Health Services Management Research **17**, 217–228

Using performance indicators to improve performance

Penelope M. Mullen

Health Services Management Centre, University of Birmingham, UK

The recent introduction of performance assessment within the UK NHS, incorporating numerous performance indicators (PIs) and league tables, has led to health care organisations facing large numbers of targets and a star rating system with associated rewards and penalties. However, there is considerable evidence that using PIs for judgement rather than learning provides perverse incentives and can prove counterproductive. Drawing on earlier PI systems which, supplemented by expert systems, were designed to promote learning and exploration particularly by encouraging analysis of interactions between different indicators, a series of 'mini case studies' is presented. These reveal interesting relationships and suggest explanations for variations in performance, areas worth exploring further and possible approaches to improving performance – approaches not apparent from individual indicators and league tables. It is concluded that presentation of PIs in a format that encourages exploration and analysis could greatly enhance the potential of the current PIs to improve NHS performance.

Introduction

Few in the UK can have failed to notice the rise of a performance assessment culture within the NHS, especially in England. A wide range of performance indicators (PIs) have been published largely in the form of league tables, produced in book format and as individual Excel tables and charts, although the format has changed somewhat in the more recent PIs published in 2003 (CHI, 2003a; 2003b). Both health care organisations that provide services (including hospitals) and those which plan and commission services (currently Primary Care Trusts, which have responsibility for securing services for populations averaging 150,000) face large numbers of targets. Performance on a subset of these targets feeds into a star rating system, from zero to three stars, with accompanying rewards and penalties.

The NHS is not alone in adopting performance assessment and targets. Many public services are subject to similar systems and there is a growing national and international literature on performance measurement and assessment. Explicitly stated objectives of performance measurement and assessment in the public sector include accountability, both to the public/consumers and to the government, and the detection and prevention of scandals. However, it often appears to be assumed not only that the primary purpose of producing and using PIs is to improve performance but also that performance assessment and the use of PIs will, apparently axiomatically, lead to performance improvement. In reality, the position appears rather more complex.

Penelope M. Mullen, *Senior Lecturer*, Health Services Management Centre, University of Birmingham, 40 Edgbaston Park Road, Birmingham B15 2RT, UK
E-mail: p.m.mullen@bham.ac.uk

In addition to noting the enabling contribution of increased computing power (Pollitt, 1986), a number of commentators have attributed the rise of performance measurement and assessment in the public sector over the past two decades to the 'new managerialism' or New Public Management (Jacobs and Manzi, 2000; Power, 1997); to the increasing complexity and diversity of government and the public sector (Chapman, 2002); and to a 'crisis of trust', especially in relation to professionals, where the resultant 'accountability takes the form of detailed control' (O'Neill, 2002; p.46). Whilst ostensibly aimed at promoting performance improvement, performance assessment under these circumstances relates more to a perceived need to control complex, decentralised organisations 'on behalf of taxpayers and consumers and against cosy cultures of professional self-regulation' (Power, 1997; p.44). Further, Chapman (2002; p.56) argues that 'controlling complex adaptive systems by imposing fixed targets has never been possible'.

Nevertheless, even if it is accepted that a primary objective of PIs and performance assessment is simply to improve performance, many problems remain. There is a long history of evidence that use of PIs provides perverse incentives and can prove counterproductive. Mullen (1985) cites a number of examples from the former Soviet Union where extensive use of indicators and targets resulted in a range of unintended outcomes, satirised in a cartoon showing a single 10-tonne nail produced to fulfil the outcome target of a nail factory. To the list of well-known consequences of perverse incentives, summarised by Smith and Goddard in 2002 – tunnel vision, measure fixation, sub-optimization, myopia, complacency, misrepresentation, misinterpretation, gaming and ossification – can be added others such as risk avoidance, for example, where publication of death rates leads to surgeons refusing to operate on riskier cases (Vass, 2002; Mulley, 1999). In addition to the many examples cited by Smith and Goddard (2002), the consequences of perverse incentives in the NHS are frequently reported in both the popular and specialist press; a recent example being where hospitals were reported to have cancelled routine operations and transferred additional resources to Accident and Emergency departments for the single week in which performance on A&E waiting times was being measured (Revill, 2003).

The literature abounds with attempts to avoid the distortions of perverse incentives. Mullen (1985) noted the tendency for indicators to multiply in attempts to overcome perverse incentives or, as Davies and Lampel (1998; p.160) put it more colourfully, 'Performance measurement is addictive, requiring larger and larger doses to get a temporary fix'. However, drawing on the experience of the Soviet Union, Nove (1958; p.6) warned that 'far from correcting the deviations . . . these other indicators either exercise a distorting pull in the same direction or introduce new distortions of their own'.

Mullen (1985) attributed perverse incentives to the use of PIs as targets in hierarchical organisations to control lower levels and subordinates, comparing this to their use for self-improvement, whether by individuals or organisations. Other authors have made similar distinctions. For example, Freeman (2002; p.134) distinguishes between the use of indicator systems as a 'summative mechanism for external accountability and verification in assurance systems' and their use as a 'formative mechanism for internal quality improvement'. Boland and Fowler (2000) present a matrix, with one dimension being control location (internal or external) and the other the resultant action (positive [supportive and beneficial] or negative [threatening or punitive]). Whilst they argue that Quadrant 1 (internal/positive) is 'the most desirable location for most public sector organisations, in terms of satisfying, in the long term, the needs of the majority of stakeholders', they conclude that in many public services, including health, PIs and performance management lie in Quadrant 4 (external/negative).

To summarise, it is argued that approaches promoting learning, investigation, explanation and self-motivation, and involving alerting, tracers, trust and respect for professional values and autonomy have positive effects on performance. On the other hand, approaches involving judgement, targets, 'imposed performance improvement', threats, sticks and carrots, 'naming and shaming' and lack of trust can have negative effects on performance.

In which Quadrant do the current NHS PIs and star system lie? Viewpoints differ. On the one hand, the Chief Executive of a zero-star hospital Trust was quoted as saying: 'The star system was used in such a destructive way in terms of the publicity. When you have newspapers claiming that your hospital is on a list of

shame you are not in a position which lends itself to an objective and contextualised discussion about the performance assessment system', whilst noting his Trust 'had received a good report from the CHI [Commission for Health Improvement] and high marks on other measures' (Carlisle, 2002). On the other hand, at the same time, a Department of Health spokesperson was quoted as stating that 'The star-rating system is not about punishing trusts. It is about improving performance and strengthening organisations which have had difficulties by putting robust management arrangements in place to make sure that those trusts get out of trouble and stay out of trouble' (Carlisle, 2002). Nevertheless, commenting on the February 2002 NHS PIs, Gibbs (2002; p.21) claims that 'the new PIs are more concerned about delivering judgements about the performance of health authorities and trusts and less about providing a learning tool than the old PIs'. So can we learn from those old PIs?

PIs in the NHS are not new. Since the early 1980s indicator systems have existed in England and elsewhere in the UK. There were two independent initiatives in the early 1980s. First, following analysis of routine data from psychiatric hospitals, Professor Yates and his team found that whilst poor performance on a single indicator did not signify performance failure, poor performance on, for example, at least four out of six selected indicators was associated with performance failure. As a result of this, they devised a microcomputer-based system which permitted the display of user-selected sets of indicators in a range of formats, which include histograms; multi-indicator, multi-speciality and some multi-year (i.e. time series) profile (percentile bar) charts; and scattergrams for a wide range of chronic and acute specialties (Yates and Davidge, 1984). The aim was not only to identify inter-authority and inter-hospital variations in performance but to help look for explanations. Second, following the earlier publication of NHS indicators in printed book format (DHSS, 1983), the DHSS produced a microcomputer-based set of indicators in 1985, which, while using slightly different types of display, also permitted inter-authority comparisons of combinations of indicators.

Following major changes in information collection in the NHS in 1987, the two systems were combined into the Health Service Indicators (HSIs), which were produced until 1996 (1994/5 data). The HSIs retained the facility for users with no specialist computer knowledge to compare performance on user-defined sets of indicators, offering a range of multi-indicator displays in addition to the single-indicator histogram. Scattergrams permitted analysis of the inter-relationships of indicators as well as their correlation with environmental factors such as socio-economic deprivation. Both the 1985 DHSS PIs and the subsequent HSIs were supplemented by expert systems (Payling *et al.*, 1987), which not only analysed atypical performance, but additionally suggested possible causal or associated factors.

Thus, the old PI and HSI systems were designed to promote learning and exploration as a basis for performance improvement, in particular by encouraging analysis of interactions between different indicators. While it was never claimed that they provided all the answers, they suggested areas worth exploring further in order to explain variations and suggested potential approaches to improving performance – approaches which are not apparent from examination of individual indicators and league tables alone. To illustrate this, and to see if we can learn from the past, a series of anonymised 'mini case studies' are presented here, which apply ideas and approaches from the earlier systems to analyse and present a range of PIs used in the current system. As well as supporting some prior hypotheses, these analyses reveal a large number of interesting relationships and potential explanations for variations in performance.

DNA (Do Not Attend) rates

Failure to attend for appointments, both with GPs and at hospital outpatient (OP) departments, has long been a cause for concern within both the health service and the media. DNA rates were included in the league tables for 2000/1, where wide variations in the OP DNA rate were found, both between Health Authorities (Figure 1) (DoH, 2002a) and between individual hospital trusts.

The existence of high DNA rates has resulted in pressure on hospitals to reduce DNA rates and led to repeated calls to impose fines or other sanctions on the apparently irresponsible 'offenders' who fail to keep appointments.

However, since the early 1990s examination of DNA rates across specialties using the provider profile charts in the HSI system had

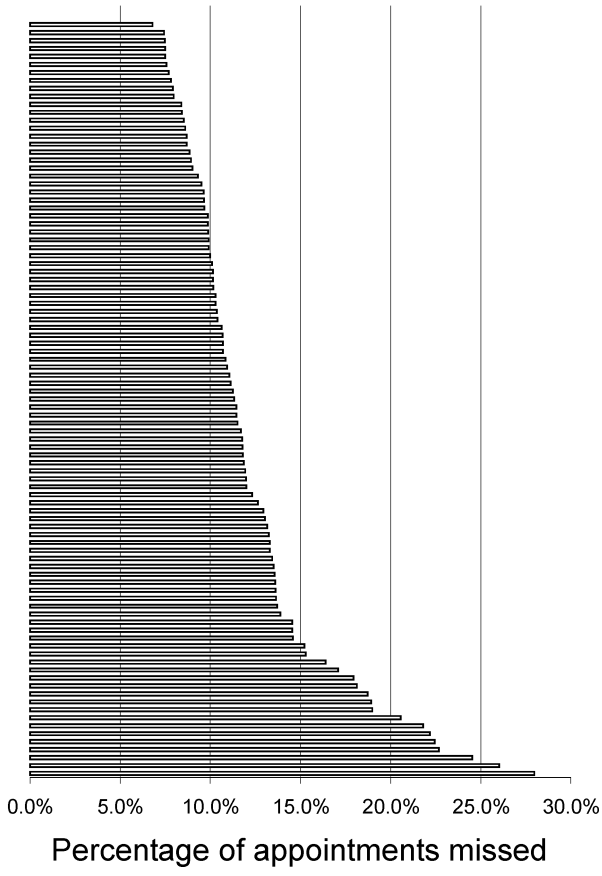

Fig. 1 *Percentage of first outpatient appointments for which patient did not attend 2000/1 (ranked by Health Authority)*

revealed, year after year, a clear contrast between hospitals located in areas scoring as privileged on socio-economic indicators and those scoring as underprivileged. Figure 2 reproduces examples of typical profile charts from the 1994/5 HSIs for two hospitals: Hospital A (LHS) is located in a relatively privileged area and has low DNA rates across specialties; Hospital B (RHS), which is located in an underprivileged area, has high DNA rates.

To determine whether there is still such a relationship, a special analysis was carried out, plotting the data from the 2000/01 league table in Figure 1 against the deprivation scores for the populations of the Health Authority areas. Figure 3 shows that a significant relationship (Pearson's correlation coefficient [r] = 0.744) exists between DNA and deprivation measured on the Underprivileged Area score (r = 0.760 using the DOE Index of Urban Deprivation). Similar relationships were found when plotting the DNA rate for each hospital trust against the deprivation scores for the areas in which the hospital is located (r = 0.5893 and r = 0.6591 respectively).

Finding such a relationship prompts a number of hypotheses as to possible causal factors, such as levels of car ownership and other transport difficulties, child care difficulties and access to telephones to cancel or change appointments. However, evidence from some earlier studies that forgetting appointments was a major cause of DNA led to the suggestion that differential use of diaries might be an explanatory factor. To examine this, Dutton (1998) carried out a small-scale study in two contrasting electoral wards on the use of memory aids in respect of appointments. She found that, whilst both groups used memory aids, all respondents in the ward with greater formal education and higher status employment routinely used calendars and diaries, compared

Indicators	VALUE	RANK	RELATIVE POSITION (as a percentile) 1.:.20.:.40.:.60.:.80.:100	NAT VALUE	VALUE	RANK	RELATIVE POSITION (as a percentile) 1.:.20.:.40.:.60.:.80.:100	NAT VALUE
CL51 Out Patient % non-attend rate - General surgery & urology	7.90	27	+	11.28	17.54	94		11.28
CL51 Out Patient % non-attend rate - General medical	7.73	13		12.54	15.61	88		12.54
CL51 Out Patient % non-attend rate - Adult MI	7.26	5		18.19	NoVal			18.19
CL51 Out Patient % non-attend rate - Trauma & orthopaedics	5.44	6		10.86	15.40	91		10.86
CL51 Out Patient % non-attend rate - ENT	11.89	32	+	14.53	19.88	92		14.53
CL51 Out Patient % non-attend rate - Ophthalmology	6.04	12		10.32	17.05	91		10.32
CL51 Out Patient % non-attend rate - Paediatrics	10.11	8		18.55	24.24	92		18.55
CL51 Out Patient % non-attend rate - Geriatrics	2.30	4		9.90	22.06	98		9.90
CL51 Out Patient % non-attend rate - Gynaecology	8.83	20		13.61	13.87	68	+	13.61

Fig. 2 *HSI 1994/95 Outpatient DNA rates ranking: Hospitals A & B*

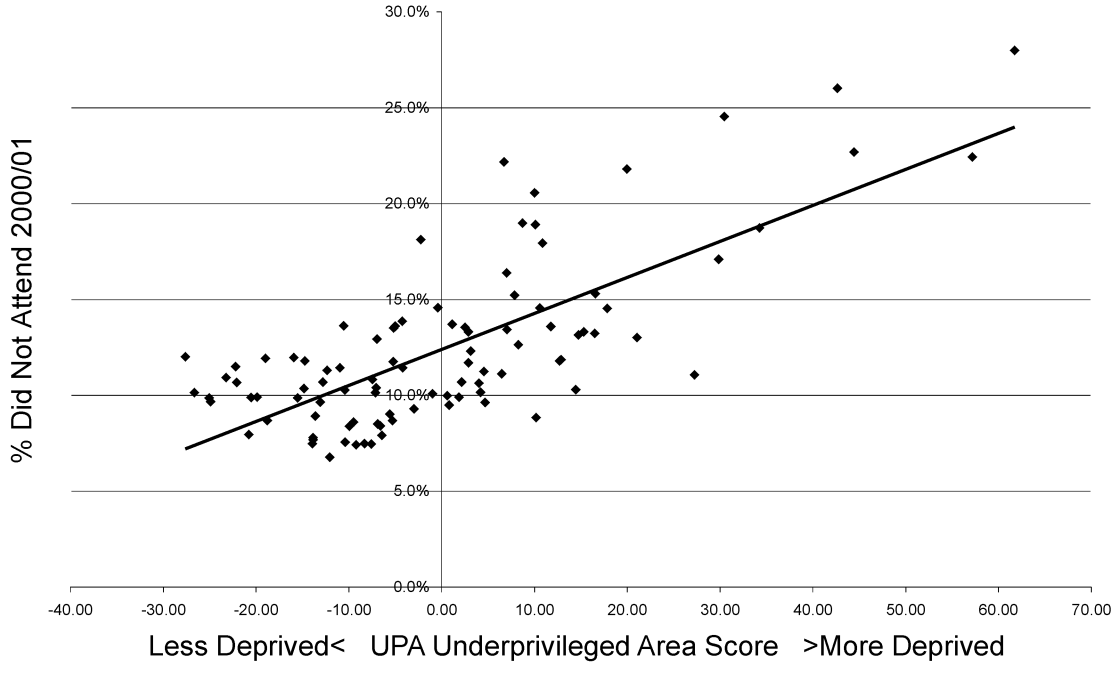

Fig. 3 *Percentage of first OP Appointments for which patient did not attend 2000/1 by Deprivation (UPA)*

with less than two-thirds of those in the lower socioeconomic status ward. This differential use (and possession) of diaries is confirmed by qualitative evidence from the evaluation of the Booked Admissions Pilot scheme that, when asked to bring their diary in order to book the date for their admission, some patients went out and bought a diary (Personal Communication [Kipping, 2000]). These findings suggest that, rather than penalising supposedly feckless patients, strategies such as those adopted by some hospitals to remind patients of their appointment shortly beforehand are more likely to prove successful and socially equitable.

Whilst very close examination of the 2000/1 DNA League Table might reveal the relationship with deprivation, especially in the presence

of a prior hypothesis, the dramatic presentation of the profile charts which, of course, emphasise highs and lows, revealed a pattern during browsing which in itself suggested the further lines of enquiry.

Surgery rates for cataracts

In the 2000/1 PI league tables, Health Authority A (HA A) ranked 79th out of 95 authorities in the age/sex standardised elective surgery rate for cataract removal, a rate selected nationally as an indicator of 'Fair Access' to services. This ranking is likely to cause concern to HA A, especially as the league tables also showed that HA A had made little improvement (increase in the rate) on this indicator since the previous year. However, as presented (using the same format as Figure 1), the league tables gave little assistance in suggesting possible lines of investigation. Thus the 1994/5 HSIs were revisited to see if they would have given any such assistance.

A user-defined profile chart (Figure 4) for that year demonstrated that HA A was in the lowest 10% of HAs for cataract surgery overall and for the 65+ age group, and in the lowest 15% for hospitalisation rate for patients over 65 years for Ophthalmology (a variable which correlates highly [r = 0.869] with the cataract rates). However, that profile chart also shows that HA A was in the top 20% (84th percentile) for the percentage of ophthalmology hospital episodes with no operation, in the lowest 20% (18th percentile) in the percentage of residents treated as day cases, was almost top (99th percentile) for length of stay for cataract removal and for ophthalmology as a whole and was one of the highest (98th percentile) in the percentage of admissions which were emergencies. At the same time HA A appeared to have a relatively low waiting time and was in the lowest 15% in the percentage waiting over 12 months for admission for ophthalmology. These findings suggest some areas for investigation.

Although HA A might in any case wish to investigate the relatively high percentage of admissions which did not result in an operation, this is unlikely to help in explaining the lower hospitalisation and cataract surgery rates, as the scattergram and associated statistics produced

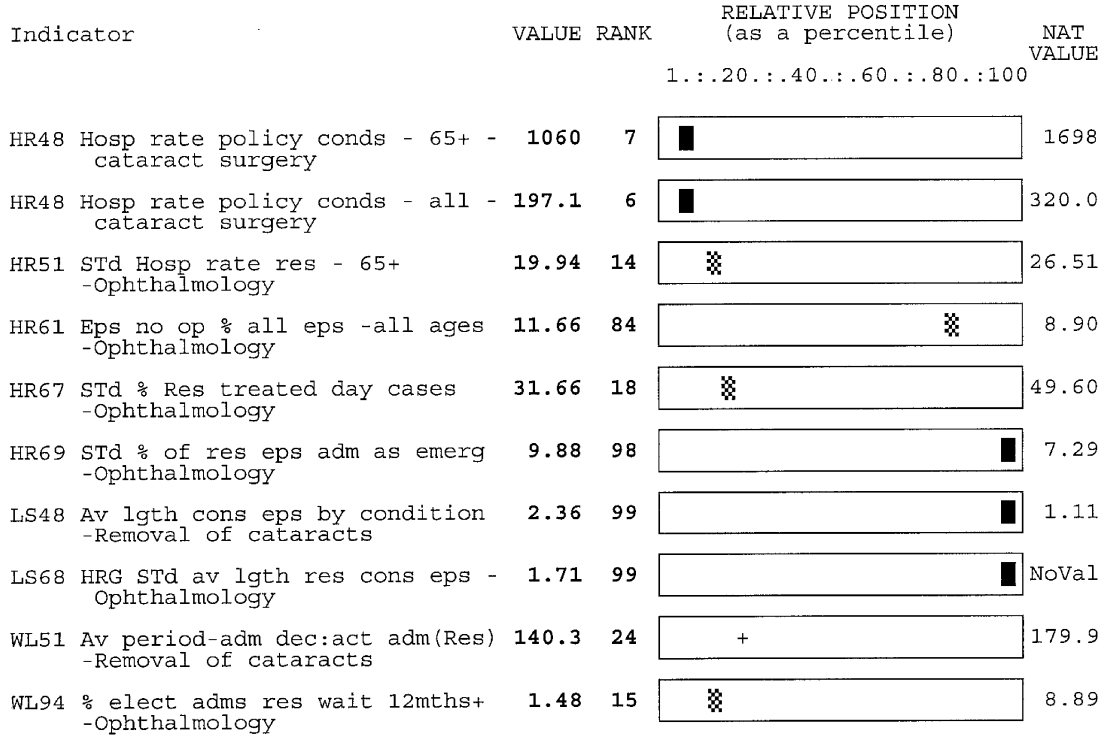

Fig. 4 *Ophthalmology and Cataract Surgery Profile Chart for Health Authority A*

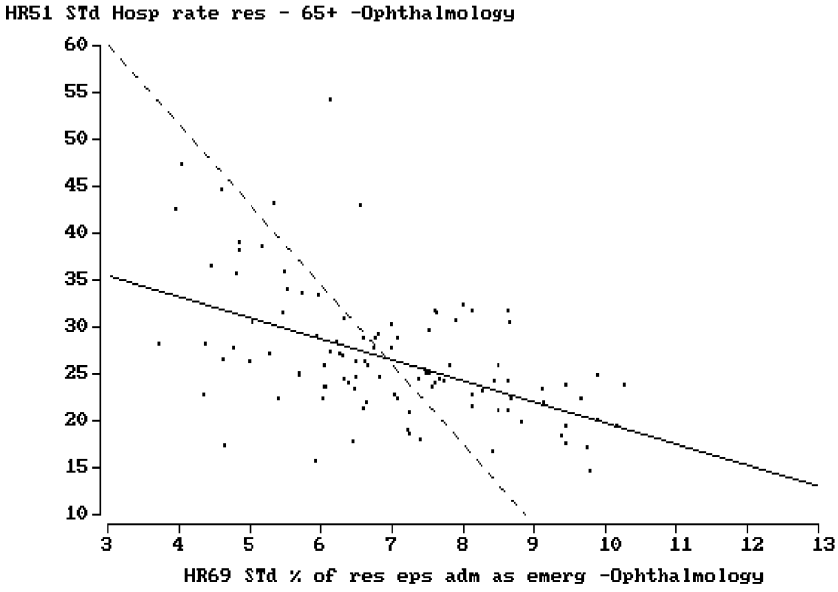

Fig. 5 *Ophthalmology Admissions/100,000 aged 65+ by Percentage Admitted as Emergencies*

using the HSI package show no significant correction (r = 0.096 and -0.101 respectively). The fairly low day-case rate and the high length of stay, which of course are highly correlated (r = -0.838), might suggest more promising lines of enquiry, as their correlation coefficients with the ophthalmology hospitalisation rate and cataract surgery rate are all around 0.35. However, as Figure 5 shows, there is a significant correlation (r = -0.512) between the percentage of cases admitted as emergencies and the hospitalisation rate for over 65s (r = -0.461 for the cataract surgery rate).

This raises questions such as whether the high emergency rate might be a cause of the low treatment rate, or whether the low treatment rate for the local population is resulting in a high emergency admission rate. Analysis also shows significant correlations between emergency admissions and the day case rate (r = -0.449) and length of stay (r = 0.514). Whilst there could be a number of reasons for such relationships, together they do suggest areas for further investigation in relation to the relatively low treatment rate in HA A.

It is worth noting that the above analyses were carried out using only the data and standard presentation features of the 1994/5 HSI package. Such analysis, taking only a few minutes and not requiring special computer skills, can suggest a number of areas for further investigation, as well as eliminating other potential lines of enquiry.

Hip replacement rate

Joint replacement is another Fair Access indicator and the rates for Joint (Hip and Knee) replacement were included in the 2000/1 PIs in league table format, again with improvement in performance measured by increase in the rates. The introduction to the Fair Access section in the printed booklet notes that 'social and economic factors may influence access to health service' but continues by stressing that patients should be treated equally regardless of their social status (DoH, 2002a). However, the format in the printed book and the downloadable Excel files does not permit easy investigation of this.

Apparently differential access to joint replacement (especially hip replacements) was demonstrated by the HSIs in the early 1990s. Figure 6 from the 1994/5 HSIs shows that the hospital rate for hip replacements for patients over 65 years reduces as deprivation increases (the greater the deprivation, the higher score on the DOE index); i.e. the more deprived the population of the Health District, the lower the NHS hip replacement rate (r = -0.51). Since it is unlikely that deprived populations have higher than average rates of private operations, the

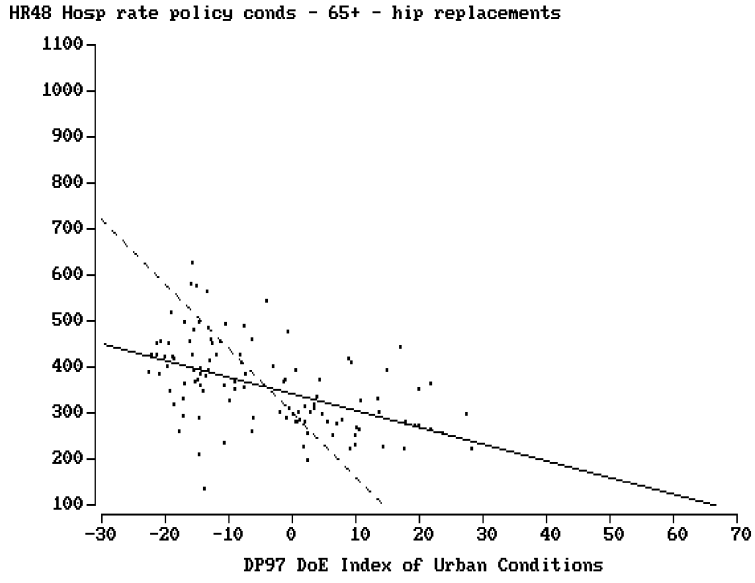

Fig. 6 *Hip Replacements/100,000 aged 65+ (1994/5) by Deprivation*

inclusion of such operations (estimated to be about 25% of the total carried out) is likely to emphasise the differential access still further. The same correlation was observed year after year and a similar, but less strong, correlation ($r = -0.355$) was observed between hospitalisation for knee replacement and deprivation.

The fact that the deprivation score can explain 25% of the variation in the NHS hospitalisation rate for hip replacement is surely worthy of further investigation if fair access is sought. One possible explanation for the statistical relationship lies in the fact that the HSI rates are not age and sex standardised; they are simply the rates for the 65+ age group. If the populations of the more deprived districts have lower survival rates into the older age groups where hip replacements are more common, this might explain the lower rate for the whole 65+ age group. However, this explanation is not supported by the current figures. The 2000/1 PI data for Hip and Knee replacement (combined), which is age and sex standardised, was plotted against the DOE deprivation index (obtained from the Public Health Common Data Set). The resulting correlation ($r = -0.4175$) shows that the differential access by deprivation still exists.

From the viewpoint of Fair Access, this finding is disappointing and worthy of further investigation. However, it is noted that the relationship using the recent PI data was found only after special analysis that merged, with some difficulty, different data sets. The comparable analysis using HSIs took seconds and was identified during browsing.

Day case rate

There is pressure on hospitals to increase the percentage of elective surgery and some other cases treated on a day case basis, thus avoiding the need for an overnight stay in hospital. In fact, the Government aims to have over 75% of all operations carried out as day cases by 2008 (DoH, 2002b).

To help demonstrate the potential scope for exploration afforded by the HSI presentations, a small selection of 27 indicators were extracted from the downloadable Excel files for the PIs published in July 2000 and after editing into the required format were imported into the Windows version of the original HSI presentation package. Browsing through this limited set of indicators yielded a number of interesting, if expected, patterns and relationships. However, it also produced a scattergram (Figure 7) that, in the light of the policy outlined above, at the very least suggests a *prima facie* case for investigation. Although the correlation between the day case rate and the rate for deaths in hospital within 30 days of surgery for 1998/9 is fairly low ($r = 0.357$), it is both positive and statistically significant. Obviously before any action is

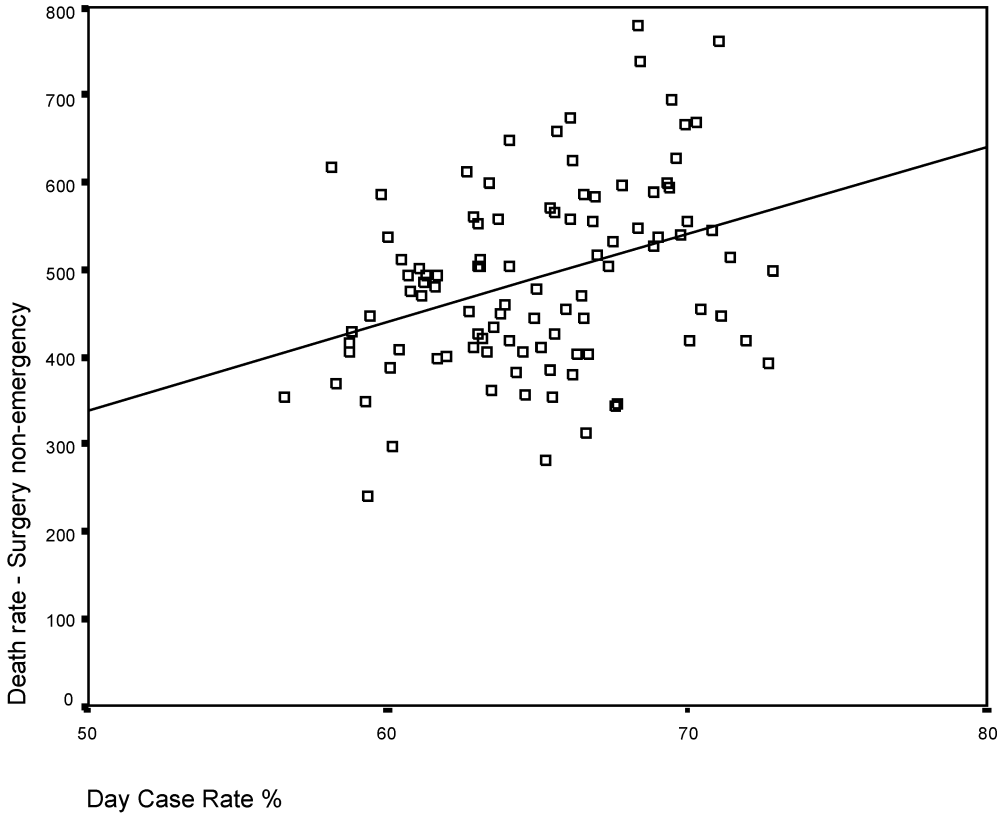

Fig. 7 *Deaths in hospital within 30 days of surgery (all ages): Rates/100,000 non-emergency admissions by Day Case Rate*

taken, the data and assumptions need to be checked and data from other years explored to determine if this is a continuing relationship – but preliminary analysis carried out with the benefit of hindsight suggests that it is not.

Accuracy of data

Both the current PIs and the earlier HSIs have been charged with data inaccuracy, although it could be argued that inaccurate data is more damaging when used for judgement and control than when used for exploration and learning. Nevertheless, the following example both illustrates the benefits of comparing data year-on-year and suggests that data may be more accurate than some might hope.

Shortly after the publication of the first set of the new PIs in 1999, the Chief Executive of a hospital Trust (Hospital C), which was shown by the indicators to have twice the national average death rate within 30 days of surgery, was reported as saying:

> 'I just don't recognise the hospital they are talking about . . . [this hospital] . . . the recipient of two beacons of excellence is an excellent hospital and the people of this town know it. These figures are nonsense. They're crackers.'
>
> (BJHCM, 1999)

To investigate this claim, ideally one would be able to turn to a time-series showing the same indicator for the same hospital over a number of years. However, not only have data collection systems changed, NHS organisations have undergone a number of reconfigurations since the early 1990s, making such an ideal comparison impossible. Nevertheless, it was possible to turn to a similar indicator (Standardised number of deaths in hospital for General Surgery for the 16–64 age group) for which data was available for the predecessors of current Hospital C for three years between 1991 and 1995.

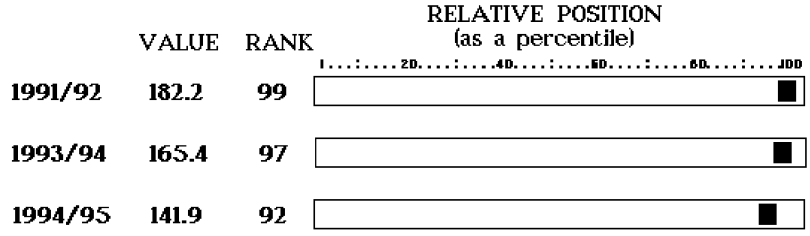

Fig. 8 *Standardised number of deaths in hospital: General Surgery—Hospital Trust C*

Figure 8, which is redrawn from a cut-and-paste of the profile bars for those three years, shows that the predecessors of Hospital C were consistently in the highest 10% on this indicator (99th, 97th and 92nd percentiles). Allowing for all the caveats of not quite comparing like with like, this consistent pattern suggests that either data recording in Hospital C is consistently out of line (e.g. using different definitions) or that, despite the Chief Executive's protestations, there really is a persistent problem which needs investigation.

Discussion

Given the claimed advantages of the HSIs and their presentation package, why were they discontinued? After all, it could be argued that the HSI package, encouraging and permitting simultaneous examination and exploration of a range of indicators as it did, effectively constituted an interactive balanced scorecard.

Firstly, it is important to note that the Windows successor to the HSI presentation package continues to be used. Inter-Authority Comparisons and Consultancy (IACC), established by Professor Yates, continues to use the software, which can accommodate comparative data at any level, as an essential tool in its research and consultancy. Until the October 2002 structural changes in the NHS, a number of health authorities and consortia of health authorities used the software for primary care indicators. The WHO Indicators for the European Region are presented in an interactive package that includes profile bar charts, box plots and scattergrams derived from the HSI package.

Secondly, it is important to consider the purpose of the PIs. An indicator system that encourages exploration, explanation and learning and raises questions rather than producing clear cut league tables and star systems might not be appropriate where the primary objective is control and accountability, especially if there is a related lack of trust. However, it is interesting to record that the expert system based on the original DHSS PIs was originally developed to assist the DHSS in its annual top-down regional performance reviews (Payling *et al*., 1987). Despite this origin, a number of regional and district health authorities realised that the expert system would assist them in analysing and improving their own performance, and the successor expert system (the HSI Analyst) was specifically designed to facilitate such use at these lower levels. The HSI Analyst did not survive the 1991 changes to the NHS that introduced the Internal Market, as the knowledge bases would have needed almost complete rewriting.

Thirdly, the HSI system was criticised for being concerned with process rather than outcome indicators. There is a certain amount of truth in this and there has been considerable effort to include a range of outcome indicators in the new PIs, although they were not absent from the HSIs. Nevertheless this criticism does not explain the discarding of the presentation package, which will accommodate any type of indicator. Furthermore there is considerable debate about the relative merits of outcome and process indicators, which is beyond the scope of this paper. Where the aim is alerting and preventing problems, process indicators can often be more useful, particularly in health care, as outcome indicators might be too late to prevent or even ameliorate problems.

Fourthly, the HSI system was criticised for containing too many indicators – at times there were many thousands; however, many individual indicators were presented in a number of specialty/age/gender variants, thus boosting the total. Even the new PIs are not immune to such criticism. According to the Director of Information Services at a hospital Trust, 'a

quick "flick through" is interesting, but detailed study becomes tedious' (Cundy, 2002; p.23). However, the appropriate number of indicators must be determined by the purpose of indicators. If the aim is judgement and control, a single indicator (for example the number of stars) or a handful (perhaps the nine Key Targets which contribute to the 2003 star ratings) might be ideal. As noted above, however, the associated perverse incentives can lead to a proliferation of indicators and targets with their own resultant distortions. If the purpose is exploration, explanation and learning in order to improve performance, a large number of indicators can be an advantage if presented and used appropriately. For this purpose, an ideal system might be one which initially presents a limited range of important indicators and then allows users to drill down to more detailed indicators and data, both to obtain more specific information about the performance being indicated (for example, does a low access rate for a condition apply to all or only some age groups?) and to explore, analyse and compare the information. Thus, with an appropriate presentation package, a large number of indicators can be an advantage rather than a disadvantage and experience has shown that even new users, within 30 minutes of being introduced to the HSI package with its thousands of indicators, asked for more in the specific area they were investigating. Further, the large comparative data set that the HSIs constituted was essential to support the HSI Analyst expert system.

Conclusion

There appear to be many reasons for the explosion in performance measurement and assessment and PIs both within the English NHS and elsewhere. While many commentators attribute this to the need to judge and control within complex organisations and to lack of trust, especially of professionals, there is long and ample evidence that using PIs as a judgemental hierarchical control tool may not improve performance and frequently produces distortions and unintended outcomes. However, assuming that the underlying aim of the PIs in the NHS is to improve performance, this paper argues that publication and presentation in a format such as that of the old HSIs, which encourages exploration and analysis to help determine the reasons for poor or outlying performance and suggest possible approaches to improving performance or explain outliers, would greatly enhance the potential of the current PIs to improve NHS performance.

References

BJHCM. Editorial: 'Six of the best' for the NHS? *British Journal of Health Care Management* 1999; **5**(7): 258

Boland, T. and Fowler, A. A systems perspective of performance management in public sector organisations. *The International Journal of Public Sector Management* 2000; **13**(5): 417–446

Carlisle, D. 'Dirty dozen' chief speaks out. *Health Service Journal* 2002; **112**(5809): 10

Chapman, J. *System failure.* London: Demos, 2002

Commission for Health Improvement. *NHS Performance Ratings: Acute trusts, Specialist trusts, Ambulance trusts 2002/2003.* Norwich: TSO, 2003a

Commission for Health Improvement. *NHS Performance Ratings: Primary care trusts, Mental health trusts, Learning disability trusts 2002/2003.* Norwich: TSO, 2003b

Cundy, A. How am I doing? A personal view of the usefulness of performance indicators. *British Journal of Healthcare Computing & Information Management* 2002; **19**(5): 23–25

Davies, H. T. O. and Lampel, J. Trust in performance indicators. *Quality in Health Care* 1998; **7**: 159–162

DHSS. *Performance Indicators National Summary for 1981* (plus accompanying regional summaries), London: Department of Health and Social Security, 1983

DoH. *NHS Performance Indicators: Health Authorities.* London: Department of Health, 2002a

DoH. *Delivering the NHS Plan,* Norwich: TSO, 2002b

Dutton, F. *A study of the strategies used to remember appointments,* Birmingham: MSc Dissertation, University of Birmingham, 1998

Freeman, T. Using performance indicators to improve health care quality in the public sector: a review of the literature. *Health Services Management Research* 2002; **15**(2): 126–137

Gibbs, R. Performance indicators: for judging or learning? *British Journal of Healthcare Computing & Information Management* 2002; **19**(5): 21–22

Jacobs, K. and Manzi, T, Performance indicators and social constructivism: conflict and control in housing management. *Critical Social Policy* 2000; **20**(1): 85–103

Mullen, P. M. Performance Indicators—Is Anything New? *Hospital and Health Services Review* 1986; **81**(4): 165–167

Mulley, A. Learning from differences within the NHS *British Medical Journal* 1999; **319**(7209): 528–530

Nove, A. The problem of success indicators in Soviet Industry. *Economica* New Series 1958; **25**(97): 1–13

O'Neill, O. *A Question of Trust.* Cambridge: CUP, 2002

Payling, L., Bowen, T. and Briggs, I. PIs become crystal clear. *Health Service Journal* 1987; **97**: 502–3

Pollitt, C. Beyond the Managerial Model: the case for broadening performance assessment in Government and the Public Services *Financial Accountability and Management* 1986; **2**(3): 155–170

Power, M. *The Audit Society: rituals of verification.* Oxford: OUP, 1997

Revill, J. Hospitals faking cuts in casualty wait times. *The Observer* (11 May 2003)

Smith, P. C. and Goddard, M. Performance management and Operational Research: a marriage made in heaven? *Journal of the Operational Research Society* 2002; **53**(3): 247–255

Vass, A. Performance of individual surgeons to be published. *British Medical Journal* 2002; **324**(7331): 189

Yates, J. M. and Davidge, M. G. Can you measure performance? *British Medical Journal* 1984; **288**: 1935–4