Star wars, NHS style

Richard M Barker, Mark S Pearce, Miles Irving

NHS trusts awarded three stars are supposed to be the best performing in the country. However, problems with the 2002-3 assessment mean that this is not necessarily true

Monitoring of performance is generally agreed to be essential for the efficient running of large publicly funded organisations such as the NHS. However, to be effective it must not only be statistically sound but also have the confidence of the organisations to which it applies. A report of the Royal Statistical Society in 2003 noted that "Done badly it [performance monitoring] can be very costly and not merely ineffective but harmful and indeed destructive."¹

The performance of NHS trusts is monitored through the star ratings system, which the Department of Health introduced in September 2001. The Commission for Health Improvement was given responsibility for assessment in 2002-3. We investigated the rating system after our trust was downgraded from three stars (the highest rating) in 2001-2 to two stars in 2002-3. Our analysis shows important shortcomings with the method of assessment.

2002-3 star ratings

In 2002-3 the performance of NHS trusts was assessed for nine key targets, together with a balanced scorecard of clinical focus (10 indicators), patient focus (19 indicators), and capacity and capability (seven indicators). Each trust's performance on the key targets was assessed on the basis that they had been "achieved, underachieved, or significantly underachieved." Trusts were given two penalty points for underachievement of a target and six points for significant underachievement. An overall score was derived by adding scores in key targets and the balanced scorecard. The highest scoring trusts were given three stars. However, trusts that had significantly underachieved in one of the key targets could not be awarded three stars, whatever the overall score.

The assessment showed that our trust had significantly underachieved on one key target, the outpatient indicator (by 0.1%), and the trust was thus awarded two stars. This was despite the trust achieving the top band of performance for all three of the clinical focus, patient focus, and capacity and capability sets of indicators, and eight out of nine of the key target indicators.

Statistical considerations

Our review of the statistical methods adopted for the star ratings shows that inappropriate criteria were used to arrive at the judgments. We obtained data on acute trusts and indicator targets from the Commission for Health Improvement and Department of Health websites (www.chi.nhs.uk/ratings/ and www.doh.gov.uk). The outpatient indicator was a composite score calculated from the number of patients waiting longer than 26 weeks at the end of each of the first three quarters of 2002-3 plus the number of patients who were still waiting longer than 21 weeks at the end of the fourth quar-

ter. To achieve the target, trusts had to have had no more than five breaches; more than 50 breaches constituted significant underachievement.

Use of quarter end figures

The size of a sample must be large enough to be representative of a population. The outpatient indicator stated that its purpose was to measure a trust's performance throughout the year. However, the commission sampled only four days, even though it had all the figures necessary to calculate performance over the whole year. With this method, a trust could have a considerable number of breaches during a quarter yet show none on the final day. The total figure for Newcastle showed that there were 1502 breaches during the year. Other trusts had more total breaches during the year yet were judged to have met the target because their figures on the four quarter end days fell within requirements. Table 1 shows the trusts that had over 1000 breaches of the target throughout the year. Of these 19 trusts, 11 achieved the target, six underachieved, and two (including Newcastle) significantly underachieved.

Absolute or relative test

In contrast to other key targets based on percentages, the outpatient key target was an absolute test—that is, it measured the number of breaches of the target irrespective of the size and activity of the trust. The number of referrals differs greatly across trusts, which range from small specialist hospitals that receive just 64 general practitioner referrals a year to large university hospitals such as Newcastle, which receives 118 000 referrals a year. Table 2 shows that a breach of the target waiting time in up to 6% of patients could lead a trust to achieve, underachieve, or significantly

Newcastle upon Tyne Hospitals NHS Trust, Freeman Hospital, Newcastle upon Tyne NE7 7DN

Richard M Barker
business and development director
Miles Irving
chairman of board

School of Clinical Medical Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne
Mark S Pearce
senior research associate

Correspondence to: M Irving
m.hirving@ncl.ac.uk

BMJ 2004;329:107-9

Freeman Hospital, Newcastle upon Tyne

Table 1 Total breaches of outpatient waiting time within 2002-3 among trusts with over 1000 breaches

Trust	No of patients waiting >26 weeks	Achieved standard	Star rating
Leeds Teaching Hospitals NHS Trust	3139	No	2
West Hertfordshire Hospital NHS Trust	2803	Yes	1
Royal Free Hampstead NHS Trust	2369	Yes	2
St George's Healthcare NHS Trust	2172	No	2
Wirral Hospital	1692	Yes	3
University Hospital of Coventry and Warwickshire NHS Trust	1671	Yes	2
Barnet and Chase Farm Hospitals NHS Trust	1660	Yes	1
Mid Yorkshire Hospitals NHS Trust	1577	No	1
Barking, Havering, and Redbridge Hospitals NHS Trust	1565	Yes	2
Hull and East Yorkshire Hospitals NHS Trust	1544	No	1
Newcastle upon Tyne Hospitals NHS Trust	1502	No	2
Princess Royal Hospitals NHS Trust	1438	Yes	1
Plymouth Hospitals NHS Trust	1344	Yes	2
North Staffordshire Hospital NHS Trust	1312	Yes	2
South Tees Hospitals NHS Trust	1288	No	2
East and North Hertfordshire NHS Trust	1225	No	1
Royal United Hospital Bath NHS Trust	1114	No	1
Peterborough Hospitals	1108	Yes	3
Norfolk and Norwich Healthcare NHS Trust	1041	Yes	2

Table 2 Analysis of breaches of outpatient waiting time for all NHS trusts

	No of outpatients waiting longer than the standard		
	0-5 (achieved)	6-50 (underachieved)	>50 (significantly underachieved)
No of trusts	156	14	6
Range of No of breaches	0-5	8-49	57-288
Range of No of new written referrals	64-13 9011	764-152 778	27 788-188 102
Possible range of % outpatient breaches	0-7.8	0.004-6.5	0.03-100

underachieve this target, depending on patient workload.

Clearly, it is easier for a small trust to meet the threshold of five or fewer breaches a year than it is for trusts with a large referral base. The Commission for Health Improvement used a relative test for other indicators. The rationale for choosing relative tests for some key targets and one patient focus indicator and numerical targets for other similar indicators is difficult to understand.

Six of the nine key target indicators specify absolute rather than percentage targets. Absolute targets may demand comparatively higher levels of service from larger trusts, although even those key targets measured in percentage terms may be easier to attain in small trusts. To investigate this, we considered data from the 150 trusts that had returned information

for all six key target indicators concerning patient numbers.

We compared the patient population in the highest ranking trusts with that in the remaining trusts for the six key targets. For all indicators, the highest ranked trusts had smaller average patient populations (table 3).

Discussion

A three star rating brings appreciable benefits to a trust. Trusts with three stars can expect to be given greater autonomy in line with the NHS commitment to devolve responsibility and decision making away from central government.

With such high stakes, the performance assessments should be beyond reproach. We have shown that this was not the case in 2002-3. Calculation of annual breaches of targets by using quarter end figures was flawed. Some trusts with many breaches of the target were judged as having achieved the indicator while others with lesser breaches were judged to have underachieved. The method might be acceptable if only end of quarter figures were available. However, the relevant data were available throughout the year.

Our analysis also shows that the methods used for the 2002-3 star ratings seem to discriminate against larger trusts. This is because absolute rather than relative measures were used. The performance of larger trusts will also be affected by factors such as the complexity of cases referred, especially when regional or national specialisation exists. In addition, capacity for specialist services is limited, and expansion is a challenge because growing requirements are difficult to predict.

Techniques used to judge the delivery of a service must have the confidence of those being judged as well as that of the government and the community. The concerns over the methods of assessment that we raise above have been reinforced by the Royal Statistical Society working party report on performance monitoring in the public sector.¹ This report identifies inadequacies in current performance management techniques, including some of those used by the Commission for Health Improvement. In particular, the society criticises targets that are the same in absolute terms for small groups as for large groups when the precision of estimation differs greatly.

Table 3 Relation between patient population and performance ranking. Data were analysed with the Mann-Whitney test

Key target	Trusts ranked joint first		Trusts ranked lower than first		P value
	No of trusts	Average patient population	No of trusts	Average patient population	
Waiting time for admission from accident and emergency	91	65 497*	59	79 548*	0.001
Cancelled operations not readmitted within 28 days	26	22 169†	124	29 193†	0.01
Inpatient waiting time	130	27 853‡	20	31 920‡	0.39
Outpatient waiting time	26	9 887‡	124	12 686‡	0.005
Total time in accident and emergency departments§	50	1 370¶	100	1 497¶	0.12
Time between GP referral and first cancer appointment	18	1 575**	132	2 023**	0.03

*Median annual total number of first attendances.

†Median annual total number of elective admissions.

‡Median annual total number of outpatients seen after written referral from general practitioner.

§Top 50 ranked trusts, as no trusts were ranked joint first. Based on weekly data.

¶Median total number of attendances for the week ending 30 March 2003.

**Median annual total number of patients seen for first outpatient appointment when urgently referred by their general practitioner with suspected cancer and whose referral was received by the NHS trust within 24 hours.

The Healthcare Commission has inherited the ratings task for 2003-4 and has recently advised that the number of breaches will now be measured as a percentage of the total number of patients a trust treats.² Our analysis shows that trusts need to monitor constantly the statistical rigour of the performance indicators used to assess them.

We thank L Parker for help with the statistical analyses and M F Laker for advice.

Contributors and sources: RMB is an executive director with extensive knowledge and experience of planning and performance management in healthcare. MSP is a Royal Statistical Society recognised chartered statistician and an experienced non-clinical epidemiologist. MI is a retired clinical academic with a background in health technology assessment and the transfer of evidence based practice into acute trust policies. All the above contributed equally to the paper. The article resulted from a detailed analysis of the reasons why a high performing trust breached a key indicator in the 2002-3 star ratings exercise.

Competing interests: RMB and MI are members of the Newcastle upon Tyne Hospitals NHS Trust Board of Directors. MSP was paid by the trust for a statistical report on the trust's performance in the star ratings exercise

Summary points

Performance of NHS trusts is reported in terms of a star rating

The methods used to assess performance in 2002-3 were flawed

Use of absolute rather than relative figures for some measures created bias against larger trusts

Use of end of quarter figures for outpatient waiting times masked large breaches in some hospitals

More statistical rigour is needed to ensure trusts' confidence in the ratings system

1 Bird SM, Cox D, Farewell VT, Goldstein H, Holt T, Smith PC. *Performance indicators: good, bad, and ugly*. London: Royal Statistical Society, 2003.

2 Healthcare Commission. *Indicator listings for acute trusts*. http://ratings.healthcarecommission.org.uk/Indicators_2004/ (accessed 11 July 2004).

(Accepted 18 April 2004)

Commentary: Performance assessment is here to stay

Anna Walker

Barker and colleagues look at how performance on one indicator out of 45 affected last year's star rating for their trust. The two specific concerns they raise (which the Commission for Healthcare Improvement discussed with them last summer) are that the target was measured at the quarter end and that it was measured using absolute rather than relative thresholds.

No one has ever argued that the commission's method was the only way to measure the target. When the commission accepted responsibility for ratings from the Department of Health, some consistency had to be maintained. The previous year, the Department of Health had assessed performance on both inpatient and outpatient targets on an absolute basis.

The absolute thresholds were retained because the target is framed in absolute terms—that is, no patient will wait longer than a specified time. Quarter end data were chosen so that they could be adjusted to exclude, for example, Welsh patients treated in English trusts, who are not subject to the same rigorous waiting times targets. The data used were the only set for which such an adjustment was made by the Department of Health in 2002-3.

Trusts had an opportunity to comment on the methods used to construct all the indicators included in ratings before they were finally published. Two or three trusts commented on the construction of this indicator, but Newcastle did not make these points until after the ratings were published.

The commission committed itself to a review after it had published its first set of ratings in 2003. We now know the result of that review, and the outpatient target will be measured this year by using month end data and with proportionate thresholds, meeting many of the concerns that the Newcastle case illustrates. This revised approach, however, would be unlikely to have

changed the result for Newcastle last year. According to its own quarter end data, calculated in proportion to overall patient numbers, Newcastle had the sixth high percentage of breaches in the country.

Future of performance ratings

Although there is always benefit in looking at the past, I am keen to look to the future on assessing the performance of NHS organisations. Everyone has always agreed that star ratings are not perfect but also that performance assessment can be beneficial. It is now time to build some agreement about a better method.

The Healthcare Commission wants a system of performance assessment that is more accessible to the public, that drives improvement in the NHS, that is seen as relevant and fair by the service and clinicians, and that is more comprehensive in how it measures organisations. If our performance assessments are going to be used to decide rewards and sanctions, we also need to work with the Department of Health to ensure that they are fit for that purpose.

This year's and next year's star ratings will look familiar. We have made and will make improvements, but some of the aspects that have been criticised will necessarily remain. As a new organisation we are beginning to consider how we will assess NHS organisations in the future. In the autumn, we will publish a consultative document about our future methods of assessment, with a view to using the new methods for ratings from 2006. We will be looking to engage with doctors, nurses, and others across the NHS about how this system might work. I hope *BMJ* readers will take the opportunity to help us design something better.

Competing interests: None declared.

Healthcare Commission,
London EC1Y 8TG
Anna Walker
chief executive
anna.walker@healthcarecommission.org.uk