

Performance league tables: the NHS deserves better

Peymané Adab, Andrew M Rouse, Mohammed A Mohammed, Tom Marshall

League tables are frequently used to depict comparative performance in sport and commerce. However, extension of their use to rank services provided by healthcare agencies has attracted resistance, criticism, and anxiety. In this article we discuss the benefits and drawbacks of league tables and suggest that an alternative technique, based on statistical process control, could be introduced in their place. We believe that this technique would have the dual advantage of being less threatening to providers of health services and would be more easily understood and correctly interpreted by patients, auditors, and commissioners of services.

League tables

For many years league tables have been used to rank the quality of goods or services provided by competing organisations. They are commonly published in the popular press and magazines, specialist journals, and the internet. These tables range from those that simply rank crude performance on indicators to those that report sophisticated comparisons of summary adjusted statistics (such as those with uncertainty intervals around the rank). The public is prepared to pay intermediaries, such as financial advisers, or consumer organisations for this information. One of the best known UK organisations is the Consumers' Association. Its main publication, *Which?*, produces several league table equivalents each month (fig 1). Nearly half a million subscribers pay £37 annually (and many more subscribe to similar organisations) to study these tables in the belief that they will then make more informed purchasing decisions.

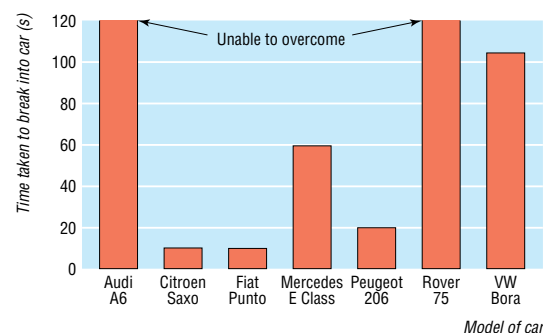


Fig 1 Time taken to break into different models of car (example of league table, adapted from *Which?* 2000;October:43 (published by the Consumers' Association, 2 Marylebone Road, London NW1 4DF))

Summary points

League tables are an established technique for displaying the comparative ranking of organisations in terms of their performance

League tables provoke anxiety and concern among health service providers for several reasons, including concerns over adjustment for case mix and the role of chance in determining their rank

Control charts, used for monitoring and control of variation in the manufacturing industry, overcome these problems by displaying performance without ranking and helping to differentiate between random variation and that due to special causes

League tables are useful for comparing quality or outputs from different systems, whereas control charts are more useful for comparison of units within a single system, such as the NHS

Control charts avoid stigmatising “poor performers” and promote the use of a systems approach to quality improvement

The popularity of such league tables suggests that they are easily interpreted and valued by subscribers, which may, in part, explain the rapidity with which they were introduced in a modified form to rank the performance of public sector and similar organisations. Examples are the schools performance tables prepared by the Department for Education and Employment, the Home Office crime statistics tables, and the NHS high level performance indicators (fig 2).

These NHS performance league tables are accompanied by explanatory notes, including the following statement: “In interpreting these types of graphs, it should be noted that if a Trust’s confidence intervals do not overlap with the England rate, it is likely that their indicator values are genuinely different from the national rate.”¹

Although these performance league tables (fig 2) resemble traditional league tables (fig 1), they have some important differences. A major difference is that the performance league tables include 95% confidence intervals around each provider’s performance score. In

Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham, B15 2TT

Peymané Adab
clinical lecturer

Andrew M Rouse
senior lecturer

Mohammed A Mohammed
researcher

Tom Marshall
clinical lecturer

Correspondence to:
P Adab
padab@bham.ac.uk

BMJ 2002;324:95-8



Details of constructing control charts appear on bmj.com

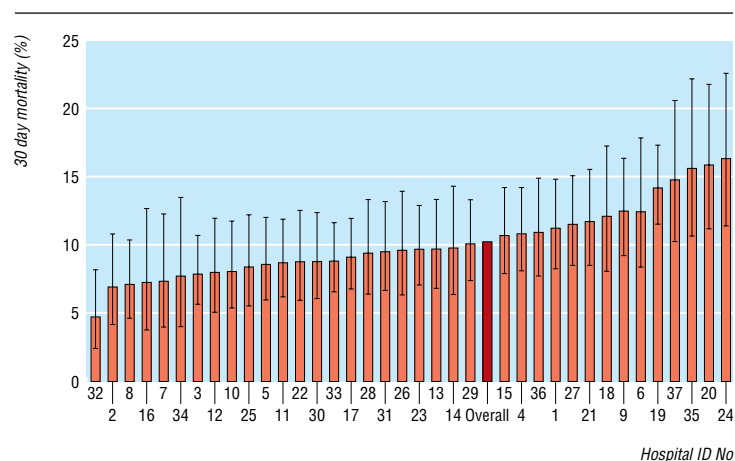


Fig 2 League table for mortality (with 95% confidence interval) in hospital within 30 days of admission for patients admitted with myocardial infarction (patients aged 35-74 years admitted to the 37 very large acute hospitals in England during 1998-9)

addition, traditional league tables usually rank goods or services produced by different production systems, whereas NHS performance league tables rank components (such as hospitals) within a single system (the NHS). We know that dissimilar production systems produce goods with very different quality characteristics. Thus, when the products from each system are compared and ranked, we have good reason to take differences in rank position at face value. We therefore have reason to believe that car security rankings noted in figure 1 do vary between models (for example, Citroens are less secure than Audis). On the other hand, NHS providers comprise a single system. All NHS hospitals recruit staff in similar ways; require each grade of staff to have almost the same training; have the same pay structure, administration, and support; and work to the same national policies. The NHS Executive does differentiate between six types of hospital based on size, whether they are acute, specialist, teaching or multiservice hospitals. It would therefore seem reasonable to consider that the 37 hospitals in figure 2 are components of a single "NHS large acute hospital system." Some of the observed differences between performance scores recorded at each hospital will undoubtedly have arisen by chance (sampling error), and rank positions cannot be taken as absolute. The confidence interval error bars are an attempt to portray the range of performance score estimates that include the true value.

Definition of NHS performance league tables

Numerous authorities support the use of performance league tables but none defines them. We define them as a technique for displaying comparative rankings of performance indicator scores of several similar providers. They are principally used when no standard against which to judge performance has been set. According to the NHS Executive, performance league tables have two purposes: firstly, to identify a few providers whose performance indicator scores are appreciably greater or lower than expected, and, secondly, to show the range in variation between providers. The Department of Health's intention in

publishing the league tables is to ensure that "where there are large and unexplained variations in performance, every effort is made to find out why, and work is put in train to bring about an early improvement."²

Pros and cons of NHS performance league tables

There is little doubt that NHS providers generally oppose the publication of league tables³ and that such publication has a negative impact on public trust and professional morale.⁴ Some arguments against their use could be levelled against any monitoring or assessment system, and two of these have considerable merit. Firstly, the value of any performance indicator depends on the quality of the data used in its calculation. Unfortunately, many NHS data have been of poor quality,⁵ partly because the NHS relies on the cooperation of health service providers and depends on their data management systems. Another justified criticism is that not all outcomes valued by society are measurable, and most NHS performance indicators have been selected on the basis of what is available and practical rather than what is meaningful.⁵ However, the weightiest arguments against NHS performance league tables are specific to them and unrelated to the data used to construct them.

Proponents of performance league tables believe their publication stimulates competition, and that, as each provider adopts "best practice," the quality of services will improve.⁶ Some published evidence seems to support this. The state of New York probably has the best established system for providing the public with information on the quality of coronary artery bypass graft services. Soon after publication of performance league tables based on providers, the risk adjusted mortality for bypass surgery declined, leading some to conclude that this was a direct result.⁷ These conclusions have been challenged, however,⁸ with one alternative explanation being that, once providers know their data will be used for comparative purposes, they may resort to "creative reporting."⁹

Performance league tables may also improve patients' choice, and proponents argue that this is necessary for an efficient market economy by encouraging consumers to seek out high ranking providers.⁹ This is largely irrelevant in Britain, however, as patients have little choice when they use the services of a doctor, clinic, or hospital. Furthermore, a systematic review of the literature on the effects of public release of performance data showed that individual consumers and purchasers don't search out, understand, or use available data.¹⁰

Performance league tables might also encourage managers and purchasers to put more emphasis on quality rather than on low unit cost. However, the published evidence suggests that the tables may lead to unintended, adverse consequences such as gaming—encouraging providers to focus on performance measures per se rather than improving the quality of care.¹¹ For example, the publication of school league tables has probably led to some schools concentrating on meeting targets at the expense of other important objectives.¹²

The strongest argument in favour of league tables is that they are one of the few aids available to health

system regulators for monitoring and ensuring the accountability of providers.⁵ Regulators can use league table rankings to identify clinicians or hospitals with a high frequency of selected adverse outcomes as a starting point for further inquiry. To our knowledge this has never been seriously pursued.

Even if publication of league tables has none of the above benefits, it might be argued that they do no harm. Unfortunately, with ranking of performance, there is an implicit assumption that providers located towards the bottom provide a worse service. A cult of naming and shaming “poor performers” has evolved,¹³ with blame often apportioned to individuals. Such apportioning of blame is generally unfair as it ignores the fact that modern medicine is a complex process and that outcomes are influenced by referral patterns, team responsibilities, and the availability of paramedical and other resources. Providers with poor performance scores inevitably argue that they treat patients with more complex problems and that insufficient adjustment has been made for their case mix. Indeed, small differences in patients’ age, sex, medical history, and social class can change performance indicator scores and a provider’s relative position in a league table.¹⁴ Unfortunately, there is little agreement among experts on the validity of various strategies for risk adjustment.⁸ As a result, some argue that, in order to avoid a poor ranking, providers may refuse to treat critically ill patients¹⁵ or may refer them to other hospitals.¹⁶

However the main reason why many dislike performance league tables is because someone must come bottom of a league. Providers realise that a performance league table produces a static snapshot of performance and that their ranking is almost entirely decided by chance. Imagine that patients are referred to a two consultant partnership. Patients are allocated to each consecutively, and, over 10 years, 100 of each consultant’s patients die. From this, most would conclude that the two consultants offer the same quality of service. In any one year, however, we would not expect the number of deaths among patients from each consultant to be exactly equal, and one would therefore be ranked above the other by chance. Since there is evidence that performance indicator rankings will be used to penalise providers and have led to hiring, firing, disciplining, and paying of bonuses to staff,⁷ 17 concerns about the capricious nature of performance ranking by providers seem to be justified.

Overcoming the problems of performance league tables

Many professionals agree that league tables are misleading, and the BMA has argued for a different method of analysing performance data.³ Several methods have been proposed, including the Monte Carlo technique and bayesian statistics³ or development of an absolute standard against which all units are measured. However, all of these techniques have their own difficulties and need specialist skills for analysis and interpretation. We believe that a more user friendly method of assessing the quality of services can be imported from industry. This method, derived from statistical process control,¹⁸ is based on the recognition that the outputs of even the most perfectly tuned production system inevitably show some variation. This

means that even under ideal conditions similar providers (doctors or trusts) will never match each other’s performance exactly, or indeed their own performance from one month to the next.

NHS performance league tables attempt to portray this variation by including 95% confidence intervals for each provider’s performance. These make performance league tables visually confusing, and the use of this level of confidence for differentiating between a significant and non-significant difference is rarely appropriate.¹⁹ It also means that when all providers are offering a similar service in a stable system about 5% (1 in 20) will always be identified as outliers. Thus, in figure 2, hospitals 32, 19, 35, 20, and 24 would be identified as outliers.

From league tables to control charts

The purpose of any monitoring system is to sort out “signals” from background “noise.”¹⁸ Engineers have long known that most of the variation detected by a monitoring system results from “common causes,” which account for the noise in a stable system. Changing any single system component cannot reduce the noise, and must be avoided if the system is to remain stable.²⁰ On the other hand, some of the variations in a system result in a “signal” that a “special cause” should be sought. With statistical process control theory,¹⁸ 20 the same data used to construct the performance league table (fig 2) can be used to create a “control chart” (fig 3), with “control limits” representing the upper and lower limits of common cause variation.

The control chart shows no ranking of providers, is visually less confusing than the league table, and the few hospital outliers (19, 35, and 32) are easy to identify. The league table shows hospital 24 to have the highest mortality and provides a strong visual suggestion that it offers the worst service. The control chart, on the other hand, does not identify hospital 24 as exceptional since its performance falls within the control limits. Its apparent high mortality can reasonably be attributed to “common cause variation,” and no specific action is needed. More importantly, the control chart shows that the high mortality at hospital 19 cannot be attributed to common cause variation and signals that those in charge should seek a special cause to explain the observation. Our experience with students, junior doctors, consultants, and managers

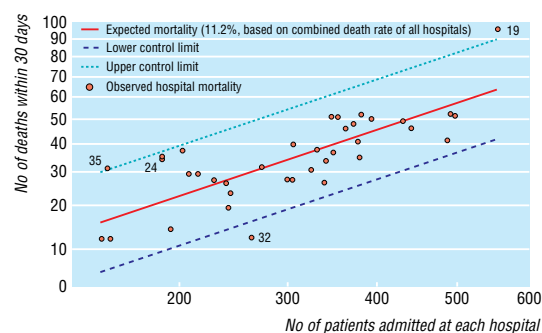


Fig 3 Control chart for number of deaths in hospital within 30 days of admission for patients admitted with myocardial infarction (patients aged 35-74 years admitted to the 37 very large acute hospitals in England during 1998-9). (Both axes show square root of values, according to method by Tukey²¹)

shows that, when presented with the league table, they do not identify hospital 19 as an outlier, whereas they invariably do so when presented with the control chart.

Both the league table and control chart show hospital 32 to have an extremely low mortality. The control chart sends out a signal that a “special cause” may be operating. Those responsible should make inquiries, paying special attention to their reporting and coding system, the hospital organisation, the equipment and facilities available, staff training, and their case mix and referral methods. If a likely explanation for the superior outcome in this hospital is identified, all other hospitals in the system should consider adapting their procedures accordingly.

When all special cause variation has been identified and addressed and subsequent monitoring shows that performance of all providers falls within the control limits, the only way to improve the quality of care would be to change the entire system. This “systems approach,” as opposed to concentrating on individuals, provides a means for improving health outcomes, and control charts have been used successfully to improve performance in both health care^{22 23} and the commercial sector.²⁴

Difference between control charts and performance league tables

The major difference between control charts and league tables is that they are based on diametrically opposed premises. The reason for constructing league tables is the implicit assumption that there is a performance difference between providers. With control charts, however, there is an explicit assumption that all providers are part of a single system and have similar performance. In the former instance a more sensitive monitoring system is needed, whereas in the latter the monitoring system should be very specific. There is no published statistical basis for league tables, but, by using the 95% level of confidence for decision making, they are more sensitive than control charts, which use the equivalent of a confidence level of 99.8%.

Conclusions

While the public has a right to know about the quality of services, it is irresponsible to provide information that is of questionable validity or difficult to comprehend.¹⁷ NHS performance league tables are difficult to comprehend and easy to misinterpret, but their publication by an official body lends them credence.²⁵ We believe control charts are easier to interpret and would provide an intuitive technique for assessing health service performance and promoting a systems approach to monitoring and improving quality.

Contributors: AMR formulated the idea for the article, prepared the first draft, and provided the examples used in figures 2 and 3. PA developed the idea, undertook the literature review, and wrote the article. MAM and TM provided insights and expertise on the role of control charts in the NHS and contributed to writing the article.

Competing interests: None declared.

- 1 NHS Executive. *NHS performance indicators: July 2000. Information in support of the publication*. London: Department of Health, 2000.
- 2 NHS Executive. *HSC 2000/023. Quality and performance in the NHS. Performance indicators: July 2000*. Leeds: Department of Health, 2001.
- 3 BMA Science Department and Health Policy and Economic Research Unit. *Clinical indicators (league tables)—a discussion document*. London: BMA, 2000.
- 4 Davies HI, Lampel J. Trust in performance indicators. *Qual Health Care* 1998;7:159-62.

- 5 Smith P. The use of performance indicators in the public sector. *J R Stat Soc A* 1990;153:53-72.
- 6 Brook RH. Health care reform is on the way: do we want to compete on quality? *Ann Intern Med* 1994;120:84-6.
- 7 Chassin MR, Hannan EL, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. *N Engl J Med* 1996;334:394-8.
- 8 Green J, Wintfeld N. Report cards on cardiac surgeons—assessing New York state's approach. *N Engl J Med* 1995;332:1229-32.
- 9 Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York state. *JAMA* 1994;271:761-6.
- 10 Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: What do we expect to gain? A review of the evidence. *JAMA* 2000;283:1866-74.
- 11 Smith P. On the unintended consequences of publishing performance data in the public sector. *Int J Public Administration* 1995;18:277-310.
- 12 Tymms P, Wiggins A. Schools' experience of league tables should make doctors think again [letter]. *BMJ* 2000;321:1467.
- 13 Anderson P. Popularising hospital performance data. *BMJ* 1999;318:1772.
- 14 Leyland AH, Boddy FA. League tables and acute myocardial infarction. *Lancet* 1998;351:555-8.
- 15 Schneider EC, Epstein AM. Influence of cardiac surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. *N Engl J Med* 1996;335:251-6.
- 16 Omoigui NA, Miller DP, Brown KJ, Annan K, Cosgrove D 3rd, Lytle B, et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation* 1996;93:27-33.
- 17 Kassirer JP. The use and abuse of practice profiles. *N Engl J Med* 1994;330:634-6.
- 18 Mohammed MA, Cheng KK, Rouse A, Marshall T, Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;357:463-7.
- 19 Sterne JAC, Davey Smith GD. Sifting the evidence—what's wrong with significance tests? *BMJ* 2001;322:226-31.
- 20 Deming WE. *Out of crisis*. Massachusetts: Massachusetts Institute of Technology, 1986.
- 21 Mosteller F, Tukey JW. The uses and usefulness of binomial probability paper. *J Am Stat Assoc* 1949;44:174-212.
- 22 Finison LJ, Finison KS. Applying control charts to quality improvement. *J Healthcare Qual* 1996;18:32-41.
- 23 Christie B. Manufacturing process helps Scottish hospital halve MRSA rates. *BMJ* 2001;322:1014.
- 24 Wood B, Williams R, Anker D, Gardener N. Performance management and improvement at the Automobile Association. *Supply Chain Pract* 2001;3:30-45.
- 25 Goldstein H, Myers K. Freedom of information: towards a code of ethics for performance indicators. *Res Intell* 1996;57:12-6. (Accepted 12 October 2001)

Corrections and clarifications

Evidence based management of nocturnal enuresis

At the end of this article by Jonathan H C Evans that appeared in the “Evidence based paediatrics” series (17 November, pp 1167-9) we stated that the author had not declared any competing interests. Unfortunately, this is wrong; he had supplied us with the following statement, which somehow did not find its way into the published article. We apologise for this.

Jonathan H C Evans has received the following from Ferring Pharmaceutical, manufacturer of desmopressin: reimbursement for attending conferences, fees for lecturing, and a consultancy fee that was paid into a research fund.

Promoting normality in childbirth

We inadvertently introduced two errors into two references at the end of this editorial by Richard Johanson and Mary Newburn (17 November, pp 1142-3), for which we apologise. In one case (reference 4) we used the wrong reference, and in the other (a Cochrane review, reference 8) we inserted the wrong CD number. The two references should read as follows:

- 4 Johanson RB, El-Timini S, Rigby C, Young PW, Jones PW. Caesarean section by choice fulfils the inverse care law. *Eur J Obstet Gynecol Reprod Biol* 2001;97:20-2.
- 8 Hodnett ED. Caregiver support for women during childbirth. *Cochrane Database Syst Rev* 2000;(2):CD000199.

Additionally, the correct name of the journal cited in references 6, 11, and 13 is *MIDIRS Midwifery Digest* [not *Midwifery Digest*].