

Performance Measurement in Social Care: A Comparison of Efficiency Measurement Methods

Paul Clarkson and David Challis

Personal Social Services Research Unit (PSSRU), Faculty of Medical and Human Sciences,
University of Manchester, UK

Performance measurement in social care is now considerably more advanced than previously. However, measurement is criticised on the basis of its presentation as neutral when, in the UK, it is part of the government's regulatory regime. However, measurement is important, especially when alternative methods may bring about different rankings of authorities to those endorsed by the recent system. This paper explores this issue through analyses of cost efficiency in English social services authorities. It concludes that the picture of authorities' performance depends on the method chosen which, it is argued, should stem from the stated aims of performance monitoring.

Introduction

Measuring performance, quality and efficiency in social care has recently achieved prominence in the UK as well as in other countries. In the UK, a major thrust in the development of this approach was the White Paper *Modernising Social Services* in 1998 (Cm 4169, 1998). Although the 'modernisation' agenda for adult social care arising from this can be criticised for its conceptual foundations, the current empirical techniques for monitoring performance have their origin here. The development of indicators and associated techniques by which the government has judged the performance of social services authorities on a national basis has been charted in a number of publications (DoH, 1999; DoH, 2002a; DoH, 2003).

The government proposals for monitoring performance, set out in the White Paper, formed part of a regulatory regime whereby local authorities were held to account for their performance (Cm 4169, 1998: Chapter 7). Thereafter, improvements in the delivery of social care were driven by the establishment of standards by which the public could appraise local services, and by which the government could reward or sanction good or bad performance. Over time, the government has established a set of key indicators with the aim of forming a comprehensive overview of social services activity across the country. These include those of the Performance Assessment Framework (PAF) and a wider set of indicators contained in the Key Indicators Graphical System (KIGS), recently disseminated as a web-based application (DoH, 2005), as well as those for public consumption such as the 'star ratings' for social services (Commission for Social Care Inspection, 2004).

Correspondence to: Paul Clarkson, Personal Social Services Research Unit, Dover Street Building, University of Manchester, Oxford Road, Manchester M13 9PL, UK. Tel: 0161 275 5674; Fax: 0161 275 5790; E-mail: Paul.Clarkson@manchester.ac.uk.

These national data collections to monitor social care performance have developed considerably since the first set of indicators in 1988 (Warburton, 1988). However, the methods for analysing these indicators to provide evidence of 'good' or 'poor' performance are less developed. It is often assumed that current methods are the only ones possible and the nature of their application can be taken for granted. However, the implementation and presentation of the approach have been questioned. One criticism of this type of measurement, for example, is that it has been used as part of a 'presentational politics' by government (Cutler and Waine, 1997). Performance measurement in UK social care, in contrast to some other countries, has been implemented in an almost exclusively top-down manner. This has meant that control and regulation characterise the way in which measurement methods are adopted, rather than their use to evaluate professional decision-making at a more local level. This has meant that the credibility of the measurement approach amongst practitioners has suffered; a problem in the NHS as well as social care (Rowan and Black, 2000).

In addition, academic debate has criticised the performance approach for its reliance on measurement *per se* and for its associations with techniques for continuous improvement drawn from business practice, traditionally considered anathema to the values of social care (Lupton, 1989; Watson, 2002). However, measurement is important, not least because without it we cannot judge whether, and in what direction, social care has improved. Viewed from the perspective primarily of measurement (rather than outright condemnation of the approach's political basis), the issue therefore becomes one of the value of the measurement approach adopted. This paper considers three different measurement methods that can be used to monitor an important aspect of the performance of English social services authorities – efficiency. It argues that each of these methods paints a different picture of an authority's position regarding the overall efficiency of its social care services. The relevance of this for the stated *aims* of performance monitoring is discussed.

Monitoring efficiency in social care

Efficiency in social care services has been much maligned in previous periods (Williams and Anderson, 1975). Often erroneously confused with purely cost cutting, its promotion has been attacked by those who see it as generating perverse incentives to reduce costs at the expense of quality services or of professionals' well-being. It is, of course, but one dimension of performance along with others, such as the responsiveness of service delivery, effectiveness, choice for users and the consistency of provision across the country (Cm 4169, 1998). It is, however, a dimension of crucial importance to those, for example social services managers, who are charged with securing services, whilst taking account of resource scarcity. However, efficiency and its monitoring are also an important part of a focus upon users; decisions, and their cost consequences, imply the use of resources in one way rather than another. Therefore, inefficiency within an authority may mean some users are being denied resources from which they could benefit and which have been put to other, often inappropriate, uses. The search for greater efficiency therefore stems from the fact of finite resources, and the need for decisions to be made regarding how best to support people within these resources.

The efficiency issue has recently received sustained interest across public services (Gershon, 2004), and concerns about the efficiency of social services authorities have

been raised. Evidence from the *Joint Reviews of Social Services* (Audit Commission/SSI, 1998) and other research evidence (Baldock, 2004) have shown large variations between authorities in the costs of providing the same service with consequent scope for efficiency improvements. For example, the unit costs of home care provided by independent agencies in the period 2000/01 varied over fourfold between the lowest cost authority (a gross hourly cost of £4.09) and the highest (a gross hourly cost of £18.67) against an average unit cost of £9.05 per hour for England (DoH, 2002a). These variations have been seen as evidence of substantial differences in the way authorities manage their services in terms of their use of resources and in their commissioning practices. However, there are other factors not taken into account in this interpretation, such as errors in the data and unexplained sources of variation arising from particular authorities' social or economic positions. Reporting such data in terms of interpreting the efficiency of social services is therefore likely to be contentious and raise issues of the appropriateness of different methods of measurement.

There are a comprehensive range of sources discussing techniques for measuring efficiency in various areas of policy (Blades *et al.*, 1987; Ferlie *et al.*, 1989). Efficiency can be defined in one of two ways: the degree to which a given combination of resources is deployed to maximise outputs, or the degree to which inputs (and thus their costs) are minimised for a given output. In social care the latter definition has often been taken as the most useful since the pursuit of cost reductions (allowing for inflation) at current output levels is the way efficiency is normally understood. It is this definition that is used in the analyses to follow.

Measuring efficiency: three methods

First, when discussing proposals to improve the efficiency of social care, it is necessary to link measurement to the objectives of services (Williams and Anderson, 1975: 28). For example, regarding services for adults and older people one objective of the UK government contained within the *National Priorities Guidance* (Department of Health, 1998) and more recent policy (Cm 6499, 2005) is to promote independence by supporting people to live at home. The efficiency with which this fairly general aim is accomplished is monitored by some of the government's PAF indicators. In particular, PAF indicator B12 measures the cost of intensive social care for adults and older people (DoH, 2003).¹ This indicator is used to assess performance year on year against the government's efficiency targets.² The stated aim of comparative analysis on such an indicator is for authorities to 'benchmark' themselves against the practices of the best-performing authorities (DoH, 2003). However, it is argued here that for authorities to be assessed on their performance in a valid way requires a clarification of the methods for achieving this aim.

Evidence of above-average costs on this indicator is interpreted as indicating poor performance and requiring further scrutiny. However, this form of measurement has been criticised on the basis that it does not compare like with like. Representatives from local authorities, especially those that are signalled as poor performers, see the recent methods for comparative analysis as unfair. Such stand-alone cost indicators do not, it is argued, take into account the particular circumstances faced by some authorities, which make it difficult to reduce their costs (Revans, 2002). Three methods are presented here, using recent data, to explore this issue of fair comparison: the present unit costs method, and

two methods that model the data to control for factors influencing overall costs – multiple regression analysis and data envelopment analysis.

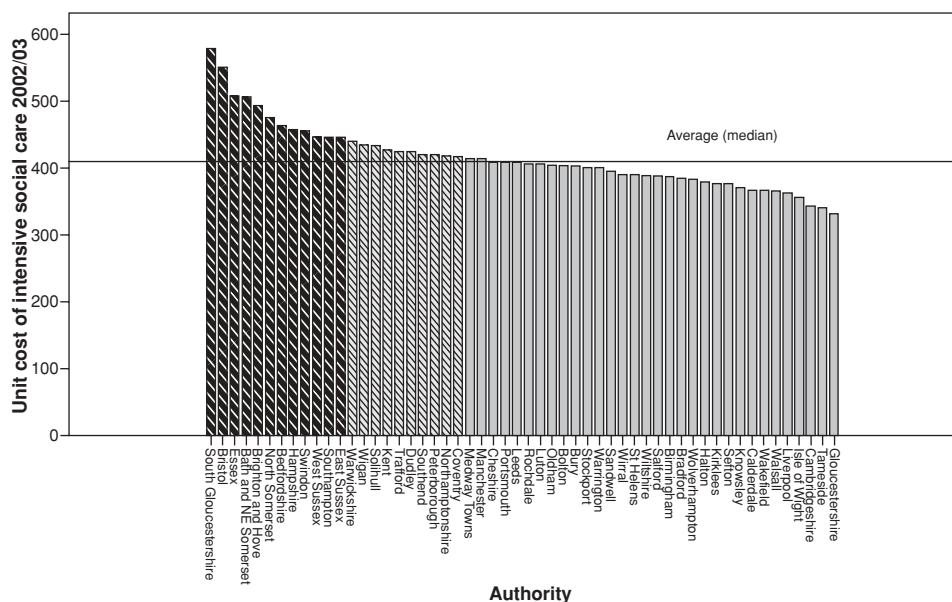
Unit costs: average performance

The recent method for judging efficiency across social services authorities investigates unit cost indices, in general and for specific services, to enquire whether an authority's costs are higher than the average (median) for its 'group' (DoH, 2003). The rationale behind this is that if costs are above average, this indicates that particular services are not being commissioned at efficient unit costs or that authorities are using services inappropriately, thus incurring higher costs than they need to. The standard of good performance for authorities is therefore costs which are average to low. Unit costs are compared within groups, which standardise for at least one shared characteristic thought to influence costs. Previously, authorities were grouped by type of authority (Metropolitan, Shire County, Unitary authority, Inner and Outer London) thought to exhibit regional variations in costs. Recently, however, cost indicators have been grouped according to 'Area Cost Adjustment (ACA) Groups', each containing a mixture of authority types. These groups reflect adjustments that have been made for the varying costs of providing services in different areas reflecting, between them, variations in average wage rates and the costs of non-domestic rates across the country (Elliot *et al.*, 1996).

Figure 1 examines the unit cost of intensive social care (PAF indicator B12) using recent KIGS data (DoH, 2005) for English ACA Group 2 authorities – a mixture of Metropolitan districts, Unitary councils and Shire counties (54 councils in total). Here authorities are banded according to the Department of Health's 'traffic light' system, signalling the extent to which authorities deviate from the average (DoH, 2003: Annex B).³ By this method, 12 authorities are classified as poor performers and are designated as raising questions about efficiency.

As stated, however, there are objections to this approach: it uses an average standard by which to judge efficient performance. However, by only standardising costs according to one input at a time (the influences characterising the cost 'group'), it cannot allow for other factors that may drive expenditure. These factors may include the population or dependency characteristics of the people served and the policies and procedures of each authority (Levitt and Joyce, 1987). The method also cannot take account of the multiple nature of such inputs characteristic of real-world operations. Social services authorities may, for example, seek efficiency improvements by attempting to secure the lowest unit costs of home care or residential care or they may, in addition, promote more balanced provision away from residential care as a way of improving their overall efficiency (DoH, 2003). Improvements to the *quality* of services are also important, such as ensuring that independent agencies are rewarded for complying with standards, which can have resource consequences (Audit Commission, 1997). In addition, some authorities may struggle to reduce their costs because they face especially difficult circumstances, such as a high level of deprivation in their population. The unit cost method provides no way of simultaneously allowing for these multiple influences.

The following two approaches attempt to address this issue by modelling the operation of social services authorities in delivering their services. Examinations of these models can bring about a fuller understanding of the factors involved in influencing costs and lead to rankings of efficiency that take into account the particular circumstances facing



Performance rankings




-  Poor - ask questions about performance (costs higher than average)
-  Acceptable – room for improvement (costs near average)
-  Good (costs average to low)

Figure 1. Average performance – the unit cost of intensive social care for English Area Cost Adjustment Group 2 authorities 2002/03.

Source: Department of Health (2005), PAF indicator B12.

management in each authority. These approaches are outlined here as alternatives to the single factor approach pursued recently and offer more sophisticated standards by which to judge performance.

Regression analysis: average expected performance

This approach models the variation in costs across authorities through the specification of a functional form for the data, estimated using regression analysis. In terms of the input orientation, above, the approach specifies a multiple regression equation that attempts to explain, as much as possible, the variation in costs in terms of the variation in outputs and a number of other explanatory variables (Knapp, 1998). In other words, the model provides the best statistical explanation for the variation in costs given the average impact, across authorities, of differences in outputs and other influences (Levitt and Joyce, 1987). This model therefore operates a standard of ‘average expected performance’ (shown by the regression line fitted to the data) against which each authority can be compared.

By comparing authorities against this standard, we can consider whether a particular authority's costs are more or less than we would expect given the conditions under which it operates.

This method demands a choice as to the factors to include in modelling costs across authorities. As an example, Table 1 sets out the variables chosen here as characterising important influences on, once again, the average gross weekly expenditure per person of intensive social care for adults and older people (PAF indicator B12), for which data were available. In routine performance evaluation, such explanatory variables could be chosen after discussion (for example, between managers) of the most likely influences on expenditure, bearing in mind those factors thought to be crucial in informing debates about performance. Ideally, however, reference should also be made to the research literature in choosing variables characterising important influences upon costs. With this in mind, Table 1 outlines the variables chosen here with the rationale for their inclusion in the model and their expected relationships to overall costs. However, in terms of a satisfactory 'fit' for the model, the criteria are those of parsimony and explanatory power; two criteria that may work against each other. That is, we require a model that, as far as possible, contains the smallest number of variables that individually relate significantly to costs but that together give the best explanation of the variation in costs across authorities (Levitt and Joyce, 1987; Knapp, 1998).

The variables chosen in the model include those that are under the direct control of management, such as the outputs of care (including their *quality*), and those not under management control, such as the characteristics of authorities and the extent of need in their populations. In making this distinction, we are including variables reflecting processes that may be changed in the short term, in pursuing greater efficiency, and also those that cannot be changed, reflecting circumstances that may prove especially difficult for authorities in containing their costs.

Table 2 presents the results of the regression analysis for the same group of authorities as above using these variables. The variables reported are those emerging as statistically significant from the model. Statistics are also presented indicating the relative strength of each of the influences on costs and the fit of the model to the observed data (Hutcheson and Sofroniou, 1999). Differences in the number of people supported by authorities are allowed for by focusing on average expenditure per person. As one example, the results show that, on average, a one percentage shift in care away from residential provision is associated with a reduction in costs of over three pounds per person, holding other factors constant. Including variables reflecting authorities' circumstances and extent of need shows that higher costs are associated with the proportion of households renting purpose-built flats (an indicator of deprivation). However, the association between costs and a need variable – the percentage of older people living alone – is negative, which is the opposite relationship to that expected. These relationships, however, are those modelled by the criterion of average performance adopted here, not necessarily those that might be considered indicative of best practice. They may, for example, be suggestive of less than optimal practices by the authorities concerned or of data errors. These factors, indicating authorities' outputs and circumstances, together account for around 27 per cent of the variation in overall costs.

Once the model has been estimated, the efficiency of a particular authority can be measured against the residuals from the regression equation (the difference between actual and expected costs). However, in order to rank authorities some strategy is needed,

Table 1 Rationale and measurement issues associated with choice of explanatory variables in regression model (dependent variable, cost of intensive social care – PAF indicator B12)

Explanatory variable	Rationale	Expected relationship to overall costs
No. of households receiving intensive home care as percentage of residential/nursing home care and intensive home care (PAF B11)	A main indicator of output reflecting the local 'balance of care' (Mooney, 1978; McCallion, 1993) assuming that older people prefer to remain at home and that this is more beneficial for their welfare (Harding, 1999)	As authorities shift the balance of care away from residential to home care, overall costs should fall
Percent of older people living alone (SN 130)	A proxy measure of need for services (Bebbington and Davies, 1980)	Authorities with a greater proportion of their older population living alone (the greatest users of social care services) will tend to incur higher costs as they will require more input from services
Number of contact hours per household (AA68)	A measure of the level of home care (Gorbach and Sinclair, 1989)	As authorities provide more intensive home care to their populations this will tend to raise costs. Intensity is also based on whether population is considered in greater need
Number of hospital admissions per head of population over 75 years – falls and hypothermia (PAF C33)	A measure of need but also health provider behaviour, indicating the extent to which services are provided outside social care and the demands made upon social services from hospital discharge (Hudson, 2002)	A higher number of admissions may mean costs rise as this is indicative of a more needy population and/or an increase in referrals from the health sector. However, costs of social care may also fall as services may be reduced once older people enter hospital
Percentage of single older people who are admitted to residential care allocated single rooms (PAF D37)	An indicator of quality conforming to standards in residential care (Centre for Policy on Ageing, 1984; DoH, 2002b)	Pursuing quality standards often raises costs as resources are committed to providing a quality service rather than a basic, less costly, one
Percent of households renting purpose-built flats (SN 123)	A proxy indicator of local deprivation (Davies <i>et al.</i> , 1971)	In more deprived areas more inputs may be needed to deliver care thus raising costs

Note: Numbers refer to numbered indicators in *Key Indicators Graphical System* (Department of Health, 2005).

Table 2 Regression analysis of average expected performance – cost of intensive social care for English Area Cost Adjustment Group 2 authorities 2002/03 – PAF indicator B12

Variable	Coefficient (β)	Standard error of coefficient	P
Constant	632.96	82.93	<0.001
Intensive home care as percentage of residential/nursing home care	-3.31	0.87	<0.001
Percent of older people living alone	-5.33	2.74	0.05
Percent of households renting purpose-built flats	5.62	2.73	0.04

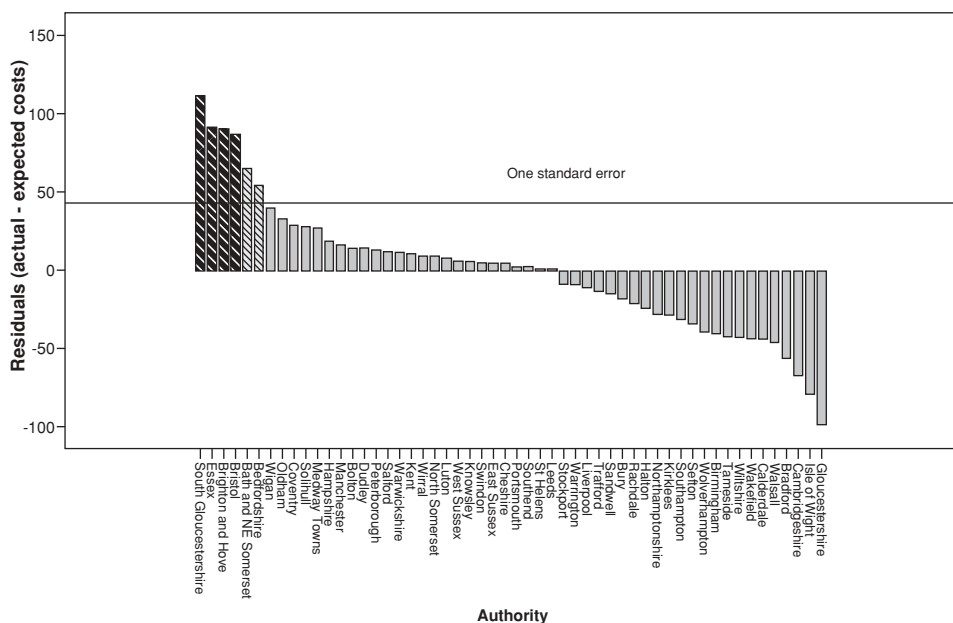
Notes: All variables are those emerging as significant from the model at $p < 0.05$. Model fit: $R^2 = 0.27$; $F = 6.05$; $P = < 0.001$; Specification tests: Normal distribution of variables confirmed from frequency histogram of residuals and normal probability plots; low degree of co-linearity confirmed from tolerance factor (> 0.02) and variance inflation factor (< 5.0) for all explanatory variables. Source: All variables, Department of Health (2005).

as in the unit cost method, for judging the extent to which particular authorities deviate, in this case from average expected performance. The degree of deviation raises questions for further enquiry. In regression analysis a statistical criterion may be employed on the distribution of residuals so as to rank authorities' performance. Here, drawing on Levitt and Joyce (1987), we take as a means for identifying particularly 'deviant' authorities those that fall outside a range on either side of expected costs of two standard errors (a measure of the 'spread' of the residuals). This follows that, with a normal distribution of residuals, we would expect 95 per cent of authorities to have actual costs within twice the standard error of the regression equation (Levitt and Joyce, 1987: 118). However, when cost minimisation is the objective, only those authorities with costs greater than expected are signalled here as inefficient. Those authorities with significantly lower costs than expected are signalled as examples of especially efficient authorities.

Of course, this is but one criterion for judging authorities' deviation from expected costs. It has the merit, however, of identifying outlier authorities whose costs differ in a statistically significant way from what would, on average, be expected given their existing practices. Figure 2 presents rankings of the authorities on these residual values using this method. By this method, only four authorities are classified as poor performers, thereby raising questions about their efficiency. Essex, for example, spends, on average, almost £91 per person more than expected, given its output and circumstances in relation to other authorities. Authorities such as Bradford and Cambridgeshire spend less than would be expected, given the factors by which all authorities are compared here.

Data envelopment analysis (DEA): best-practice performance

This approach calculates best-possible performance, in relative terms, not by specifying a form for the data but by comparing the output profile of each authority with a set of authorities with the same output profiles, but with different values of the input (cost) variable (Stone, 2002). Put simply, the approach analyses data to search all authorities in order to answer the question: what combination of outputs actually produced by



Performance rankings

- Poor - ask questions about performance (costs greater than two standard errors)
- Acceptable – room for improvement (costs greater than one but less than two standard errors)
- Good (costs less than one standard error or lower than expected)

Figure 2. Average expected performance – cost of intensive social care residual values from regression model for English Area Cost Adjustment Group 2 authorities 2002/03.
 Source: All variables, Department of Health (2005).

authorities with different influences entail the least cost? The technique was originally described by Charnes and colleagues (1973) drawing on the influential work of Farrell (1957) on the efficiency of US agricultural production. It is a mathematical technique that has been used in a number of applications, from health care to measuring the efficiency of education authorities and police forces (Sherman, 1984; Levitt and Joyce, 1987; Ozcan *et al.*, 1992; Ehreth, 1994; Jacobs, 2001; Thanassoulis, 1995). The only applications of the technique to UK social care have been by Gibbs and Smith (1989), who studied value for money in nursing homes and Jiménez *et al.* (2003), who investigated productivity changes among a sample of English county councils, following the community care reforms of the early 1990s. A description of the input-minimising model can be found in Boussofiene *et al.* (1991).

The technique constructs a best-practice frontier from the actual operation of authorities, comparing each authority with the standard of the best actually achieved, rather than a theoretical notion of perfect efficiency (Farrell, 1957: 255). The approach uses

the technique of linear programming to search all authorities (characterised as Decision Making Units or DMUs) offering at least as good a level of output as the one being observed in order to find the most efficient authorities: those with the lowest costs for that output level. Each observed authority, in turn, is compared against the best-practice frontier formed by these authorities. In our input orientation, each authority is then given an efficiency score indicating its minimum possible costs as a percentage of its actual costs, given outputs. By using the frontier to establish best practice, this method also enables targets to be set for each inefficient authority, signalling the degree to which their outputs and practices can be changed to emulate those of the most efficient authorities.

This method, unlike regression, can take account of multiple inputs as well as multiple outputs but can still allow for environmental influences outside the control of management, effectively treating them as either fixed inputs or as favourable ('more is better') or unfavourable ('more is worse') outputs (Stone, 2002). Different efficiency frontiers operate for constant (CRS) and variable returns to scale (VRS) (Banker *et al.*, 1984). For social care, it is likely that variable returns to scale will operate, in particular that authorities will exhibit decreasing returns to scale; that is, as inputs increase there is a rise in output but at a diminishing rate due to added bureaucracy and other factors.

A number of issues emerge in relation to using this technique for performance evaluation, one of the most important being how to rank authorities by this method. By definition, using this method, all authorities lying on the efficiency frontier are considered efficient and all those away from the frontier are relatively inefficient. Consequently, inefficient authorities cannot be identified as those falling outside a certain range, since diagnostic tests on residuals are not available (Burgess and Wilson, 1993). Techniques for detecting outliers when using this method have been employed (Andrews and Pregibon, 1978; Wilson, 1993; Dusansky and Wilson, 1994). However, as well as the fact that these techniques employ technical methods unfamiliar to those conducting performance analyses in this setting, they are also used primarily to detect extreme or influential observations to correct sources of error rather than to rank units on their performance; they thus may be inappropriate as a means of ranking authorities in terms of their degree of performance. A simpler and more suitable method for ranking authorities using DEA follows that described by Edvardsen *et al.* (2003) in their study of home care and nursing provision in Norwegian municipalities. In that study, all authorities calculated as inefficient in DEA are classified as poor performers as they all show relative room for improvement in terms of conserving their costs. Efficient authorities, however, may be classified as either *active peers*, who provide a reference point for at least one inefficient authority by providing targets for them to meet, or as *self evaluators*, who are not referenced by any other authority and who are thus 'alone in the crowd' (Edvardsen *et al.*, 2003: 15). Self evaluators may be considered as supreme examples of best-practising authorities without comparison elsewhere; they do not act as benchmarks for any other authority, having unrivalled practices that are *sui generis*.

Figure 3 shows rankings of the authorities considered here using the DEA method to calculate efficiency scores under the assumption of VRS and employing the above classification.⁴ This ranking uses the same variables as those in the above regression model, treating those variables not under management control (need and deprivation indices) as 'unfavourable' outputs thought likely to raise costs. Here, authorities are compared on their efficiency ratios, giving the minimum possible costs as a proportion of actual costs for each authority. These scores therefore indicate the potential for resource

Table 3 Comparison of rankings for those authorities identified as inefficient by recent national performance monitoring (English Area Cost Adjustment Group 2 authorities 2002/03)

Authority	Unit cost average	Regression analysis	DEA
South Gloucestershire	1	1	–
Bristol	2	4	1
Essex	3	2	3
Bath and NE Somerset	4	5	2
Brighton and Hove	5	3	4
North Somerset	6	–	11
Bedfordshire	7	6	–
Hampshire	8	–	26
Swindon	9	–	5
West Sussex	10	–	6
Southampton	11	–	7
East Sussex	12	–	8

Notes: Rankings, 1 = most inefficient (most room for improvement); – identified as relatively efficient in the model.

in purpose-built flats (6 per cent rather than its existing 8 per cent). In this way, this method signals scope for individual authorities to improve their efficiency to match the practices of the best performing, either by changing their outputs or by drawing attention to specific circumstances that may be improved to enable them to conserve costs.

Comparing the methods

The above techniques of efficiency measurement are compared in Table 3, which contrasts the rankings of those authorities classified as inefficient by the government's recent performance monitoring scheme under the assumptions of each method. Only four authorities – Bristol, Essex, Bath and NE Somerset, and Brighton and Hove – are classified as inefficient by all three methods. Six authorities, classified as inefficient by recent national performance monitoring, are ranked as relatively efficient using regression analysis, whereas two of these authorities are classified as efficient using DEA. The ranking of authorities is inconsistent when compared across methods; no authority would achieve the same position on efficiency if the results of each technique were to be used in forming league tables of performance. In general, DEA identifies more authorities as inefficient than either of the other two methods, with the unit cost method being rather indiscriminate in identifying 'acceptable' performers, with most authorities being classified as acceptable or good.

Conclusions

This analysis has focused on the recent method for assessing comparative efficiency between English social services authorities and contrasted this with alternative methods

that rely on modelling the operation of these authorities. These additional methods provide a contrast to the relatively simple, and some would say *simplistic*, analysis of performance data arising from the performance agenda in UK social care. Commentators such as Smith (1990) and Goldstein and Spiegelhalter (1996) have already drawn attention to the limitations inherent in average rankings of performance data and 'league tables' in other areas such as health and education, where only single unadjusted indicators have been used. The analyses presented here confirm their suspicions that such raw indicators offer only a partial view of authorities' activity. Recent methods are open to the charge of failing to take account of other influences and drivers of performance, which may lead to contaminated judgements from both government and the authorities themselves: an issue not just for social care but for other areas such as education (Wiggins and Tymms, 2002).

The results of the analyses confirm that an authority's position on a ranking of performance depends on the method chosen. Moreover, they suggest limitations using the average as a criterion by which to judge performance. Taking an average standard by which to judge efficient performance is fundamentally limited as, in most spheres of activity, there will always be units (authorities, hospitals, teams or individuals) which lie away from the average. Since variation is inevitable, ascribing the status of 'poor performers' to those units towards the end of the distribution misrepresents the nature of data and may lead to erroneous conclusions (Adab *et al.*, 2002). For more effective monitoring, it is more relevant, drawing on the industrial predecessors of performance measurement, to use methods that identify 'special causes' and distinguish these from 'common causes', or random variation in the system (Shewhart, 1931; Deming, 1986). In these methods, a statistical criterion is used (such as identifying units lying outside a range of three standard deviations) and the practice of outliers may then be interrogated to form judgements as to the reasons for inefficient performance. Regression analysis, used in the above examination, similarly uses a statistical criterion to set a standard of efficiency; only those authorities lying outside a range of expected performance are signalled as raising concern. In contrast, the DEA method seeks out those authorities able to act as yardsticks for best practice; it concentrates on the actual use of inputs and outputs in the authorities being observed and determines best practices for others to follow. The choice as to which of these methods is most useful in examining social care performance depends on the purpose of monitoring authorities in this way. It is clear, however, that different methods produce very different rankings of performance.

This analysis is intended to be more than a sterile technical exercise. Studies in other settings have compared performance measurement methods, such as the use of single ratios (as in the recent English PAF), regression analysis and DEA (Nunamaker, 1983; Giuffrida and Gravelle, 1999; Jacobs, 2001). Other modelling approaches are also available such as stochastic frontier analysis and multi-directional efficiency analysis (MEA) (Hougaard *et al.*, 2004). However, ideally, the precise aims of performance monitoring should underlie the basis of any such comparisons. The rationale for performance measurement in social care, as stated in the *Modernising Social Services* White Paper, was for social services authorities to 'drive up their standards to match those of the best' and to 'secure continuous improvements in performance . . . to deliver services which bear comparison to the best' (Cm 4169, 1998: paras 7.3 and 7.14). Recent performance measurement approaches in other settings, and in other countries, have articulated a similar basis for measurement (Brook, 1994; Mor *et al.*, 2003). If such best-practice benchmarking (Camp, 1989) is the foundation on which performance

measurement is to take place, then the use of an average standard by which to assess activity is inappropriate. As Ehreth (1994) has observed, 'an average is not an optimum', and attempts to assess authorities' against the average cannot hope to uncover practices indicative of the best-performing authorities. Only the DEA method attempts to measure efficient performance with reference to the best performing (under current constraints), and it is the only method able to offer targets for authorities to improve their practice.

There are, of course, a number of difficulties to address before these more sophisticated analytical tools can be employed in the UK social care setting. Much depends on the quality of the routine data available, and the degree of care taken in building models of the care process operating across authorities (Smith and Street, 2003). The precise position of authorities on a ranking of performance will depend, to a large extent, on the choice of variables included in the model: issues of minimising sources of error such as missing or incomplete data or a lack of data on some factors considered important in judgements of efficiency become crucial. A judicious choice of relevant variables is therefore required. In this study, although relevant explanatory variables were used, drawn from research, the degree of variance explained by the final regression model was modest. This reflected, in part, the difficulty in obtaining data reflecting all possible sources of variation in costs, such as those indicating health authority behaviour, labour market pressures and service quality. Data reflecting all these processes are unavailable in national packages. Additionally, the relationship of one variable (need) to costs was opposite to that expected, which may reflect data error or less than favourable practices within authorities. Also, applying these models in practice (particularly DEA) may be perceived by authorities as a rather technical affair, which may meet with resistance as there is a need to be clear about the reasons for their classification as efficient or inefficient.

However, important as these arguments are, they should not be allowed to obscure the central issue as explored here. If the need for more effective performance measurement is accepted, then its aims must be clearly specified and methods adopted which are closely aligned to these aims. This requires a more methodical approach than that recently adopted in the UK.

Acknowledgements

The Personal Social Services Research Unit receives funding from the Department of Health. The views expressed in this paper are those of the authors and not necessarily those of the Department of Health.

Notes

1 The definition of this indicator is the 'average gross weekly expenditure per person on supporting adults and older people in residential and nursing care and providing intensive home care'.

2 A recent efficiency target for 2002/03 for example was a 2.5 per cent improvement in efficiency at a national level (DoH, 2003). Evidence from the indicator used here can contribute to an assessment of the progress of authorities on this target.

3 The thresholds for these bandings are determined from 2001/02 distributions of authorities' costs (uprated by inflation). Thus, because these thresholds were fixed at a certain point it is possible for authorities to move into a higher band as they seek improvements year on year.

4 The DEA models were calculated using Warwick Windows DEA Version 1.02 (Thanassoulis and Emrouznejad, 1996).

References

- Adab, P., Rouse, A.M., Mohammed, M.A., and Marshall, T.** (2002), 'Performance league tables: the NHS deserves better', *British Medical Journal*, 324, 95–98.
- Andrews, D.F. and Pregibon, D.** (1978), 'Finding the outliers that matter', *Journal of the Royal Statistical Society, Series B*, 40, 1, 85–93.
- Audit Commission** (1997), *The Coming of Age: Improving Care Services for Older People*, London: Audit Commission.
- Audit Commission/Social Services Inspectorate (SSI)** (1998), *Messages for Managers: Learning the Lessons from Joint Reviews of Social Services*, London: Audit Commission/SSI.
- Baldock, J.** (2004), 'Deficiencies in British social care services and the (efficiency) consequences of an administrative rather than a democratic politics of community care', in M. Knapp, D. Challis, J. Fernandez and A. Netten (eds), *Long-Term Care: Matching Resources and Needs*, Aldershot: Ashgate, pp. 35–44.
- Banker, R.D., Charnes, A., and Cooper, W.W.** (1984), 'Some models for estimating technical and scale inefficiencies in data envelopment analysis', *Management Science*, 30, 1078–1092.
- Bebbington, A.C. and Davies, B.P.** (1980), 'Territorial need indicators: a new approach part II', *Journal of Social Policy*, 9, 4, 433–462.
- Blades, C.A., Culyer, A.J., and Walker, A.** (1987), 'Health service efficiency: appraising the appraisers – a critical review of economic appraisal in practice', *Social Science and Medicine*, 25, 5, 461–472.
- Boussofiene, R.G., Dyson, R.G., and Thanassoulis, E.** (1991), 'Applied data envelopment analysis', *European Journal of Operational Research*, 52, 1–15.
- Brook, R.H.** (1994), 'Health care reform is on the way: do we want to compete on quality?', *Annals of Internal Medicine*, 120, 84–86.
- Burgess, J.F. and Wilson, P.W.** (1993), 'Technical efficiency in veterans administration hospitals', in H.O. Fried, C.A.K. Lovell and S.S. Schmidt (eds), *The Measurement of Productive Efficiency*, Oxford: Oxford University Press, pp. 335–351.
- Camp, R.C.** (1989), *Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance*, Milwaukee, Wisconsin: ASQC Quality Press.
- Centre for Policy on Ageing** (1984), *Home Life: A Code of Practice for Residential Care*, London: Centre for Policy on Ageing.
- Charnes, A., Cooper, W.W., and Rhodes, E.** (1978), 'Measuring the efficiency of decision making units', *European Journal of Operational Research*, 2, 429–444.
- Cm 4169** (1998), *Modernising Social Services*, London: The Stationery Office.
- Cm 6499** (2005), *Independence Well-being and Choice*, London: The Stationery Office.
- Commission for Social Care Inspection** (2004), *Performance Ratings for Social Services in England*, London: Commission for Social Care Inspection.
- Cutler, T. and Waive, B.** (1997), *Managing the Welfare State*, Oxford: Berg.
- Davies, B.P., Barton, A.J., McMillan, I.S., and Williamson, V.K.** (1971), *Variations in Services for the Aged*, London: G. Bell & Sons.
- Deming, W.E.** (1986), *Out of the Crisis*, Cambridge, MA: Massachusetts Institute of Technology.
- Department of Health** (1998), *Modernising Health and Social Services: National Priorities Guidance 1999/00–2001/02*, London: Department of Health.
- Department of Health** (1999), *A New Approach to Social Services Performance: Consultation Document*, London: Department of Health.
- Department of Health** (2002a), *Key Indicators Graphical System 2002*, London: Department of Health.
- Department of Health** (2002b), *Care Homes for Older People: National Minimum Standards*, London: The Stationery Office.

- Department of Health** (2003), *Social Services Performance Assessment Framework Indicators 2002–2003*, London: Department of Health.
- Department of Health** (2005), 'Performance information comparison system – key indicators graphical system', Department of Health, London, <http://nww.pics.nhs.uk/Ardentia/portal/jsp/index.jsp>, 22 July.
- Dusansky, R. and Wilson, P.W.** (1994), 'Technical efficiency in the decentralized care of the developmentally disabled', *The Review of Economics and Statistics*, 76, 2, 340–345.
- Edvardsen, D.F., Forsund, F.R., and Kittelsen, S.A.C.** (2003), 'Far out or alone in the crowd: classification of self-evaluators in DEA', Working Paper 2003: 7, University of Oslo, Oslo.
- Ehreth, J.L.** (1994), 'The development and evaluation of hospital performance measures for policy analysis', *Medical Care*, 32, 6, 568–587.
- Elliot, R.F., McDonald, D., and MacIver, R.** (1996), 'Review of the Area Cost Adjustment', Department of the Environment, University of Aberdeen.
- Farrell, M.** (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society, Series A*, 120, 107–127.
- Ferlie, E., Challis, D., and Davies, B.** (1989), *Efficiency-Improving Innovations in Social Care of the Elderly*, Avebury: Gower.
- Gershon, P.** (2004), *Releasing Resources to the Front Line: Independent Review of Public Sector Efficiency*, London: HM Treasury.
- Gibbs, I. and Smith, P.** (1989), 'Private nursing homes: providing good value?', *Public Money and Management*, Spring, 55–59.
- Giuffrida, A. and Gravelle, H.** (1999), 'Measuring performance in primary care: econometric analysis and DEA', Discussion Papers in Economics No. 1999/36, University of York, York.
- Goldstein, H. and Spiegelhalter, D.J.** (1996), 'League tables and their limitations: statistical issues in comparisons of institutional performance', *Journal of the Royal Statistical Society, Series A*, 159, 385–443.
- Gorbach, P. and Sinclair, I.** (1989), 'Monitoring the home help service: clues to improving performance from analysing data in a computerised client information system', *Research, Policy and Planning*, 7, 24–30.
- Harding, T.** (1999), 'Enabling older people to live in their own homes', in M. Henwood and G. Wistow (eds), *With Respect to Old Age: Long Term Care – Rights and Responsibilities, Research Volume 3 – Community Care and Informal Care*, London: The Stationery Office, pp. 37–50.
- Hougaard, J.L., Kronborg, D., and Overgård, C.** (2004), 'Improvement potential in Danish elderly care', *Health Care Management Science*, 7, 225–235.
- Hudson, B.** (2002), 'Interprofessionality in health and social care: the Achilles' heel of partnership?', *Journal of Interprofessional Care*, 16, 7–17.
- Hutcheson, G. and Sofroniou, N.** (1999), *The Multivariate Social Scientist*, London: Sage Publications.
- Jacobs, R.** (2001), 'Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis', *Health Care Management Science*, 4, 103–115.
- Jiménez, J.S., Chaparro, F.P., and Smith, P.** (2003), 'Evaluating the introduction of a quasi-market in community care', *Socio-Economic Planning Sciences*, 37, 1–13.
- Knapp, M.** (1998), 'Making music out of noise: the cost function approach to evaluation', *British Journal of Psychiatry Supplementum*, 36, 7–11.
- Levitt, M.S. and Joyce, M.A.S.** (1987), *The Growth and Efficiency of Public Spending*, Cambridge: Cambridge University Press.
- Lupton, C.** (1989), 'Measuring performance in local authority social services departments', *Assignment*, 7, 20–23.
- McCallion, G.M.** (1993), 'Planning care for elderly people using the balance of care model', *Health Services Management Research*, 6, 218–228.
- Mooney, G.H.** (1978), 'Planning for balance of care of the elderly', *Scottish Journal of Political Economy*, 25, 149–164.

- Mor, V., Angelelli, J., Gifford, D., Morris, J., and Moore, T.** (2003), 'Benchmarking and quality in residential and nursing homes: lessons from the US', *International Journal of Geriatric Psychiatry*, 18, 258–266.
- Nunamaker, T.** (1983), 'Measuring routine nursing service efficiency: a comparison of cost per day and data envelopment analysis models', *Health Services Research*, 28, 2, 183–205.
- Ozcan, Y.A., Luke, R.D., and Haksever, C.** (1992), 'Ownership and organisational performance: a comparison of technical efficiency across hospital types', *Medical Care*, 30, 9, 781–794.
- Revans, L.** (2002), 'Star mapping', *Community Care*, 27 June–3 July, 32–34.
- Rowen, K. and Black, N.** (2000), 'A bottom-up approach to performance indicators through clinician networks', in J. Appleby and A. Harrison (eds), *Health Care UK – The King's Fund Review of Health Policy Spring 2000*, London: King's Fund, pp. 42–46.
- Sherman, D.H.** (1984), 'Hospital efficiency measurement and evaluation: empirical test of a new technique', *Medical Care*, 22, 922–935.
- Shewhart, W.A.** (1931), *Economic Control of Quality of Manufactured Product*, New York: Van Nostrand.
- Smith, P.** (1990), 'The use of performance indicators in the public sector', *Journal of the Royal Statistical Society, Series A*, 153, 53–72.
- Smith, P. and Street, A.** (2003), *Measuring the Efficiency of Public Services: The Limits of Analysis*, York: Centre for Health Economics, University of York.
- Stone, M.** (2002), 'How not to measure the efficiency of public services (and how one might)', *Journal of the Royal Statistical Society, Series A*, 165, 3, 405–434.
- Thanassoulis, E.** (1995), 'Assessing police forces in England and Wales using data envelopment analysis', *European Journal of Operational Research*, 87, 641–657.
- Thanassoulis, E.** (2001), *Introduction to the Theory and Application of Data Envelopment Analysis*, Dordrecht: Kluwer Academic Publishers.
- Thanassoulis, E. and Emrouznejad, A.** (1996), *Warwick Windows DEA User's Guide*, Warwick: University of Warwick.
- Warburton, R.W.** (1988), *Key Indicators of Local Authority Social Services: A Demonstration Package*, London: Social Services Inspectorate, Department of Health.
- Watson, D.** (2002), 'A critical perspective on quality within the personal social services: prospects and concerns', *British Journal of Social Work*, 32, 877–891.
- Wiggins, A. and Tymms, P.** (2002), 'Dysfunctional effects of league tables: a comparison between English and Scottish primary schools', *Public Money and Management*, 22, 43–48.
- Williams, A. and Anderson, R.** (1975), *Efficiency in the Social Services*, Oxford: Basil Blackwell.
- Wilson, P.W.** (1993), 'Detecting outliers in deterministic nonparametric frontier models with multiple outputs', *Journal of Business and Economic Statistics*, 11, 319–323.