128	0.162	0.969
ELECTEPS	4972	1.000	0.432	0.000	3.137
DCRATE	4461	0.383	0.083	0.138	0.618
MATCOST	4967	0.981	0.387	0.000	5.595
PSYCOST	4967	0.994	0.521	0.065	6.190
NEEDAC	4972	0.000	0.745	-2.070	2.673
NEEDNAC	4972	0.000	0.103	-0.202	0.559

Table 3: Coefficient estimates[1]

	OLS models		ML models		MVML model	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_{0t}$	$\hat{\beta}_{1t}$
SMR064	0.996 (0.004)	0.297 (0.008)	0.992 (0.005)	0.349 (0.003)	0.992 (0.005)	0.352 (0.003)
SMR674	1.000 (0.004)	0.235 (0.007)	0.997 (0.004)	0.242 (0.003)	0.997 (0.004)	0.241 (0.003)
SIR074	0.989 (0.004)	0.403 (0.006)	0.990 (0.004)	0.373 (0.002)	0.990 (0.004)	0.373 (0.002)
EMOLD	1.018 (0.016)	0.226 (0.028)	1.019 (0.052)	0.210 (0.001)	1.019 (0.015)	0.206 (0.006)
DEATHS	1.000 (0.021)	0.215 (0.039)	1.002 (0.023)	7.502 (0.701)	1.002 (2.291)	7.228 (0.695)
WTSURG	1.009 (0.013)	-0.128 (0.022)	1.003 (0.014)	0.028 (0.003)	0.995 (0.014)	0.027 (0.003)
WTRADIO	0.928 (0.008)	-0.003 (0.014)	0.930 (0.007)	-0.001 (0.002)	0.930 (0.007)	-0.001 (0.002)
WTLONG	7.274 (0.172)	-1.671 (0.310)	7.221 (0.187)	0.403 (0.042)	7.220 (0.187)	0.426 (0.042)
GPACCS	0.525 (0.006)	-0.072 (0.010)	0.530 (0.005)	-0.086 (0.002)	0.530 (0.005)	-0.086 (0.002)
ELECTEPS	1.122 (0.025)	-0.249 (0.046)	1.134 (0.027)	-0.063 (0.008)	1.134 (0.027)	-0.061 (0.008)
DCRATE	0.389 (0.006)	-0.012 (0.010)	0.383 (0.005)	0.003 (0.001)	0.384 (0.005)	0.003 (0.001)
MATCOST	0.984 (0.022)	0.111 (0.040)	0.976 (0.021)	0.197 (0.006)	0.976 (0.021)	0.198 (0.006)
PSYCOST	0.995 (0.019)	2.016 (0.245)	0.976 (0.020)	3.725 (0.062)	0.977 (0.020)	3.678 (0.061)

[1] Standard errors in parenthesis

Table 4: Correlation of health authority effects from the multivariate multilevel model [1]

	smr064	smr6574	sir074	emold	deaths	wtsurg	wtradio	wtlong	gpaccs	Electeps	dcrate	matcost
Smr6574	0.193*											
Sir074	-0.738*	0.073										
Emold	-0.056	-0.006	-0.169*									
deaths	-0.032	0.326*	0.036	0.365*								
wtsurg	0.158	-0.049	-0.068	-0.080	-0.209*							
wtradio	-0.167*	-0.021	0.207*	-0.064	0.019	0.132*						
wtlong	0.197*	-0.100	-0.110	-0.081	-0.234*	0.945*	0.057					
gpaccs	-0.371*	0.475*	0.675*	0.008	0.309*	0.073	0.021*	0.036				
electeps	0.140	-0.173*	-0.051	0.017	-0.086	0.264*	0.160	0.304*	0.033			
Dcrate	0.134	-0.209*	-0.089	0.037	0.027	0.382*	-0.147	0.360*	-0.111	0.167*		
matcost	0.047	-0.128	0.055	0.100	-0.143	0.194*	0.252*	0.159*	0.025	-0.024	-0.164	
psycost	0.077	-0.111	0.134	0.019	-0.311*	0.264*	0.089	0.239*	0.165*	0.077	0.025	0.192*

[1] Correlation coefficients with an asterix are significant at 5% level

Table 5: Estimates of health authority effects

	OLS models	ML models			MVML model		
	σ_u^2	$\sigma_{u_o}^2$	$\sigma_{u_o}^2 + \sigma_{u_e}^2$	ρ_u	$\sigma_{u_{ot}}^2$	$\sigma_{u_{ot}}^2 + \sigma_{u_{et}}^2$	ρ_u
SMR064	0.003	0.308	2.311	0.133	0.317	2.318	0.137
SMR674	0.003	0.179	2.088	0.086	0.179	2.088	0.086
SIR074	0.002	0.214	0.731	0.293	0.215	0.732	0.294
EMOLD	0.044	3.943	10.635	0.371	3.950	10.644	0.371
DEATHS	0.083	9.231	16.818	0.549	9.255	16.841	0.550
WTSURG	0.026	0.032	0.042	0.762	0.034	0.044	0.762
WTRADIO	0.011	0.009	0.013	0.692	0.009	0.013	0.695
WTLONG	5.326	6.229	8.904	0.700	6.254	8.929	0.700
GPACCS	0.005	0.005	0.013	0.385	0.005	0.013	0.361
ELECTEPS	0.117	0.129	0.239	0.540	0.129	0.239	0.540
DCRATE	0.005	0.005	0.007	0.714	0.005	0.007	0.729
MATCOST	0.088	0.077	0.137	0.562	0.077	0.137	0.559
PSYCOST	0.065	0.069	0.190	0.363	0.068	0.189	0.358

Table 6: Absolute differences in rankings between OLS, ML and MVML models

	OLS – ML			ML – MVML			OLS – MVML		
	AVRG	STDEV	MAX	AVRG	STDEV	MAX	AVRG	STDEV	MAX
SMR064	21.5	18.2	79	11.1	10.5	47	24.9	23.6	101
SMR6574	6.5	5.6	33	12.7	10.3	52	14.6	11.6	58
SIR074	17.4	17.2	85	6.2	6.4	31	19.6	17.8	81
EMOLD	6.6	6.7	37	2.4	3.0	21	7.9	7.6	38
DEATHS	19.3	14.9	66	1.9	2.0	11	20.3	15.5	68
WTSURG	19.9	15.2	65	0.9	1.7	21	20.1	15.2	67
WTRADIO	8.5	9.9	51	1.2	1.1	5	8.5	10.0	50
WTLONG	21.4	17.0	77	0.7	0.9	4	21.7	17.3	79
GPACCS	7.3	5.7	30	2.8	2.5	11	7.9	6.2	31
ELECTEPS	21.7	17.9	91	4.9	11.0	100	20.2	16.9	73
DCRATE	7.3	5.8	32	0.5	0.9	5	7.3	5.8	32
MATCOST	9.4	8.7	48	0.5	0.8	4	9.5	8.8	48
PSYCOST	21.8	21.0	109	1.7	1.7	9	21.7	20.9	108

Figure 1: **SMR064 - Standardized mortality ratio (ages 0-64)**

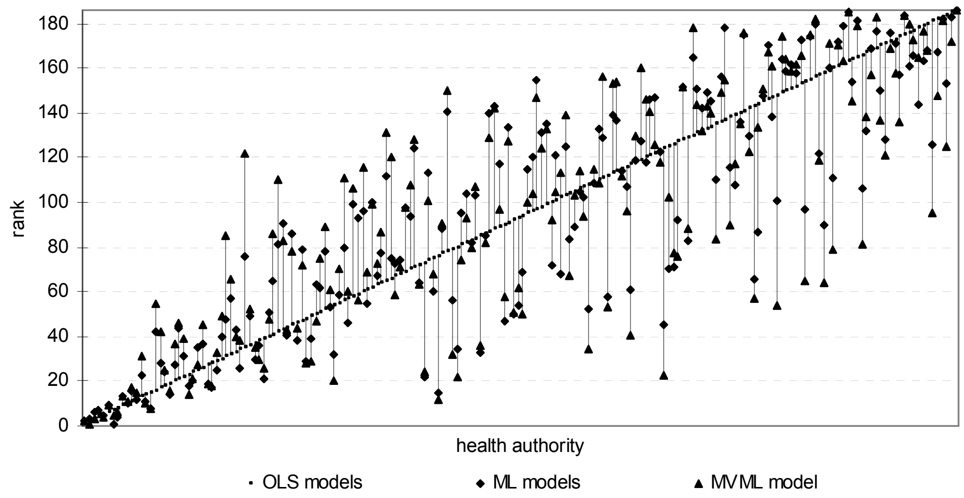

Figure 2: **WTRADIO - Waiting times for radiotherapy**

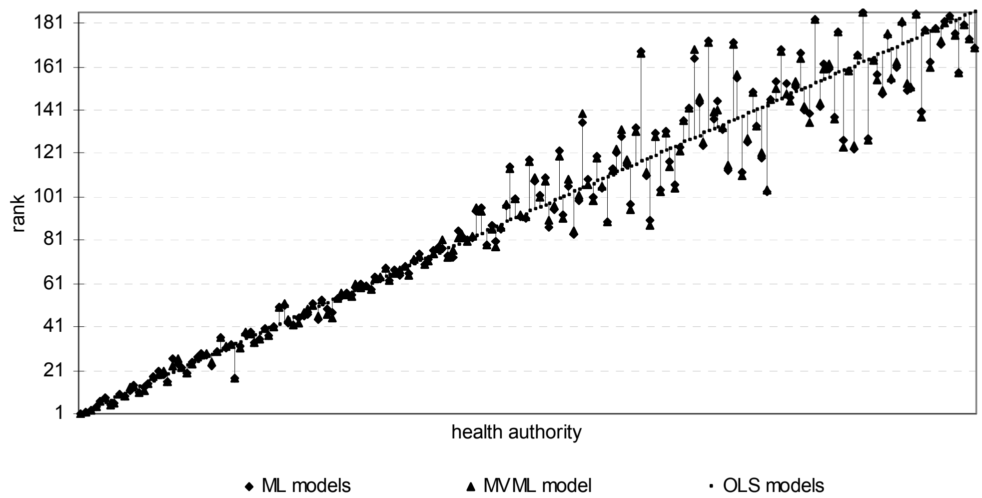

Figure 3: **PSYCOST - Psychiatry costs**