

Impact of star performance ratings in English acute hospital trusts

Russell Mannion, Huw Davies¹, Martin Marshall²

Centre for Health Economics, University of York; ¹Centre for Public Policy & Management, University of St Andrews; ²National Primary Care Research & Development Centre, University of Manchester, UK

Objective: To explore some of the impacts of star performance ratings in acute hospital trusts in England.

Methods: A multiple case study design was used which incorporated purposeful sampling of 'low' and 'high' performing trusts using the star rating system. In each case study site, data collection comprised semi-structured interviews and documentary analysis. Between eight and 12 senior managers and senior clinicians were interviewed in each organisation.

Results: There was a general view that the star ratings as presently constituted did not represent a rounded or balanced scorecard of their own organisation's performance and a widespread belief that the information used to calculate the ratings was often incomplete and inaccurate. The star ratings were viewed by some managers as useful, in that they gave added weight to their trust's modernisation agenda. In addition to driving beneficial change, the ratings were also sometimes reported to have inadvertently induced a range of unintended and dysfunctional consequences, including tunnel vision and a distortion of clinical priorities, bullying and intimidation, erosion of public trust and reduced staff morale, and ghettoisation.

Conclusions: Set in the context of an international body of research, this study highlights some important gaps in knowledge and failings in current policy and practice. In particular, the many dysfunctional consequences of publishing star ratings indicate a need for a re-examination of performance management policies.

Journal of Health Services Research & Policy Vol 10 No 1 2005: 18–24

© The Royal Society of Medicine Press Ltd 2005

Introduction

Health systems around the world are placing increasing demands on health care organisations to deliver improvements in the performance and quality of the services they provide. In many countries this aim is manifest increasingly in the development of external, publicly mandated reporting systems that seek to hold providers to account and (most especially in the USA) to provide market information to enable consumers and group purchasers to make informed choices.^{1,2} Such policies are premised on the assumption that the provision of comparative quantitative data will deliver genuine improvements in health care quality and performance, even though there is an absence of hard evidence on the benefits and risks of public disclosure and little in-depth understanding of how these data are perceived, received and acted upon in provider organisations.^{1,3–5}

Russell Mannion PhD, Senior Research Fellow, Centre for Health Economics, University of York, York YO10 5DD, UK. **Huw Davies PhD**, Professor of Health Care Management & Policy, Centre for Public Policy & Management, University of St Andrews, St Andrews, UK. **Martin Marshall MD**, Professor of General Practice, National Primary Care Research & Development Centre, University of Manchester, Manchester, UK.

Correspondence to: RM.

Policy context

In September 2001, the UK government published the first star ratings for National Health Service (NHS) organisations in England that provide acute hospital services. At the time, this initiative was heralded as a step towards fulfilling a stated government commitment to provide patients and the general public with comprehensive, easily understandable information on the performance of their local health services. The second set of annual star ratings was expanded to cover all acute hospitals, specialist hospitals and ambulance services. In July 2003, an independent body – the Commission for Health Improvement (CHI) – was given responsibility for publishing annual performance data, including star ratings, for all NHS organisations in both acute and primary care. Following the abolition of CHI in April 2004 a new regulator – the Commission for Healthcare Audit and Inspection – was handed responsibility for publishing these data. It should be noted that the star performance rating system is but one approach to assessing NHS performance. Clinical governance reviews, National Service Framework reviews, national and local audits and performance monitoring by strategic health authorities are also used to monitor the performance and quality of NHS organisations.

Categorising hospital performance

The star rating system categorised hospitals into four groups:⁶ three-star organisations were judged to have the highest level of performance; two-star organisations were judged to be performing well overall but not achieving the same consistently high level of performance as three-star hospitals; one-star organisations were judged to have some cause for concern, particularly regarding key government targets; and zero-star organisations were judged to have the lowest level of performance against key government targets.

The criteria used to judge performance fell into five areas. The scores derived from each key area from the process described above were then combined in a complex six-step process to produce the trust star ratings (see Table 1).

It is difficult to overestimate the importance of star ratings. In addition to facilitating accountability to patients and the public, they also serve as an important tool for concentrating management attention on key strategic priorities and national targets. They are the metrics used to determine access to a range of 'earned autonomy' freedoms and rewards, including eligibility to operate more independently of central government (so-called 'foundation status'), and in extremis are the criteria used to justify franchising the management of a persistently 'failing' organisation.¹ Thus, the star ratings are the fulcrum around which the current NHS performance management system operates and it would therefore appear essential to obtain a balanced understanding of how these data affect provider behaviour, for good or ill. Of course, such initiatives to improve health care systems through the public release of broad-brush quality measures are not confined to the UK. The USA, Australia, New Zealand, Norway, Sweden, Italy and others are experimenting with similar approaches.⁷

Methods

Outline of case study theory and methods

As part of a large multi-method study of organisational dynamics and performance in the NHS, we carried out six in-depth case studies of acute hospitals in England.⁸ Case studies are typically used to explore the interplay of all organisational variables and thereby provide a more holistic or 'thick description' of a situation. In case study research, generalisation does not depend on conventional statistical logic. Where data are drawn from individual or multiple settings, inferences from these settings to other contexts depend on the adequacy of the theory (theoretical inference) rather than adherence to the conventional technical rules of statistical inference.⁹

Selection of case study sites

We adopted a multiple case study design, which incorporated purposeful sampling of 'low' and 'high' performing trusts using the Department of Health star rating system. Trusts at either end of the performance spectrum are likely to offer sharper contrasts in terms of their experience and ambition for culture change and may therefore provide more valuable insights and perspectives than case studies based on a sample of middling performers. Given the current policy focus in the NHS on 'turning around' under-performing organisations, we decided to select a sample of four 'low' performing hospital trusts (zero or one star in 2000/01) and two 'high' performing hospital trusts (those awarded three stars) (Table 2). The sample was drawn from across England at a time when 20% of acute trusts in England had been awarded zero or one star and 20% had been awarded three stars.

Table 1 Key areas, criteria and scoring for star rating system

Performance areas	Criteria	Scoring system
Key government targets	Mostly relate to access and waiting list targets	Each target scored as 'achieved', 'under-achieved' or 'significantly under-achieved'
Performance indicators with a clinical focus	Include indicators relating to mortality rates, re-admission rates and rates of successful discharge to home following hip fractures and strokes	Judged as 'significantly above average', 'above average', 'average', 'below average', 'significantly below average'. Five points awarded to highest level and one to lowest. Indicator scores combined (unweighted) to produce overall score for clinical care. Scores presented with confidence intervals
Performance indicators with a patient focus	Includes indicators relating to waiting times for inpatient treatment, outpatient appointments, casualty and cancer treatment, delayed discharge, cancelled operations and patient experience of care (from surveys)	As above. Scores banded in percentiles
Performance indicators with a capacity and capability focus	Includes indicators relating to data quality, staff opinions (from survey), junior doctor hours and sickness absence rates	As above. Scores banded in percentiles
CHI reviews	Largely qualitative assessment of clinical governance systems across seven components: risk management; clinical audit; research and education; patient involvement; information management; staff involvement; education, training and development	Judged as zero star if significant weakness found by CHI review in five or more of the seven areas. Judged as three star if significant strengths in all areas and no significant weaknesses in any area

CHI, Commission for Health Improvement.

Table 2 Background characteristics of the six hospital trusts

Hospital trust	A	B	C	D	E	F
Measured performance	Low	Low	Low	Low	High	High
Hospital type	District general	District general (merged)	District general	Teaching	District general	District general
Location	Town	City	Town	City	City	Town
Income 2000/01 (£ million)	Less than 100	Greater than 100	Less than 100	Greater than 100	Greater than 100	Less than 100
Number of beds	501–1000	501–1000	Under 500	501–1000	501–1000	Under 500

Primary selection of subjects, data collection and analysis

In each site, data collection comprised semi-structured interviews and documentary analysis (e.g. CHI reports and internal clinical governance reports). Between eight and 12 key managers and senior clinicians were interviewed in each organisation in 2002–2003. The analysis is based on 61 interviews. For all participating trusts, these included the chief executive and the medical director and a range of the following staff: deputy chief executive; trust chair; director of finance; director of human resources; clinical directors; director of nursing; and business managers.

The interviews covered various questions related to organisational dynamics, including perceptions and experience of performance indicators and measures, and the impact of the star ratings on the organisation, department and individuals. (A full list of interview topics is available from the authors.) The interviews lasted on average 60 minutes, were audio-taped, transcribed and analysed using content analysis with the aid of the computer package 'Atlas ti'. At least two members of the project team read each of the transcripts and the emerging themes from the preliminary analysis were discussed at project group meetings. In the light of these discussions, transcripts were subject to further investigation and corroboration, including a search for 'disconfirming' cases.

Results

Our findings are structured around three key emergent themes.

Do hospital staff view the star ratings as fair assessments of their organisation's performance?

Within the context of eliciting information about the current performance management culture, we received information on the views of staff regarding the star ratings. These responses are grouped under four headings: coverage of the measures; sensitivity to local factors; accuracy and misrepresentation; and measure fixation.

Coverage of the measures

There was a general view across the trusts that the star ratings as presently constituted did not represent a 'rounded' or 'balanced scorecard' of their own organisation's performance. In particular, staff in the 'low' performing trusts felt that many areas of excellent practice within their organisation, especially those relating to clinical practice, were either under-valued or missed completely by the ratings. Several staff reported that if these dimensions of performance were taken into account then their organisation would not have received such a low rating. The clearest example of this is trust D, which had been classified as under-performing but was rated as one of the highest performing trusts in the country in a similar exercise conducted by an independent commercial company.¹⁰

We got one star but we also got a very favourable CHI review. On the Dr Foster scale we were 9 stars out of 10, one of the best in the UK.

This lack of coherence was assumed to be due to the increased weight accorded to clinical indicators in the methodology used by the latter.¹¹

Sensitivity to local factors

A strong theme emerging from the interviews was the view that the star ratings were used in a 'mechanical' way to classify organisational performance. In particular, we heard complaints that they did not take sufficient account of the many local contingencies and mitigating factors that might help explain variations in the measured performance of hospitals. Some of the reported mitigating factors related to long-standing problems around organisational capacity (e.g. physical layout and infrastructure; financial deficits) and other were identified as external factors (e.g. performance of the local social services and other agencies within the local health economy) which were said to impact deleteriously on the performance of the trust, but were reckoned to be largely beyond the control of the hospital's management:

We're fairly pissed off with getting one star because we thought it was very petty, we understand the reasons why – the hospital is not big enough to deliver capacities, we don't have enough beds. (Trust D)

It is extremely crude (the star rating system) and that's its weakness. One of the things we didn't do well on was the cancer lists . . . one of the reasons why we are in that situation is that we are unusual for a district hospital in that we have nine Calman Hine [cancer services] speciality working groups here. Whereas many other places will only have three or four, so we are attracting more cancer patients. (Trust C)

Accuracy and misrepresentation of data

There was a widespread belief that the data used to calculate the star rating were often incomplete and inaccurate. Indeed, several staff employed at Trust A reported that their organisation had purposefully manipulated and misrepresented data in order to improve their rating:

If the information that came up was right . . . the method by which that information was arrived at wasn't looked at. It was about presenting the right image. We've got to be seen to be delivering, even if we don't . . . Now my guess is that, if the information had been accurate, we would be a no-star trust, not a one-star. (Trust A)

One of the things on which the trust was measured was 12-hour trolley waits. We reported no over-12-hour trolley waits, when in fact we had them. (Trust A)

Measure fixation and gaming

Staff across most of the trusts reported a number of examples of where their organisation had used a number of ruses to improve their measured performance. These included cancelling operations on the evening before the operation was scheduled to take place, so that that these cancellations were not recorded on the actual day, or re-classifying trolleys as beds on recording forms. Indeed, some staff at Trust D were of the opinion that their poor rating was to a large extent due to the accuracy of their reporting which may have placed their organisation at a disadvantage when compared to trusts that had not been as truthful in their data collection and reporting:

I was just at a theatre meeting today [where] we were discussing that cancellations on day of operation is a bad mark, a cross in the box there, [and we] don't want to be on [Secretary of State for Health] Mr Milburn's list do we? So what do we do? We cancel them the night before so they are not recorded as cancellations on the same day. So it is like trolley wait figures, where you put a letter above the trolley and it becomes a bed. So there are many ways of getting around these things which usually are not in the patients' best interests. (Trust A)

What beneficial organisational responses appear to be prompted by the ratings?

Interviewees were asked how they used performance data to identify and promote change. In the two 'high' performing trusts, the star ratings were being used to align internal performance management and reporting systems with key national targets. Similarly, in the four

'low' performing trusts, the new management regimes were using the ratings to guide the development of new performance management and reporting systems. In this respect the ratings served as a mechanism for transmitting important priorities from central government and helped to direct and concentrate front-line resources on those aspects of performance deemed by central policy-makers and their political masters to be important.

The star ratings were also viewed by some managers as useful in that they gave added weight to their trust's modernisation agenda and, in particular, helped to lever changes to entrenched clinical practice and traditional modes of working. Some staff, particularly those in the 'low' performing trusts, reported that the ratings (and CHI visits) had been welcomed because they had illuminated dysfunctional senior management that had hitherto remained hidden or unchallenged, and their publication had signalled to the outside world a pressing need for external support and intervention. In addition, the two 'high' performing organisations reported that their three-star status had had a very positive affect on the morale of their staff:

I wholeheartedly embrace this concept [star ratings] because it gives us a bit of backbone and a bit of bite to say 'why are you not doing this?' (Trust A)

The place got a buzz out of being a three star because everyone wanted to work in what was perceived to be the best organisation. Everyone wants to play for Manchester United, no one wants to play for Torquay . . . we got some good press. (Trust F)

What unintended and dysfunctional responses were attributed to the ratings?

In addition to driving beneficial change, star ratings were sometimes reported to have inadvertently induced a range of unintended and dysfunctional consequences for organisations and staff. Those problems directly relating to the use (or abuse) of data have already been addressed above. In addition, we found evidence of: tunnel vision and a distortion of clinical priorities; bullying and intimidation; erosion of public trust and reduced staff morale; ghettoisation; and the creation of disincentives for improvement.

Tunnel vision and distortions to clinical priorities

It was reported that the cultural shift towards meeting the external performance agenda had focused attention on areas of performance that were measured to the exclusion of other important but unmeasured areas. Many clinicians reported that their clinical priorities had been altered to meet short-term waiting targets. For example, in Trust F, it was reported that the 13-week waiting target for children's services had forced the trust to concentrate on children referred to it by doctors

rather than other professionals, even though the clinical needs of the patients may be very similar:

I manage women's and children's services and apart from our contribution to the electives we don't really impact on the star system. So as a result it seems that, a lot of the time, what we do as a directorate is not important to the management ... because we are not starred ... [what we do is] not going to help us get a star. (Trust D)

Anyone doing surgery will tell you that it is ridiculous to have to do the 15-month waiters for things like circumcisions, which are not in any way a medical priority, and it is ridiculous to make surgeons do them to meet the targets while actually not doing other cases that should be done sooner. (Trust C)

Bullying, intimidation, stress and anxiety

The pressures to meet the performance targets were reported to have led to various degrees of bullying, intimidation and harassment of staff in the apparently under-performing trusts. Given that largely we interviewed senior and middle managers it is not clear whether or to what extent this situation was replicated with staff lower down the hierarchy. Bullying and intimidation was identified as a particularly serious problem in Trust B where there was a strong feeling that the emphasis on delivering measurable improvements in performance in order to turn around the organisation had contributed to a culture of bullying. In addition, we heard reports that in the 'high' performing trusts, the desire to retain their stellar status was creating a demanding climate for staff which could sometimes extend to discomfiting levels of pressure and coercion:

The culture's changed. It's not so much about if we achieve all these stars we can become a foundation hospital – which is what I'd like to hear; it's 'so and so's job is on the line' or, 'if you don't do this it's curtains for you'. (Trust B)

It's acute pressure that had never been there before to achieve targets. And I think it's a sort of panic reaction – people start shouting and screaming at everyone else. 'Do it! Do it! Do it!'. Well we can't. (Trust B)

Erosion of public trust and reduced staff morale

The rating system may also inadvertently damage staff morale and contribute to eroding public trust in health care providers. Case study B (a 'low' performer) showed clearly how public confidence in the trust had been damaged by the publication of (and hostile press reaction to) a poor star rating. It also revealed the extent of the pressures that staff labour under in trying to cope with the loss of public confidence in the services they provide. In particular, there were reports that staff at trust B had been subjected to a hostile local media campaign that had whipped up adverse public reaction:

It was devastating [trust being awarded zero stars]. It was the worst thing that could have happened, [and] the CHI report was bad enough. It was unfair – what drags people down is the unfairness of it. Even CHI was surprised, and didn't

expect the media reaction ... We knew our jobs were on the line, but the stars were the last straw and hit right down to the workforce – whereas bad reports usually hit senior management upwards. (Trust B)

[The public perception was] 'you go to [Trust B] and you die!' We had people on the wards demanding the self-discharge forms and getting crushed in the rush to leave! It was just awful. Nurses demanding changing rooms because they didn't want to go outside the trust [in uniform] because they were being accosted in the streets ... And in the shops, people were saying 'God, you don't work for that place do you? How many have you killed today?' (Trust B)

Ghettoisation

It was reported that, because of its negative impact on the reputation of low star trusts, the ratings may be having a differential impact on the abilities of 'high' and 'low' performing trusts to attract and retain high-quality staff. Whereas all trusts experienced difficulties recruiting staff in some clinical specialities, it was clear that the 'low' performing trusts were reporting more serious problems. Staff attributed this in part to the fact that a high performance rating was 'attractive' in that it signalled to potential recruits the impression that the trust was a 'good' organisation to work for. In contrast, 'low' performing trusts reported that a poor star rating contributed to their problems as many health professionals would be reluctant to join an organisation that had been publicly classified as under-performing:

I think it [the star rating] gives a very negative view to staff, because if they are working for a one star organisation then it affects the sort of staff who want to come and work for you, but it also makes people who are currently employed here feel that they are working for a third class organisation. (Trust F)

If I look back since we've been awarded three star status, 7–8 months ago, I cannot honestly say that we have recruited staff because we are a three star organisation. It might be a factor, only a factor though. (Trust E)

Disincentives to improve performance

There was some evidence that star ratings may have a perverse impact on the performance of acute trusts, particularly those seen to be performing well. In part this was a consequence of the perception that more resources (both financial and the support of the Modernisation Agency) were being put into low-rated organisations. It also related to a feeling that three star organisations had fewer opportunities to work collaboratively with their strategic health authority:

The one area we accept that we've got to do better is on cancer waits. Then the league tables come out and we get less money than everyone else and people are saying to me in the corridor 'shouldn't we be two star next year if the money's coming out on that basis?' (Trust F)

A disbenefit of being a three star trust is that we were told that our proposals for action on initiative would go straight

to the Department of Health. It did and we got nothing. I know for a fact if it had gone through our regional office ours would have been a high priority. (Trust F)

Discussion

Relating the findings to the literature on performance rating systems

There is a growing body of evidence evaluating stakeholder perceptions of public reporting systems, the use that is made of performance ratings, and their impact on processes and outcomes of care.^{2,3,12–15}

Most of the empirical work has been conducted in the USA. This evidence relates to reporting systems, which are usually somewhat different in nature from the UK's star ratings (presenting comparative data for discrete areas of performance for each institution, rather than aggregated into a single score) and designed with a different purpose (consumer choice in the USA, performance management in the UK). The most detailed evaluation of public reporting in the UK examined the publication of the Scottish Clinical Research and Audit Group data.^{4,12,15,16} This evaluation took place at a time that the data were not being published specifically for performance management purposes, so again its relevance to the star rating system is uncertain. Therefore, most of the existing evidence should be interpreted with an element of caution in the context of the impact of England's star rating system.

Much of the evidence relates to patient, consumer and public perceptions and use of published information. By and large, patients show little interest in comparative information, often don't understand it and are sometimes quite suspicious of it.² Some of the evidence also describes the perceptions of and use made by purchasers on behalf of the public; they too fail to make use of the information, though there is some evidence that when the right data are presented in the right way then they can be encouraged to respond.² The body of evidence describing the role played by hospital staff – the focus of this study – is less well developed.

The most notable contrast between the findings of this study relating to the star rating system and those reporting other systems is the apparent high profile of the star ratings. Previous studies have suggested that hospital staff are often unaware of, or ignore, performance ratings.^{4,5,17} In this study, however, all staff seem to be highly engaged with the information. This is perhaps to be expected given our focus on senior managers rather than front-line clinical staff. However, it might also reflect the effectiveness of the communication and dissemination strategy linked to the reporting system, or possibly the comprehensibility and appeal of such a stark and simple way of presenting the data. As far as the latter explanation is concerned, there is certainly evidence that summary scores are more understandable to non-specialist audiences,^{18,19} even though they are less useful and less acceptable to an informed audience.

The historical context may be important here; star ratings are relatively new and there is evidence that initial responses are often not sustained once the novelty or profile of the reports decline.^{20,21} This finding has implications for the future of the star rating system, if the aim is to maintain an interest in the data.

The largely negative response of many hospital staff to the star ratings is congruent with the literature. We know that staff are inclined to criticise or dismiss quality reports, and that staff working in apparently poorly performing hospitals are more critical than those working in apparently highly performing organisations.^{4,5,22–24}

Whilst the initial public response to performance ratings is often negative, the finding that organisations then use the published reports to improve their internal data collection systems and processes has been described in previous studies.^{4,23} This finding suggests that the initial criticisms should not necessarily be taken at face value. The suggestion from this study is that hospitals use public reports as a lever to influence staff behaviour (sometimes positively to motivate them and sometimes negatively to bully them).

The unintended and often negative consequences of the star rating system came across loud and clear. Considerable attention has been paid to the issue of unintended responses to performance management initiatives in the public sector generally in the UK^{25,26} and in the NHS.^{27,28} Nevertheless, little attention has been paid to the range of negative responses that might be seen in the highly performance managed context of the NHS.^{29–31}

Research issues arising

The study reported here labours under some significant limitations. It was small-scale, on a small sample of trusts, and explored the dynamics over relatively short time-scales soon after the introduction of star ratings. In addition, we focused on senior managers' responses and were thus unable to triangulate their views with those from front-line staff on some of the phenomena reported.

These findings, set in the context of an international body of research, highlight a number of important gaps in knowledge that are in need of more research:

- To what extent do global measures, such as star ratings, capture a valid and reliable assessment of overall performance?
- What is the most appropriate unit of analysis for rating health services' performance?
- What is the most appropriate time period to be covered by performance ratings?
- How can performance assessments be presented in ways that both capture attention and garner credibility?
- How can organisations be encouraged and facilitated to make more productive use of performance data?

- How can organisations integrate internal quality improvement activities (such as clinical governance) with external assessments of performance?
- What dysfunctional consequences are being stimulated within organisations, how important are they, and how can they be mitigated?

The complex and dynamic nature of the phenomena under study suggest that research in this area will need to exhibit a number of features. It will need to be naturalistic, taking place in real-world settings and making careful note of the mediating role of contexts. It should be multi-method and multi-disciplinary in design, drawing on quantitative and qualitative traditions, including detailed ethnographic and discourse analytical approaches. As the phenomena of interest are essentially dynamic (performance and change), longitudinal study will offer important insights over cross-sectional designs. Finally, to provide better opportunities for theoretical transference and generalisability, the research will need sound conceptual underpinnings rather than relying on simple empiricism.

Acknowledgements

We thank all those from the NHS and associated organisations who gave so freely of their time and insights during every stage of the project.

References

1. Mannion R, Goddard M. Performance measurement and improvement in health care. *Journal of Applied Health Economics and Health Policy* 2002; 1: 13–23
2. Marshall MN, Shekelle PG, Davies HTO, Smith PC. Public reporting on quality in the United States and the United Kingdom. *Health Affairs* 2003; 22: 134–148
3. Marshall MN, Shekelle PG, Leatherman S, Brook RH. Dying to know: public release of comparative data in health care. London: Nuffield Trust, 2000
4. Davies HTO. Public release of performance data and quality improvement: internal responses to external data by US health care providers. *Quality in Health Care* 2001; 10: 104–110
5. Mannion R, Goddard M. Impact of published clinical outcomes data: case study in NHS hospital trusts. *BMJ* 2001; 323: 260–263
6. Department of Health (www.doh.gov.uk)
7. Smith P, ed. Measuring up: improving health system performance in OECD countries. Paris: OECD, 2002
8. Mannion R, Davies HTO, Marshall MN. Cultures for performance in health care: evidence on the relationships between organisational culture and performance in the NHS. A report for the Department of Health. York: University of York, 2003
9. Yin R. Case study research: design and methods. Beverly Hills, CA: Sage, 1984
10. drFoster organisation (www.dr.foster.co.uk)
11. Snelling I. Do star ratings really reflect hospital performance? *Journal of Health Organization and Management* 2003; 17: 194–209
12. Mannion R, Goddard M. The impact of performance measurement in the NHS. Report 1: Empirical analysis of the impact of public dissemination of the Scottish Clinical Resource and Audit Group data. York: Centre for Health Economics, University of York, 2000
13. Schneider EC, Lieberman T. Publicly disclosed information about the quality of health care: response of the US public. *JAMA* 2001; 10: 96–103
14. Dranove D, Kessler D, McClellan M, Satterthwaite M. Is more information better? The effects of report cards on health care providers. Cambridge: National Bureau of Economic Research, 2002
15. Mannion R, Goddard M. Public disclosure of comparative clinical performance data: lessons from the Scottish experience. *Journal of Evaluation in Clinical Practice* 2003; 9: 277–286
16. Mannion R, Goddard M. General practitioners' assessments of hospital quality and performance. *Clinical Governance: An International Journal* 2004; 9: 42–47
17. Schneider EC, Epstein AM. Use of public performance reports. *Journal of the American Medical Association* 1998; 279: 1638–1642
18. Hibbard JH, Slovic P, Peters E, Finucane ML. Strategies for reporting health plan performance information to consumers: evidence from controlled studies. *Health Services Research* 2002; 37: 291–313
19. Vaiana ME, McGlynn EA. What cognitive science tells us about the design of reports for consumers. *Journal of Medical Research* 2002; 59: 3–35
20. Luce JM, Thiel GD, Holland MR, Swig L, Currin SA, Luft HS. Use of risk-adjusted outcome data for quality improvement by public hospitals. *Western Journal of Medicine* 1996; 164: 410–414
21. Mukamel DB, Mushlin AI. Quality care information makes a difference: an analysis of market share and price changes after publication of the New York State cardiac surgery mortality reports. *Journal of Medical Care* 1998; 36: 945–954
22. Berwick DM, Wald DL. Hospital leader's opinions of the HCFA mortality data. *Journal of the American Medical Association* 1990; 263: 247–249
23. Dziuban SW, McIluff JB, Miller SJ, Dal Col RH. How a New York cardiac surgery program uses outcomes data. *Annals of Thoracic Surgery* 1994; 58: 1871–1876
24. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care: a survey of cardiovascular specialists. *New England Journal of Medicine* 1996; 337: 251–256
25. Smith P. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 1995; 18: 277–310
26. Mannion R, Goddard M. The impact of performance measurement in the NHS. Report 2: Performance measurement: a cross-sectoral study. York: Centre for Health Economics, University of York, 2000
27. Smith P, Mannion R, Goddard M. Performance management in health care: information, incentives and culture. HM Treasury series on public services productivity. London: HM Treasury, 2003
28. Goddard M, Mannion R, Smith P. Enhancing performance in health care: a theoretical perspective on agency and the role of information. *Health Economics* 2000; 9: 95–107
29. 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. *Journal of the American Medical Association* 2001; 283: 1866–1874
30. Mannion R, Davies HTO. Reporting health care performance: learning from the past; prospects for the future. *Journal of Evaluation in Clinical Practice* 2002; 8: 215–228
31. Coulter H, Fitzpatrick R, Davis L. Patient and public perspectives on health care performance. Report prepared for the Commission for Health Improvement, 2002